

NAAM en VOORNAAM STUDENT: (VOLUIT EN IN HOOFDLETTERS)

STUDENTENNUMMER (= R-nummer):

Algemene afspraken voor een correct examenverloop

Voor de start van het examen

Plaats alle persoonlijke spullen, met inbegrip van uitgeschakelde multimediatoestellen (GSM, smartphone, smartwatch, ...), pennenzak en (hand)tas buiten handbereik.

Tijdens het examen

Leg examenkaart en studentenkaart zichtbaar bij jou op de tafel.

Toegelaten materiaal: zie onderaan dit formulier.

Examenopgave en -papier

Gebruik alleen officieel examenpapier uitgedeeld op het examen.

Schrijf op elk blad duidelijk je naam; schrijf leesbaar en verzorgd.

Maak de opgavenbundel NIET los.

Alle ontvangen bladen terug indienen, inclusief kladbladen.

Informatie betreffende onregelmatigheden

Elke onregelmatigheid of fraude (elk gedrag van een student in het kader van een examen waardoor de student het vormen van een juist oordeel omtrent de kennis, het inzicht en/of de vaardigheden van hemzelf of van andere studenten geheel of gedeeltelijk onmogelijk maakt of poogt te maken) wordt gemeld aan en beoordeeld/gesanctioneerd door de examencommissie.

Naam docent: Stefan Van Gulck
OPO / OLA-naam: **Kansrekenen en Beschrijvende Statistiek (HBN65B)**
Datum: dinsdag 16 januari 2018
Tijdstip: van 8u30 tot 11u30

Bijlage: formularium (2 bladzijden)

Vakspecifieke richtlijnen

1. De mondelinge toelichting van de taak Beschrijvende Statistiek telt mee voor 4 punten.
2. De kansvragen in deze bundel tellen mee voor de overige 16 punten. De puntenverdeling wordt per vraag vermeld.
3. Je mag gebruik maken van een (grafisch) rekentoestel en het formularium in bijlage.
4. Wees volledig en nauwkeurig in jouw antwoorden. Schrijf alle tussenstappen in jouw redenering op en definieer elke gebeurtenis/toevalsvariabele die je gebruikt. Louter de finale oplossing opschrijven volstaat niet als antwoord.
5. Schrijf al jouw antwoorden op deze bundel. Maak de pagina's uit de bundel niet los. Losse pagina's worden niet verbeterd. Bereid jouw oplossingen voor op kladpapier.
6. Vermijd tussenafroonden bij berekeningen. Rapporteer jouw resultaten ofwel exact ofwel met minstens vier beduidende cijfers.

Score:

Vraag 1 (1.5 + 1 = 2.5 punten)

Twee machines produceren eenzelfde goed. Van deze goederen wordt 40% door machine 1 geproduceerd en 60% door machine 2. De productietijden van goederen zijn normaal verdeeld, voor beide machines. Het gemiddelde en de standaardafwijking van de productietijden van goederen geproduceerd door machine 1 bedragen 45 sec en 8 sec, respectievelijk. Het gemiddelde en de standaardafwijking van de productietijden van goederen geproduceerd door machine 2 bedragen 40 sec en 6 sec, respectievelijk.

- Bereken het gemiddelde en de standaardafwijking van de productietijden van alle goederen (geproduceerd door machines 1 en 2).
- De productietijd van een goed bedroeg meer dan 43 sec. Wat is de kans dat dit goed door machine 1 werd geproduceerd?

