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

Afdeling: Propedeuse KW/MBW 2019-2020
Examinator: Dr. J.B.M. Melissen
Datum: 5 juni 2020, duur tentamen: 2 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. **Lever de antwoorden in op het geprinte antwoordformulier (zet je naam erop), de berekeningen en uitleg op gelinieerd papier.**
7. **De opgaven dienen na afloop van het tentamen ingeleverd te worden.**

Dit tentamen bestaat uit vier opgaven (30, 25, 25, 20 punten). Score = Puntentotaal/10

Opgave 1 (Totaal 30 punten)

In een landelijke enquête wordt aan een grote groep volwassenen gevraagd hoe vaak ze in hun leven zijn verhuisd. Dit aantal kan worden beschreven door een kansvariabele $\underline{k}$. Verder is ook gevraagd of de geënquêteerde van plan is om binnen een jaar te verhuizen. Dit aantal wordt beschreven door een kansvariabele $\underline{l}$ met $\underline{l} = 0$: geen verhuisplannen, $\underline{l} = 1$: wel verhuisplannen. De kansen worden hieronder gegeven:

	$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$P(\underline{l} = l)$
$l = 0$	0,009	0,012	0,017	0,035	0,032	0,007	0,112
$l = 1$	0,171	0,318	0,213	0,115	0,058	0,013	0,888
$P(\underline{k} = k)$	0,180	0,330	0,230	0,150	0,090	0,020	

1a [4pt]. Leg uit hoe je aan de kansen in de onderste regel van de tabel kunt zien dat de kansfunctie $f(k)$ van $\underline{k}$ goed gedefinieerd is (twee eigenschappen).

1a [4pt]. De onderste regel van de tabel is de kansfunctie $f(k)$ van $\underline{k}$ want $f(k) = P(\underline{k} = k)$

1. De kansen liggen tussen 0 en 1 — 2pt
2. De som van de kansen is 1 — 2pt

1b [4pt]. Geef in een tabel de cumulatieve kansfunctie $F(k) = P(\underline{k} \leq k)$.

k	0	1	2	3	4	5
$P(\underline{k} \leq k)$	0,18	0,51	0,74	0,89	0,98	1,00

1c [8pt]. Bereken de verwachtingswaarde, de variantie en de standaarddeviatie van $\underline{k}$. Maak geen gebruik van het statistisch menu van de grafische rekenmachine (hooguit ter controle).

$E(\underline{k}) = 0 \cdot 0{,}18 + 1 \cdot 0{,}33 + 2 \cdot 0{,}23 + 3 \cdot 0{,}15 + 4 \cdot 0{,}09 + 5 \cdot 0{,}02 = 1{,}70$ **3pt**

$E(\underline{k}^2) = 0^2 \cdot 0{,}18 + 1^2 \cdot 0{,}33 + 2^2 \cdot 0{,}23 + 3^2 \cdot 0{,}15 + 4^2 \cdot 0{,}09 + 5^2 \cdot 0{,}02 = 4{,}54$ **3pt**

$Var(\underline{k}) = E(\underline{k}^2) - E(\underline{k})^2 = 1{,}65$ **2pt**

$\sigma = \sqrt{Var(\underline{k})} = 1{,}2845$ **2pt**

1d [6pt]. Wat is de kans dat een geënquêteerde **meer dan drie** keer is verhuisd? Bereken de kans dat er onder 125 ondervraagden minstens 16 personen zijn die meer dan 3 keer zijn verhuisd (Gebruik geen benadering).

$P(\underline{k} > 3) = P(\underline{k} = 4) + P(\underline{k} = 5) = 0{,}11.$ **2pt**

De kans dat er onder 125 ondervraagden minstens 16 personen zijn die meer dan 3 keer zijn verhuisd is

$1 - \text{binomcdf}(125\,, 0.11\,, 15) = 0{,}2988$ **4pt**

1e [8pt]. Bereken de kansen

$P(\underline{k} < 2), P(\underline{l} = 1), P(\underline{k} < 2 \text{ én } \underline{l} = 1), P(\underline{k} < 2 \text{ óf } \underline{l} = 1), P(\underline{k} < 2 \mid \underline{l} = 1), P(\underline{l} = 1 \mid \underline{k} < 2)$

$P(\underline{k} < 2) = 0{,}18 + 0{,}33 = 0{,}51$ **1pt**

$P(\underline{l} = 1) = 0{,}888$ **1pt**

$P(\underline{k} < 2 \text{ én } \underline{l} = 1) = 0{,}171 + 0{,}318 = 0{,}489$ **1pt**

