



Faculteit Militaire Wetenschappen

| Gegevens student | |
|-------------------|--|
| Naam: | |
| Peoplesoftnummer: | |
| Klas: | |
| Handtekening: | |

(Her)Tentamen

| Algemeen | | | |
|--------------|-----------------------------------|------------------|-------------|
| Vak: | Statistiek (deel 2) -- derde kans | Vakcode: | STA#2 |
| Datum: | 23 mei 2025 | Tijdsduur: | 10:00-13:00 |
| Examinator: | Dr. ir. D.A.M.P. Blom | Aantal pagina's: | 4 |
| Peer-review: | Dr. J.B.M. Melissen | Aantal opgaven: | 4 |

| Algemene instructies |
|---|
| <ul style="list-style-type: none">- Alle antwoorden dienen gemotiveerd te worden. Indien u een deelopgave niet kunt oplossen en het antwoord in vervolgvragen nodig hebt, probeer uit te gaan van een redelijke fictieve waarde.- U mag een grafische rekenmachine gebruiken zonder CAS (computer algebra systeem) gebruiken.- Antwoorden, in welke vorm dan ook, mogen de zaal niet verlaten.- Vermeld op elk antwoordvel je naam, Peoplesoft-nummer en maak een nummering van je antwoordvellen.- Iedere vorm van mobiele (potentiële) datadragers (telefoon, smartwatch, etc) of andere vormen om te frauderen (bv. communicatieapparatuur) zijn niet toegestaan gedurende de gehele duur van het tentamen en mogen ook niet in het lokaal meegebracht worden of zijn uitgeschakeld en ingeleverd.- Schrijf leesbaar ter voorkoming van misverstanden bij de beoordeling van uw werk. Indien uw antwoord niet leesbaar is, wordt uw antwoord fout gerekend.- Toiletbezoek tijdens het tentamen vindt enkel plaats na toestemming van de examinerator.- Lever bij het verlaten van de zaal, kladpapier, tentamenopgaven en andere tentamen gerelateerde documenten in bij de examinerator. |

Opgave 1 (25 punten) De Koninklijke Militaire Academie onderzoekt of een nieuw type gevechtstraining op basis van virtual reality de reactietijd van soldaten onder crisissomstandigheden significant verbetert. Soldaten worden ondergedompeld in een realistische VR-simulatie waarin ze snel moeten reageren en vijandelijke dreigingen moeten zien te identificeren.

Om de effectiviteit te testen, meten instructeurs de reactietijd (in milliseconden) van 18 cadetten na 6 weken VR-training en vergelijken dit met de gemiddelde reactietijd van 420 ms bij conventionele trainingsmethoden. Uit de steekproef blijkt dat de gemiddelde reactietijd na de VR-training gelijk is aan 390 milliseconden met een standaardafwijking van 40 milliseconden.

- 1a [5pt]** Bereken op grond van deze steekproef een 95%-betrouwbaarheidsinterval voor de gemiddelde reactietijd van een cadet. Rond de grenzen van dit interval af op gehele milliseconden, zodanig dat de betrouwbaarheid gewaarborgd blijft.

Uitwerking

- 1b [3pt]** De instructeurs willen de bewering dat de VR-training de reactietijd van cadetten significant verlaagt statistisch verantwoord toetsen. Formuleer de bijbehorende nul- en alternatieve hypothese van deze hypothesetoets.

Uitwerking

- 1c [8pt]** Voer de hypothesetoets uit. Bepaal de toetsuitslag door het berekenen van een kritiek gebied op basis van de gegeven steekproef van 18 cadetten met significantieniveau $\alpha = 0,02$.

Uitwerking

- 1d [4pt]** Leg in eigen woorden uit wat de uitslag van deze toets betekent voor de gemiddelde reactietijd van een cadet na VR-training.

Uitwerking

1e [5pt] Stel nu dat de instructeurs niet de reactietijd van een individuele cadet willen meten, maar van een team van zes cadetten. De teamscore wordt bepaald door de som te nemen van de reactietijden van de zes cadetten in het team. Elk team wil dus een zo laag mogelijke teamscore bereiken.

