1.5 Totaalbeeld

Samenvatten

Je hebt nu het onderwerp **Statistiek voor wiskunde B** 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

- populatie, steekproef — aselect, representatief — statistische variabele — kwalitatief, nominaal, ordinaal — kwantitatief, discreet, continu
- frequentietabel — absolute, relatieve, cumulatieve frequenties — klassenindeling, klassegrenzen, klassebreedte — beelddiagram, staafdiagram, cirkeldiagram — histogram, frequentiepolygoon, cumulatief frequentiepolygoon
- centrummaten: modus, mediaan, gemiddelde — spreidingsmaten: spreidingsbreedte, kwartielafstand, standaardafwijking — boxplot
- normale verdeling — vuistregels normale verdeling — normale kansen
- gemiddelde en standaardafwijking som en verschil van kansvariabelen (stochasten) — wortel-n-wet
- hypothese toetsen, nulhypothese, alternatieve hypothese — significantieniveau, kritieke gebied — linkszijdige, rechtszijdige, tweezijdige toets

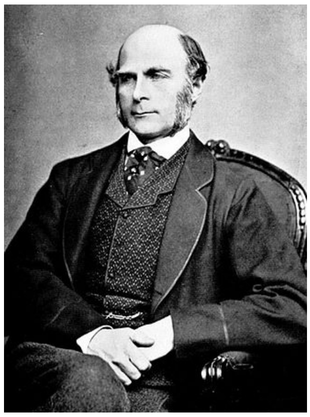
Activiteitenlijst

- verschillende soorten statistische variabelen onderscheiden — een aselecte, representatieve steekproef herkennen — valkuilen bij statistisch onderzoek herkennen
- data ordenen in frequentietabellen, al dan niet met een klassenindeling — data ordenen in geschikte diagrammen afhankelijk van de soort statistische variabele
- centrum- en spreidingsmaten berekenen — boxplots maken en gebruiken
- normale verdeling herkennen en karakteriseren met gemiddelde en standaardafwijking — vuistregels voor de normale verdeling toepassen — kansen berekenen bij normale verdelingen — gemiddelde, standaardafwijking, grenswaarde berekenen bij normaal verdeelde kansvariabelen
- gemiddelde en standaardafwijking som en verschil van kansvariabelen (stochasten) berekenen — de wortel-n-wet gebruiken
- bij normaal verdeelde kansvariabelen een hypothese toetsen — met behulp van het significantieniveau het kritieke gebied vaststellen

Achtergronden

Met de toepassing van de kansrekening op het verzekeringswezen ontstond ook voor het eerst de behoefte tot het bijhouden van bevolkingsgegevens. En zo werd een klimaat geschapen voor de ontwikkeling van de statistiek. In 1662 publiceerde de Brit John Graunt zijn ‘Natural and Political Observations’, waarin een eerste statistische analyse voorkwam van de wekelijkse lijst van sterfgevallen in en om London, de zogenaamde ‘Bills of mortality’. En in 1693 kwam de astronoom Edmond Halley met een levensverwachtingstabel gebaseerd op de sterftecijfers van de stad Breslau.

In de achttiende eeuw zette zich deze statistische traditie verder voort. Daarnaast vonden in de kansrekening de nodige ontwikkelingen plaats: de normale verdeling werd bedacht door de Britse wiskundige De Moivre en door de Belg Adolphe Quetelet in het begin van de negentiende eeuw toegepast op de sociale statistiek. De echte start van de mathematische statistiek is echter toe te schrijven aan Francis Galton en vond plaats aan het einde van de negentiende eeuw.