$P(\underline{k} < 2 \text{ óf } \underline{l} = 1) = (0{,}009 + 0{,}012) + (0{,}213 + 0{,}115 + 0{,}058 + 0{,}013) = 0{,}420$ **1pt**

$P(\underline{k} < 2 \mid \underline{l} = 1) = \frac{P(\underline{k}<2 \text{ én } \underline{l}=1)}{P(\underline{l}=1)} = \frac{0{,}489}{0{,}888} = 0{,}5507$ **2pt**

$P(\underline{l} = 1 \mid \underline{k} < 2) = \frac{P(\underline{l}=1 \text{ én } \underline{k}<2)}{P(\underline{k}<2)} = \frac{0{,}489}{0{,}51} = 0{,}9588$ **2pt**

Opgave 2 (Totaal 25 punten) De reistijd van sergeant Buldermans van zijn woonplaats naar zijn kazerne kan worden beschouwd als een normaal verdeelde kansvariabele $\underline{t}$ met μ = 25 minuten en σ = 4 minuten (enkele reis).

2a. [5pt] Bereken de kans dat hij op een willekeurige ochtend meer dan 35 minuten reistijd nodig heeft.

$$P(\underline{t} > 35) = \text{normalcdf}(35\,,\ 10^{10}\,,\ 25\,,\ 4) = 0{,}0062.$$

2b. [5pt] Bereken hiermee de kans dat hij in een jaar (240 werkdagen) minstens één keer meer dan 35 minuten reistijd nodig heeft.

Elke ochtend geldt dat de reistijd meer dan 35 minuten is met (succes)kans 0,0062. Volgens de binomiale verdeling is de kans dat dit minstens één keer per 240 ochtenden gebeurt:

$$P(\underline{k} \geq 1) = 1 - \text{binomcdf}(240\,, 0.0062\,, 0) = 0{,}7752.$$

Als je middagen ook meeneemt met dezelfde kans dan is het

$$1 - \text{binomcdf}(480\,, 0.0062\,, 0) = 0{,}9495.$$

2c. [5pt] Hoe groot is de kans dat gedurende een week (5 dagen) de **reistijd elke ochtend** meer dan 27 minuten is? Neem aan dat de reistijden onafhankelijk zijn en dat de verdeling elke dag hetzelfde is.

Voor één dag is $P(\underline{t} > 27) = \text{normalcdf}(27,\ 10^{10}\ ,\ 25\ ,\ 4) = 0{,}308537$. De kans dat dit op vijf dagen gebeurt is $0{,}308537^5$ =0,002796.

2d. [5pt] Hoe groot is de kans dat in een week (5 dagen) de **gemiddelde reistijd** 's ochtends meer dan 27 minuten is? Neem weer aan dat de reistijden onafhankelijk zijn en dat de verdeling elke dag hetzelfde is.

De gemiddelde reistijd is normaal verdeeld met μ = 25 minuten en σ = $4/\sqrt{5}$ minuten. De kans dat het gemiddelde groter is dan 27 minuten is

$$P(\underline{t} > 27) = \text{normalcdf}\left(27,\ 10^{10}\ ,\ 25\ ,\ 4/\sqrt{5}\right) = 0{,}1318$$

2e. [5pt] Het vertrektijdstip is gemiddeld 7:32 met $\sigma = 4$ minuten. Hoe groot is de kans dat hij later dan 8.00 aankomt?

Aankomst = vertrek + reistijd $\sim N(7{:}32\ ,\ 4) + N(0{:}25\ ,\ 4) = \mathrm{N}\left(7{:}57\ ,\ \sqrt{4^2 + 4^2}\right)$ is normaal verdeeld.

$$\begin{aligned} P(\underline{t} > 8{:}00) &= \text{normalcdf}(8{:}00\ ,\ 10^{10}\ ,\ 7{:}57\ ,\ 5.6569) = (\text{t.o.v. } 7{:}00) \\ &= \text{normalcdf}(60\ ,\ 10^{10}\ ,\ 57\ ,\ 5.6569) = 0{,}7021. \end{aligned}$$

BIVAK


EN DAN PAKKEN WE NU ALLEMAAL ONS ETENSPANNETJE, CANTEEN CUP EN NOODRANTSOEN
TUUT TUUT
UHH... DACHT HET NIET, SERGEANT
ETEN BEZORGD .NL
B.

