Uitwerkingen inleveropdrachten reader statistiek voor de minor

Inleveropdracht A

Een stochast is een **kansvariabele** waarvan de uitkomst louter en alleen afhangt van het **toeval**. Bij een stochast hoort een kansverdeling. Dat kan een **discrete** kansverdeling maar ook **continu** kansverdeling zijn.

Opgave 1

In **opgave 1** gaat het over het gewicht van theezakjes. We noemen het gewicht van een willekeurig theezakje X. De stochast X is **normaal verdeeld** met een **gemiddelde** van 2,2 gram en een **standaarddeviatie** van 0,5 gram.

Je kunt je dan allerlei dingen af gaan vragen:

- In een doosje zitten 15 theezakjes. Wat is de kans dat een doosje minder dan 28 gram thee bevat?
- Wat is de kans dat het gemiddelde gewicht van 20 theezakjes minder is dan 1,9 gram?
- Wat is de kans dat een doosje van 15 theezakjes meer dan 10 theezakjes bevat die elk meer dan 2,1 gram bevatten?

Daarbij gaat het eigenlijk om de vraag wat er gebeurt met de **som** of het **gemiddelde** van een aantal stochasten? Kan je iets zeggen over het gemiddelde? Of de standaarddeviatie?

Er geldt:

$$E(X_1 + X_2 + \ldots + X_n) = E(X_1) + E(X_2) + \ldots + E(X_n)$$

- De **verwachtingswaarde** van de 'som van n (onafhankelijke) stochasten' is gelijk aan de som van de afzonderlijke verwachtingswaarden.

$$Var(X_1 + X_2 + \ldots + X_n) = Var(X_1) + Var(X_2) + \ldots + Var(X_n)$$

- De **variantie** van de 'som van n (onafhankelijk) stochasten' is gelijk aan de som van de afzonderlijke varianties.

Bij **opgave a.** kijk je naar 15 theezakjes. Deze **nieuwe** kansvariabele X is de som van 15 keer dezelfde stochast. De verwachtingswaarde is 15·2,2=33 gram. De variantie is gelijk aan $15\cdot(0{,}5)^2=3{,}75$. De standaarddeviatie is $\sqrt{3{,}75}$.

X: totale gewicht
X~normaal verdeeld met $\mu = 33$ gram en $\sigma = \sqrt{3{,}75}$ gram.
Gevraagd: P(X<28)$\approx 0{,}05$.

Waarschijnlijk had je de standaarddeviatie van X anders uitgerekend:

$\sigma(X) = \frac{1}{2} \cdot \sqrt{15}$ vanwege de **wortel-n-wet**. Dat is (uiteraard) hetzelfde! Ga maar na!

Eerst het kwadraat nemen, dan met 15 vermenigvuldigen en dan de wortel trekken is hetzelfde als meteen vermenigvuldigen met $\sqrt{15}$.

Bij **opgave b.** gaat om het gemiddelde van 20 zakjes. Deze nieuwe kansvariabele $\overline{X}$ heeft als verwachtingswaarde 2,2. Eerst doe je 20 keer de verwachtingswaarde en daarna deel je door 20 omdat het over het gemiddelde gaat.

Bij de standaarddeviatie gaat het weer op dezelfde manier als bij de som. De standaarddeviatie van de som van 20 zakjes is $\sigma \cdot \sqrt{20}$ (de wortel n-wet). Voor het gemiddelde deel je door 20. Dus de standaarddeviatie van $\bar{X}$ is gelijk aan $\frac{\sigma}{\sqrt{20}}$.

$\bar{X}$: gemiddelde gewicht van 20 zakjes
$\bar{X}$~normaal verdeeld met $\mu = 2{,}2$ gram en $\sigma = \frac{0{,}5}{\sqrt{20}}$ gram.
Gevraagd: P($\bar{X} < 1{,}9$)$\approx 0{,}004$

Bij **opgave c.** gaat het om iets anders. Bij een gegeven kans dat één theezakje een gewicht heeft van meer dan 2,1 gram heb je bij een doosje van 15 zakjes te maken met een **binomiaal kansprobleem**. De kans dat een zakje een gewicht heeft van meer dan 2,1 gram bereken je zo:

X: gewicht van 1 theezakje
X~normaal verdeeld met $\mu = 2{,}2$ gram en $\sigma = 0{,}5$ gram
Gevraagd: P(X>2,1) ≈ 0,5793

Y: aantal zakjes met een gewicht van meer dan 2,1 gram
Y~binomiaal verdeeld met n=15 en p=0,5793
Gevraagd: P(Y>10) = 1 – P(Y≤10) ≈ 0,172

Bij **opgave d** heb je te maken met een **geometrische verdeling** (zie hoofdstuk 3). De vraag is naar minder dan 20 zakjes. Dat is lastig want het kan dan in 1, 2, 3, 4, … , 19 zakjes. Die zou je dan allemaal apart moeten uitrekenen. De kan op 20 zakjes of meer is makkelijker. Je moet dan 19 keer een zakje van minder dan 3 gram.

