Hoofdstuk 7: Statistische gevolgtrekkingen voor distributies

7.1 Het gemiddelde van een populatie

Standaarddeviatie van de populatie en de steekproef

In het vorige deel is bij de significantietoets uitgegaan van een bekende σ. In de praktijk kennen we de standaarddeviatie van de populatie vaak niet. De standaarddeviatie van de steekproef (s) wordt dan gebruikt om σ te schatten.

De t-distributie

We gebruiken de t-distributie (in plaats van de z-distributie) wanneer we de standaarddeviatie van de populatie niet kennen.

- Wanneer de standaarddeviatie van een statistiek uit de data wordt geschat, dan wordt het resultaat de (geschatte) *standaardfout* van de statistiek genoemd. De standaardfout van het steekproefgemiddelde is: $SE_M = s/\sqrt{n}$.
- Het gestandaardiseerde steekproefgemiddelde wordt ook wel weergegeven met de één-steekproef-z-toets: $z=(\bar{x}-\mu)/(\sigma/\sqrt{n})$. Deze toets is de normaalverdeeld: N(0,1). Wanneer we $(\sigma/\sqrt{n})$ vervangen door $(s/\sqrt{n})$, dan is er echter geen sprake meer van een normaalverdeling. Er is nu een *t-distributie* ontstaan.
- Als er een SRS van grootte n uit een normaalverdeelde populatie N(μ,σ) wordt getrokken, dan is de één-steekproef-t-toets: $t=(\bar{x}-\mu)/(s/\sqrt{n})$. Dit is een *t-distributie* met n-1 *vrijheidsgraden (k).*
- Bij elke vrijheidsgraad hoort een andere t-distributie. De dichtheidscurves van de t-distributies lijken qua vorm op die van de bekende normaal verdeelde curve. De piek is echter lager en de staarten staan wat hoger. Er is dus meer spreiding aanwezig. Dit komt omdat de standaarddeviatie van de populatie niet gebruikt wordt in de formule. De standaarddeviatie van de steekproef zorgt voor meer spreiding. Tabel D geeft kritische t*-waarden voor de t-distributies. Bij het gebruiken van de tabel moet gekeken worden naar de bijbehorende vrijheidsgraden.

Betrouwbaarheidsintervallen voor t-distributies

Als je een SRS van grootte n uit een populatie met een onbekend gemiddelde (μ) trekt, dan is het betrouwbaarheidsinterval (C) voor μ:

- $\bar{x} \pm t^*(s/\sqrt{n})$. In dit verband is t* de waarde voor de t(n-1) dichtheidscurve met gebied C tussen –t* en t*. In deze formule staat $t^*(s/\sqrt{n})$ voor de *foutenmarge.*

De t-toets

Het toetsen van significantie met een t-toets lijkt erg op het toetsen van significantie met de z-toets. De t-waarde wordt verkregen door:

- $t= (\bar{x}- \mu_0)/(s/\sqrt{n})$. De bijbehorende p-waarde kan opgezocht worden in tabel D achterin het boek.

- We kunnen er, net zoals bij de z-toets, voor kiezen om eenzijdig of tweezijdig te toetsen. Als er geen vermoeden over de richting van het effect bestaat, dan is het altijd beter om een tweezijdige toets uit te voeren.

Gematchte paren en de t-distributie
Bij een gematchte paren-onderzoek vormen deelnemers paren. Vervolgens wordt er naar de onderzoekresultaten binnen elk paar gekeken. De onderzoeker kan bijvoorbeeld twee vormen van therapie uitproberen op de verschillende leden van een paar, om te kijken of er verschillen in uitkomsten optreden. We willen elk lid van een paar dus met het andere lid vergelijken. Er wordt voor elk paar (of elk individu) een verschilscore berekend. Deze scores worden gebruikt als data. Tot slot kunnen met deze informatie t-betrouwbaarheidsintervallen en t-significantietoetsen uitgevoerd worden. Zo een onderzoek wordt vaak ook gedaan als randomisatie niet mogelijk is.

