Statistiek 2

Week 4: Toetsen met binomiale verdeling (fracties)

Toetsen (Boek/reader 9.1)

Bij toetsen wordt een uitspraak getest, vaak met betrekking tot een parameter van een kansverdeling. De uitspraak wordt getoetst aan de hand van een steekproef.
De **nulhypothese H_0** is een uitspraak (over een parameter). Bij het toetsen wordt gekeken of deze uitspraak met een bepaalde betrouwbaarheid waar is of niet. Als de uitspraak niet waar wordt bevonden, dan wordt de ontkenning van de uitspraak als waar aangenomen. De ontkenning wordt de **alternatieve hypothese H_1** genoemd. Je voert een steekproef uit en kijkt hoe waarschijnlijk de uitkomst is onder de aanname dat de nulhypothese waar is.

Voorbeeld: Je hebt een dobbelsteen waarvan je vermoedt dat hij (of zij, of het) vals is. Je formuleert de hypotheses als volgt:

H_0: De kans op het gooien van een 6 is $p = \frac{1}{6}$ (d.w.z., de dobbelsteen is eerlijk, wat 6 gooien betreft)

H_1: $p \neq \frac{1}{6}$ (d.w.z., de dobbelsteen is oneerlijk).

Om dit te testen doe je een steekproef: Je gooit de dobbelsteen 30 keer. Je verwacht ongeveer 30/6 = 5 keer een zes maar je krijgt 10 keer een zes. Is dit zo extreem dat je kunt concluderen dat de dobbelsteen vals?

Toetsen Voorbeeld Aanpak 1

H$_0$: De dobbelsteen is eerlijk.
H$_1$: De dobbelsteen is oneerlijk.

Je gooit de dobbelsteen 30 keer en je krijgt 10 keer een zes. Kun je H$_0$ verwerpen, d.w.z., kun je concluderen dat de dobbelsteen inderdaad vals is?

De toetsingsgrootheid $\underline{k}$ is het aantal zessen uit 30 worpen. Als H$_0$ waar is verwacht je ongeveer 30/6 = 5 keer een zes te gooien. Je wilt H$_0$ verwerpen als $\underline{k}$ daar teveel van afwijkt. Je kunt hier naar boven of naar beneden afwijken, dus je toetst tweezijdig en je wilt de twee **grenswaarden** g_1 en g_1 van het acceptatiegebied bepalen.

Als het aantal zessen in de steekproef ($\underline{k}$) tussen g_1 en g_1 ligt (met grenzen erbij) neem je aan dat H$_0$ waar is (acceptatiegebied). Als het aantal kleiner dan g_1 is of groter dan g_2, dan wijkt het teveel af van 5 en verwerp je H$_0$ (kritiek gebied), je neemt aan dat H$_1$ waar is.

Toetsen Voorbeeld Aanpak 1

De grenswaarden moet je zo kiezen, dat de kans klein is dat je ten onrechte H_0 verwerpt, d.w.z., als H_0 waar is (de dobbelsteen is eerlijk) moet de kans klein zijn dat je op grond van je steekproef concludeert dat H_1 waar is (de dobbelsteen is oneerlijk). Dit heet een **fout van de eerste soort**:

$$P(\, H_1 \,|\, H_0 \,) \;\leq\; \alpha$$

De acceptabele foutkans (onbetrouwbaarheid) α is gegeven of kies je van tevoren, bv. $\alpha = 0.05$ (of $\alpha = 0.01$)
g_1 en g_1 moeten vervolgens zo worden gekozen dat

$$P(\, H_1 \,|\, H_0 \,) = P\left(\underline{k} < g_1 \text{ of } \underline{k} > g_2 \,|\, H_0\right) \leq 0{,}05$$

Hierbij voldoet $\underline{k}$ aan een binomiale verdeling met $n = 30$ en $p = \frac{1}{6}$ omdat we uitgaan van H_0.
We kiezen nu g_1 en g_1 zo dat de linker- en de rechteroverschrijdingskansen elk maximaal de helft zijn van de gekozen onnauwkeurigheid (hier 0,05):

$$P\left(\underline{k} < g_1 \,|\, H_0\right) \leq 0{,}025 \;\text{ en }\; P\left(\underline{k} > g_2 \,|\, H_0\right) \leq 0{,}025$$

Het vinden van deze g_1 en g_1 is lastig, omdat het om discrete variabelen gaat. (Het boek gebruikt hiervoor een benadering met de normale verdeling met continuïteitscorrectie, maar dat is nog lastiger).

Toetsen Voorbeeld Aanpak 1

We gaan toch proberen om de vergelijkingen

$$P(\underline{k} < g_1 |\ H_0) \leq 0{,}025 \text{ en } P(\underline{k} > g_2 |\ H_0) \leq 0{,}025$$

op te lossen met de grafische rekenmachine, hoewel dat eigenlijk niet kan omdat k discreet is.

