Tentamen Statistiek deel 1 MBW/KW (finale-finale kans)

Afdeling: Propedeuse MBW/KW 2022-2023 en 2021-2022
Examinator: Dr. J.B.M. Melissen, T. Zijlstra MSc.
Datum: vrijdag 28 juli 2023 09:00 – 12:00, duur tentamen: 3 uur

1. **Alle antwoorden moeten gemotiveerd worden**!
2. Rond eindantwoorden (kommagetallen) af op *vier* decimalen, tenzij anders vermeld.
3. Boeken, reader en aantekeningen mogen worden geraadpleegd.
4. De aanwezigheid van *communicatieapparatuur* is niet toegestaan.
5. 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!
6. **De opgaven dienen na afloop van het tentamen ingeleverd te worden.**

Dit tentamen bestaat uit vier opgaven van elk 25 punten. Score = Puntentotaal/10

Opgave 1 [25pt].

Voor een kansvariabele $\underline{k}$ is de kansfunctie gegeven in de volgende tabel:

k	0	1	2	3	4	5	6	7	8	9
$P(\underline{k} = k)$	0,04	0,03	0,12	0,24	0,18	0,17	0,12	0,00	0,06	0,04

1a [4pt]. Toon aan dat $f(k) = P(\underline{k} = k)$ een goed gedefinieerde kansfunctie is.

De kansvariabele $\underline{k}$ is discreet, dus als je de getallen in de onderste rij bij elkaar optelt moet dit optellen tot 1. Inderdaad geldt dat $0{,}04 + 0{,}03 + 0{,}12 + 0{,}24 + 0{,}18 + 0{,}17 + 0{,}12 + 0{,}00 + 0{,}06 + 0{,}04 = 1$. Tevens geldt dat alle kansen groter dan of gelijk aan 0 zijn, waarmee ook aan de tweede voorwaarde is voldaan.

1b [7pt]. Bereken de verwachtingswaarde $\mu(\underline{k})$ en de standaarddeviatie $\sigma(\underline{k})$.

$$\mu(\underline{k}) = \sum_k kP(\underline{k} = k) = 0 * 0{,}04 + 1 * 0{,}03 + \cdots + 9 * 0{,}04 = 4{,}12$$

$$\begin{aligned}\sigma^2(\underline{k}) &= \sum_k \left(k - \mu(\underline{k})\right)^2 P(\underline{k} = k) \\ &= (0 - 4{,}12)^2 * 0{,}04 + (1 - 4{,}12)^2 * 0{,}03 + \cdots + (9 - 4{,}12)^2 * 0{,}04 \\ &= 4{,}2256\end{aligned}$$

$$\sigma(\underline{k}) = \sqrt{4{,}2256} \approx 2{,}0556$$

1c [4pt]. Bereken $P(1 < \underline{k} < 4)$.

$$P(1 < \underline{k} < 4) = P(2 \le \underline{k} \le 3) = P(\underline{k} = 2) + P(\underline{k} = 3) = 0{,}12 + 0{,}24 = 0{,}36$$

Een discrete uniforme kansverdeling wordt bepaald door een kleinste geheel getal a en een grootste gehele waarde b. Er geldt dan dat $f(k) = P(\underline{k} = k) = \frac{1}{b-a+1}$ (constant) voor alle k met $a \le k \le b$ en $f(k) = 0$ voor alle overige waarden van k.

Voor deze discrete uniforme kansverdeling geldt dat $\mu = \frac{a+b}{2}$ en $\sigma = \sqrt{\frac{(b-a+1)^2-1}{12}}$

1d [7pt]. Men wil de uniforme verdeling vinden die het best past bij de kansverdeling uit de tabel en doet dit door de waarden van a en b zo te bepalen dat de gemiddelde waarden van de twee verdelingen aan elkaar gelijk zijn en ook voor de standaarddeviaties.
Bereken uit de gelijkheid van de verwachtingswaarden de waarde van $a + b$ en uit de gelijkheid van de standaarddeviaties de waarde van $b - a$. Los hieruit vervolgens de waarden van a en b op (tel de twee vergelijkingen op en trek ze af) en rond beide waarden af op een geheel getal.