Figuur 1

Francis Galton (1822 - 1911) was een neef van Charles Darwin, de bioloog die de evolutietheorie opzette. Hij paste statistische methoden toe bij de analyse van sociale gegevens en erfelijke eigenschappen. Hij dacht dat de normale verdeling de mate van variatie van fysieke eigenschappen aangaf. Hij werkte met het begrip standaarddeviatie als maat voor de spreiding van de normale verdeling. In tegenstelling tot Quetelet dacht hij niet zozeer in termen van ‘foute afwijkingen van het juiste gemiddelde’, maar als noodzakelijke verscheidenheid in het licht van de evolutietheorie.
Galton ontdekte een maat voor de correlatie (dat is de mate waarin een verband bestaat) tussen statistische variabelen (zoals bijvoorbeeld lengte en gewicht): de correlatiecoëfficiënt, een getal tussen - 1 en 1. Als die correlatiecoëfficiënt de waarde 0 had was er geen enkel verband tussen de variabelen.
Galton paste zijn statistische theorieën vooral toe op de eugenetica, de studie van erfelijke eigenschappen.

Testen

Opgave 1

Bekijk de frequentieverdeling van de weeklonen van 65 werknemers van een bedrijf.

loon (€)	*aantal*
$500- < 600$	8
$600- < 700$	10
$700- < 800$	16
$800- < 900$	14
$900- < 1000$	10
$1000- < 1100$	5
$1100- < 1200$	2
totaal	65

Tabel 1

a Bereken de relatieve frequenties bij deze tabel.

b Maak een frequentiepolygoon bij de tabel.

Het bedrijf neemt vijf extra werknemers in dienst. Zij krijgen een weekloon van € 835,00; € 1156,00; € 1345,00; € 1567,00 en € 1714,00.

c Pas de frequentietabel aan voor de zeventig werknemers.

d Teken een relatieve frequentiepolygoon van de zeventig werknemers met een klassenbreedte van 200.

Opgave 2

In een fabriek verpakt een machine in kleine zakjes poedermelk voor in de koffie. Elk van die zakjes hoort 3 gram melkpoeder te bevatten. De fabrikant heeft zijn machine zo afgesteld dat het vulgewicht van deze zakjes normaal is verdeeld met een gemiddelde van 3,1 gram en een standaardafwijking van 0,06 gram.

a Hoeveel procent van de zakjes melkpoeder die deze machine produceert is te licht?

De fabrikant voldoet hiermee niet aan de richtlijnen van de Europese Unie. Die schrijven voor dat niet meer dan 1% van de zakjes poedermelk minder dan 3 gram mag bevatten.

b De fabrikant besluit om iets meer melkpoeder in de zakjes te doen. Op welk gemiddelde vulgewicht moet hij de machine instellen om aan de richtlijn van de Europese Unie te voldoen? Ga ervan uit dat de standaardafwijking van de verdeling van de vulgewichten hetzelfde blijft.

c Je koopt een doosje met daarin twintig zakjes van het melkpoeder dat nog het oorspronkelijke gemiddelde van 3,1 gram heeft.

Hoeveel gram melkpoeder verwacht je gemiddeld per zakje in het doosje met twintig zakjes? En welke standaardafwijking hoort daarbij?

d Hoe groot is de kans dat je in totaal minder dan $20 \cdot 3 = 60$ gram melkpoeder hebt gekocht?

Licht je antwoord toe met behulp van je kennis over de normaalkromme.

Opgave 3

Bloeddruk wordt gemeten in mm Hg (spreek uit: millimeter kwik).
Bij een groep van duizend mannen is de bloeddruk normaal verdeeld met een gemiddelde van 128,5 mm Hg en een standaardafwijking van 12,5 mm Hg.

a Maak een normaalkromme bij de bloeddrukverdeling van deze groep mannen. Verdeel de oppervlakte onder de normaalkromme in zes delen en noteer in ieder deel het juiste percentage.

b Hoeveel procent van de mannen heeft naar schatting een bloeddruk van minder dan 141 mm Hg?

c Hoeveel procent van de mannen heeft naar schatting een bloeddruk die meer dan twee keer de standaardafwijking afwijkt van de gemiddelde bloeddruk?

d Maak op basis van de normaalkromme een schatting van het percentage mannen dat een bloeddruk heeft van meer dan 150 mm Hg en vergelijk dat met het daadwerkelijke percentage mannen met een dergelijke bloeddruk.