Robuustheid
De uitkomsten van de één-steekproef-t-toets zijn helemaal juist wanneer de populatie normaal verdeeld is. In werkelijkheid is geen enkele populatie precies normaal verdeeld. De bruikbaarheid van de t-distributie hangt in de praktijk daarom vooral af van hoe niet-normaal verdeeld een populatie is. Een distributie die niet erg door een niet-normaal verdeelde populatie wordt beïnvloed, wordt *robuust* genoemd.

- Een statistische procedure is *robuust* wanneer de benodigde kansberekeningen niet worden beïnvloed als niet aan de voorwaarden voor die kansberekeningen wordt voldaan. De voorwaarde voor de t-distributie is dus dat de populatie normaal verdeeld is, wat vaak niet het geval blijkt te zijn. Toch is de t-distributie robuust.

Robuustheid van de t-distributie
De t-distributie is behoorlijk robuust tegen niet-normaliteit van de populatie, behalve als er uitbijters zijn of als er een sterke afwijking naar links of rechts aanwezig is. Grote steekproeven verbeteren de accuraatheid van de p-waarden wanneer de populatie niet normaal verdeeld is. Dit is waar om de volgende twee redenen:

- Allereerst is de steekproevendistributie van een steekproefgemiddelde $\bar{x}$ van een grote steekproef bijna normaal verdeeld. Normaliteit van individuele observaties is niet erg belangrijk wanneer de steekproef groot genoeg is.
- Als de steekproefgrootte n groter wordt, dan zal de steekproefstandaarddeviatie s een betere schatter zijn van σ, of de populatie nou wel of niet normaal verdeeld is. Het is slim om de t-distributie pas te gebruiken als de steekproef 15 of meer deelnemers heeft. Als er sprake is van een hele kleine steekproef, dan is de aanname dat de data van een SRS afkomstig belangrijker dan de aanname dat de populatiedistributie normaalverdeeld is.

Steekproefgrootte en de t-distributie
- Als de steekproef minder dan 15 deelnemers bevat, dan kan de t-distributie alleen gebruikt worden wanneer de data bijna normaal verdeeld is. Als de data niet normaal verdeeld is en er uitbijters aanwezig zijn, dan dient de t-distributie niet gebruikt te worden.
- Als de steekproef minstens 15 deelnemers heeft, kan dan de t-distributie gebruikt worden, behalve als er sprake is van uitbijters of een sterke afwijking naar links of rechts.

- De t-distributie kan zelfs gebruikt worden voor distributies met een sterke afwijking naar links of rechts als de steekproef 40 of meer deelnemers bevat.

Power van de t-toets
De power van een toets meet het vermogen van de toets om afwijkingen van de nulhypothese vast te stellen. De precieze berekening van de power van de t-toets is wat complexer, omdat er rekening gehouden moet worden met het feit dat de standaarddeviatie van de steekproef gebruikt moet worden om de standaarddeviatie van de populatie te schatten. Vaak wordt de precieze power niet berekend, maar wordt er genoegen genomen met een benadering van de power. Deze berekening lijkt erg op die van de z-toets:

- Bepaal een standaarddeviatie, significantieniveau, de een- of tweezijdigheid van de toets en een alternatieve hypothese. Het is altijd beter om een waarde van de standaarddeviatie te gebruiken die iets groter is dan wat we zouden verwachten.
- Noteer bij welke gebeurtenis de nulhypothese afgewezen zal worden (in termen van $\bar{x}$).
- Vind de kans op deze gebeurtenis als de alternatieve hypothese waar zou zijn. Voor een voorbeeld zie bladzijde 415.