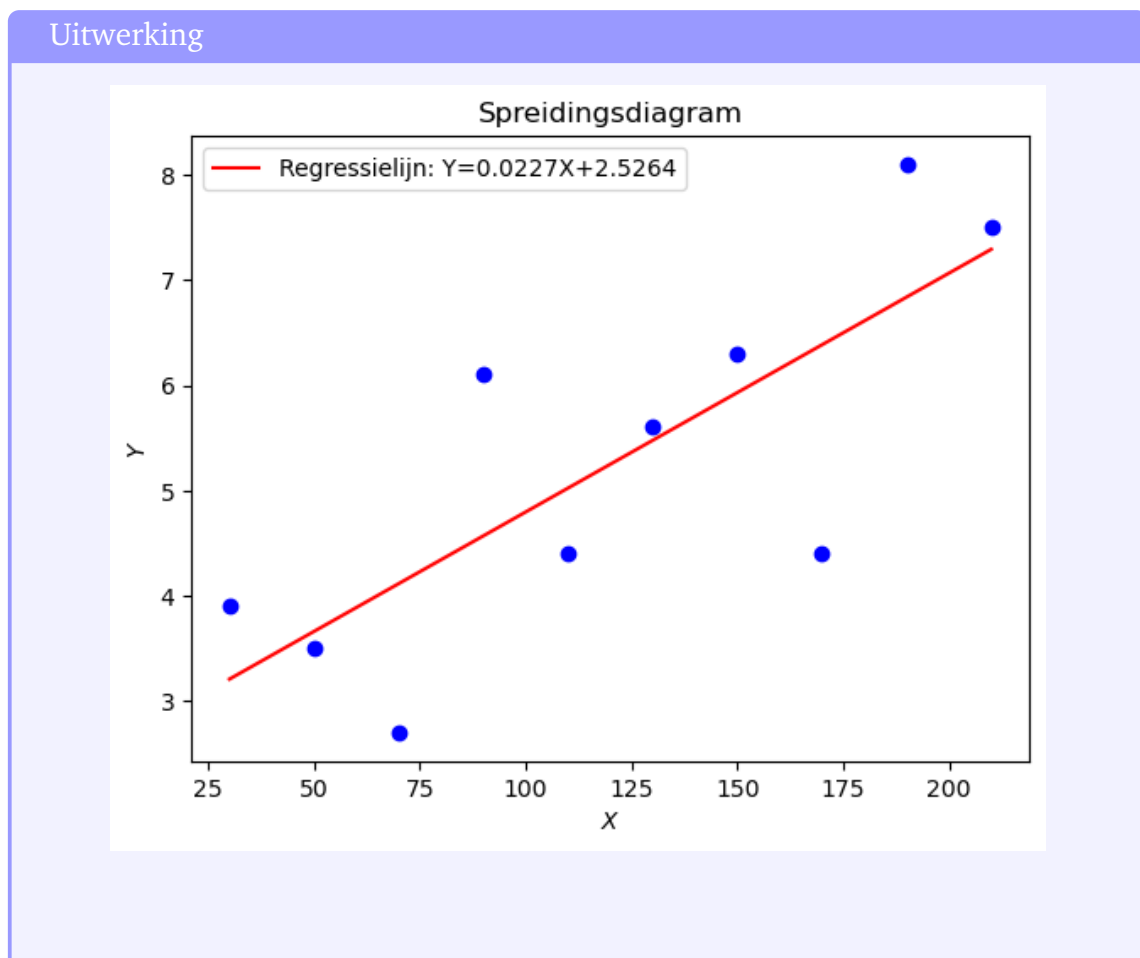
Op dit moment staat het record op x milliseconden. Voor welke waarde van x is de kans op verbreking van het record gelijk aan 5%?

Uitwerking

Opgave 2 (30 punten) Een luchtmacht onderzoekt de nauwkeurigheid van GPS-geleide raketten in een omgeving van elektronische oorlogsvoering. De raketten worden gericht op een vijandelijke luchtmachtbasis. Tijdens de vlucht worden de raketten “gedesoriënteerd” omdat de vijandelijke troepen gebruik maken van GPS jammers om precisiewapens te verstoren. Onderzocht wordt nu wat het verband is tussen de afstand (in km) tussen de eerste locatie waar de raket gejamd wordt en het doelwit (X) en de afstand (in km) tussen het doelwit en het daadwerkelijke inslagpunt van de raket (Y). Van tien raketten worden data verzameld over deze twee variabelen.

| | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| X | 30 | 50 | 70 | 90 | 110 | 130 | 150 | 170 | 190 | 210 |
| Y | 3.9 | 3.5 | 2.7 | 6.1 | 4.4 | 5.6 | 6.3 | 4.4 | 8.1 | 7.5 |

2a [4pt] Teken de gegevens uit bovenstaande tabel in een spreidingsdiagram.



