Tentamen Statistiek MBW/KW (deel 1, eerste kans) Uitwerking

Afdeling: Propedeuse MBW/KW 2021-2022
Examinator: Dr. J.B.M. Melissen, Dr. R.J. Nijboer
Datum: vrijdag 10 juni 2022 09:00 – 12:00, duur tentamen: 3 uur

1. Lever de antwoorden in op het geprinte antwoordformulier en zet je naam erop. De berekeningen en uitleg graag op gelinieerd papier inleveren. Er zijn reserveformulieren aanwezig voor als je fouten hebt gemaakt.
2. **Alle antwoorden moeten gemotiveerd worden**!
3. Rond eindantwoorden (kommagetallen) af op *vier* decimalen, tenzij anders vermeld.
4. Boeken, reader en aantekeningen mogen worden geraadpleegd.
5. De aanwezigheid van *communicatieapparatuur* is niet toegestaan.
6. Het gebruik van een (grafische) rekenmachine met statistische programmatuur en het raadplegen van de bijbehorende handleiding is toegestaan. Het *statistische* gebruik van deze rekenmachine is bij een aantal onderdelen ingeperkt. Let op de aanwijzingen!
7. **De opgaven dienen na afloop van het tentamen ingeleverd te worden.**

Dit tentamen bestaat uit vier opgaven (30, 30, 20, 20 punten). Score = Puntentotaal/10

Opgave 1 (30 punten).

Voor het tentamen Statistiek voor MBW/KW deel 1 beschrijft de continue kansvariabele $\underline{x}$ het cijfer dat een student haalt. Het cijfer is een reëel getal tussen 0 en 10. De kansverdeling wordt hopelijk bepaald door de volgende kansdichtheidsfunctie:

$$f(x) = \begin{cases} \dfrac{3}{2500}x^2(10-x) & \text{als} \quad 0 \le x \le 10 \\ 0 & \text{elders} \end{cases}$$

1a [4pt]. Laat door berekening (met de GR) en redenering zien dat de bovenstaande functie voldoet aan de eisen voor een kansdichtheidsfunctie.

De eerste eis voor een kansdichtheidsfunctie is dat de functiewaarden altijd $f(x) \ge 0$ zijn. Dat is duidelijk als $x < 0$ en als $x > 10$ omdat dan $f(x)$ =0. Als $0 \le x \le 10$ dan zijn zowel x^2 als $10 - x \ge 0$, dus daar is ook $f(x) \ge 0$. **1pt**
Verder moet de oppervlakte onder de grafiek van $f(x)$ (de totale kans) gelijk zijn aan 1:

$$\int_{-\infty}^{\infty} f(x)dx = \int_0^{10} \frac{3}{2500}x^2(10-x)dx = \text{FnInt}\left(\frac{3}{2500}x^2(10-x), x, 0, 10\right) = 1. \qquad \textbf{3pt}$$

1b [8pt]. Bereken de verwachtingswaarde en de standaarddeviatie van het behaalde cijfer.

$$E(\underline{x}) = \int_{-\infty}^{\infty} \boldsymbol{x} f(x)dx = \int_{0}^{10} \frac{3}{2500}\boldsymbol{x^3}(10-x)dx = \text{FnInt}\left(\frac{3}{2500}x^3(10-x), x, 0{,}10\right) = 6.$$

3pt

$$\text{Var}(\underline{x}) = \int_{-\infty}^{\infty} (\boldsymbol{x}-\boldsymbol{6})^{\boldsymbol{2}} f(x)dx = \text{FnInt}\left((x-6)^2 \frac{3}{2500}x^2(10-x), x, 0{,}10\right) = 4. \quad \textbf{3pt}$$

$$\sigma = \sqrt{\text{Var}(\underline{x})} = 2. \qquad \textbf{2pt}$$

1c [6pt]. Bereken het percentage studenten dat naar verwachting slaagt voor dit tentamen, d.w.z. het percentage studenten dat een 5,5 of hoger haalt.

De kans dat een student slaagt voor het tentamen is

$$P(\underline{x} \geq 5{,}5) = \int_{5{,}5}^{10} \frac{3}{2500}x^2(10-x)dx = \text{FnInt}\left(\frac{3}{2500}x^2(10-x), x, 5.5{,}10\right) = 0{,}60902.$$

$\mathbf{2+2+2pt}$

Het slaagpercentage is dus 60,90%.