Statistiek voor populaties die niet normaal verdeeld zijn
Wat te doen als een populatie niet normaal verdeeld is en als de steekproef klein is? Er zijn in dat geval drie alternatieven:

- Soms geeft een niet-normaal verdeelde distributie een goede beschrijving van de data. Er zijn veel niet-normaal verdeelde modellen voor data en er zijn ook statistische procedures beschikbaar voor deze modellen.
- Het is ook mogelijk om te proberen de afwijkende data te transformeren in normaal verdeelde data. Er zal dan geen sprake zijn van een perfect normaal verdeelde distributie, maar de distributie zal wel zo normaal verdeeld mogelijk zijn. Deze transformatie kan gemaakt worden met het *logaritme.* Daarmee wordt de rechterstaart van een distributie aangepakt.
- Tot slot kan een *distributievrije* methode gebruikt worden. Zo een methode heeft niet als aanname dat de populatie normaal verdeeld moet zijn. Distributievrije procedures worden ook wel *nonparametrische procedures* genoemd.

De tekentest (The sign test)
Een voorbeeld van de makkelijkste en meest gebruikte nonparametrische procedure is de *tekentest.*

- Negeer bij een tekentest de paren met een verschil van nul: het aantal gebeurtenissen n is de telling van de overige paren. De teststatistiek is de telling X van paren met een positief verschil. P-waarden voor X zijn gebaseerd op de binomiale *B(n,1/2)* distributie.

De tekentest toetst in feite de hypothese dat de mediaan van de verschillen nul is. Als p de kans is dat een verschil positief is, dan is p 0.5 als de *mediaan* 0 is. De nulhypothese is dat de populatiemediaan 0 is en de alternatieve hypothese is dat de populatiemediaan groter dan 0 is. De tekentest maakt geen gebruik van echte verschillen, er wordt alleen gekeken of scores gestegen zijn. Omdat er niet naar inhoudelijke verschillen wordt gekeken, is de tekentest veel minder sterk dan de t-toets. Voor een voorbeeld zie bladzijde 438.

7.2 Twee gemiddelden vergelijken

Twee groepen

Het doel van statistische gevolgtrekkingen is vaak het vergelijken van responsen in twee groepen. Elke groep wordt als een aparte steekproef uit een populatie gezien. De responsen van de groepen zijn onafhankelijk van elkaar. Bij twee aparte steekproeven is er geen sprake van gematchte paren en statistische procedures om twee steekproeven te vergelijken verschillen dan ook van statistische procedures om gematchte paren te vergelijken. We bestuderen twee onafhankelijke steekproeven en dus ook twee afzonderlijke populaties. Dezelfde (afhankelijke) variabele wordt gemeten voor beide steekproeven.

- We kunnen de variabele x_1 noemen voor de eerste populatie en x_2 voor de tweede populatie. Dit omdat de variabele verschillende distributies in de twee populaties kan hebben.
- Het gemiddelde van de eerste populatie noemen we μ_1 en het gemiddelde van de tweede populatie noemen we μ_2.
- De standaarddeviatie van de eerste populatie is σ_1 en voor de tweede populatie is dat σ_2.

De z-toets voor twee steekproeven

De nulhypothese die we willen onderzoeken is of het gemiddelde van beide populaties hetzelfde is, dus: $H_{0:}\ \mu_1 = \mu_2$. Dit kunnen we uitzoeken door middel van twee steekproeven met gemiddelden $\bar{x}_1$ en $\bar{x}_{2.}$ Deze steekproeven zijn van grootte n_1 en n_2. De bijbehorende steekproefstandaarddeviaties zijn s_1 en s_2. We schatten het verschil tussen de populatiegemiddelden dus door middel van het verschil tussen de steekproefgemiddelden. Grote steekproeven zijn nodig om kleine verschillen vast te stellen. De z-toets voor twee steekproeven wordt als volgt gevonden:

- Vind eerst: ($\bar{x}_1$ en $\bar{x}_2$)- (μ_1 - μ_2).
- Trek daarna de wortel uit ($\sigma^2_1/\ n_1+\sigma^2_2/\ n_2$).
- Deel de eerste uitkomst door de tweede uitkomst.

De z-toets voor twee steekproeven heeft de N(0,1)-steekproevendistributie. De z-toets voor twee steekproeven wordt echter zelden gebruikt, aangezien de standaarddeviaties van de populaties zelden geweten worden. In de praktijk wordt daarom veel vaker de t-toets voor twee steekproeven gebruikt.

