5.2 Toetsen van hypothesen

Inleiding

De inhoud van een fles cola is ongeveer 1,5 liter. Omdat de fabrikant volgens Europese richtlijnen niet te veel klanten mag teleurstellen moet hij zijn flessen vullen met een volume dat normaal is verdeeld met gemiddeld 1530 mL en een standaardafwijking van 18 mL. Nu bevat minder dan 5% van zijn flessen te weinig cola. Hoe kan deze fabrikant nagaan of zijn flessen aan deze norm voldoen?

Figuur 1

Je leert in dit onderwerp

- de begrippen hypothese toetsen, nulhypothese, alternatieve hypothese, kritieke gebied, significantieniveau;
- linkszijdige, rechtszijdige, tweezijdige toetsen uitvoeren.

Voorkennis

- de normale verdeling met zijn vuistregels over het gemiddelde en de standaardafwijking;
- kansen en grenswaardes berekenen met de normale verdeling;
- de wortel-n-wet gebruiken.

Verkennen

Opgave V1

De inhoud van een fles cola is ongeveer 1,5 liter. Omdat de fabrikant volgens Europese richtlijnen niet te veel klanten mag teleurstellen moet hij zijn flessen vullen met een volume dat normaal is verdeeld met gemiddeld 1530 mL en een standaardafwijking van 18 mL. Nu bevat minder dan 5% van zijn flessen te weinig cola.

a Reken na dat inderdaad 5% van zijn flessen te weinig cola bevat.

b Hoe kan deze fabrikant nagaan of zijn flessen aan deze norm voldoen? Met andere woorden: of ze inderdaad een gemiddelde volume van 1530 mL hebben met een standaardafwijking van 18 mL.

Uitleg 1

Op een fles frisdrank staat dat de inhoud 1,5 liter is. Natuurlijk zal de inhoud nooit precies 1,5 liter zijn. De vulmachine is zo afgesteld dat het vulgewicht V normaal is verdeeld met een gemiddelde van $\mu = 1530$ mL is en een standaardafwijking van $\sigma = 18$ mL. Minder dan 5% van de flessen bevat nu te weinig frisdrank.

Figuur 2

De fabrikant controleert regelmatig de afstelling van zijn vulmachine door in een steekproef van 25 flessen de gemiddelde inhoud te meten. De fabrikant voert een hypothesetoets uit.

De nulhypothese H_0 is: $\mu = 1530$ mL.
De alternatieve hypothese H_1 is: $\mu \neq 1530$ mL.

Ongelijk aan 1530 mL betekent dat het gemiddelde zowel groter als kleiner dan 1530 kan zijn. Dan voer je een tweezijdige toets uit. Je kunt ook een eenzijdige toets uitvoeren. In het geval van H_1: $\mu > 1530$ mL, spreek je van een rechtszijdige toets, in het geval van H_1: $\mu < 1530$ mL, van een linkszijdige toets.

Vooraf heeft de fabrikant een beslissingsvoorschrift opgesteld, bijvoorbeeld: als het gemiddelde volume van een fles uit de steekproef kleiner is dan 1525 mL of groter dan 1535 mL, klopt de nulhypothese niet. Vulgemiddeldes die minder zijn dan 1525 mL of meer dan 1535 mL vormen het kritieke gebied. Valt het gemiddelde volume in de steekproef in het kritieke gebied, dan gaat de fabrikant er niet meer van uit dat de vulmachine de flessen vult met een gemiddelde van 1530 mL. Hij verwerpt daarmee de nulhypothese en accepteert de alternatieve hypothese. Het gevolg is dat hij de machine bij moet stellen.

Er is een kans dat de fabrikant de nulhypothese ten onrechte verwerpt. Die kans heet het significantieniveau α. Toevallig kan het gemiddelde volume in de steekproef een keer bijzonder klein of bijzonder groot zijn. De fabrikant zal van tevoren een afweging gemaakt hebben over de waardes die in het kritieke gebied komen: zijn het er te veel, dan is de kans dat hij ten onrechte de vulmachine anders afstelt groot. Zijn het er te weinig, dan is de kans dat hij zijn vulmachine ten onrechte niet aanpast weer groot.