We willen waardes van a en b vinden waarvoor geldt dat:

1. De gemiddeldes hetzelfde zijn: $\frac{a+b}{2} = 4{,}12 \rightarrow a + b = 8{,}24$
2. De standaarddeviaties hetzelfde zijn:

$$\sqrt{\frac{(b-a+1)^2-1}{12}} = 2{,}0556 \rightarrow \frac{(b-a+1)^2-1}{12} = 4{,}2256 \rightarrow (b-a+1)^2 - 1$$
$$= 50{,}7072 \rightarrow (b-a+1)^2 = 51{,}7072 \rightarrow (b-a+1) = 7{,}1908$$
$$\rightarrow b - a = 6{,}1908$$

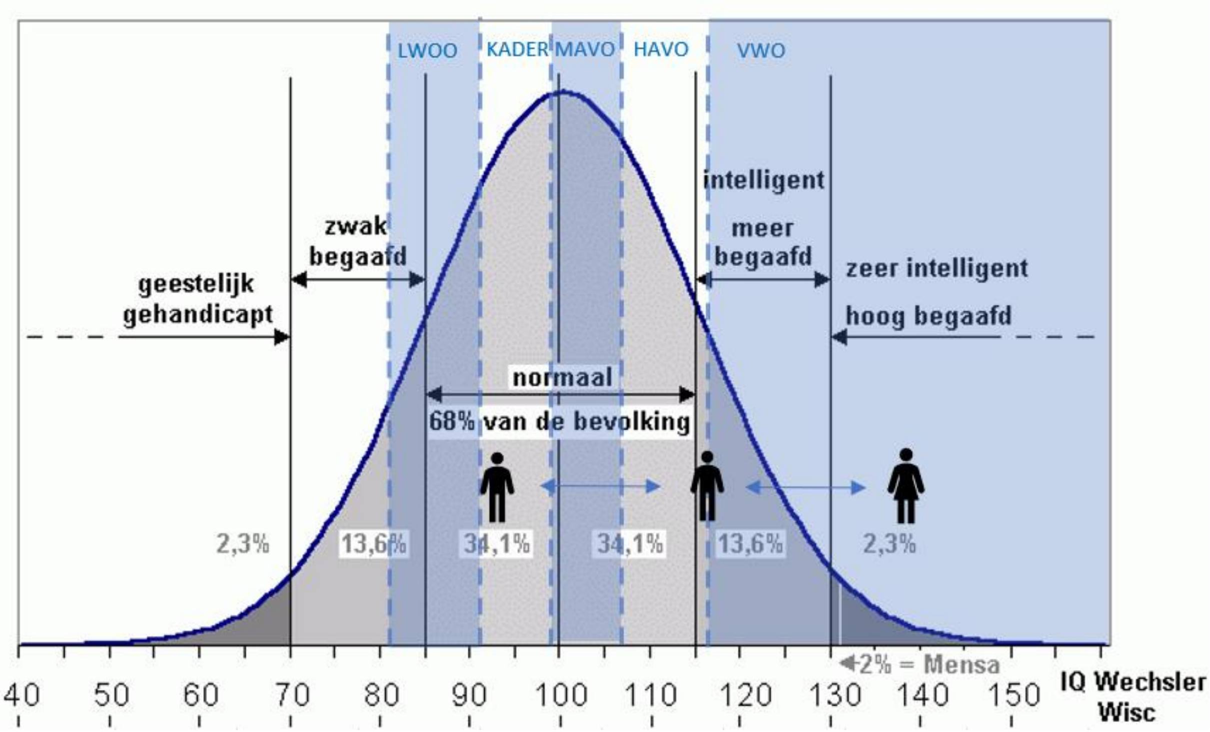
Deze vergelijkingen optellen geeft:

$$2b = 8{,}24 + 6{,}1908 = 14{,}4308 \rightarrow b = 7{,}2154 \rightarrow a + 7{,}2154 = 8{,}24 \rightarrow a = 1{,}0246$$

Na afronden krijgen we $a = 1$ en $b = 6$.

1e [3pt]. Bereken $P(\underline{k} = 5{,}12)$.
Merk op dat $\underline{k}$ een discrete kansvariabele is met alleen gehele getallen als mogelijke waardes. De kans dat $\underline{k}$ een niet-geheel getal aanneemt is dus gelijk aan 0.

