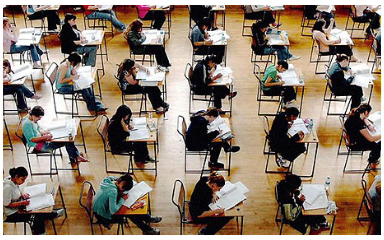
6.4 Bijzondere toetsen

Inleiding

Je hebt nu kennis gemaakt met binomiale toetsen en met het toetsen van het gemiddelde van een normale verdeling. Er bestaan nog veel andere soorten toetsen, afhankelijk van het type kansverdeling dat er achter zit. Maar ook een zogenaamde tekentoets vereist een speciale aanpak, hoewel het daarbij gewoon om een binomiale verdeling gaat. Je vergelijkt dan twee sets gegevens met elkaar, bijvoorbeeld het cijfer voor het SE (schoolexamen) en het CE (centraal examen).



Figuur 1

Je leert in dit onderwerp

- de tekentoets gebruiken;
- het verschil van twee normaal verdeelde stochasten toetsen.

Voorkennis

- binomiale en normale toetsen uitvoeren.

Verkennen

Opgave V1

De inspectie voor het onderwijs vergelijkt van een bepaalde school de cijfers voor wiskunde B van het SE (schoolexamen) en het CE (centraal examen).

a Kun je een manier te bedenken om vast te stellen of de CE-cijfers significant afwijken van de SE-cijfers?

b Waarom zou de inspectie daarin zijn geïnteresseerd?

Uitleg 1

Bij statistisch onderzoek wordt vaak onderzoek gedaan met meerdere statistische variabelen. Deze worden dan vergeleken. Hier wordt een situatie bekeken waarbij onderzoek wordt gedaan met paren waarnemingen.

In de voedselindustrie wordt gewerkt met smaakpanels, groepen mensen die bijvoorbeeld beoordelen of een gerecht smaakvoller wordt als het anders wordt klaargemaakt. De leden van het smaakpanel proeven hierbij beide gerechten en geven aan of het ze het eerste of het tweede gerecht smaakvoller vinden. Er zijn drie mogelijke uitkomsten: het tweede gerecht is smaakvoller, van gelijke smaak of minder smaakvol. Deze uitkomsten worden bijvoorbeeld $+$, 0 en $-$ genoemd.

Figuur 2

Bij onveranderde smaak is de verwachting dat er evenveel $+$ als $-$ zal zijn. Daarom kun je hier toetsen met de binomiale toets op proportie:

X is het aantal $+$ (binomiaal verdeeld)
$H_0 : p = 0{,}5$
$H_1 : p > 0{,}5$

Deze toets wordt tekentoets genoemd. Deze toets werkt dus zelfs bij sommige kwalitatieve variabelen.

Opgave 1

Bekijk **Uitleg 1**.

a Waarom heet deze wijze van toetsen een tekentoets?

b Waarom moet daarbij altijd als nulhypothese $p = 0{,}5$ worden gehanteerd?

Opgave 2

Bekijk **Uitleg 1**. Stel dat er bij de smaaktest van de 15 keer 10 keer + uit is gekomen.

a Waarom mag je nu niet concluderen dat er 5 keer − uit is gekomen?

b Veronderstel dat er toch 5 keer − uit kwam. Mag je er nu met een significantie van 10% van uitgaan dat de nieuwe bereidingswijze smaakvollere gerechten tot gevolg heeft?

Uitleg 2

Bij de normale toets werd onderzoek gedaan naar de juistheid van een bewering over het gemiddelde. Het gaat hier over één statistische variabele.

Maar bij statistisch onderzoek wordt vaak onderzoek gedaan met meerdere stochastische variabelen. Deze worden dan vergeleken. Bijvoorbeeld een onderzoek waarbij onderzocht wordt of de gemiddelden van twee variabelen gelijk zijn. Ga hierbij uit van de volgende veronderstellingen:

- De gemiddelden van beide populaties zijn bekend.
- Beide variabelen zijn normaal verdeeld.
- De standaardafwijkingen van beide verdelingen zijn bekend.

Altijd is het verschil van twee normale verdelingen weer een normale verdeling.