Opgave 1

Gebruik de gegevens uit **Uitleg 1**.

a De hypothesetoets in de uitleg heet een tweezijdige hypothesetoets.
Leg uit waarom.

b Waarom is de steekproevenverdeling normaal verdeeld?

c Bereken in vier decimalen de kans dat het gemiddelde volume van de steekproef van 25 flessen frisdrank in het kritieke gebied valt dat de fabrikant bepaald heeft.

Opgave 2

Gebruik de gegevens uit **Uitleg 1**.

a Schets de normaalkromme van de steekproevenverdeling en geef daarin het kritieke gebied aan.

b Beargumenteer met statistische argumenten of je deze grenswaardes van het kritieke gebied zou aanbevelen aan de fabrikant of juist niet.

Uitleg 2

Een fabrikant beweert dat zijn vulmachine flessen frisdrank vult met een gemiddelde van 1530 mL en een standaardafwijking van 18 mL. Het vulgewicht is normaal verdeeld. Hij krijgt echter steeds meer klachten van klanten die vinden dat het gemiddelde lager ligt.

De fabrikant doet daarom een steekproef van 25 flessen met het volgende beslissingsvoorschrift: als de kans op het gemiddelde volume van een fles frisdrank in zijn steekproef kleiner is dan 5% stelt hij zijn vulmachine opnieuw in. Dit betekent:

H_0: $\mu = 1530$
H_1: $\mu < 1530$

Dit is een linkszijdige hypothesetoets met een significantieniveau α van 5%.

Het gemiddelde volume in de steekproef van de fabrikant blijkt 1519 mL te zijn. Je kunt nu op twee manieren deze linkszijdige hypothesetoets verder uitvoeren.

Manier 1

- Bereken de grenswaarde g van het kritieke gebied met behulp van $\alpha = 0{,}05$:
  $\text{P}\left(\overline{V} < g | \mu = 1530 \text{ en } \sigma = \frac{18}{\sqrt{25}}\right) = 0{,}05$
  geeft een grenswaarde van ongeveer 1524 mL.
- Omdat 1519 < 1524 ligt het steekproefgemiddelde in het kritieke gebied.
- Conclusie: de nulhypothese wordt verworpen, de vulmachine moet bijgesteld worden.

Manier 2

- Bereken de kans op hoogstens het steekproefgemiddelde. Deze kans is
  $\text{P}\left(\overline{V} < 1519 | \mu = 1530 \text{ en } \sigma = \frac{18}{\sqrt{25}}\right) \approx 0{,}0011$.
- Omdat 0,0011 < 0,05 ligt het steekproefgemiddelde in het kritieke gebied.
- Conclusie: de nulhypothese wordt verworpen, de vulmachine moet bijgesteld worden.

De fabrikant trekt deze conclusie met een betrouwbaarheid van 95%. Door een significantieniveau van 5% te gebruiken is de kans dat zijn nulhypothese wel degelijk juist is en dat hij zijn vulmachine dus voor niets bijstelt gelijk aan 5%.

Opgave 3

Gebruik de gegevens uit **Uitleg 2**.

a Bereken zelf de grenswaarde van het kritieke gebied voor de hypothesetoets.

b Bereken zelf de kans op hoogstens het steekproefgemiddelde van de hypothesetoets.

Opgave 4

Gebruik de gegevens uit **Uitleg 2**.
De klanten voeren doen een eigen onderzoek.
Zij gebruiken een significantieniveau van 1% en trekken ook een steekproef van 25 flessen.

a Geef de nulhypothese en de alternatieve hypothese van deze hypothesetoets.

Het gemiddelde volume van de flessen in de steekproef van de klanten is 1521 mL.

b Voer de hypothesetoets uit van de klanten op basis van de grenswaarde van het kritieke gebied.

c Welke conclusie trekken de klanten? Hoe groot de betrouwbaarheid van deze conclusie?

Door het significantieniveau van 1% hebben de klanten in principe een kans van 1% om H_0 ten onrechte te verwerpen.

d Hoe groot is in vier decimalen de kans dat de vulmachine van de fabrikant toch goed afgesteld staat bij een gemiddeld steekproefvolume van maximaal 1521 mL?