Eerst de berekening van g_1. We lossen op met de MATH-SOLVER

$$P(\underline{k} < g_1 |\ H_0) = P(\underline{k} \leq g_1 - 1 |\ H_0) = \mathbf{binomcdf}\left(\boldsymbol{n = 30\ , p = \frac{1}{6}, k = g_1 ? - 1}\right) = \mathbf{0{,}025}$$

Dit levert $g_1 = 2{,}0000$. We ronden af en controleren $g_1 = 2$:

$$\text{binomcdf}\left(n = 30\ , p = \frac{1}{6}, k = \mathbf{2} - 1\right) = 0{,}029 \leq ?\ 0{,}025$$

Dit klopt niet helemaal. We moeten dus een kleinere waarde kiezen: $g_1 = 1$:

$$\text{binomcdf}\left(n = 30\ , p = \frac{1}{6}, k = \mathbf{1} - 1\right) = 0{,}00421 < 0{,}025$$

Die klopt, dus $\boldsymbol{g_1 = 1}$ voldoet en is de grootst mogelijke gehele waarde.

Toetsen Voorbeeld Aanpak 1

Hetzelfde doen we voor de vergelijking $P(\underline{k} > g_2 | H_0) \leq 0{,}025$

We lossen op met de MATH-SOLVER:

$$P(\underline{k} > g_2 | H_0) = 1 - P(\underline{k} \leq g_2 | H_0) = 1 - \text{binomcdf}\left(n = 30\,, p = \frac{1}{6}, k = g_2?\right) = 0{,}025$$

$$\mathbf{binomcdf}\left(\boldsymbol{n = 30\,, p = \frac{1}{6}, k = g_2?}\right) \boldsymbol{= 0,975}$$

Dit levert $g_2 = 9{,}0000$. We controleren dit:

$$\text{binomcdf}\left(n = 30\,, p = \frac{1}{6}, k = 9\right) = 0{,}9803 \geq ?\, 0{,}975$$

Dit klopt. Is een kleinere waarde mogelijk: $g_2 = 8$?

$$\text{binomcdf}\left(n = 30\,, p = \frac{1}{6}, k = 8\right) = 0{,}9494 \geq ?\, 0{,}975$$

Nee, die is te klein, dus $\boldsymbol{g_2 = 9}$ is de kleinst mogelijke gehele waarde die voldoet.

Als het aantal zessen $g_1 = 1, 2, \ldots, 9 = g_2$ is, verwerpen we H_0 niet en is de bewijslast niet genoeg om met 95% betrouwbaarheid te zeggen dat de dobbelsteen vals is.
Als het aantal zessen 0, of 10, 11, …, 30 is (zoals nu) kunnen we dat wel zeggen.

Toetsen Voorbeeld Aanpak 1

Een berekening m.b.v. de p-waarde, i.p.v. met het kritieke gebied berekent de overschrijdingskans:

$\underline{k}$ heeft een verwachte waarde van $30 \cdot \frac{1}{6} = 5$. De steekproef ligt $|10-5| = 5$ hier vanaf.

De p-waarde is (omdat het om een tweezijdige toets gaat) de kans dat de afwijking van de verwachte waarde 5 is, of meer:

$p = P\left(\left|\underline{k} - 5\right| \geq 5 \middle| H_0\right) = P\left(\underline{k} \leq 0 \middle| H_0\right) + P\left(\underline{k} \geq 10 \middle| H_0\right)$. Nu geldt:

$$P\left(\underline{k} \leq 0 \middle| H_0\right) = \text{binomcdf}\left(n = 30\,, p = \frac{1}{6}, k = 0\right) = 0{,}00421$$

$$P\left(\underline{k} \geq 10 \middle| H_0\right) = 1 - P\left(\underline{k} \leq 9 \middle| H_0\right) = 1 - \text{binomcdf}\left(n = 30\,, p = \frac{1}{6}, k = 9\right) = 0{,}01971$$

Dus $p = 0{,}00421 + 0{,}01971 = 0{,}02392$.

Deze waarde is kleiner dan de gekozen $\alpha = 0{,}05$, dus H_0 wordt verworpen, de dobbelsteen is vals met een betrouwbaarheid van 95%.

De betrouwbaarheid is zelfs $100(1-p) = 97{,}608 > 97{,}5\%$.

Toetsen Voorbeeld Aanpak 2

We nemen nu (enkelzijdige toets)

$$\mathrm{H}_0 \colon p \leq \frac{1}{6}$$

$$\mathrm{H}_1 \colon p > \frac{1}{6}$$

Dan betekent H_0 weliswaar niet meer: De dobbelsteen is niet vals, maar als het je lukt om H_0 te verwerpen is H_1 waar en kun je concluderen dat de kans op een zes gooien met deze dobbelsteen groter is dan 1/6.