Oplossing van vraag 1:

- Stel dat T de productietijd is van een willekeurig goed, M_1 de gebeurtenis waarbij dit goed geproduceerd wordt door machine 1 en M_2 de gebeurtenis waarbij dit goed geproduceerd wordt door machine 2. We weten dat $P[M_1] = 0.4$, $P[M_2] = 0.6$, $T|M_1 \sim \mathcal{N}(45, 8^2)$ en $T|M_2 \sim \mathcal{N}(40, 6^2)$. Hieruit volgt dat

$$\mu_T = E[T] = E[T|M_1]P[M_1] + E[T|M_2]P[M_2] = 45 \times 0.4 + 40 \times 0.6 = 42 \text{ sec} \quad [0.5]$$

en

$$\begin{aligned} E[T^2] &= E[T^2|M_1]P[M_1] + E[T^2|M_2]P[M_2] \\ &= (\text{Var}[T|M_1] + E[T|M_1]^2)P[M_1] + (\text{Var}[T|M_2] + E[T|M_2]^2)P[M_2] \\ &= (8^2 + 45^2) \times 0.4 + (6^2 + 40^2) \times 0.6 = 1817.2 \text{ sec}^2, \quad [0.5] \end{aligned}$$

$$\text{zodat } \sigma_T = \sqrt{\text{Var}[T]} = \sqrt{1817.2 - 42^2} = \sqrt{53.2} \cong 7.294 \text{ sec.} \quad [0.5]$$

- De gevraagde voorwaardelijke kans is

$$\begin{aligned} P[M_1|T > 43] &= \frac{P[T > 43|M_1]P[M_1]}{P[T > 43|M_1]P[M_1] + P[T > 43|M_2]P[M_2]} \\ &= \frac{0.5987 \times 0.4}{0.5987 \times 0.4 + 0.3085 \times 0.6} \cong 0.5640. \quad [1] \end{aligned}$$

Vervolg oplossing van vraag 1:

Vraag 2 (1 + 1 + 1 + 1.5 = 4.5 punten)

Een motorclub bestaat uit 20 leden jonger dan 30 jaar, 43 leden tussen 30 en 50 jaar, en 27 leden ouder dan 50 jaar. We kiezen lukraak en zonder terugleggen 3 leden. Noteer met X het aantal gekozen leden dat jonger is dan 30 jaar en met Y het aantal gekozen leden dat ouder is dan 50 jaar.

- Wat is de waarde van $P(X = 1)$, de kans dat precies 1 lid jonger dan 30 jaar wordt gekozen?
- Wat is de waarde van $P(X = 1 \text{ of } Y = 2)$, de kans dat precies 1 lid jonger dan 30 jaar wordt gekozen of precies 2 leden ouder dan 50 jaar worden gekozen?
- Geef de waarden van $E(X)$ en $\text{var}(X)$, de verwachte waarde en de variantie van het aantal gekozen leden jonger dan 30 jaar.
- Bereken $\rho_{X,Y} = \text{corr}(X, Y)$, de correlatie tussen het aantal gekozen leden jonger dan 30 jaar en het aantal gekozen leden ouder dan 50 jaar in deze selectie.

Oplossing van vraag 2:

- X is hypergeometrisch verdeeld. De gevraagde kans is

$$P[X = 1] = \frac{\binom{20}{1} \binom{70}{2}}{\binom{90}{3}} = \frac{80}{1958} \cong 0.4111. \text{ [1]}$$

- De wet van Boole (veralgemeende somregel) toont dat

$$\begin{aligned} P[X = 1 \text{ of } Y = 2] &= P[X = 1] + P[Y = 2] - P[X = 1 \text{ en } Y = 2] \\ &= \frac{\binom{20}{1} \binom{70}{2} + \binom{27}{2} \binom{63}{1} - \binom{20}{1} \binom{43}{0} \binom{27}{2}}{\binom{90}{3}} \\ &= \frac{1921}{3560} \cong 0.5396. \text{ [1]} \end{aligned}$$

- Omdat $X \sim \mathcal{HG}(3, 20, 70)$ geldt

$$E[X] = 3 \times \frac{20}{90} = \frac{2}{3} \text{ [0.5]} \quad \text{en} \quad \text{Var}[X] = 3 \times \frac{20}{90} \times \frac{70}{90} \times \frac{90-3}{90-1} = \frac{406}{801}. \text{ [0.5]}$$