Theorie en voorbeelden

Om te onthouden

Statistische methodes kunnen worden gebruikt om een bewering over een populatie te controleren. Dit heet **een hypothese toetsen**.

De **nulhypothese** H_0 is de gangbare bewering (bijvoorbeeld op grond van voorgaand onderzoek). De **alternatieve hypothese** H_1 is een bewering die de nulhypothese bestrijdt.

Stel, toevalsvariabele X is normaal verdeeld. Er wordt beweerd dat het gemiddelde $\mu(X) = \mu$ is, waarin μ een bepaalde waarde is.
Iemand anders vertrouwt het gemiddelde niet en vermoedt (bijvoorbeeld): $\mu(X) < \mu$.

Je hebt dan:
$H_0 : \mu(X) = \mu$
$H_1 : \mu(X) < \mu$

Dit wordt getoetst met een steekproef van grootte n. Je bepaalt dan het gemiddelde in de steekproef en kijkt of de afwijking van μ significant is. Het steekproefgemiddelde $\mu(\overline{X})$ is normaal verdeeld met $\mu(\overline{X}) = \mu$ en $\sigma(\overline{X}) = \frac{\sigma(X)}{\sqrt{n}}$.

Opmerking: Dit is vanwege de **centrale limietstelling** ook zo als X niet normaal verdeeld is, mits de steekproef voldoende groot is.

Bij de alternatieve hypothese hoort een **kritiek gebied** dat aangeeft waar de afwijking van $\mu(\overline{X})$ zo groot is dat je de nulhypothese verwerpt. Dat kritieke gebied bepaal je op grond van een vooraf vastgesteld **significantieniveau α** . Het significantieniveau kies je voordat je de toets uitvoert, bijvoorbeeld $\alpha = 10\%$ of $\alpha = 5\%$.

Als de kans op een steekproefgemiddelde van $\mu(\overline{X})$ kleiner is dan α, verwerp je H_0 en accepteer je H_1.

Afhankelijk van de situatie, zijn er drie mogelijkheden voor de alternatieve hypothese:

- een **rechtszijdige toets** waarbij H_0 getoetst wordt tegen $H_1 : \mu(\overline{X}) > \mu(X)$.
- een **linkszijdige toets** toetst H_0 tegen $H_1 : \mu(\overline{X}) < \mu(X)$.
- een **tweezijdige toets** toetst H_0 met $H_1 : \mu(\overline{X}) \neq \mu(X)$.

Voorbeeld 1

Een groothandel verkoopt pakken hagelslag met een gemiddeld gewicht van 255 gram en een standaardafwijking van 4 gram. Het gewicht van een pak hagelslag is normaal verdeeld.

De fabrikant van de pakken hagelslag vermoedt dat zijn pakken tegenwoordig te veel hagelslag bevatten en dat is nadelig voor hem.
Hij besluit een hypothesetoets te doen met een significantieniveau van 5%.
De fabrikant neemt een steekproef van 15 pakken hagelslag.
Voer de hypothesetoets uit en geef het kritieke gebied.

Antwoord

Deze hypothesetoets heeft betrekking op de normaal verdeelde toevalsvariabele G, het gewicht van een pak hagelslag.

De fabrikant stelt als nulhypothese en alternatieve hypothese:
H_0: $\mu(G) = 255$ gram
H_1: $\mu(G) > 255$ gram

Het is dus een rechtszijdige hypothesetoets: het kritieke gebied ligt rechts van de grenswaarde.

Het gemiddelde gewicht $\overline{G}$ van de steekproevenverdeling is normaal verdeeld (omdat G normaal verdeeld is). De grenswaarde g van het kritieke gebied is te berekenen met:

$$\text{P}\left(\overline{G} > g | \mu = 255 \text{ en } \sigma = \frac{4}{\sqrt{15}}\right) = 0{,}05$$

Je vindt met de grafische rekenmachine $g \approx 256{,}7$ gram.

Het kritieke gebied bestaat uit alle gewichten die groter zijn dan 256,7 gram.