Laten we kijken of we dat kunnen aantonen. We zoeken nu één grenswaarde g zodat $\underline{k} \leq g$ overeenkomt met H_0 en $\underline{k} > g$ met H_1.

We kunnen g weer berekenen, want $P(\underline{k} > g) \leq 0{,}05$ is hetzelfde als $P(\underline{k} \leq g) > 0{,}95$ voor de binomiale verdeling.

Nu is er wel een probleem, want H_0 betekent nu niet meer dat we uit kunnen gaan een vaste binomiale verdeling met $p = \frac{1}{6}$, maar het moet nu gelden voor alle binomiale verdelingen met $p \leq \frac{1}{6}$.

Toetsen Voorbeeld Aanpak 2

Gelukkig is het zo dat de binomiale verdeling met $p = \frac{1}{6}$ de *worst case* situatie is. Als het daarvoor waar is geldt het automatisch voor alle binomiale verdelingen met $p \leq \frac{1}{6}$. Als je p kleiner maakt schuift de top van de binomiale verdeling voor H_0 (die ligt bij np) naar links, waardoor $P(\underline{k} > g)$ alleen maar kleiner wordt, dus beter voldoet aan $P(\underline{k} > g) \leq 0{,}05$.

We lossen g op uit de fout van de eerste soort: $P(\underline{k} > g |\, H_0) \leq 0{,}05$ ofwel

$$\text{binomcdf}\left(n = 30\,, p = \frac{1}{6}, k = g?\right) = 0{,}95$$

Dit levert $g = 8{,}9999$, afgerond $g = 9$. We controleren dit:

$$\text{binomcdf}\left(n = 30\,, p = \frac{1}{6}, k = 9\right) = 0{,}9803 \geq 0{,}95$$

Dit klopt. Is een kleinere waarde mogelijk: $g = 8$?

$$\text{binomcdf}\left(n = 30\,, p = \frac{1}{6}, k = 8\right) = 0{,}9494 \geq 0{,}95$$

Dit is een twijfelgeval, hij klopt net niet, maar als je afrondt op twee decimalen klopt het wel: $g = 8$ of $g = 9$. In dit geval maakt het niet uit, want 10 zessen uit de 30 liggen steeds in het kritieke gebied.

Toetsen Voorbeeld Aanpak 2

Een berekening m.b.v. de p-waarde, i.p.v. met het kritieke gebied levert als overschrijdingskans

$$p = P(\underline{k} \geq 10 |\, H_0) = 1 - \text{binomcdf}\left(n = 30\,, p = \frac{1}{6}, k = 9\right) = 0{,}01971$$

Deze waarde is kleiner de $\alpha = 0{,}05$, dus H_0 wordt verworpen.

Conclusie: De kans op een zes is bij deze dobbelsteen groter dan 1/6 met een betrouwbaarheid van $100(1 - p) \geq 98\%$.

Toetsen Voorbeeld Aanpak 3

De conclusie is nu wel wat beter geworden, we weten nu in ieder geval dat de kans op zes gooien groter is dan 1/6, maar dat zou ook wel 0,1669 kunnen zijn, dat zegt nog niet veel.
Misschien kunnen we eens proberen:

$$\mathrm{H}_0\colon p \leq \frac{1}{4}$$

$$\mathrm{H}_1\colon p > \frac{1}{4}$$

Kunnen we dat aantonen? We doen dat nu met de p-waarde:

$$p = P(\underline{k} \geq 10 | H_0) = 1 - \text{binomcdf}\left(n = 30\,, p = \frac{1}{4}, k = 9\right) = 0{,}1966$$

Dit blijkt te hoog gegrepen. Deze waarde is groter dan de $\alpha = 0{,}05$, dus H_0 kan niet nu worden verworpen, we kunnen niet met 95% betrouwbaarheid constateren dat de kans op een zes is bij deze dobbelsteen groter is dan 1/4.

De beste waarde voor p die we wel kunnen aantonen met 95% betrouwbaarheid kun je oplossen uit

$$1 - \text{binomcdf}(n = 30\,, p?\,, k = 9) = 0{,}05$$

Dit levert $p = 0{,}1933$, dus $p > 0{,}19$ lukt wel.

Toetsen Voorbeeld 9.7

Voorbeeld 9.7 Een farmaceutisch bedrijf beweert dat ze een geneesmiddel heeft dat in minstens 99% van de gevallen doeltreffend is. Een universiteit is gevraagd onafhankelijk onderzoek naar deze claim te doen. Men heeft 200 willekeurige gebruikers van het middel bevraagd en 8 gaven aan dat het bij hen niet werkte. Men wil aan de hand van deze resultaten met een betrouwbaarheid van 99% toetsen of de fabrikant gelijk heeft.