1d [6pt]. Bereken met het antwoord van 1c de kans dat van 90 studenten minstens de helft niet slaagt voor het tentamen. Als je je antwoord van 1c niet vertrouwt, ga dan in je berekening uit van een slaagpercentage van 60%.

De slaagkans per student is 0,6090. Dat betekent dat je te maken hebt met een binomiale verdeling. **2pt**

De kans dat van 90 studenten er minstens 45 niet slagen is de kans dat er hoogstens $90-45=45$ **wel** slagen:

$$P(\underline{k} \leq 45) = \text{binomcdf}(n=90, p=0.6090, k=45) = 0{,}02311. \ \mathbf{1+3+2pt}$$

Of , andere manier: De kans dat van 90 studenten er minstens 45 niet slagen is 1 min de kans dat er maximaal 44 niet slagen (gebruik de kans op niet slagen: $p = 1 - 0.6090 = 0{,}391$):

$$1 - \text{binomcdf}(n=90, p=1-0.6090, k=44) = 0{,}02311.$$

Voor 60% is het

$$P(\underline{k} \leq 45) = \text{binomcdf}(n=90\,, p=0.60\,, k=45) = 0{,}03465.$$

1e [6pt]. Bereken met de antwoorden van 1b de kans dat het gemiddelde cijfer van 90 studenten minstens 5,5 is. Als je je antwoord van 1b niet vertrouwt, ga dan in je berekening uit van een slaagpercentage van $\mu = 5{,}9$ en een standaarddeviatie van $\sigma = 1{,}9$.

Het **gemiddelde** tentamencijfer van 90 studenten gedraagt zich volgens de centrale limietstelling als normaal verdeeld met $\mu = 6$ en $\sigma = \frac{2}{\sqrt{90}} = 0{,}2108$, **2pt**
dus de kans dat het gemiddelde minimaal 5,5 is, is

$$\text{normalcdf}\left(5.5\,, 10^{10}\,, \mu = 6\,, \sigma = 2/\sqrt{90}\right) = 0{,}99115 \qquad \mathbf{4pt}$$

Voor de waarden $\mu = 5{,}9$ en $\sigma = 1{,}9$ is dat

$$\text{normalcdf}\left(5.5\,, 10^{10}\,, \mu = 5.9\,, \sigma = 1.9/\sqrt{90}\right) = 0{,}97710$$

Opgave 2 (30 punten)

De ICT helpdesk van de faculteit in Den Helder wordt door één persoon bemand. Elke werkdag (8 uur) wordt er gemiddeld 12 keer een beroep op de helpdesk gedaan.

2a [6pt]. Bereken met behulp van de Poissonverdeling de kans dat er in een bepaald uur 0, 1, 2, 3 of meer dan 3 calls zijn en vul je resultaten in in kolom 2 van de tabel van het antwoordmodel.

$\mu = \frac{12}{8} = 1{,}5$ calls per uur ($\mu = \frac{8}{12}$ of $\mu = 12$: **-2pt**)
0 calls: $p = \text{poissonpdf}(1.5{,}0) = 0{,}2231$
1 call: $p = \text{poissonpdf}(1.5{,}1) = 0{,}3347$
2 calls: $p = \text{poissonpdf}(1.5{,}2) = 0{,}2510$
3 calls: $p = \text{poissonpdf}(1.5{,}3) = 0{,}1255$
$>$ 3 calls: $p = 1 - \text{poisson}\mathbf{cdf}(1.5{,}3) = 0{,}0656$ $\mathbf{1 + 1 + 1 + 1 + 2pt}$

$\boldsymbol{k}$	$P(\underline{k} = k)$	t **minuten over**	**Kans geen tijd over**	$p \cdot t$
0	0,2231			
1	0,3347			
2	0,2510			
3	0,1255			
≥4	0,0656			
totaal				

De duur van de calls is normaal verdeeld met een gemiddeld van 22 minuten en een standaarddeviatie van 10 minuten. Bij een gegeven aantal calls in een uur kan worden bepaald wat de gemiddelde totale duur van deze calls is en wat de bijbehorende standaarddeviatie is (rekenregel of centrale limietstelling).