- Analoog aan c. vinden we dat $E[Y] = 9/10$ en $\text{Var}[Y] = 5481/8900$. Voorts is

$$\begin{aligned} E[XY] &= 1P[X = 1, Y = 1] + 2P[X = 1, Y = 2] + 2P[X = 2, Y = 1] \\ &= \frac{\binom{20}{1} \binom{27}{1} \binom{43}{1} + 2 \binom{20}{1} \binom{27}{2} \binom{43}{0} + 2 \binom{20}{2} \binom{27}{1} \binom{43}{0}}{\binom{90}{3}} = \frac{36}{89}. \text{ [0.5]} \end{aligned}$$

Hieruit volgt dat $\text{Cov}[X, Y] = E[XY] - E[X]E[Y] = -87/445 \text{ [0.5]}$ en $\text{Corr}[X, Y] = \text{Cov}[X, Y] / \sqrt{\text{Var}[X]\text{Var}[Y]} \cong -0.3499 \text{ [0.5]}$.

Vervolg oplossing vraag 2:

Uiteraard kan ook de volledige kanstabel

| $x \backslash y$ | 0 | 1 | 2 | 3 |
|------------------|--------|--------|--------|--------|
| 0 | 0.1050 | 0.2075 | 0.1285 | 0.0249 |
| 1 | 0.1537 | 0.1977 | 0.0598 | 0 |
| 2 | 0.0965 | 0.0437 | 0 | 0 |
| 3 | 0.0097 | 0 | 0 | 0 |

gebruikt worden. Deze tabel kan eenvoudig ingegeven worden in het rekentoestel a.d.h.v. de formule

$$P[X = x, Y = y] = \frac{\binom{20}{x} \binom{27}{y} \binom{43}{4-x-y}}{\binom{90}{4}}.$$

Een student die deze werkwijze volgt leest de gevraagde kansen en kengetallen eenvoudig af van het rekentoestel.

Vraag 3 (1 + 0.5 + 1 + 1 + 1.5 + 1 = 6 punten)

De bedieningstijd T (in minuten) aan een loket van een willekeurige klant kan gemodelleerd worden met de volgende kansdichtheid:

$$f_T(t) = \begin{cases} at(5-t)^2 & \text{als } 0 \leq t \leq 5, \\ 0 & \text{anders.} \end{cases}$$

Veronderstel dat de bedieningstijden van verschillende klanten onafhankelijk zijn.

- Bepaal het voorschrift van de cumulatieve verdelingsfunctie $F_T(t)$ zonder de constante a te benoemen.
- Gebruik je antwoord op deelvraag a. om de waarde van de constante a te vinden.
- Wat is de kans dat er meer dan 100 klanten bij het loket moeten langskomen alvorens de vierde klant toekomt waarvoor de bedieningstijd meer dan 4 minuten bedraagt?
- Bereken de verwachte waarde $E(T)$ en de variantie $\text{var}(T)$.
- Op een dag komen er 50 willekeurige klanten langs het loket. Bereken *benaderend* de kans dat de totale bedieningstijd van deze klanten minstens 110 minuten duurt.
- Wat is het antwoord op deelvraag d. indien het aantal klanten Poisson verdeeld is met verwachte waarde 50 (en onafhankelijk is van de bedieningstijden)?

Oplossing van vraag 3:

Deze vraag kan deels opgelost worden met $T/5 \sim \text{Beta}(\alpha = 2, \beta = 3)$.

- Uit $t(5-t)^2 = 25t - 10t^2 + t^3$ volgt dat

$$F_T(t) = \begin{cases} 0 & \text{als } t < 0, \\ \frac{at^2(150 - 40t + 3t^2)}{12} & \text{als } 0 \leq t \leq 5, \quad [1] \\ 1 & \text{als } t > 5. \end{cases}$$