De t-toets voor twee steekproeven

De t- toets voor twee steekproeven wordt als volgt gevonden:

- Bereken eerst ($\bar{x}_1$ en $\bar{x}_2$)
- Trek daarna de wortel uit ($s^2_1/\ n_1+s^2_2/\ n_2$).
- Deel de eerste uitkomst door de tweede uitkomst.

Deze toets heeft echter geen t-distributie. Een t-distributie vervangt de N(0,1)-distributie alleen wanneer een enkele standaarddeviatie (σ) in een z-toets wordt vervangen door een steekproefstandaarddeviatie (s). Bij de t-toets worden echter beide standaarddeviaties (σ_1 en σ_2) vervangen door s_1 en s_2. Toch kunnen we de t-distributie voor twee steekproeven benaderen door een *benadering met de vrijheidsgraden (k)* te maken. We gebruiken deze benadering om waarden voor t* te vinden voor betrouwbaarheidsintervallen en om p-waarden te vinden voor significantietoetsen. Hoe maken we deze benadering? Dit kan op twee manieren. Hieronder zullen die manieren beschreven worden.

- Gebruik een waarde van k die met de data berekend is. Vaak is dit geen heel getal. Vaak maken computerprogramma's gebruik van deze manier. De berekende vrijheidsgraden zijn in dit geval minstens even groot als de kleinste van n_1-1 en n_2 -1. Het is echter ook zo dat de berekende vrijheidsgraden nooit groter zijn dan n_1+n_2 -2.
- Bereken n_1-1 en n_2-1 en kies de kleinste uitkomst. Deze manier is het gemakkelijkst wanneer handmatig te werk wordt gegaan.

Eerst moet de t- toets voor twee steekproeven dus berekend worden. Vervolgens moeten de vrijheidsgraden per steekproef gevonden worden. De kleinste moet gekozen worden om gebruik te maken van tabel D achterin het boek.

Het t-betrouwbaarheidsinterval voor twee steekproeven

Stel: we trekken een SRS van grootte n_1 uit een normaalverdeelde populatie met een onbekende μ_1 en we trekken ook een onafhankelijke SRS van grootte n_2 uit een andere normaalverdeelde populatie met een onbekende μ_2. In dat geval wordt het betrouwbaarheidsinterval als volgt berekend:

- Eerst moet $\bar{x}_1$-$\bar{x}_2$ berekend worden.
- Daarna moet t* de wortel uit (s^2_1/ n_1+s^2_2/ n_2) berekend worden.
- De gehele formule is: $\bar{x}_1$-$\bar{x}_2$ ± t* de wortel uit (s^2_1/ n_1+s^2_2/ n_2).

Wat de standaarddeviaties van de populaties zijn is dus niet van belang voor deze formule. Voor een voorbeeld zie bladzijde 435.

Robuustheid van berekeningen op basis van twee steekproeven

De t-toets op basis van twee steekproeven is robuuster dan de t-toets die op een enkele steekproef is gebaseerd. Wanneer beide steekproeven van dezelfde grootte zijn en de bijbehorende populaties dezelfde vorm hebben, dan zijn de waarden uit de t-tabel behoorlijk accuraat, als de steekproeven maar minimaal uit 5 deelnemers bestaan. Het is daarom aan te raden om steekproeven van gelijke grootte te kiezen als dat mogelijk is. De t-toets voor twee steekproeven zijn het meest robuust tegen niet-normaliteit in dit geval. Wanneer de twee populatiedistributies verschillende vormen hebben, zijn er grotere steekproeven nodig.

Kleine steekproeven

De power van significantietoetsen is vaak klein wanneer gebruik gemaakt wordt van kleine steekproeven. De foutenmarge dat bij de betrouwbaarheidsintervallen hoort is juist erg groot. Ondanks deze moeilijkheden kunnen we toch belangrijke conclusies trekken op basis van kleine steekproeven. De steekproefgrootte heeft invloed op de p-waarde van een toets. Een effect dat niet significant is op een specifiek significantieniveau, kan dat wel worden als een grotere steekproef uit de populatie getrokken wordt.