X: gewicht van 1 theezakje
X~normaal verdeeld met $\mu = 2{,}2$ gram en $\sigma = 0{,}5$ gram
Gevraagd: P(X>3) ≈ 0,0548

Y:aantal keren een zakje proberen
Y~geometrische verdeling met p=0,0548 (kans op succes)
Gevraagd: P(X<20) = 1 – P(X$\geq$ 20) = 1 – $(0{,}9452)^{20}$ ≈ 0,676

Bij **opgave e.** gaat het om de stochast Y met $\mu = 2{,}2$ gram en $\sigma = \frac{0{,}5}{\sqrt{n}}$ gram.
Y: aantal zakjes in een doosje
Y~normaal verdeeld met $\mu = 2{,}2$ gram en $\sigma = \frac{0{,}5}{\sqrt{n}}$ gram.
Gevraagd: n zodat P(Y >2) ≥0,99
We stellen eerst vast wel z-waarde er hoort bij 0,01. Dat geeft $\phi(z) = 0{,}01 \Rightarrow z \approx -2{,}3263$

Gebruik dan $z = \frac{Y-\mu}{\sigma}$. Dit geeft ingevuld $-2,3263 = \frac{2-2,2}{\sigma} \Rightarrow \sigma \approx 0,0860$

We weten nu dat $0,0860 = \frac{0,5}{\sqrt{n}} \Rightarrow n = 34$

Dus de fabrikant moet minimaal 34 theezakjes in een doosje doen.

Opgave 2

De tekentoets. Tellen geeft 8 keer plus, 3 keer min en 1 nul.

Het aantal plussen bij n=11.

H_0: p=0,5
H_1: p>0,5
$\alpha = 0,05$

Onder H0 berekenen we de kans op 8 plussen bij n=11.
Dit is een binomiaal kansprobleem, dus:

X: aantal plussen
p=0,5 (onder H0)
Gevraagd: $P(X \geq 8)$

$P(X \geq 8) = 1 - P(X \leq 7) \approx 0,113$

Deze kans is (veel) groter dan 0,05. We kunnen H_0 dus niet verwerpen. Het is niet statistisch aangetoond dat de tweede toets beter gemaakt is.

Inleveropdracht B

Opgave 1

a. $(26)_3 = \frac{26!}{(26-3)!} = 26 \cdot 25 \cdot 24 = 15.600$

b. $\binom{26}{3} = \frac{26!}{3! \cdot (26-3)!} = 2600$

c. $15.600 = 3! \cdot 2600$

Die 3! Is precies het aantal permutaties dat je kan maken van 3 letters.

Opgave 2

a. 26^3 = 17.576

b. $\binom{26\text{-}1\text{+}3}{3} = \binom{28}{3} = 3276$

c. Het verband tussen 17.576 en 3.276 is lastig. Die 17.576 bestaat uit 15.600 permutaties met 3 verschillende letters (zie boven). De rest, 1.976 bestaat uit permutaties waarin 2 of 3 keer dezelfde letter voorkomt. De 1.976 bestaat uit 26 permutaties met 3 dezelfde letters en 1950 permutaties met 2 dezelfde letters. (3 × 26 × 25).
15.600 + 1950 + 26 = 17.576

De 3.276 verschillende herhalingscombinaties bestaat uit 2.600 combinaties met 3 verschillende letters (zie boven), 650 combinaties met 2 dezelfde letters en 26 combinaties met 3 dezelfde letters.
2.600 + 650 + 26 = 3.276

Opgave 3

In de opgave was gegeven:

- **De kans dat het morgen regent is 0,5 en de kans dat je in de file komt te staan is 0,8.**

Wat is nu de kans dat als je morgen met de auto gaat het regent en dat je in de file staat?
Je zou kunnen denken dat die kans gelijk is aan 0,5·0,8=0,4 is. Maar is dat wel zo?
Laten we eens naar onderstaande tabel kijken:

		regen		
		ja	nee	
file	ja	...	...	0,8
	nee	...	...	0,2
		0,5	0,5	1,0

Dit is (feitelijk) alles wat je zeker weet.