- Omdat $1 = F_T(5) = 625a/12$, vinden we dat $a = 12/625$ [0.5].
- We weten dat $P(T > 4) = 1 - F_T(4) = 1 - 608/625 = 17/625 = 0.0272$ [0.25].
Als $Y_r \sim \mathcal{NB}(r, p)$ en $X_n \sim \mathcal{Bi}(n, p)$, dan $P(Y_4 > 100) = P(X_{100} \leq 3) \cong 0.7106$ [0.75].
- Uit $t^2(5-t)^2 = 25t^2 - 10t^3 + t^4$ en $t^3(5-t)^2 = 25t^3 - 10t^4 + t^5$ volgt dat

$$E[T] = \frac{12}{625} \left[\frac{25}{3} t^3 - \frac{10}{4} t^4 + \frac{1}{5} t^5 \right]_0^5 = 2, [0.5]$$
$$E[T^2] = \frac{12}{625} \left[\frac{25}{4} t^4 - \frac{10}{5} t^5 + \frac{1}{6} t^6 \right]_0^5 = 5.$$

We vinden dus ook $\text{Var}[T] = E[T^2] - E[T]^2 = 1$. [0.5]

Vervolg oplossing van vraag 3:

- e. De totale bedieningstijd is $S_{50} = T_1 + \dots + T_{50}$ (i.i.d. termen) [0.25], zodat $E[S_{50}] = 50E[T] = 100$ en $\text{Var}[S_{50}] = 50\text{Var}[T] = 50$ [0.5]. Uit de centrale limietstelling volgt dat $S_{50} \approx \mathcal{N}(100, 50)$ [0.25]. De gevraagde kans is $P[S_{50} \geq 110] \cong 0.07865$ [0.5].
- f. Nu is $S_N = T_1 + \dots + T_N$ de totale bedieningstijd, met $N \sim \mathcal{P}\sigma(50)$ onafhankelijk van (T_1, T_2, \dots) . We gebruiken nu de eigenschappen $E[S_N] = E[N]E[T]$ en $\text{Var}[S_N] = E[N]\text{Var}[T] + \text{Var}[N]E[T]^2$ in $S_N \approx \mathcal{N}(E[S_N], \text{Var}[S_N])$ [0.25]. Omdat $E[N] = \text{Var}[N] = 50$, vinden we $E[S_N] = 100$ en $\text{Var}[S_N] = 250$ [0.5]. Hieruit volgt dat $P[S_N \geq 110] \cong 0.2635$ [0.25].

Vraag 4 (3 punten)

Toon aan dat het quotiënt van twee onafhankelijke standaardnormaal verdeelde toevalsvariabelen een standaard-Cauchy verdeelde toevalsvariabele is.

In symbolen: als $Z_1 \sim \mathcal{N}(\mu = 0, \sigma^2 = 1)$ en $Z_2 \sim \mathcal{N}(\mu = 0, \sigma^2 = 1)$ onafhankelijk zijn, dan geldt $X = Z_1/Z_2 \sim \text{Cauchy}(x_0 = 0, \gamma = 1)$.

Oplossing van vraag 4:

Het uitwerken van de verdelingsfunctie van X toont voor elke $x \in \mathbb{R}$ dat

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{\infty} P(Z_1/Z_2 \leq x | Z_2 = y) f_Z(y) dy \\ &= \int_{-\infty}^0 (1 - F_Z(xy)) f_Z(y) dy + \int_0^{\infty} F_Z(xy) f_Z(y) dy. \end{aligned} \quad [1]$$

Onafhankelijkheid tussen Z_1 en Z_2 werd in de tweede gelijkheid gebruikt. Afleiden naar x levert

$$f_X(x) = \int_{-\infty}^0 (-y) f_Z(xy) f_Z(y) dy + \int_0^{\infty} y f_Z(xy) f_Z(y) dy. \quad [0.5]$$

De symmetrie $f_Z(z) = f_Z(-z)$ toont dat

$$f_X(x) = 2 \int_0^{\infty} y f_Z(xy) f_Z(y) dy = \frac{1}{\pi} \int_0^{\infty} y e^{-y^2(1+x^2)/2} dy. \quad [1]$$

Substitutie $t = y^2(1+x^2)/2$ (dus $t'(y) = y(1+x^2)$) geeft tenslotte

$$f_X(x) = \frac{1}{\pi(1+x^2)} \int_0^{\infty} e^{-t} dt = \frac{1}{\pi(1+x^2)}. \quad [0.5]$$

Vervolg oplossing van vraag 4: