

5.5 Totaalbeeld

Samenvatten

Je hebt nu het onderwerp **Populatie en steekproeven** doorgewerkt. Er moet een totaalbeeld van deze leerstof ontstaan...

Ga na, of je al de bij dit onderwerp horende begrippen kent en weet wat je er mee kunt doen. Ga ook na of je de activiteiten die staan genoemd kunt uitvoeren. Maak een eigen samenvatting!

Begrippenlijst

- verklarende statistiek, statistische cyclus — populatie, steekproef — aselekt, representatief — steekproevenverdeling — centrale limietstelling
- hypothese toetsen — nulhypothese, alternatieve hypothese — kritiek gebied — significantieniveau
- populatiegemiddelde — steekproefgemiddelde — betrouwbaarheidsinterval
- populatieproportie — steekproefproportie — betrouwbaarheidsinterval

Activiteitenlijst

- een aselekte, representatieve steekproef herkennen — weten dat een steekproevenverdeling normaal is
- een hypothese toetsen met een afgesproken significantieniveau — links-, rechts-, tweezijdig toetsen
- een betrouwbaarheidsinterval gebruiken om een populatiegemiddelde te schatten
- een betrouwbaarheidsinterval gebruiken om een populatieproportie te schatten — een hypothesetoets op een populatieproportie toepassen

Achtergronden

In de **inductieve statistiek** probeer je om aan de hand van een steekproef informatie omtrent de gehele populatie te krijgen. Maar zo krijg je alleen beperkte informatie. De inductieve statistiek geeft methoden om daarmee uitspraken over de populatie als geheel te doen. Bekende methoden zijn hypothesen toetsen, schattingsmethoden, correlatie en regressie.

Francis Galton (1822—1911) richtte aan het Londense University College een leerstoel in de eugenetica op. De wiskundigen die deze leerstoel bezetten hebben veel voor de ontwikkeling van de mathematische statistiek betekend. Zij ontwikkelden vooral de methoden van statistische toetsing. **Karl Pearson** (1857—1936) bedacht de chi-kwadraat-toets waarmee een antwoord kon worden gegeven op de vraag hoe goed een theoretische verdeling past bij de gevonden gegevens. Zijn opvolger **Ronald Fischer** (1890—1962) en zijn volgelingen ontwikkelden methoden die geschikt zijn voor kleine steekproeven en vonden diverse verdelingen die juist voor die situatie geschikt zijn. Verder formuleerden zij de principes van het hypothese toetsen en vonden een techniek die bekend werd als de variantieanalyse. De variantieanalyse draaide om het met wiskundige methoden scheiden van 'echte effecten' en 'fouten'. Als een experiment een echt effect oplevert, blijkt uit de methode hoe sterk dit effect is in verhouding tot de fout.

Vanaf de jaren '20 van de vorige eeuw werd de statistiek voor wiskundigen een steeds volwaardiger onderwerp van onderzoek, waardoor de methoden sterk werden verfijnd en een exactere onderbouwing kregen. In 1928 publiceerden **Jerzy Neyman** (1894—1981) en **Egon Pearson** (1895—1980) (zoon van Karl Pearson) enkele geschriften waarin begrippen als 'fout van de tweede soort' en 'betrouwbaarheidsinterval' werden ingevoerd. In die tijd begon ook de industrie steeds meer de statistische methoden toe te passen, met name bij kwaliteitscontrole. Bovendien werd er gezocht naar steeds betere methoden om goede representatieve steekproeven te nemen.

Vanaf 1939 werd door **Abraham Wald** (1902—1950) de statistische beslissingstheorie ontwikkeld. Hierin werd de statistiek opgevat als een spel met de natuur als tegenstander. Hoewel dit een zeer algemene theorie is wordt hij tegenwoordig door heel veel statistici gebruikt.

Testen

Opgave 1

Zijn de volgende steekproeven aselect en/of representatief? Beargumenteer je antwoord.

- a Om de gemiddelde leeftijd van concertbezoekers te meten wordt aan de bezoekers van een klassiek concert gevraagd hoe oud ze zijn.
- b Een restaurantketen onderzoekt hoe gasten de hygiëne van hun restaurants ervaren. Hiervoor stellen ze de gasten bij één van hun restaurants vragen met betrekking tot de hygiëne.
- c Om het gewicht van de gemiddelde Nederlandse tiener te bepalen worden de gewichten van de leerlingen op jouw school gebruikt.

Opgave 2

Een fabrikant beweert dat een vulmachine pakken hagelslag vult met een gemiddeld gewicht van 351 gram en een standaardafwijking van 6,4 gram.

De inspectie vermoedt dat het gemiddelde van een pak hagelslag lager is. Om dit te toetsen met een significantieniveau van 5% neemt de inspectie een steekproef van 50 pakken hagelslag. Het gemiddelde gewicht van de steekproef is 349 gram.