(4pt)

2b [8pt] Bereken met behulp van de tabel hierboven Pearson's correlatiecoëfficiënt. Bepaal of er sprake is van een lineaire correlatie en leg in woorden uit wat de betekenis is van de grootte en het teken van de correlatiecoëfficiënt.

Uitwerking

Hiervoor moeten we eerst een tabel uitschrijven met de grootheden \bar{X} , \bar{Y} , \overline{XY} , $\overline{X^2}$ en $\overline{Y^2}$.

| X | Y | XY | X^2 | Y^2 |
|--|-----|------|-------|--------|
| 30 | 3.9 | 117 | 900 | 15.21 |
| 50 | 3.5 | 175 | 2500 | 12.25 |
| 70 | 2.7 | 189 | 4900 | 7.29 |
| 90 | 6.1 | 549 | 8100 | 37.21 |
| 110 | 4.4 | 484 | 12100 | 19.366 |
| 130 | 5.6 | 728 | 16900 | 31.36 |
| 150 | 6.3 | 945 | 22500 | 39.69 |
| 170 | 4.4 | 748 | 28900 | 19.36 |
| 190 | 8.1 | 1539 | 36100 | 65.61 |
| 210 | 7.5 | 1575 | 44100 | 56.25 |
| $\bar{X} = 120 \quad \bar{Y} = 5.25 \quad \overline{XY} = 704.9 \quad \overline{X^2} = 17700 \quad \overline{Y^2} = 30.3590$ | | | | |

De correlatiecoëfficiënt van Pearson is gelijk aan

(4pt)

$$\begin{aligned}
 r &= \frac{\overline{XY} - \bar{X} \cdot \bar{Y}}{\sqrt{(\overline{X^2} - \bar{X}^2) \cdot (\overline{Y^2} - \bar{Y}^2)}} \\
 &= \frac{704.9 - 120 \cdot 5.25}{\sqrt{(120^2 - 17700) \cdot (5.25^2 - 30.359)}} \\
 &= \frac{74.9}{96.065} \\
 &\approx 0.7797.
 \end{aligned}$$

(3pt)

Deze correlatiecoëfficiënt duidt op een sterk positieve correlatie, oftewel wanneer een raket van grotere afstand gejamd wordt zal de raket ook verder van zijn doelwit inslaan.

(1pt)

2c [8pt] Bereken de regressielijn $Y = aX + b$ met behulp van de tabel hierboven, en geef

de interpretatie van a en b aan in dit scenario.

Uitwerking

De coëfficiënten a en b van de regressielijn kunnen worden berekend aan de hand van twee vergelijkingen. Allereerst hebben we

$$a = \frac{\overline{XY} - \overline{X} \cdot \overline{Y}}{\overline{X^2} - \overline{X}^2} = \frac{704.9 - 120 \cdot 5.25}{17700 - 120^2} \approx 0.0227 \quad (3\text{pt})$$

Vervolgens geldt dan dat

$$b = \overline{Y} - a \cdot \overline{X} = 5.25 - 0.0227 \cdot 120 \approx 2.5264 \quad (2\text{pt})$$

De regressielijn behorende bij deze steekproef is dus gelijk aan $Y = 0.0227X + 2.5264$. (1pt)

Je kunt deze regressielijn interpreteren als volgt: voor elke km dat de raket verder van het doelwit af jamming ervaart, zal de afwijking van de raket ten opzicht van het doelwit stijgen met 0.0227 km. Verder is in het geval van geen jamming de afstand tussen doelwit en inslagpunt 2.5264 km (bijvoorbeeld door weersfenomenen of rekenfouten tijdens lancering). (2pt)

2d [2pt] Geef een statistisch verantwoorde voorspelling van de afstand tot het doelwit waarop een raket inslaat die op 45 kilometer van het doelwit voor het eerst GPS jamming ondervindt.

Uitwerking

Je kunt de zojuist uitgerekend regressielijn gebruiken om een voorspelling te vinden. Vul in $X = 45$, dit geeft $Y = 0.0227 \cdot 45 + 2.5264 \approx 3.5477$ km. (1pt)

De raket zal naar verwachting zo ongeveer op iets meer dan 3.5 km van het doelwit inslaan. (1pt)

2e [8pt] Bepaal een 95%-voorspellingsinterval voor de afstand tussen doelwit en inslagpunt als die op 120 kilometer van het doelwit voor het eerst GPS-jamming ondervindt.