De kans op regen en file kan best 0,5 zijn. Je krijgt dan:

		regen		
		ja	nee	
file	ja	**0,5**	0,3	0,8
	nee	0,0	0,2	0,2
		0,5	0,5	1,0

...dat zou betekenen dat je bij regen altijd in de file staat. Dat betekent niet dat als je in de file staat het altijd regent!
Maar de kans op file en regen zou ook best 0,3 kunnen zijn. Je krijgt dan:

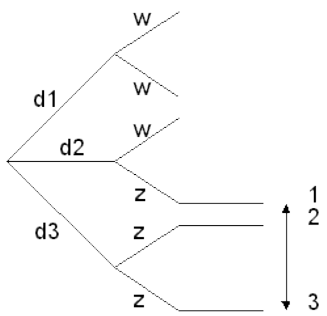
		regen		
		ja	nee	
file	ja	0,3	0,5	0,8
	nee	0,2	0,0	0,2
		0,5	0,5	1,0

Kortom 'bijna' alles kan. Je kunt de vraag dus niet beantwoorden. Het is goed 'denkbaar' dat regen invloed heeft op de kans op files. Maar zeker weten doe je dat niet....
Je hebt hier te maken met **voorwaardelijke kans**.

Bij veel kansproblemen heb je hier niets mee te maken, omdat je vaak uit mag gaan van de veronderstelling dat opeenvolgende gebeurtenissen geen invloed op elkaar uitoefenen. Als je bijvoorbeeld met twee dobbelstenen gooit, zal het aantal ogen van de ene dobbelsteen geen invloed hebben op het aantal ogen van de andere dobbelsteen.

Wanneer geldt dat de kans 'dat als je morgen met de auto gaat, het regent en dat je in de file staat' **wel** gelijk is aan $0{,}5 \cdot 0{,}8 = 0{,}4$? In dat geval zijn 'in de file staan' en 'het regent' onafhankelijk! Of anders geformuleerd: 'of het nu regent of niet, de kans op file blijft even groot' of ook 'of ik nu in de file sta of niet de kans op regen blijft even groot'.

Opgave 4



Bij 2 van de 3 mogelijkheden zit er nog een zwarte bal in de doos de kans is dus $^2/_3$

Of ook:

$$\text{Er geldt}: P(A|B) = \frac{P(A \text{ en } B)}{P(B)}$$

$$P(\text{pak zwart} \mid \text{zwart blijft over}) = \frac{P(\text{pak zwart EN zwart blijft over})}{P(\text{zwart blijft over})} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

Inleveropdracht C

Hier onder zie je de resultaten van 12 leerlingen op twee toetsen:

leerling	1	2	3	4	5	6	7	8	9	10	11	12
1e toets	7	4	4	4	8	6	6	8	7	8	3	6
2e toets	8	8	4	7	7	9	9	5	9	6	7	9

- Onderzoek of je met een significantieniveau van 5% aannemelijk kan maken dat de tweede toets **gemiddeld** beter gemaakt is dan de eerste toets.

Uitwerking

We gaan eerst maar 's kijken naar het verschil tussen toets 1 en 2.

leerling	1	2	3	4	5	6	7	8	9	10	11	12
1e toets	7	4	4	4	8	6	6	8	7	8	3	6
2e toets	8	8	4	7	7	9	9	5	9	6	7	9
Verschil	1	4	0	3	-1	3	3	-3	2	-2	4	3

We kunnen nu gaan kijken of het gemiddelde van dit 'verschil' **significant** afwijkt van 0. Als de tweede toets niet beter gemaakt is dan zou het verschil niet significant af moeten wijken van 0.

Dat kan met de t-toets (met 95% betrouwbaarheid):

H_0: μ=0
H_1: μ>0

$$\bar{x} \approx 1{,}417 \text{ en } s_X \approx 2{,}392$$
$$v = 12 - 1 = 11$$
$$t_{11} = 1{,}796$$

$$\bar{x} - z \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z \cdot \frac{\sigma}{\sqrt{n}}$$
$$1{,}417 - 1{,}796 \cdot \frac{2{,}392}{\sqrt{12}} < \mu < 1{,}417 + 1{,}796 \cdot \frac{2{,}392}{\sqrt{12}}$$
$$0{,}18 < \mu < 2{,}66$$

Met 95% betrouwbaarheid kunnen we zeggen dat het 'echte' gemiddelde groter is dan nul.
We verwerpen H_0.
De tweede toets is beter gemaakt.