Stel dat het aantal sterftegevallen door griep per gemeente normaal verdeeld is.
In 2015 gaf een steekproef $\mu_{15} = 20$ en $\sigma_{15} = 2{,}3$.
In 2016 gaf een steekproef $\mu_{16} = 21$ en $\sigma_{16} = 3{,}0$.

Verschillen deze gemiddelden significant?
Als de gemiddelden niet zouden verschillen, kun je uitgaan van verdeling van het verschil V met $\mu_V = 0$ en

$\sigma_V = \sqrt{2{,}3^2 + 3{,}0^2} \approx 3{,}8$.

Het verschil tussen de gemiddelden is hier $\mu_{16} - \mu_{15} = 1$.
De overschrijdingskans van dit verschil is $\text{P}(V > 1) \approx 0{,}3957 > 0{,}05$.
Dat wil zeggen dat er geen significant verschil is tussen de aantallen sterftegevallen van 2015 en 2016.

Deze toets heet een verschiltoets voor gemiddelden.

Opgave 3

Bekijk **Uitleg 2**.

a Waaraan is zichtbaar dat hier een verschiltoets moet worden toegepast en dus geen hypothesetoets zoals eerder in dit hoofdstuk beschreven?

b Wat is het toegepaste significantieniveau van deze toets?

c Het verschil tussen de gemiddelden is hier $\mu_{16} - \mu_{15} = 1$. Wat moet je doen als je μ_{16} en μ_{15} in deze berekening verwisselt?

Theorie en voorbeelden

Om te onthouden

Soms wil je bij statistisch onderzoek twee verschillende populaties vergelijken door hypothesetoetsen toe te passen. Hier worden twee soorten toetsen genoemd.

De **tekentoets** is geschikt om twee populaties te vergelijken door middel van steekproeven met bij elkaar horende paren van waarnemingen. Deze waarnemingen moeten te vergelijken zijn met begrippen als meer/minder, groter/kleiner, kleuriger/minder kleurig, enzovoort. De eventuele waarden van de waarnemingen worden niet gebruikt.

Met tekens (bijvoorbeeld +, − en 0) wordt aangegeven hoe en of een paar waarnemingen van elkaar verschilt. Als de populaties niet verschillen, zal het aantal + en − gelijk zijn en daarmee zal de kans op een + 0,5 zijn. Je kunt dan toetsen met een binomiale verdeling met $p = 0{,}5$.

De **verschiltoets voor gemiddelden** wordt gebruikt om uitspraken over het verschil tussen de gemiddelden van twee populaties steekproeven te doen. De populatie hoeft niet normaal verdeeld zijn. Deze toets is gebaseerd op de volgende uitgangspunten:

- Steekproefgemiddelden zijn normaal verdeeld.
- Het verschil van twee normale verdelingen is ook weer normaal verdeeld.

Als de populaties niet verschillen, zullen de gemiddelden ervan ook niet verschillen. Het (normaal verdeelde) verschil van de gemiddelden heeft hier dus altijd een gemiddelde van 0. Je toetst met een normale verdeling of het verschil significant van 0 afwijkt.

Voorbeeld 1

De inspectie voor het onderwijs vergelijkt van een bepaalde school de cijfers voor wiskunde B van het SE (schoolexamen) en het CE (centraal examen). In de tabel vind je de gegevens van een klas van 19 leerlingen.

leerling	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
SE-cijfer	6,0	6,7	5,8	7,1	5,4	6,5	8,8	6,9	7,9	5,1	6,1	6,1	6,4	7,4	5,9	6,2	7,1	6,8	6,3
CE-cijfer	6,4	6,3	5,2	6,5	5,4	6,1	9,0	6,8	7,5	5,6	6,0	6,5	6,0	6,5	6,0	6,6	7,0	6,6	6,4

Tabel 1

De inspectie besluit om een tekentoets toe te passen met een significantieniveau van 5%. Een + geeft aan dat de leerling het CE beter heeft gemaakt, een − dat het CE minder is gemaakt. Mag de inspectie op grond hiervan concluderen dat het CE slechter is gemaakt?

Antwoord

leerling	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
teken	+	−	−	−	0	−	+	−	−	+	−	+	−	−	+	+	−	−	+

Tabel 2

Een + geeft aan dat de leerling het CE beter heeft gemaakt, een − dat het CE minder is gemaakt. Een 0 geeft aan dat de twee resultaten niet verschillen. Dit paar waarnemingen wordt niet meegeteld. De reden is dat er anders een (te) grote kans op de fout ‘H_1 is juist, maar wordt niet geaccepteerd’ kan optreden.