2b [6pt]. Bereken voor 0, 1, 2, 3 en meer dan 3 calls in een uur hoeveel minuten er in dat uur gemiddeld overblijven waarin de helpdesker niet bezig is met het oplossen van calls, onder de aanname dat er van het vorige uur geen calls meer afgewerkt hoeven te worden. Vul de resultaten in in kolom 3 van de tabel van het antwoordmodel. $\mathbf{1+1+1+1+2pt}$

$\boldsymbol{k}$	$\boldsymbol{P(\underline{k}=k)}$	**$\boldsymbol{t}$ minuten over**	**Kans geen tijd over**	$\boldsymbol{p \cdot t}$
0	0,2231	60		
1	0,3347	38		
2	0,2510	16		
3	0,1255	0 (-6)		
≥4	0,0656	0 (-28)		
totaal				

2c 6[pt]. Bereken voor 0, 1, 2, 3 en 4 calls in een uur de kans dat het hele uur wordt besteed aan het oplossen van deze calls. Vul de resultaten in in kolom 4 van de tabel van het antwoordmodel.

0 calls: $p = 0$
1 call: $p = \text{normalcdf}(60, 10^{10}, 22{,}10) = 0{,}00007237$
2 calls: $p = \text{normalcdf}\left(60, 10^{10}, 44{,}10\sqrt{2}\right) = 0{,}1289$
3 calls: $p = \text{normalcdf}\left(60, 10^{10}, 66{,}10\sqrt{3}\right) = 0{,}6355$
4 calls: $p = \text{normalcdf}\left(60, 10^{10}, 88{,}10\sqrt{4}\right) = 0{,}9192$
Het gaat om de som van de tijden, dus de standaarddeviatie moet worden vermenigvuldigd met de wortel uit het aantal tijden dat is opgeteld. $\mathbf{1+1+1+1+2pt}$

$\boldsymbol{k}$	$\boldsymbol{P(\underline{k}=k)}$	**$\boldsymbol{t}$ minuten over**	**Kans geen tijd over**	$\boldsymbol{p \cdot t}$
0	0,2231	60	0	
1	0,3347	38	0,00007237	
2	0,2510	16	0,1289	
3	0,1255	0	0,6355	
≥4	0,0656	0	0,9192	
totaal				

2d [6pt]. Bereken de verwachtingswaarde van het aantal minuten dat de helpdesker in een bepaald uur niet bezig is met het afwerken van calls onder de aanname dat er geen werk van het vorige uur ligt. Gebruik hiervoor kolom 5 van de tabel van het antwoordmodel.

k	$P(\underline{k} = k)$	t minuten over	Kans geen tijd over	$p \cdot t$
0	0,2231	60	0	13,386
1	0,3347	38	0,00007237	12,7186
2	0,2510	16	0,1289	4,016
3	0,1255	0	0,6355	0
≥4	0,0656	0	0,9192	0
totaal				30,1206

Neem het product van de tijd die over is (kolom 3) en de kans daarop (kolom 2)

$\mathbf{1 + 1 + 1 + 1 + 2pt}$

2e [2pt]. Leg uit waarom het antwoord van 2d niet klopt met wat je zou verwachten op grond van de gemiddelde waarde: $60 - 1{,}5 \cdot 22 = 27$ minuten.

Gemiddeld heeft de helpdesk 12/8=1,5 calls per uur, dus er blijven in een uur gemiddeld $60 - 1{,}5 \cdot 22 = 27$ minuten over. Dat dit minder is dan 30,1206 komt omdat in de berekening in 2d niet de tijd is meegenomen dat de helpdesker nog werkt aan calls van het vorige uur. $\mathbf{2pt}$

2f [4pt]. Bereken met behulp van de negatief exponentiële verdeling de kans dat na een call de volgende call tussen 30 en 60 minuten daarna komt.