De gepoolde t-toets voor twee steekproeven

Er is één situatie waarin een t-toets om twee gemiddelden te vergelijken precies een t-distributie heeft. Bij deze situatie wordt meer gewicht gegeven aan de grotere steekproef, want deze bevat meer informatie. We zeggen dan dat we de t-toets poolen. Het resultaat is:

- $s^2_p = (n_1-1)s^2_1+(n_2-1)s^2_2/\ n_1+n_2-2$. De uitkomst wordt ook wel de *gepoolde schatter van σ²* genoemd. Als we hier de wortel uittrekken, dan hebben we de gepoolde standaarddeviatie gevonden. Deze kunnen we weer gebruiken om een betrouwbaarheidsinterval te berekenen:
- $\bar{x}_1-\bar{x}_2 \pm t^*s_p \sqrt{\ }\ (s^2_1/\ n_1+s^2_2/\ n_2)$.
- Om de nulhypothese te toetsen (die beweert dat de gemiddelden van beide populaties hetzelfde is), kan de gepoolde standaarddeviatie ook gebruikt worden: $\bar{x}_1 - \bar{x}_2/\ s_p \sqrt{\ }\ 1/n_1+1/n_2$.

Het nadeel van t-toetsen die op twee steekproeven gebaseerd zijn, is dat ze uitgaan van de aanname dat de twee onbekende populatiestandaarddeviaties gelijk aan elkaar zijn. Deze aanname is moeilijk te verifiëren. Het poolen van t-toetsen is daarom risicovol.

7.3 Spreiding binnen de populatie

Spreiding

De belangrijkste kenmerken van een distributie zijn het middenpunt en de spreiding. We gebruiken onder andere de F-toets om de spreiding van twee normaal verdeelde populaties te beschrijven. De F-toets en andere statistische procedures die standaarddeviaties gebruiken zijn, in tegenstelling tot de t-toetsen, erg gevoelig voor niet-normaal verdeelde distributies. Dit gebrek aan robuustheid vermindert niet wanneer er grotere steekproeven geselecteerd worden. Het is niet altijd duidelijk wat een significante F-toets nou precies bewijst. Is een significante F-toets bewijs voor het feit dat er sprake is van ongelijke populatiespreidingen? Of zegt een significante F-toets alleen maar dat er sprake is van niet-normaalverdeelde populaties?

De F-toets

Stel: we trekken twee onafhankelijke SRS-en van grootte n_1 en n_2 uit normaalverdeelde populaties. We vinden dan twee varianties: s^2_1 en s^2_2. De F-toets is dan:

- $F= s^2_1/s^2_2$.
- De F-distributie heeft n_1-1 en n_2-1 vrijheidsgraden.

De F-distributies worden bepaald door twee vrijheidsgraden die samengaan met de steekproefvarianties. Deze vrijheidsgraden zijn de noemer en teller van de F-ratio. Deze F-ratio wordt ook wel afgekort met F(j,k). J wordt ingevuld met het aantal vrijheidsgraden in de noemer en k wordt ingevuld met het aantal vrijheidsgraden in de teller. Op basis van deze gegevens kan tabel E achterin het boek gebruikt worden om p-waarden te vinden. De F-ratio is altijd groter dan 1, omdat de grootste variantie gedeeld moet worden door de kleinste variantie. Dus: F= grotere s²/ kleinere s². De p-waarden uit de tabel moeten *verdubbeld* worden, omdat F-toetsen altijd tweezijdige toetsen zijn.

Robuustheid

- Concluderend kan gezegd worden dat t-distributies behoorlijk robuust zijn tegen niet-normaalverdeelde populatiedistributies. T-distributies zijn vooral robuust wanneer de populatiedistributies symmetrisch zijn en wanneer de groottes van de steekproeven gelijk zijn.
- De F-toets en andere methoden die gebruik maken van varianties zijn niet robuust en daarom is het gebruik van deze methoden niet aan te raden.