Opgave 4

Uit onderzoek van het gemengde boerenbedrijf bleek het houden van kippen een belangrijke rol te spelen bij het tot stand komen van het inkomen van deze boeren. Daarom werd de boeren gevraagd naar het aantal kippen op hun bedrijf.

aantal kippen	*aantal bedrijven*	*aantal kippen*	*aantal bedrijven*
1 – 10	5	101 – 110	123
11 – 20	12	111 – 120	101
21 – 30	19	121 – 130	85
31 – 40	24	131 – 140	79
41 – 50	33	141 – 150	60
51 – 60	52	151 – 160	43
61 – 70	69	161 – 170	21
71 – 80	75	171 – 180	9
81 – 90	108	181 – 190	4
91 – 100	120	191 – 200	2

Tabel 2

a Met welk type variabele heb je hier te maken?

b Teken een cumulatieve relatieve frequentiepolygoon.

c Schat de mediaan en de beide kwartielen. Teken een boxplot bij deze gegevens.

d Je kunt het gemiddelde en de standaardafwijking schatten met behulp van de klassenmiddens en de frequentietabel. Laat zien hoe dat gaat en geef je antwoord in één decimaal nauwkeurig.

e Hoe kun je aan de cumulatieve relatieve frequentiepolygoon uit b zien dat het aantal kippen voor gemengde bedrijven redelijk normaal verdeeld is?

Opgave 5

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 Wat voor soort toets doet de inspectie?

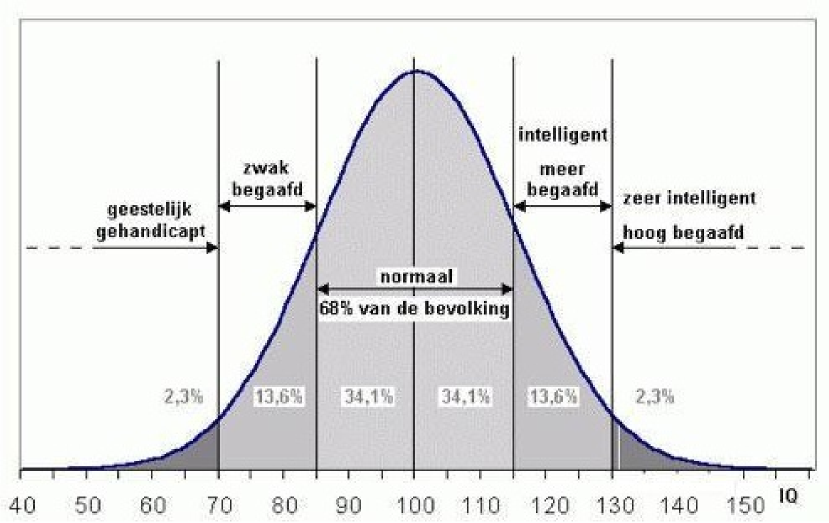
b Welke de conclusie trekt de inspectie?

c Is de conclusie hetzelfde als er een significatieniveau van 1% wordt gebruikt?

Toepassen

Opgave 6: Intelligentiequotiënt

In het begin van de vorige eeuw werd er veel waarde gehecht aan het zogenaamde **intelligentie-quotiënt (IQ)** van met name kinderen. In 1904 werd de psycholoog **Alfred Binet (1857 - 1911)** door de Franse overheid gevraagd een test te ontwerpen om 'slimme' van 'domme' kinderen te onderscheiden.