Met de kansfunctie $F(t) = 1 - e^{-\mu t}$van de negatief exponentiele verdeling is dit

$$F(60) - F(30) = (1 - e^{-1{,}5 \cdot 1}) - (1 - e^{-1{,}5 \cdot 0{,}5}) = e^{-1{,}5 \cdot 0{,}5} - e^{-1{,}5 \cdot 1} = 0{,}2492$$

(neem de tijd en de μ in uren) $\mathbf{2 + 1 + 1pt}$

Opgave 3 (20 punten)

De bloedbank Sanquin regelt in Nederland de afname van bloed bij donoren. Belangrijk bij bloedtransfusie zijn de bloedfactoren. De belangrijkste zijn A en B, die hebben te maken met twee soorten suikers die aanwezig kunnen zijn op de bloedcellen. Iemand kan type O hebben (de factoren A en B ontbreken), type A (alleen factor A aanwezig), type B (alleen type B aanwezig) of AB (beide factoren A en B aanwezig). Daarnaast is er de bloedfactor Rhesus D (RhD) op de bloedplaatjes die positief of negatief kan zijn (aanwezig of niet). In de onderstaande tabel zie je de percentages van de bloedfactoren, zoals die in Nederland voorkomen (Sanquin, 2022)

	O	**A**	**B**	**AB**	**Totaal**
RhD-	6,8%	6,4%	1,3%	0,5%	15,0%
RhD+	38,2%	36,6%	7,7%	2,5%	85,0%
Totaal	45,0%	43,0%	9,0%	3,0%	100%

3a [5pt]. Bereken de kans dat iemand's bloed minstens één van de factoren A of B bevat.

$0{,}43 + 0{,}09 + 0{,}03 = 0{,}55.$ $\mathbf{3 + 2pt}$

3b [5pt]. Bereken de kans dat iemand bloed heeft met minstens één van de factoren A of B in zijn bloed, als bekend is dat hij/zij RhD- is.

Conditionele kans:

$$\frac{P((A \text{ of } B) \text{ en } RhD+)}{P(\,RhD+)} = \frac{0{,}064+0{,}013+0{,}005}{0{,}15} = 0{,}5467 \qquad \mathbf{3 + 2pt}$$

Niet ieder type bloed geschikt is geschikt voor transfusie bij ieder patiënt. Iedereen is allergisch voor factoren A, B of RhD die hij/zij zelf niet heeft. In de onderstaande tabel zie je voor welke bloedgroepen antistoffen aanwezig zijn.

Bloedgroep	O	A	B	AB	RhD+	RhD-
Antistoffen voor	A en B	B	A			RhD

Iemand met bloedgroep A- (factor A en geen RhD) kan bijvoorbeeld geen bloed krijgen dat factoren B of RhD bevat, dus alleen bloedgroep O of A.

3c [4pt]. Bereken de kans dat een willekeurige persoon bloed kan ontvangen van een willekeurige andere persoon. Beschouw daarvoor alleen de factoren A en B, maar niet RhD. Maak eerst een overzicht van welke bloedgroepen een persoon met bloedgroep O, A, B of AB kan ontvangen. Gebruik hiervoor de tabel in het antwoordmodel.

Bloedgroep	**Kans hierop**	**Kan ontvangen**	**Kans hierop**	**Kans totaal**
O	0,45	O	0,45	0,2025
A	0,43	O, A	0,88	0,3784
B	0,09	O, B	0,54	0,0486
AB	0,03	O, A, B, AB	1,00	0,0300
Som				0,6595

3d [6pt]. Bereken de kans dat in een groep van acht willekeurige personen alle acht combinaties van bloedfactoren voorkomen (O, A, B, AB alle vier met RhD+ of RhD-).

$8! \cdot 0{,}068 \cdot 0{,}064 \cdot 0{,}013 \cdot 0{,}005 \cdot 0{,}382 \cdot 0{,}366 \cdot 0{,}077 \cdot 0{,}025 = 3{,}0697 \cdot 10^{-5}$.

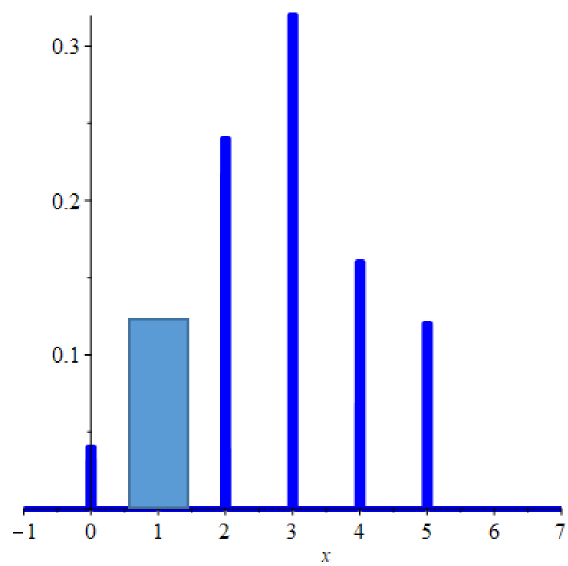
6pt (8! Vergeten: -2pt)