- a Waarom mag je ervan uitgaan dat de steekproevenverdeling normaal verdeeld is? Is het daarbij belangrijk dat het gewicht van een pak hagelslag normaal verdeeld is?
- b Wat voor soort toets doet de inspectie?
- c Wat is de conclusie van de inspectie?
- d Is de conclusie hetzelfde als er een significantieniveau van 1% wordt gebruikt?

Opgave 3

Bij een marktonderzoek wordt gekeken naar de belangstelling voor elektrische auto's onder particulieren in Nederland. In een aselecte steekproef worden 1660 mensen benaderd en hiervan zeggen 917 particulieren dat ze de overstap naar een elektrische auto serieus overwegen.

- a Bereken het 99%-betrouwbaarheidsinterval voor de populatieproportie van mensen die geïnteresseerd zijn in een elektrische auto naar aanleiding van de steekproef voor de particuliere markt in Nederland.
- b Men wil de foutmarge beneden de 3% hebben en er komt nog een aanvullend aselect onderzoek onder 340 particulieren. Hiervan zeggen 162 particulieren dat ze de overstap overwegen. Wat zijn de gevolgen voor het 99%-betrouwbaarheidsinterval?
- c Wat zijn de gevolgen voor het 99%-betrouwbaarheidsinterval als de steekproefgrootte nog groter gaat worden en als \hat{p} nagenoeg gelijk blijft?

Opgave 4

Op het etiket van een pot jam staat dat het percentage rietsuiker 12,4% is. Om dit te controleren wordt het percentage rietsuiker van 20 willekeurige potten jam bepaald en in een tabel gezet.

Je mag aannemen dat het percentage rietsuiker normaal verdeeld is.

- a Formuleer de hypothesen in deze situatie.
- b Schat met deze resultaten de standaardafwijking in het percentage rietsuiker. Rond af op drie decimalen.

11,9	11,1	11,4	12,6
11,6	11,4	11,7	11,8
11,8	11,8	12,4	12,5
11,7	12,3	11,3	12,2
11,9	12,6	11,7	11,8

Tabel 1

Deze waarde mag je bij de rest van deze opgave blijven gebruiken.

- c Toets je hypothese met een significantieniveau $\alpha = 0,10$.
- d Bij welk gemiddeld percentage zou de nulhypothese verworpen worden?

Opgave 5

Een opdrachtgever eist van een fabrikant dat het gemiddelde gewicht werkzame stof in een partij van 2000 flessen met reinigingsmiddel 12 gram is.

De fabrikant test dit door middel van een steekproef. Hij vindt een betrouwbaarheidsinterval met een breedte van 2 gram voldoende klein en hij gaat uit van een betrouwbaarheid van 98%.

De fabrikant weet dat de standaardafwijking van het gewicht van de werkzame stof bij zijn productieproces 4 gram is.

De fabrikant wil weten hoe groot de steekproefomvang moet zijn om aan deze eisen te kunnen voldoen.

Bereken de minimaal benodigde steekproefomvang.

Toepassen

Opgave 6: Noord/Zuidlijn Amsterdam

Op de website van de Dienst Onderzoek en Statistiek van de gemeente Amsterdam stond in december 2008 het volgende:

“In totaal hebben 379 Amsterdammers meegedaan aan de telefonische enquête van november. Op de vraag of men voor of tegen de aanleg van de Noord/Zuidlijn is, zegt ongeveer de helft van respondenten (49%) voor de aanleg te zijn. Het percentage voorstanders ligt hoger dan bij de vorige meting: in september was 38% voorstander. Het percentage voorstanders is hiermee terug op het niveau van 2006 toen 48% voorstander was. Het effect van de gebeurtenissen rond de Vijzelgracht lijkt te zijn vervolgen in de mening van Amsterdammers.”

- Welke uitspraak over het aantal voorstanders in december 2008 van de Noord/Zuidlijn kon men toen met 95% betrouwbaarheid doen?
- In september kon men met een betrouwbaarheid van 95% beweren dat het aantal voorstanders tussen de 37% en de 39% lag. Hoeveel mensen zijn er toen ondervraagd?

Opgave 7: Pokergames

Bij Pokergames kun je chips inzetten en winnen. Big blind (BB) is een inzet van een stapel van bijvoorbeeld 1000 chips zonder dat de kaarten gezien zijn. De winrate (WR) is het aantal BB dat je per 100 handen ($BB/100$) wint. In pokergames wordt je winrate bijgehouden. Het is een identificatie hoeveel chips je wint, en daarmee of je een winnende speler bent of niet. In pokergames wordt ook het aantal handen bijgehouden dat je gespeeld hebt. En de standaarddeviatie σ .