Opgave 5

Gebruik de gegevens uit **Voorbeeld 1**.

a Schets het kritieke gebied bij een steekproefomvang van 15 pakken hagelslag in de normaalkromme van de steekproevenverdeling.

b Stel dat de fabrikant een steekproef van 10 pakken zou doen. Welk kritiek gebied krijg je dan?

Opgave 6

Gebruik de gegevens uit **Voorbeeld 1**. De fabrikant heeft het significantieniveau verlaagd naar 2,5% en heeft een steekproef van 15 pakken hagelslag getrokken. Het gemiddelde gewicht van een pak hagelslag in deze steekproef blijkt 256,4 gram te zijn.

a Ligt het steekproefgemiddelde in het kritieke gebied?

b Wat zal de fabrikant doen op basis van zijn beslissingsvoorschrift bij deze steekproefuitslag?

Voorbeeld 2

Een groothandel verkoopt pakken hagelslag met een gemiddeld gewicht van 255 gram en een standaardafwijking van 4 gram. Het gewicht van de pakken is normaal verdeeld.
De fabrikant van de pakken hagelslag wil niet te veel, maar ook niet te weinig hagelslag in de pakken voor de groothandel stoppen.
Regelmatig trekt hij daarom een steekproef van 20 pakken hagelslag en voert daarmee een hypothesetoets uit met een significantieniveau van 5%.

Wat is de conclusie die de fabrikant uit deze hypothesetoets zal trekken als het steekproefgemiddelde 253,75 gram is?

Antwoord

De hypothesetoets die de fabrikant telkens uitvoert, is een tweezijdige toets van toevalsvariabele G, het gewicht van een pak hagelslag.

Er geldt:
H_0: $\mu(G) = 255$ gram
H_1: $\mu(G) \neq 255$ gram

Het steekproefgemiddelde $\overline{G}$ is kleiner dan het gemiddelde van de nulhypothese: als het al in het kritieke gebied ligt, dan ligt het in het linkerdeel van dit tweezijdige kritieke gebied.

$$\mathrm{P}\left(\overline{G} < 253{,}75 | \mu = 255 \text{ en } \sigma = \frac{4}{\sqrt{20}}\right) \approx 0{,}0811$$

Het significantieniveau is 5%, maar omdat het een tweezijdige toets is, vergelijk je dit met 2,5% = 0,025.
Omdat de kans op een steekproefgemiddelde van hoogstens 253,75 gram groter is dan 2,5% ligt het steekproefgemiddelde niet in het kritieke gebied.
De fabrikant verwerpt de nulhypothese daarom niet.

Opgave 7

Gebruik de gegevens uit **Voorbeeld 2**.

a Bereken de grenswaarde van het linkerdeel van het kritieke gebied bij de hypothesetoets.

b Wat is de betekenis van dit kritieke gebied voor de hypothesetoets?

c Bereken ook de grenswaarde van het rechterdeel van het kritieke gebied.

d Klopt de conclusie van de fabrikant op basis van een steekproefgemiddelde van 253,75 gram ook als je niet naar de kans op dit steekproefgemiddelde kijkt, maar checkt of dit steekproefgemiddelde wel of niet in het kritieke gebied ligt?

Opgave 8

Gebruik de gegevens uit **Voorbeeld 2**. Stel, de fabrikant vindt de kosten van te volle hagelslagpakken geen probleem meer en daarom wil hij zijn vulproces alleen nog maar toetsen op te weinig inhoud. Hij blijft bij het toetsen gebruikmaken van een steekproef van 20 pakken en behoudt het significantieniveau van 5%.

a Geef de nulhypothese en de alternatieve hypothese van de nieuwe hypothesetoets van de fabrikant en bepaal het bijbehorende kritieke gebied.

b Vergelijk het oorspronkelijke kritieke gebied en het nieuwe kritieke gebied met elkaar.

Opgave 9

Bij het uitvoeren van statistische hypothesetoetsen kan de conclusie fout zijn, zelfs als het onderzoek helemaal goed wordt uitgevoerd. Welke twee foute conclusies zijn er te trekken?