Figuur 2

Binet ontwierp een intelligentietest waarmee hij de intelligentieleeftijd van een kind vaststelde. Wanneer de intelligentieleeftijd wordt gedeeld door de werkelijke leeftijd (en met 100 vermenigvuldigd) dan krijg je het IQ. Dit IQ is normaal verdeeld met een gemiddelde dat op 100 is gesteld. Zo is de Stanford-Binet intelligentieschaal ontwikkeld. Bekijk de normale verdeling van de IQ's zoals Binet die aantrof.

a Welk gemiddelde en welke standaardafwijking heeft het IQ volgens de Stanford-Binet-schaal?

b Is het IQ afhankelijk van je leeftijd?

c Hoeveel procent van de mensen heeft een minder dan normale intelligentie?

d Hoeveel procent van de mensen heeft een IQ van boven de 140?

e Welk IQ heeft de meest intelligente 10% van de bevolking volgens de Stanford-Binet-schaal?

Opgave 7: Zwangerschap

De zwangerschapsduur bij mensen is normaal verdeeld met een gemiddelde van 266 dagen en een standaarddeviatie van 16 dagen.

a Bij een premature geboorte wordt het kind minstens drie weken voor het gemiddelde geboren. Hoe groot is de kans daar op?

b In 7% van de gevallen duurt een zwangerschap zo lang dat de geboorte moet worden ingeleid. Vanaf welke zwangerschapsduur gebeurt dit dus?

c In 1973 beweerde een vrouw dat ze 310 dagen zwanger was geweest omdat op de dag van de bevalling haar man al 310 dagen als marinier van huis was. Hoe groot is de kans op een zwangerschap van minstens 310 dagen?

Examen

Opgave 8: Antropometrie

Een ontwerp moet niet alleen mooi, maar ook functioneel zijn. Bij veel ontwerpen wordt daarom rekening gehouden met de maten van het menselijk lichaam. Ontwerpers maken daarom vaak gebruik van antropometrietabellen. Dit zijn tabellen waarin het gemiddelde en de standaardafwijking van allerlei afmetingen van het menselijk lichaam staan. Al deze lichaamsmaten zijn (bij benadering) normaal verdeeld.

a Bij het inrichten van een optimale werkplek houdt men rekening met de lichaamsmaten. Een bureaustoel heeft precies de goede zithoogte als de zithoogte gelijk is aan de knieholtehoogte van een persoon plus 30 mm voor de schoenzool. Van een bureaustoel is de zithoogte verstelbaar van 436 tot 516 mm. De knieholtehoogte is gemiddeld 464 mm met een standaardafwijking van 40 mm.
Bereken voor hoeveel procent van de mensen deze stoel op precies de goede zithoogte ingesteld kan worden.

Bij bovenstaande vraag is geen onderscheid gemaakt tussen mannen en vrouwen. In werkelijkheid staan in antropometrietabellen de lichaamsmaten voor mannen en vrouwen apart vermeld. Zie bijvoorbeeld de gegevens voor lichaamslengte in mm in tabel 1.

	man gemiddeld	man standaardafwijking	vrouw gemiddeld	vrouw standaardafwijking
lichaamslengte in mm	1817	83	1668	67

Tabel 3

Vaak maakt men voor een gemengde groep toch gebruik van één normale verdeling. Dit is dan een vrij ruwe benadering. Het gemiddelde en de standaardafwijking van deze normale verdeling berekent men met behulp van de volgende formules:

$$\overline{x}_g = a_m \cdot \overline{x}_m + a_v \cdot \overline{x}_v$$

$$s_g^2 = a_m \cdot s_m^2 + a_v \cdot s_v^2 + a_m \cdot a_v \cdot (\overline{x}_m - \overline{x}_v)^2$$

waarbij

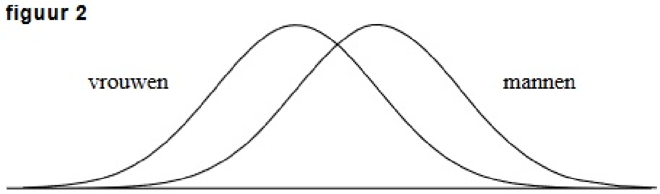
- $\overline{x}_g$ het gemiddelde is van de gemengde groep;
- $\overline{x}_m$ en $\overline{x}_v$ het gemiddelde is van de mannen, respectievelijk de vrouwen;
- s_g de standaardafwijking is van de gemengde groep;
- s_m en s_v de standaardafwijking is van de mannen, respectievelijk de vrouwen;
- a_m en a_v het aandeel mannen in de groep en het aandeel vrouwen. Er geldt: $a_m + a_v = 1$.