- Veronderstel dat een pokergame een $WR = 3,5BB/100$ aangeeft met $\sigma = 40$ na 30000 handen. Je kunt zeggen dat deze 30000 handen een steekproef vormen. Alles wordt dus uitgedrukt per 100 handen. Dus $n = \frac{30000}{100} = 300$. Een hand is de 5 beste kaarten van een speler per spelletje. Wat kun je nu zeggen over de WR voor de eerstvolgende hand met een betrouwbaarheid van 95%?
- Hoeveel handen moet je bij deze winrate en σ spelen om met 95% zekerheid geen verlies te behalen?

Examen

Opgave 8: Basketballen

Jaarlijks controleert de materiaalcommissaris of de ballen van Flits voldoen aan de eisen die de basketbalbond stelt. Deze zijn:

De omtrek van de bal mag niet minder bedragen dan 75 cm en niet meer dan 78 cm. Het gewicht mag niet minder zijn dan 600 g en niet meer dan 650 g.

Bij zo'n controle komt hij tot de ontdekking dat het gewicht van de ballen klopt, maar dat de omtrek van 15 ballen niet in orde is. Omdat hierbij ook een redelijk aantal nieuwe ballen is, stelt hij zich in verbinding met de leverancier: het bedrijf Balfa. Dit bedrijf beweert dat het dagelijks 125 ballen produceert, waarvan de omtrek normaal verdeeld is met een gemiddelde van 76,5 cm en een standaarddeviatie van 0,70 cm. Neem aan dat deze gegevens juist zijn.

- a Toon aan dat men kan verwachten dat 4 ballen in de dagproductie niet voldoen aan de eisen die de bond stelt aan de omtrek.
- b Bereken in procenten nauwkeurig de kans dat in een aselechte steekproef van 5 door Balfa gemaakte ballen, elke bal voldoet aan de eisen die de bond stelt aan de omtrek.

Op grond van de eigen gegevens beweert de verkoper van Balfa dat gemiddeld hoogstens één op de twintig ballen niet aan alle eisen van de bond voldoet. De materiaalcommissaris heeft zo zijn twijfels. Zij spreken met elkaar af de bewering van de verkoper te toetsen door middel van een aselechte steekproef van 15 stuks bij een significantieniveau van 5%. Indien het resultaat de verkoper in het ongelijk stelt, krijgt Flits de 15 nieuwe ballen uit de steekproef gratis.

X is het aantal ballen in de steekproef dat niet voldoet aan de eisen van de bond.

- c Bereken de kleinste waarde van X waarbij Flits de ballen gratis krijgt.

(bron: examen wiskunde A vwo 1990, tweede tijdvak)

Opgave 9: Rookgedrag van leerlingen

Sinds de jaren tachtig meet het Trimbos-instituut regelmatig via een enquête het gebruik van alcohol, drugs en tabak in aselechte, representatieve steekproeven onder alle leerlingen van het voortgezet onderwijs. Ook werd de leerlingen in de enquête gevraagd naar hun leeftijd (in jaren), hun geslacht (jongen, meisje), en hun schoolniveau (vmbo, havo, vwo).

Aan de enquête van 2015 deden 6714 leerlingen mee in de leeftijd van 12 tot en met 16 jaar. In deze groep is onder andere gekeken naar de lifetime-prevalentie van roken. Hieronder staat wat dit begrip betekent:

lifetime-prevalentie van roken = het percentage van de leerlingen dat rookt of ooit gerookt heeft in zijn of haar leven.

steekproefomvang	6714
aantal dat rookt of ooit gerookt heeft	1544
lifetime-prevalentie	23%

Tabel 2

In de tabel zie je dat van de leerlingen in de steekproef 23%, bijna een kwart, rookt of ooit gerookt heeft. Op basis van bovenstaande gegevens kun je het 95%-betrouwbaarheidsinterval voor de lifetime-prevalentie van roken berekenen.


Bereken dit 95%-betrouwbaarheidsinterval. Rond de percentages in je antwoord af op gehele getallen.

(bron: examen wiskunde A havo in 2018, tweede tijdvak)



© 2024

Deze paragraaf is een onderdeel van het Math4All wiskundemateriaal.

Math4All stelt het op prijs als onvolkomenheden in het materiaal worden gemeld en ideeën voor verbeteringen in de content of dienstverlening kenbaar worden gemaakt. Klik op  in de marge bij de betreffende opgave. Uw mailprogramma wordt dan geopend waarbij het emailadres en onderwerp al zijn ingevuld. U hoeft alleen uw opmerkingen nog maar in te voeren.

Email: f.spijkers@math4all.nl

Met de Math4All Foliostraat kunnen complete readers worden samengesteld en toetsen worden gegenereerd. Docenten kunnen bij a.f.otten@math4all.nl een gratis inlog voor de maatwerkdienst aanvragen.