Verwerken

Opgave 10

Een firma die batterijen levert voor rekenmachines, beweert dat die batterijen geschikt zijn om zo'n apparaat gemiddeld 3600 uur te laten werken. Ze gaan ervan uit dat die levensduur normaal is verdeeld met een standaarddeviatie van 600 uur.

De leverancier van die rekenmachines is bang dat de levensduur van de batterijen gemiddeld korter is en daarom gaan ze dit toetsen. Ze kiezen aselect 75 rekenmachines en stoppen in elk apparaat een aselect gekozen batterij van deze firma.

Zij nemen voor de grenswaarde van het kritieke gebied van hun hypothesetoets een gemiddelde levensduur van 3350 uur.

a Stel de hypothesetoets op.

b Het steekproefgemiddelde is 3400 uur. Welke conclusie over de gemiddelde levensduur van de batterijen zal de leverancier van de rekenmachines trekken?

Opgave 11

Volgens een wetenschappelijk tijdschrift is het gewicht van 17-jarige meisjes normaal verdeeld met een gemiddelde van 54,2 kg en een standaarddeviatie van 4,7 kg. Om deze bewering te toetsen wordt van 200 aselect gekozen 17-jarige meisjes het gewicht bepaald. Als significantieniveau is 5% gekozen.

a Zal dit een eenzijdige of tweezijdige toets worden?

b Beschrijf de hypothesetoets.

c Welke grenswaarde(s) heeft het kritieke gebied?

d Het steekproefgemiddelde is 54,7 kg. Welke conclusie trek je?

Opgave 12

Volgens de informatie op een pakje drinkyoghurt zou dit gemiddeld 12,5 gram suiker bevatten. Een onderzoeksbureau beweert dat er in werkelijkheid veel meer suiker in de pakjes zit.

De leverancier van de pakjes besluit een steekproef van 50 pakjes te nemen. De pakjes uit de steekproef bevatten gemiddeld 16,4 gram suiker.

Neem aan dat de hoeveelheid per pakje normaal verdeeld is met een standaardafwijking van 3,1 gram.

Onderzoek of dit resultaat voldoende aanleiding is om de informatie die op de pakjes staat te verwerpen. Neem een significantieniveau van 1%.

Opgave 13

Vacuüm verpakte vleeswaren mogen maximaal 0,022% natriumnitriet bevatten. De keuringsdienst van waren toetst dit percentage, omdat men denkt dat het gemiddelde percentage natriumnitriet boven 0,022% ligt. Je mag aannemen dat het natriumnitrietpercentage normaal verdeeld is.

a Formuleer de nulhypothese.

b Is de toets eenzijdig of tweezijdig? Formuleer ook de alternatieve hypothese.

Bekijk 25 meetresultaten:

0,0219	0,0226	0,0225	0,0225	0,0216
0,0219	0,0220	0,0216	0,0229	0,0226
0,0214	0,0219	0,0226	0,0220	0,0212
0,0225	0,0223	0,0215	0,0221	0,0223
0,0224	0,0215	0,0228	0,0223	0,0223

Tabel 1

c Toets met behulp van deze steekproef of de keuringsdienst gelijk heeft. Neem een significantieniveau van 5%. Gebruik hierbij de standaardafwijking van deze meetresultaten.

Opgave 14

In een medisch laboratorium worden voortdurend cholesterolgehaltes in bloedmonsters bepaald. Het cholesterolgehalte is normaal verdeeld. De gebruikte apparatuur wordt elk uur gecontroleerd met behulp van een ijkmonster. Hiervan is bekend dat het gemiddelde 175 mg per 100 mL zou moeten zijn. De controlemetingen aan het ijkmonster leveren op: 168, 170, 188, 170, 174, 190, 188 en 171. Is er met een significantie van $\alpha = 0{,}01$ reden om aan te nemen dat de meetapparatuur niet goed meer werkt?

Gebruik de standaardafwijking van de controlemetingen als schatting voor de standaardafwijking van de populatie.