Opgave 2 [25pt]



Het Intelligentie Quotiënt (IQ) is een getal dat een indicatie geeft van de intelligentie van een persoon. Er wordt aangenomen dat het IQ normaal is verdeeld met een gemiddelde van 100 en een standaarddeviatie van 15. Het IQ stijgt tot de leeftijd van 18 jaar en neemt daarna niet meer toe.
Voor het VWO wordt een gemiddelde IQ waarde gehanteerd van minimaal 116, voor Havo een waarde van 107, voor VMBO-tl/Mavo een minimum IQ van 100, voor VMBO-kl een waarde van 92 en het LWOO (LeerWeg Ondersteunend Onderwijs) vanaf de waarde 82.
We nemen aan dat alle kinderen het basisonderwijs afronden en dat hun IQ zich na het basisonderwijs gedraagt volgens de normale verdeling van het IQ.
Onder IQ=70 wordt gesproken van een geestelijke handicap, tussen 70 en 85 van zwak begaafd, tussen 85 en 115 van normaal begaafd, tussen 115 en 130 van meer begaafd en daarboven van hoog begaafd.

2a [3pt]. Wat is de kans dat een zwak begaafd kind een advies LWWO krijgt? Neem aan dat de IQ grenzen van hierboven strikt worden aangehouden.
Laat $\underline{x}$ de kansvariabele zijn van het IQ van een kind. Voor een zwakbegaafd kind geldt: $70 \leq \underline{x} \leq 85$. Verder krijgt een kind een LWWO advies als geldt $\underline{x} \geq 82$. We bekijken dus de kans $P\left(70 \leq \underline{x} \leq 85 \; en \; \underline{x} \geq 82\right) = P(82 \leq \underline{x} \leq 85)$. Omdat verder geldt $\mu = 100$ en $\sigma = 15$, geldt dat

$$P\left(82 \leq \underline{x} \leq 85\right) = normalcdf(82, 85, 100, 15) = 0{,}0436$$

2b [3pt]. Bereken de kans dat iemand die in het vervolgonderwijs (vanaf LWOO) terecht komt een VWO advies krijgt.
We willen de kans $P\left(\underline{x} \geq 116 \,\middle|\, \underline{x} \geq 82\right)$ berekenen.

Hiervoor geldt: $P(\underline{x} \geq 116 \mid \underline{x} \geq 82) = \frac{P(\underline{x} \geq 116 \text{ en } \underline{x} \geq 82)}{P(\underline{x} \geq 82)} = \frac{P(\underline{x} \geq 116)}{P(\underline{x} \geq 82)} =$
$\frac{normalcdf(116,10^{10},100,15)}{normalcdf(82,10^{10},100,15)} = \frac{0{,}1431}{0{,}8849} = 0{,}1617$

2c [3pt]. Leg uit waarom er ongeveer evenveel zwak begaafden als meer begaafden zijn.
De groep zwakbegaafden is gedefinieerd als de groep mensen met IQ tussen $\mu - 2\sigma$ en $\mu - \sigma$, terwijl de groep meerbegaafden tussen $\mu + \sigma$ en $\mu + 2\sigma$ ligt. Aangezien het IQ van mensen bij benadering een normale verdeling volgt (symmetrische verdeling!), zijn er ook bij benadering evenveel zwakbegaafden als meer begaafden.

2d [4pt]. Om te voorkomen dat personen met lage intelligentie gevaarlijke wapens hanteren of ermee in aanraking komen eist het Amerikaans Congres van het Pentagon (Code of Federal Regulations, 2015) dat geen rekruten worden aangenomen uit de 10% van de bevolking met het laagste IQ en slechts een beperkt aantal uit de laagste 10 tot 30%. Bereken welke IQ grenzen hierbij horen.

Voor de grenswaarde g die hoort bij het IQ van de laagste 10% moeten we kijken naar de vergelijking $P(\underline{x} \leq g) = 0.10$. Dit berekenen we door middel van $InvNorm(0.10, 100, 15) \approx 80.7767$. Dus de 10% van de bevolking met het laagste IQ heeft een IQ van maximaal deze waarde. Voor de laagste 30% hebben we dan de grens $InvNorm(0.30, 100, 15) \approx 92.1340$.

In het artikel “Are Blonds Really Dumb” van Jay Zagorsky (*Economics Bulletin*, Volume 36, Issue 1, pages 401-410, 2016) laat de auteur zien dat blonde vrouwen juist een hoger IQ hebben dan vrouwen met een andere haarkleur:

Table 1: IQ Categorized by Hair Color.

Type of Individual	Mean IQ	Standard Deviation	Median IQ	Percent of Group	Number Respondents
Blonde Hair White Women	103.2	12.8	102.7	20.7%	597
Brown Hair White Women	102.7	13.8	102.9	73.0%	2,205
Red Hair White Women	101.2*	13.2	100.5	3.8%	118
Black Hair White Women	100.5**	13.4	101.4	2.5%	77

2e [3pt]. Bereken het (gewogen) gemiddelde IQ van alle onderzochte vrouwen uit de tabel.

Het gewogen gemiddelde IQ van alle onderzochte vrouwen uit de tabel is gelijk aan
$$\frac{597 * 103{,}2 + 2205 * 102.7 + 118 * 101.2 + 77 * 100.5}{597 + 2205 + 118 + 77} = 102{,}6840$$

2f [4pt]. Bereken de standaarddeviatie in het gemiddelde IQ van 597 vrouwen uit deze groep (neem aan dat de standaarddeviatie in het IQ van één vrouw 15 is).

We bekijken eht gemiddelde van $\underline{x}_1, \ldots, \underline{x}_{597}$, waarbij elke $\underline{x}_i$ normaal verdeeld is met gemiddelde 102,8, en standaarddeviatie 15. Het gemiddelde van 597 vrouwen is dan normaal verdeeld met gemiddelde 102,6840 en standaarddeviatie $\frac{15}{\sqrt{597}} \approx 0{,}6139.$

2g [3pt]. Bereken met deze gegevens de kans dat het gemiddeld IQ van 597 vrouwen 103,2 of meer is, uitgaande van het in 2e berekende (gewogen) gemiddeld IQ van alle vrouwen.

We willen de kans $P(\overline{\underline{x}} \geq 103{,}2)$ berekenen, waarbij $\overline{\underline{x}}$ volgens de centrale limietstelling normaal verdeeld is met gemiddelde 102,6840 en standaarddeviatie 0,6139 (zoals berekend in 2f). Hiervoor vinden we

$$P(\overline{\underline{x}} \geq 103{,}2) = \text{normalcdf}(103{,}2\ , 10^{10}, 102.6840\ , 0.6139) = 0{,}2003$$

2h [2pt]. Wat kun je hieruit concluderen?

De kans dat deze waarde voor het gemiddelde IQ van blonde vrouwen wordt gevonden is vrij groot, waardoor deze ook wel aan het toeval kan worden toegeschreven. Als de kans heel klein was geweest, was het onwaarschijnlijk geweest dat het gemiddelde IQ van een blonde vrouw hetzelfde zou zijn als het gemiddelde IQ van alle vrouwen.

Opgave 3 (25 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.

Iemands bloed heeft minstens een van de factoren A of B als geldt dat die persoon bloedgroep A, B of AB heeft: de kans hierop is gelijk aan $0{,}43 + 0{,}09 + 0{,}03 = 0{,}55.$

$\mathbf{3 + 2pt}$

3b [5pt]. Bereken de kans dat iemand met RhD- bloed heeft met factoren B.

Conditionele kans:

$$\frac{P(B \text{ en } RhD-)}{P(\ RhD-)} = \frac{0{,}013+0{,}005}{0{,}15} = 0{,}36 \qquad \mathbf{3 + 2pt}$$

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)

3e [5pt]. Bereken de kans dat in een groep van acht willekeurige personen minstens de helft bloedgroep O heeft.
Het aantal personen $\underline{k}$ uit een groep van 8 personen dat bloedgroep O heeft is een kansvariabele die een binomiale verdeling volgt met $n = 8$ (aantal personen in de groep) en $p = 0{,}45$ (de kans dat een persoon bloedgroep O heeft). Minstens de helft bloedgroep O correspondeert dus met de kans $P(\underline{k} \geq 4) = 1 - P(\underline{k} \leq 3) = 1 - binomcdf(8;\ 0{,}45; 3) \approx 0{,}5230$.

Opgave 4 [25pt]

In een militair depot in oorlogstijd worden materialen aangevoerd door middel van 12-tonners in 20ft zeecontainers. Trucks keren terug met een lege container en materiaal uit de volle containers wordt gesorteerd en opgeslagen in het depot. Tegelijkertijd arriveren er lege viertonners die lading komen halen om naar locaties in de buurt van de operaties te vervoeren. Voor het gemak nemen we aan dat 12-tonners 12 ton lading aanvoeren en dat viertonners 4 ton lading afvoeren.
De 12-tonners arriveren met een gemiddelde van μ_{12} voertuigen per uur, van de viertonners arriveren er gemiddeld μ_4 per uur. De aankomsten van viertonners en van 12-tonners kunnen elke 24 uur met een eigen Poissonverdeling worden beschreven. Per 24 uur zijn deze waarden vast, maar elke dag kunnen de waarden veranderen.

4a [7pt]. Op een bepaalde dag (24 uur) arriveren er zeven 12-tonners en er arriveren gemiddeld $\mu_4 = 0{,}9$ viertonners per uur. Wat is de kans dat de geleverde lading van deze 24 uur door de arriverende viertonners kan worden afgevoerd? Houd hierbij geen rekening met transactietijd in het depot (uitladen, opslaan, orderpicken, inladen).
In deze 24 uur arriveert er 7×12 = 84 ton lading, hiervoor zijn 7×12/4 = 21 viertonners nodig.
2pt
De viertonners arriveren met een gemiddeld aantal van 0,9×24 = 21,6 viertonners per 24 uur. **2pt**
De kans dat er minstens 21 viertonners arriveren is **3pt**

$$P(\underline{k} \geq 21) = 1 - P(\underline{k} \leq 20) = 1 - \text{poissoncdf}(21.6\,, 20) = 0{,}5802$$

4b [5pt]. Er is net een viertonner binnengekomen. Bereken de kans dat de volgende viertonner binnen een uur binnenkomt. **5pt, onjuiste μ: -1pt**

Deze kans is $F(1) = 1 - e^{-0,9 \cdot 1} = \mathbf{0,5934}$.

Ofwel, voor één uur geldt: $P(\underline{k} \geq 1) = 1 - \text{Poissoncdf}(0.9\,,0) = \mathbf{0,5934}$

Niet helemaal juiste aanpak is: $\mathbf{poissonpdf(0.9\,,1)} = 0{,}365$ **max 3pt**

4c [8pt]. De commandant van het depot verwacht voor de volgende 24 uur gemiddeld $\mu_{12} = 0{,}35$ 12-tonners per uur. Hij wil met 99% zekerheid weten wat het maximale aantal 12-tonners is dat die dag kan arriveren. Bereken dat aantal.

Hij wil de grootste k weten zodanig dat $P(\underline{k} \leq k) \geq 0{,}99$. Dat betekent dat

$$\text{Poissoncdf}(0.35 * 24 = 8.4\,, k) \geq 0{,}99$$

De grootst mogelijke k is dan 16, want

$$\text{Poissoncdf}(0.35 * 24\,, 15) = 0{,}9875$$
$$\text{Poissoncdf}(0.35 * 24\,, 16) = 0{,}9941$$

4d [5pt]. De commandant wil vervolgens weten wat de minimale μ_4 voor die 24 uur moet zijn zodat hij met 99% zekerheid de aangevoerde lading ook weer kan afvoeren in het geval van de situatie in 4c. Bereken deze μ_4. Als je niet zeker bent van je antwoord in 4c, ga dan uit van 17 12-tonners.

Voor het afvoeren van de lading van 16 12 tonners zijn 48 viertonners nodig. De kleinste μ_4 waarvoor de kans minstens 99% is dat er minstens 48 viertonners zullen zijn voldoet aan

$$1 - \text{Poissoncdf}(24\mu_4\,, 47) = 0{,}99$$

De solver levert:

$$\mu_4 = 2{,}7321$$

(65,6 per 24 uur)