Opgave 3 (Totaal 25 punten) In het kader van verscherpte grensbewaking in COVID-19 tijden voert de Koninklijke Marechaussee op lokale wegen in de grensregio's 100%-controles uit. Voor een bepaalde dag wordt een inzet gepland waarbij gedurende drie uur alle voertuigen op een vanwege veiligheidsoverwegingen niet nader aan te duiden weg in Limburg gecontroleerd moeten worden. Gebaseerd op gegevens van de RDW is de verwachting dat zich gedurende de controle gemiddeld 26,5 voertuigen per uur aandienen en er wordt aangenomen dat dit volgens een Poissonverdeling zal gebeuren. Uit eerdere inzetten is gebleken dat de benodigde tijd per voertuig uniform is verdeeld tussen 3 en 15 minuten. Elke controle wordt uitgevoerd door een team van twee marechaussees.

Preventief fouilleren van foutparkeerders op Eindhoven Airport
(Waarom draagt de linkshandige marechaussee zijn horloge links?).

3a. [4pt] Bereken hoeveel teams er gemiddeld nodig zijn om deze controles uit te voeren.

3a. Gemiddeld 26,5 voertuigen per uur, gemiddelde controletijd $(3 + 15)/2 = 9$ minuten per voertuig, dus gemiddeld $26{,}5 * 9 = 238{,}5$ minuten controletijd per uur nodig. 238,5/60 = 3,975 teams nodig, dus 4.

3b. [6pt] Bereken de kans dat zich gedurende de controletijd van 3 uur meer dan 100 voertuigen aandienen.

3b. Gedurende 3 uur is het verwachte aantal voertuigen $\mu = 26{,}5*3 = 79{,}5$. Volgens de Poissonverdeling is

$$P(\underline{k} > 100) = 1 - \text{poissoncdf}(\mu = 79.5\,, k = 100) = 0{,}01132$$

3c. [6pt] Neem aan dat elk team direct van start kan gaan en vervolgens continu bezig is met controles. Bereken hoeveel tijd een team nodig heeft om met 97% zekerheid 20 controles te kunnen uitvoeren. Maak hiervoor gebruik van een geschikte benadering op grond van de centrale limietstelling en de parameters van de uniforme verdeling.

3c. De tijd die nodig is wordt bepaald door de som van 20 tijden getrokken uit de uniforme verdeling tussen 3 en 15 minuten. Daarvan is de gemiddelde waarde (3+15)/2 = 9 minuten en de standaarddeviatie is $\frac{(15-3)}{\sqrt{12}} = 3{,}4641$ per controle.
Volgens de centrale limietstelling geldt voor de totale tijd van 20 controles een normale verdeling met $\mu = 20 \cdot 9 = 180$ minuten en $\sigma = \sqrt{20} \cdot 3{,}4641 = 15{,}4919$.
De tijd waarin de controles met 97% zekerheid kunnen worden gedaan is
$z = \text{invNorm}(0.97\,, 180\,, 15.4919) = 209{,}1371$. De gevraagde tijd is 209,1372 minuten = 3,3856 uur.
Ofwel: $\mu + z\sigma$, waarin $z = \text{invNorm}(0.97) = 1{,}8808$. De gevraagde tijd is 180 + 1,8808 X 15,4919 = 209,1372 minuten = 3,3856 uur.

3d. [4pt] Uit de voorgaande berekeningen volgt dat een team met 97% zekerheid 20 voertuigen kan controleren binnen 3,5 uur, en dat er met 98,9% zekerheid binnen 3 uur niet meer dan 100 voertuigen zijn om te contoleren.
Bereken met behulp van deze gegevens de kans dat vijf teams voldoende zijn om binnen 3,5 uur de controles uit te voeren, dus (voor het gemak): de kans dat er in 3 uur niet meer dan 100 voertuigen komen én dat elk van de teams daarvan 20 auto's binnen 3,5 uur kan controleren. Dit is een worst-case inschatting, in werkelijkheid is de kans groter.

3d. 0,989 * $0{,}97^5$ = 0,8493.

3e. [5pt] Hoe groot is de kans dat het, na het aanhouden van een voertuig, minimaal 5 minuten duurt voordat het volgende voertuig arriveert? Maak gebruik van de negatief exponentiële verdeling.

3e. De tijd tussen twee voertuigen wordt beschreven door exp(λ), waarbij de tijd in minuten is en λ het aantal verwachte voertuigen per minuut, dus $\lambda = \frac{26{,}5}{60} = 0{,}4417$. De kansfunctie is

$$F(t) = P(\underline{t} \le t) = 1 - e^{-\lambda t}$$

$$P(\underline{t} > 5) = 1 - P(\underline{t} \le 5) = 1 - F(5) = 1 - (1 - e^{-0{,}4417 \cdot 5}) = e^{-0{,}4417 \cdot 5} = 0{,}10987 \ldots$$

