Onderzoekspracticum 2, aantekeningen college 2

Afhankelijke en onafhankelijke steekproeven

In een experiment worden vaak twee groepen vergeleken. De steekproefgemiddelden van deze groepen kunnen met elkaar worden vergeleken aan de hand van statistisch toetsen. Steekproeven kunnen afhankelijk of onafhankelijk zijn.

De volgende kenmerken horen bij een afhankelijke steekproef:

- Er zijn gematchte paren. Elke proefpersoon is gekoppeld aan een partner in de andere groep. Bijvoorbeeld broer en zus of mensen met dezelfde eigenschap(pen).
- Ook een herhaalde meting valt onder een afhankelijke steekproef. De proefpersoon wordt aan zichzelf gematcht. Dit is bijvoorbeeld het geval bij een voor- en nameting.
- De aantallen in beide groepen zijn altijd gelijk.
- De onderzoeker verkrijgt N (aantal paren) stukjes onafhankelijke informatie.

De volgende kenmerken horen bij een onafhankelijke steekproef:

- Er wordt gebruik gemaakt van twee aselecte steekproeven waartussen geen verband bestaat. Bijvoorbeeld een controlegroep en een experimentele groep.
- De aantallen in de groepen kunnen ongelijk zijn.
- De onderzoeker verkrijgt $n_1 + n_2$ (n is steekproefgrootte) stukjes onafhankelijke informatie.

Afhankelijke steekproeven: toetsen van verschil in gemiddelden

μ_d is het verschil tussen het gemiddelde van steekproef 1 en het gemiddelde van steekproef 2. Dit wordt ook wel het gemiddelde verschil genoemd en kun je noteren als $\mu_d = \mu_2 - \mu_1$. De nulhypothese en de alternatieve hypothese noteren we als volgt:
H_0: $\mu_d = 0$ (want in dat geval geldt dat $\mu_1 = \mu_2$)
H_a: $\mu_d \neq 0$ (dit is alleen bij tweezijdig toetsen. Bij enkelzijdig toetsen geldt $\mu_d > 0$ of $\mu_d < 0$)

De t-toets voor afhankelijke steekproeven wordt ook wel de gepaarde t-toets genoemd:

$$t = \frac{\overline{d} - \mu_d}{S_d / \sqrt{N}}$$

Hierin staat $\overline{d}$ voor het gemiddelde verschil tussen de steekproeven. Voor μ_d kun je 0 invullen, omdat we bij experimenten altijd uitgaan van de nulhypothese. S_d is de standaarddeviatie van het verschil en N is de steekproefgrootte. Let op: N is het aantal paren. De gevonden t-waarde kan worden opgezocht in Tabel D (*Introduction to the Practice of Statistics* van Moore, McCabe en Craig) bij het juiste aantal vrijheidsgraden N-1. Bij tweezijdig toetsen verdubbel je de bijbehorende p-waarde. Als $p < \alpha$, dan kun je de nulhypothese verwerpen.

Het kan ook zijn dat je eenzijdig toetst, maar dat het gemiddelde verschil de andere kant uitvalt dan je had verwacht. Je vindt dan bijvoorbeeld geen gemiddeld verschil van -3, maar van 3. Dit wordt een contra-intuïtief resultaat genoemd. De oplossing: gebruik niet p, maar 1-p. In plaats van bijvoorbeeld $0{,}005 < p < 0{,}01$, wordt je p-waarde dan $0{,}99 < 0 < 0{,}995$.

Afhankelijke steekproeven: betrouwbaarheidsinterval t-toets

Het C%-betrouwbaarheidsinterval van μ_d kan als volgt worden bepaald:

$$\overline{d} \pm t^* \frac{s_d}{\sqrt{N}}$$

Zoek in Tabel D (*Introduction to the Practice of Statistics* van Moore, McCabe en Craig) de t-waarde op bij het juiste betrouwbaarheidsinterval en het juiste aantal vrijheidsgraden (N-1).
De conclusie die bij een 95%-betrouwbaarheidsinterval hoort, luidt: in de populatie ligt het gemiddelde verschil tussen de …(getal) en …(getal) met 95% zekerheid.

Aannamen gepaarde t-toets
Een aanname van de t-toets is dat er een normaalverdeling van de populatie is. De robuustheid van de t-toets houdt in hoe goed de t-toets bestand is tegen schending van deze aanname. Er zijn een aantal vuistregels voor het gebruik van de t-toets:

- N< 15: geen t-toets gebruiken bij uitbijters of duidelijke niet-normaliteit.
- N: 15 – 39: geen t-toets gebruiken bij uitbijters of duidelijke scheefheid.
- N ≥ 40: (bijna) altijd gebruik maken van de t-toets.

Of data normaal verdeeld zijn, kan worden bepaald door middel van een normaalkwantielplot, waarbij de verwachte waarden worden uitgezet tegen de geobserveerde waarden. De data worden als normaal beschouwd wanneer alle punten bij benadering op een rechte lijn liggen.