Stel dat p de kans op mislukken is. De volgende toets is opgesteld:

H_0: $p \leq 0{,}01$(claim is correct)

H_1: $p > 0{,}01$ (claim is niet correct).

De toetsingsvariabele is $\underline{k}$, het aantal mislukkingen in de steekproef van 200, $\underline{k}$ is binomiale verdeeld met $n =$ 200 en alle $p \leq 0{,}01$. Het is weer voldoende om de *worst case* situatie $p = 0{,}01$ te bekijken.

Er moet gelden $P(\underline{k} > g|p = 0{,}01) \leq 0{,}01$, ofwel $P(\underline{k} \leq g|p = 0{,}01) = \mathrm{binomcdf}(200\,,0.01\,,g?) \geq 0{,}99$

Oplossen levert: $g = 6$. Controleer:

$$\mathrm{binomcdf}(200\,,0.01\,,g = 6) = 0{,}995704$$

OK, maar kan g kleiner?

$$\mathrm{binomcdf}(200\,,0.01\,,g = 5) = 0{,}983977$$

Nee, deze waarde is kleiner dan 0,99.

De kleinste g die voldoet is dus $g = 6$. Het aantal mislukkingen is 8, dat is groter dan de kritieke waarde $g = 6$, dus in het kritieke gebied, dus H_0 verwerpen, dus de claim is niet correct (met 99% betrouwbaarheid).

Toetsen Terminologie

In het eerste voorbeeld hebben we een **tweezijdige** toets gezien, zowel een afwijking naar links als naar rechts wordt in de berekeningen meegenomen

Het groene gebied tussen de twee grenzen (met de grenzen erbij) heet het **acceptatiegebied**, daarin wordt H_0 geaccepteerd. Dat is ook wel het voorspellingsgebied van de toetsingsvariabele behorend bij de gekozen betrouwbaarheid. Het rode gebied buiten de grenzen heet het **kritieke gebied**, hier wordt H_0 verworpen.
Het tweede voorbeeld was een **enkelzijdige toets**, alleen afwijkingen aan een kant zijn relevant.

We hebben ook gezien dat er sprake kan zijn van **een samengestelde nulhypothese**.

$$\mathrm{H}_0: p \leq \frac{1}{6}$$

Daarbij gaat het niet om één vaste parameterwaarde, maar om een combinatie van meer waarden. Dat maakt het lastiger. Gelukkig kun je je in de praktijk vaak beperken tot één ***worst-case* waarde**.

Toetsen

Keuze van α. Bij het opstellen van een toets moet je een betrouwbaarheid kiezen. Die moet niet te klein zijn, anders is de kans dat je ten onrechte H_0 verwerpt niet klein (fout van de eerste soort), dus dan denk je dat je iets hebt aangetoond, maar er is een aardige kans dat dat niet zo is.
Je moet hem ook niet te groot kiezen, want dan wordt het acceptatiegebied groter dus H_0 moeilijker te verwerpen. Het kan dan voorkomen dat H_1 niet wordt aangenomen, terwijl hij wel waar is. Dat heet een **fout van de tweede soort**.

Enkelzijdig of tweezijdig toetsen? Dit hangt af van de conclusie die je wilt kunnen trekken. Als uit de formulering blijkt dat het gaat om het wel of niet overschrijden van een boven- of ondergrens is de toets enkelzijdig. Als het gaat om een significante afwijking van een gegeven waarde aan te tonen met een steekproef, dan is de toets tweezijdig.

Toetsen Stappenplan

1. Formuleer een nulhypothese H_0 en een alternatieve hypothese H_1. Uitgangspunt is dat je H_0 wilt verwerpen, dus stel eerst H_1 op voor de conclusie die je zou willen trekken.

2. Kies de waarde van de betrouwbaarheid $1 - \alpha$, ofwel de toegestane fout van de eerste soort α.

3. Bepaal de steekproefgrootte (in opgaven vaak gegeven) en de toetsingsgrootheid.

4. Bepaal het kritieke gebied. Hiervoor heb je de grens/grenzen nodig zoals gebruikt in H_0, de kansverdeling van de toetsingsgrootte en de steekproefgrootte.

5. Bepaal de toetsingsgrootte met een steekproef (of bereken de overschrijdingskans, afhankelijk van de opgave, of je eigen voorkeur als de methode niet is gespecificeerd) en of

- de waarde in het kritieke gebied ligt (of overschrijdingskans kleiner dan α): Dan H_0 verwerpen
- de waarde niet in het kritieke gebied ligt (of overschrijdingskans groter dan α): Dan H_0 accepteren

6. Formuleer de conclusie in heldere bewoording (niet alleen dat je H_0 hebt geaccepteerd of verworpen, maar wat het betekent).