- X is het aantal minnen (CE slechter) in de steekproef ($n = 18$). X is binomiaal verdeeld.
- $\text{H}_0 : p = 0{,}5$ (altijd bij de tekentoets)
- $\text{H}_1 : p > 0{,}5$

$\text{P}(X \geq 11 | n = 18 \text{ en } p = 0{,}5) \approx 0{,}2403 > 0{,}05$

De inspectie mag op grond hiervan deze conclusie niet trekken.

Opgave 4

Gebruik de gegevens uit **Voorbeeld 1**.

a Waarom is $n = 18$ en niet $n = 19$?

b Reken na dat $\text{P}(X \geq 11 | n = 18 \text{ en } p = 0{,}5) \approx 0{,}2403$.

c Stel dat het CE-cijfer van leerling 5 een 5,2 was geweest. Had dan de inspectie met een significantieniveau van 10% mogen concluderen dat het CE slechter is gemaakt?

Voorbeeld 2

Twee groepen hebben dezelfde toets gemaakt. De scores van beide groepen zijn normaal verdeeld. Groep A heeft een gemiddelde score gehaald van 6,4 met een standaardafwijking van 0,7. Groep B heeft een gemiddelde score gehaald van 7,9 met een standaardafwijking van 0,9.

Als je een onbetrouwbaarheidsdrempel van 10% neemt, kun je dan zeggen dat de scores van groep B significant beter zijn dan die van groep A?

Antwoord

Noem V het verschil van beide scores.

V is normaal verdeeld met een gemiddelde van $7{,}9 - 6{,}4 = 1{,}5$ en een standaardafwijking van $\sqrt{0{,}7^2 + 0{,}9^2} = \sqrt{1{,}3}$.

Wanneer de scores van groep B hetzelfde zouden zijn als die van groep A, zou het verschil 0 moeten zijn. De nulhypothese is daarom H_0: $\mu_V = 0$ (altijd) en de alternatieve hypothese luidt H_1: $\mu_V > 0$.

$\text{P}\left(V > g | \mu_V = 0 \text{ en } \sigma_V = \sqrt{1{,}3}\right) < 0{,}1$ geeft $g \approx 1{,}46$.

Omdat $\mu_V = 7{,}9 - 6{,}4 = 1{,}5 > 1{,}46$ is de afwijking voldoende om te concluderen dat de cijfers van groep B significant beter zijn.

Opgave 5

Gebruik de gegevens uit **Voorbeeld 2**.

a Reken na dat $g \approx 1{,}46$.

b Als je een onbetrouwbaarheidsdrempel van 5% hanteert, kun je dan ook concluderen dat de cijfers van groep B significant beter zijn dan die van groep B?

c Stel dat er een fout is gemaakt en dat groep B gemiddeld een 7,8 heeft gehaald. De standaardafwijking blijft hetzelfde. Welke conclusie trek je nu bij een onbetrouwbaarheidsdrempel van 10%?

Opgave 6

Twee machines vullen pakken met suiker. De machines horen de pakken met een gelijke hoeveelheid suiker te vullen. Om te controleren of dit zo is, worden van beide machines steekproeven genomen.

Bij machine A is het gemiddelde gewicht van de pakken suiker in de steekproef 1002 gram met een standaardafwijking van 2,5 gram. Bij machine B is het gemiddelde gewicht van de pakken suiker in de steekproef 1008 gram met een standaardafwijking van 3,6 gram.

Verschillen de gemiddelde vulgewichten van machine A en machine B significant van elkaar? Neem een significantie van 10%.

Verwerken

Opgave 7

De ondernemingsraad van een bedrijf beweert dat het ziekteverzuim op afdeling A significant hoger is dan op afdeling B. De raad legt de directie het volgende overzicht voor over het percentage ziekteverzuim:

maand	jan	feb	mrt	apr	mei	jun	jul	aug	sep	okt	nov	dec
afd.A	9	9	8	10	12	13	12	12	10	11	8	12
afd.B	7	10	9	8	11	11	7	9	9	10	10	7

Tabel 3

De directie besluit hierop een tekentoets toe te passen met een significantieniveau van 5%.

a Beschrijf de tekentoets, geef de nulhypothese, de alternatieve hypothese, de steekproefgrootte en de onbetrouwbaarheidsdrempel.

b Onderzoek of de ondernemingsraad gelijk krijgt.