Overige aannamen zijn dat de proefpersonen gematcht zijn (zo niet dan gebruik je een andere toets) en dat ze aselect getrokken zijn. Aselect trekken is echter doorgaans niet haalbaar, waardoor resultaten vaak beperkt generaliseerbaar zijn.

Afhankelijke steekproeven: effectgrootte
Significante resultaten geven wel de waarschijnlijkheid, maar niet de grootte aan van een effect. Bovendien geldt: hoe groter de steekproef, hoe eerder een significant effect. Effectgroottes daarentegen zijn niet afhankelijk van de steekproefgrootte en geven juist wel aan hoe groot een effect is. De effectgrootte bij een gepaarde t-toets kun je berekenen door middel van een gestandaardiseerde maat, namelijk Cohen's d:

$$d = \frac{|\bar{x}_1 - \bar{x}_2|}{s_d / \sqrt{2(1-r)}}$$

In deze formule is r de correlatie tussen paren of herhaalde metingen. Cohen's d geeft aan hoe groot het gemiddelde verschil tussen de groepen is in relatie tot de standaarddeviatie van het verschil. Er zijn enkele vuistregels voor de Cohen's d effectgrootte:

- Kleiner dan 0.2: het effect is verwaarloosbaar
- 0.2 tot 0.4: er is een klein effect
- 0.5 tot 0.7: er is een gemiddeld effect
- 0.8 en groter: er is een groot effect

Let op: 0.77 valt bijvoorbeeld onder gemiddeld effect. Vanaf 0.8 is een groot effect. Of je tevreden bent met de een bepaalde effectgrootte is afhankelijk van de context.

Toetsingsschema gepaarde t-toets
1. Formuleer je onderzoeksvraag.
2. Ga de aannamen af en ga na of eraan is voldaan. Zo niet, ga dan na wat de gevolgen zijn.
3. Stel je hypothesen op.
4. Kies de juiste toets, in dit geval t-toets, en bepaal alpha α (vaak 0.05)
5. Voer je berekening uit.
6. Zoek de bijbehorende p-waarde op.
7. Neem een beslissing: vergelijk p met α. Als p < α dan verwerp je de nulhypothese.
8. Bereken de effectgrootte door middel van Cohen's d.
9. Trek je conclusie en beschrijf deze inhoudelijk. Geef de richting aan van het effect dat je gevonden hebt en plaats kanttekeningen als je (één van) de aannamen geschonden hebt.

Onafhankelijke steekproeven: toetsen van verschil in gemiddelden
Het verschil in gemiddelden tussen twee steekproeven kan berekend worden door de z-toets. De gemiddelden van steekproef 1 en steekproef 2 zijn dan $\bar{x}_1$ en $\bar{x}_2$, en we gaan ervan uit dat de standaarddeviaties van de populaties (σ_1 en σ_2) ook bekend zijn. Wat we dan onderzoeken is of beide populatiegemiddelden van elkaar verschillen. De nulhypothese die hierbij hoort is H_0: $\mu_1 = \mu_2$ of $\mu_1 - \mu_2 = 0$. De z-toets voor twee gemiddelden is:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

In de praktijk zijn de standaarddeviaties van de populaties echter vrijwel nooit bekend, maar de standaarddeviaties van de steekproeven (s_1 en s_2) wel. Daarom wordt vaak de t-toets gebruikt:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

In dit geval heeft t echter geen t-verdeling. Het exacte aantal vrijheidsgraden kun je namelijk niet bepalen. Wel kun je deze op twee manieren benaderen, met behulp van een t(k) verdeling, waarin k een benadering is van het aantal vrijheidsgraden:

1. Neem de kleinste waarde van n_1-1 en n_2-1. Dit is veel minder nauwkeurig dan de tweede manier, maar deze methode mogen we gebruiken.
2. Gebruik software, bijvoorbeeld SPSS, om het aantal vrijheidsgraden te benaderen. Deze software gebruikt hiervoor een uitgebreide formule.

Onafhankelijke steekproeven: betrouwbaarheidsinterval t-toets
Het C%-betrouwbaarheidsinterval van μ_1-μ_2 wordt op de volgende manier berekend:

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Onafhankelijke steekproeven: effectgrootte
Bij onafhankelijke steekproeven, wordt Cohen's d op de volgende manier berekend:

$$d = \frac{|\bar{x}_1 - \bar{x}_2|}{s_p}$$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

S_p is in dit geval de samengestelde standaarddeviatie.

Toetsingsschema t-toets bij onafhankelijke steekproeven
Het toetsingsschema is hetzelfde als bij de gepaarde t-toets. Wel is er een verschil in aannamen. Evenals bij de gepaarde t-toets moet de afhankelijke variabele normaal verdeeld zijn en moeten de proefpersonen aselect getrokken worden, maar in dit geval moet dit *onafhankelijk* van elkaar gebeuren. Proefpersonen worden niet gematcht.