Of, als je moeilijk wilt doen met de kansdichtheidsfunctie:

$$f(t) = \lambda e^{-\lambda t}$$

$$P(\underline{t} > 5) = 1 - P(\underline{t} \le 5) = 1 - \int_0^5 \lambda e^{-\lambda t} dt =$$

$$1 - \text{fnInt}(0.4417 * e^{-0.4417 * t}\,, t\,, 0\,, 5) = 0{,}10987$$

Opgave 4 (totaal 20 punten)

9600 tests afgenomen sinds openen corona-nummer GGD

Via het landelijke nummer voor een coronatest, maandagmorgen geopend, zijn al bijna 16.000 afspraken gemaakt. Er zijn intussen bijna 9600 tests afgenomen en 94 mensen positief bevonden, maar nog niet alle uitslagen zijn bekend. (Bron: RTL Nieuws 3 juni 2020)

4a. [2pt] Bereken op grond van deze gegevens een schatting voor de fractie Nederlanders die op het moment besmet zijn met het coronavirus.

4a. 94/9600 = 0,009792.

4b. [2pt] Bereken op grond van deze fractie hoeveel van de 17,3 miljoen Nederlanders naar schatting besmet zijn met het Coronavirus.

4b. 94/9600 * 17300000 = 169.399.

4c. [2pt] Is de waarde uit 4b. waarschijnlijk een ondergrens, een bovengrens of is dat niet duidelijk? Geef hiervoor argumenten.

4c. Niet duidelijk. Nog niet alle uitslagen zijn bekend, dus het percentage kan in werkelijkheid hoger zijn. Het gaat om mensen die zelf om een toets vragen, dus ook mensen met klachten, dus het percentage in deze groep kan juist hoger zijn dan in de rest van Nederland, maar als er veel gezonde mensen bij zijn, die alleen maar duidelijkheid willen kan het juist weer lager zijn dan in de rest van Nederland. De ernstige gevallen zitten hier niet bij, die liggen al in het ziekenhuis. Dan gaat het ook nog om een steekproef, waardoor het werkelijke percentage kan afwijken.

Is echt maar 5,5 procent van de Nederlanders besmet geweest met corona?

Zo'n 5,5 procent van de Nederlandse bloeddonoren heeft antistoffen aangemaakt tegen het coronavirus. Dat werd vanochtend duidelijk. (Bron: RTL Nieuws 3 juni 2010)

4d. [5pt] Bereken op grond van deze 5,5% hoeveel van de 80 cadetten, adelborsten en burgers die dit tentamen doen besmet is geweest met het coronavirus. Bereken ook de standaarddeviatie in deze waarde.

4d. We gaan uit van een binomiale verdeling voor $n = 80$ personen met een vaste “slaag”-kans $p = 0{,}055$.
Dan is de verwachtingswaarde voor het aantal “besmet geweesten”: $np = 0{,}055 * 80 = 4{,}4$
Met een standaarddeviatie van $\sqrt{np(1-p)} = 2{,}03911$

4e. [4pt] Bereken de kans dat van de 80 studenten er 2, 3, 4, 5, 6 of 7 besmet zijn geweest.

4d. $\mathrm{binomcdf}(80\,, 0.055\,, 7) - \mathrm{binomcdf}(80\,, 0.055\,, 1) = 0{,}8659$

4f. [5pt] De betrouwbaarheid van een meting neemt in het algemeen toe als de grootte van de steekproef groter wordt. In Breda wordt in een teststation zo lang doorgegaan met het verzamelen van testresultaten dat de onzekerheid (σ) in aantal besmette personen nog maar 10% is van het aantal besmetten (μ). Hierbij wordt uitgegaan van een vaste besmettingskans voor de gehele populatie van Breda en dat de binomiale verdeling van toepassing is.
Deze situatie blijkt op te treden bij testpersoon 1935. Bereken de bijbehorende (schatting van de) besmettingsgraad.

4d. Als er n personen worden getest en er bestaat een besmettingskans van p, dan geldt volgens de binomiale verdeling dat $\mu = np$ en $\sigma = \sqrt{np(1-p)}$. Dat betekent dat

$$\frac{\sigma}{\mu} = \frac{\sqrt{np(1-p)}}{np} = \sqrt{\frac{1-p}{np}}$$

Omdat n in de noemer staat zie je dat deze relatieve nauwkeurigheid kleiner wordt naarmate n groter wordt.
Als deze waarde 0,1 is bij $n = 1935$, dan geldt dat

$$\frac{1-p}{1935p} = 0{,}1^2 = 0{,}01,$$

dus

$$1 - p = 19{,}35p,$$

dus

$$p = \frac{1}{20{,}35} = 0{,}0491.$$

=== EINDE TENTAMEN ===