Toepassen

Opgave 15: Conditietest

Een Canadese gymnastiekdocent traint regelmatig jongens van 14 jaar om hun conditie te verbeteren. De gemiddelde score van deze leeftijdscategorie is 8,0 en de standaardafwijking is 2,0. De docent is van mening dat deze training daadwerkelijk helpt. Om dat na te gaan laat hij na een aantal trainingen 132 jongens van 14 jaar de conditietest doen. Het resultaat is dat deze jongens een gemiddelde score van 8,43 hebben gehaald.

Onderzoek of deze gymnastiekdocent op grond van dit resultaat gelijk krijgt.
Neem als significantieniveau 5%.

(bron: examen vwo wiskunde A in 2009, tweede tijdvak)

Opgave 16: Fout van de tweede soort

Als het significantieniveau van een hypothesetoets gelijk is aan 5% wil dat zeggen dat je bij deze toets minder dan 5% kans hebt om de nulhypothese onterecht te verwerpen.

Als je de nulhypothese ten onrechte verwerpt, ga je ervan uit dat de toevalsvariabele van de populatie die je onderzoekt een ander gemiddelde heeft dan in de nulhypothese wordt gesteld: H_0 is waar maar je gaat vanaf nu uit van H_1. Dit wordt ook wel de fout van de eerste soort genoemd.

Het is ook mogelijk om de nulhypothese ten onrechte niet te verwerpen: H_1 is eigenlijk waar maar je blijft uitgaan van H_0. Dit wordt de fout van de tweede soort genoemd.

a Waarom zou normaal gesproken een fout van de eerste soort ernstiger worden gevonden dan een fout van de tweede soort?

b Waarom kun je een fout van de eerste soort nooit helemaal uitsluiten?

Testen

Opgave 17

Volgens de fabrikant is het gewicht G (in gram) van zijn pakken suiker normaal verdeeld met $\mu = 1002$ en $\sigma = 3$.

Omdat de Consumentenbond veel klachten heeft binnengekregen waarin wordt gemeld dat de pakken suiker van deze fabrikant te weinig suiker bevatten, wordt er door hen getwijfeld aan dit gemiddelde. De Consumentenbond stelt dat het gemiddelde gewicht van de pakken suiker van deze fabrikant lager ligt dan 1002 gram.

Ze onderzoeken de bewering van de fabrikant met behulp van een hypothesetoets.

a Formuleer (in woorden) de nulhypothese die hier getoetst moet worden.

De Consumentenbond stelt het beslissingsvoorschrift op voor de hypothesetoets.

De kans dat het gemiddelde gewicht van de pakken suiker in een steekproef van 10 stuks lager is dan 1000 gram, moet kleiner zijn dan 1%.

b Bepaal het kritieke gebied van deze hypothesetoets.

c Is de eis van de Consumentenbond mogelijk als er een steekproef van 10 pakken suiker wordt getrokken? En als er een steekproef van 20 pakken suiker wordt getrokken?

d Leg uit wat de invloed van de steekproefomvang is op het beslissingsvoorschrift van een hypothesetoets en leg uit hoe dat komt. Gebruik daarbij het resultaat uit c.

Opgave 18

Op een pak melk staat: “Het natuurlijk vetgehalte van melk – zoals die van de koe komt – varieert van 3,7% tot 4,3%”.

Volle melk wordt in de fabriek altijd afgeroomd tot 3,5%.

Een consumentenorganisatie besluit na te gaan of volle melk 3,5% vet bevat. In een aselecte steekproef van 20 pakken volle melk vindt ze de percentages uit de tabel.

percentage	frequentie
3,445– < 3,455	1
3,455– < 3,465	0
3,465– < 3,475	4
3,475– < 3,485	3
3,485– < 3,495	3
3,495– < 3,505	4
3,505– < 3,515	2
3,515– < 3,525	2
3,525– < 3,535	1

Tabel 2

a Laat met behulp van normaal waarschijnlijkheidspapier zien dat de resultaten van deze proef reden geven om aan te nemen dat het vetgehalte van pakken melk normaal verdeeld is.

b Schat met behulp van de steekproef de standaardafwijking van het vetgehalte van pakken melk.

c Toets met significantieniveau 0,05 of de consumentenorganisatie op grond van de steekproef de bewering op het pak kan verwerpen.