Opgave 4 (20 punten)

Een discrete kansvariabele $\underline{k}$ kan worden benaderd door een continue variabele $\underline{x}$ waarbij de staafjes in het staafdiagram van de kansfunctie van $\underline{k}$ worden vervangen door balkjes ter breedte 1, zie het voorbeeld in de bijgaande figuur. In het voorbeeld is $P(\underline{k} = 1) = 0{,}12$ en het staafje ter hoogte 0,12 is vervangen door een balkje van $x = 0{,}5$ tot $x = 1{,}5$ met breedte 1 en hoogte 0,12. De oppervlakte van dit balkje is 0,12. Daardoor is

$$P(0{,}5 \leq \underline{x} \leq 1{,}5) = P(\underline{k} = 1)$$

waardoor de (aansluitende) balkjes de grafiek van een kansdichtheidsfunctie voor $\underline{x}$ vormen.



We gaan dit toepassen voor de discrete uniforme kansverdeling op de waarden 1, 2, 3, 4 en 5. Daarvoor geldt

$P(\underline{k} = k) = 0{,}20$ als $k = 1, 2, 3, 4$ of 5, en $P(\underline{k} = k) = 0$ voor andere waarden van k.

Let op: dit is niet het voorbeeld van het plaatje!

4a [6pt]. De zojuist beschreven benaderende variabele $\underline{x}$ heeft een uniforme kansdichtheidsfunctie tussen de waarden a en b. Bepaal de waarden van a en b, en ook de maximale en de minimale hoogte van de grafiek van de kansdichtheidsfunctie van $\underline{x}$.

Het meest linkse balkje begint bij $a = 1 - 0{,}5 = 0{,}5$ en het meest rechtse eindigt bij $b = 5 + 0{,}5 = 5{,}5$. De hoogte van de grafiek (alle balkjes zijn even hoog) is 0,2 tussen 0,5 en 5,5 en 0 daarbuiten, dus de maximale hoogte is 0,2 en de minimale hoogte is 0. **2+2+1+1pt**

4b [6pt]. Bereken de gemiddelde waarde en de standaarddeviatie van $\underline{k}$. Gebruik niet de GR, hooguit ter controle.

De gemiddelde waarde van $\underline{k}$ is $\mu = 1 \cdot 0.2 + 2 \cdot 0.2 + 3 \cdot 0.2 + 4 \cdot 0.2 + 5 \cdot 0.2 = 3$. **2pt**
De variantie van $\underline{k}$ is

$$\mathrm{Var}(\underline{k}) = E\left(\left(\underline{k} - \mu\right)^2\right) =$$
$$= (1-3)^2 \cdot 0.2 + (2-3)^2 \cdot 0.2 + (3-3)^2 \cdot 0.2 + (4-3)^2 \cdot 0.2 + (5-3)^2 \cdot 0.2 = 2.$$

De standaarddeviatie van $\underline{k}$ is $\sigma = \sqrt{\mathrm{Var}(\underline{k})} = \sqrt{2} = 1{,}4142.$ **4pt**

4c [6pt]. Bepaal de gemiddelde waarde en de standaarddeviatie van $\underline{x}$.
Een uniforme verdeling op het interval $[0{,}5\,,5{,}5]$ heeft als gemiddelde waarde $\mu = \frac{0{,}5+5{,}5}{2} = 3$ en standaarddeviatie $\sigma = \frac{5{,}5-0{,}5}{\sqrt{12}} = 1{,}4434$ (zie formuleblad). **3+3pt**

4d [2pt]. Leg uit waarom de standaarddeviatie van $\underline{x}$ groter is dan die van $\underline{k}$.
De staafjes van $\underline{k}$ worden uitgesmeerd tot balkjes van $\underline{x}$, waardoor de verdeling breder wordt, waardoor de standaarddeviatie iets groter is.