Een groep bestaat uit 40% mannen en 60% vrouwen, dus $a_m = 0{,}40$ en $a_v = 0{,}60$. Men kan op twee manieren berekenen hoeveel procent van deze groep langer is dan 185 cm:

- met behulp van één normale verdeling voor de gemengde groep en de hierboven gegeven formules voor het gemiddelde en de standaardafwijking;
- zonder gebruik te maken van deze formules, met behulp van de aparte gegevens voor mannen en vrouwen. De uitkomsten van beide berekeningswijzen zullen in het algemeen verschillen.

b Bereken op beide manieren hoeveel procent van deze groep langer is dan 185 cm.

Voor sommige lichaamsafmetingen geldt dat het gemiddelde voor mannen en vrouwen verschillend is, maar de standaardafwijking gelijk. We noemen deze standaardafwijking s. Er geldt dus: $s_m = s_v = s$.
In de figuur hieronder zie je een schets van de verdelingskrommen die bij zo'n situatie horen. De gemengde groep (mannen en vrouwen samen) heeft een grotere spreiding dan elke groep afzonderlijk. Als je in de figuur de grafiek voor de gemengde groep zou tekenen, zou deze breder zijn dan de grafieken voor mannen en vrouwen afzonderlijk. Er geldt dus: $s_g > s$.



Figuur 3

De formule voor s_g^2 kan dan geschreven worden als:

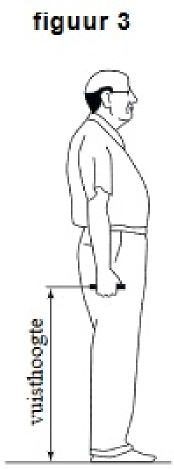
$$s_g^2 = a_m \cdot s^2 + a_v \cdot s^2 + a_m \cdot a_v \cdot (\overline{x}_m - \overline{x}_v)^2$$

Ook zonder de figuur, dus alleen aan de hand van de formule voor s_g^2, kun je met een redenering aantonen dat in dit geval $s_g > s$.

c Geef die redenering.

Voor sommige doeleinden wordt ook onderscheid gemaakt tussen oudere mensen (70 jaar en ouder) en jongere mensen (20 tot 60 jaar). Zie de figuur. De TU Delft heeft in 1998 uitgebreid antropometrisch onderzoek gedaan bij oudere mensen. Hierbij is onder andere de vuisthoogte gemeten, zie figuur 3.

De vuisthoogte is van belang voor bijvoorbeeld koffers en tassen op wieltjes. Omdat oudere mensen gemiddeld minder lang zijn dan jongere mensen, verwacht men dat de vuisthoogte van oudere mannen kleiner is dan die van mannen van 20 tot 60 jaar. De vuisthoogte van mannen van 20 tot 60 jaar is gemiddeld 817 mm met een standaardafwijking van 47 mm. Bij een steekproef van 128 mannen van 70 jaar en ouder was de gemiddelde vuisthoogte 761 mm. Dit steekproefresultaat (761 mm) was ruim voldoende aanleiding om te concluderen dat de vuisthoogte van mannen van 70 jaar en ouder kleiner is dan die van mannen van 20 tot 60 jaar.



Figuur 4

d Bereken bij een steekproef van 128 mannen van 70 jaar en ouder tot welke waarde van het steekproefresultaat men deze conclusie nog kan trekken. Neem een significantieniveau van 5%.

(bron: examen vwo wiskunde A in 2010, tweede tijdvak)

Email: f.spijkers@math4all.nl