Opgave 8

De diameters van machinaal geproduceerde bouten en de bijbehorende moeren zijn normaal verdeeld: de diameter van de moer is normaal verdeeld met een gemiddelde van 8,10 mm en een standaarddeviatie van 0,05 mm. De diameter van de bout is normaal verdeeld met een gemiddelde van 8,05 mm en een standaardafwijking van 0,03 mm. De bouten passen in de moeren als het verschil in diameter van de moer en de bout minder dan 0,02 mm is. Er wordt regelmatig gecontroleerd of de machines die deze bouten en moeren maken niet moeten worden bijgesteld, omdat te veel moeren niet op de bouten passen. Wekelijks wordt een steekproef van 100 bouten en moeren getest.

a Waarom is hier sprake van een tweezijdige toets?

b Stel de nulhypothese en de alternatieve hypothese op.

c Welke standaardafwijking moet er worden gehanteerd? Waarom speelt nu ook de wortel-n-wet ($\sqrt{n}$-wet) een rol?

d Voer de toets uit met een significantieniveau van 5%. Bij welk gemiddelde verschil in de steekproef worden de machines bijgesteld?

Opgave 9

In een laboratorium worden twee geneesmiddelen voor dezelfde ziekte getest op muizen die men kunstmatig aan deze ziekte laat lijden. Ze worden met één van beide middelen behandeld.
Elke dag wordt bijgehouden hoeveel dieren er genezen zijn. De helft van de muizen kreeg geneesmiddel A toegediend, de andere helft geneesmiddel B.
De resultaten staan in deze tabel.

dagnummer	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
middel A	2	5	4	6	3	3	6	3	2	2	2	4	6	9	4	2	3	2	4	9
middel B	3	8	6	9	2	4	8	5	5	2	5	5	3	11	8	4	5	0	5	1

Tabel 4

Onderzoekers in dit laboratorium toetsen de mening dat beide middelen even goed werken met een onbetrouwbaarheidsdrempel van 5%. Er wordt een tekentoets uitgevoerd.

a Stel een nulhypothese en een alternatieve hypothese op.

b Stel vast of beide middelen op grond van de resultaten in deze test inderdaad even goed werken binnen de gegeven betrouwbaarheidseis.

Opgave 10

Bekijk de resultaten van 19 leerlingen voor het SE en het CE.

leerling	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
SE-cijfer	6,0	6,7	5,8	7,1	5,4	6,5	8,8	6,9	7,9	5,1	6,1	6,1	6,4	7,4	5,9	6,2	7,1	6,8	6,3
CE-cijfer	6,4	6,3	5,2	6,5	5,4	6,1	9,0	6,8	7,5	5,6	6,0	6,5	6,0	6,5	6,0	6,6	7,0	6,6	6,4

Tabel 5

Eerder is hier een tekentoets mee uitgevoerd, waaruit bleek dat de resultaten niet significant verschilden. Omdat hier waarden bekend zijn, kan ook gebruik worden gemaakt van een verschiltoets. Ga na of uit deze toets met een significantieniveau van 5% volgt dat het SE beter is gemaakt dan het CE.

Toepassen

Opgave 11: Voetlengtes

Open het bestand **Voetlengtes van 100 mannen en 100 vrouwen**. De gangbare opvatting is dat mannen gemiddeld grotere voeten hebben dan vrouwen. Je wilt deze opvatting toetsen met een significantieniveau van 5% met behulp van de gegevens in dit bestand.

a Kun je met deze meetgegevens een tekentoets uitvoeren?

b Je toetst het verschil in voetlengtes van mannen en vrouwen. Maakt het verschil of je twee willekeurige groepen mannen en vrouwen onderzoekt of een groep van 100 echtparen?

c Voer de toets uit. Wordt de hierboven gedane uitspraak bevestigd?

Testen

Opgave 12

Je hebt in dit onderdeel met twee soorten toetsen kennis gemaakt. Beschrijf hoe deze toetsen in elkaar zitten en onder welke omstandigheden je ze toepast.