Uitwerking

Het voorspellingsinterval is $[Y_0 - t \cdot s_f; Y_0 + t \cdot s_f]$, waarbij $Y_0 = 3.5477$ de puntschatting gemaakt bij de vorige opgave. Verder geldt dat t de t -waarde is die hoort bij een betrouwbaarheid van 95% met $df = n - 2 = 10 - 2 = 8$ vrijheidsgraden. Bij een betrouwbaarheid van 95% is de linkeroverschrijdingskans $0.95 + 0.05/2 = 0.975$ en

(1pt)

$$t = \text{InvT}(\text{opp} = 0.975; df = 8) \approx 2.3060$$

Verder geldt voor de standaardafwijking van de errorterm:

(1pt)

$$\begin{aligned} s_\epsilon &= \sqrt{\frac{n}{n-2} \cdot (\overline{Y^2} - a \cdot \overline{XY} - b \cdot \overline{Y})} \\ &= \sqrt{\frac{10}{8} \cdot (30.3590 - 0.0227 \cdot 704.9 - 2.5264 \cdot 5.25)} \\ &\approx 1.0963 \end{aligned}$$

Verder geldt

(2pt)

$$\begin{aligned} u &= \frac{1}{n} \cdot \left(1 + \frac{(X_0 - \overline{X})^2}{\overline{X^2} - \overline{X}^2}\right) \\ &= \frac{1}{10} \cdot \left(1 + \frac{(45 - 120)^2}{17700 - 120^2}\right) \\ &\approx 0.2705 \end{aligned}$$

De standaardafwijking van de forecasting term is dan

(2pt)

$$s_f = s_\epsilon \cdot \sqrt{u + 1} \approx 1.2357$$

Dit levert een interval op van $[Y_0 - t \cdot s_f; Y_0 + t \cdot s_f] = [0.6982; 6.3972]$. Merk op dat dit interval zeer onnauwkeurig is en laat zien dat de voorspellende waarde van de regressielijn zeer beperkt is.

(1pt)

Opgave 3 (20 punten) Binnen het NAVO-bondgenootschap wordt over een periode van zes maanden door een internationaal cybersecurity team inlichtingen uitgewisseld over cyberaanvallen gericht op de Baltische staten: Estland, Letland en Litouwen. Hierbij worden de cyberaanvallen van drie belangrijke types gemonitord, namelijk phishing aanvallen, injecties van malware en DDoS-aanvallen (Distributed Denial-of-Service). De verzamelde data ziet er als volgt uit:

| Land | Phishing | Malware | DDoS | Totaal |
|---------------|----------|---------|------|--------|
| Estland | 120 | 200 | 140 | 460 |
| Letland | 150 | 170 | 180 | 500 |
| Litouwen | 100 | 220 | 160 | 480 |
| Totaal | 370 | 590 | 480 | 1440 |

Het cybersecurity team wil testen of de verdeling van het type inkomende cyberaanvallen significant verschilt tussen de drie Baltische staten.

3a [3pt] Formuleer de nulhypothese H_0 en de alternatieve hypothese H_1 van de bijbehorende hypothesetoets.

Uitwerking

3b [10pt] Voer een chi-kwadraattoets voor onafhankelijkheid uit met significantieniveau $\alpha = 0,05$ en bereken de bijbehorende p -waarde.

Uitwerking

3c [7pt] Kunnen we concluderen dat de verdeling van het type inkomende cyberaanvallen significant verschilt voor de drie Baltische staten?

Uitwerking

Opgave 4 (25 punten) Een marinefregat opereert in een gebied waar regelmatig vijandelijke onderzeeboten actief zijn. Het fregat gebruikt hiervoor sonar en magnetische sensoren om onderzeeboten te detecteren. De bemanning wil analyseren hoe vaak ze per dag succesvolle detecties uitvoeren en of hun waarnemingen een Poisson-verdeling volgen.

Gedurende een periode wordt het aantal gedetecteerde vijandelijke onderzeeboten geteld.

| Aantal gedetecteerde onderzeeboten | Aantal dagen |
|------------------------------------|--------------|
| 0 | 40 |
| 1 | 80 |
| 2 | 100 |
| 3 | 85 |
| 4 | 35 |
| 5 | 15 |
| 6 | 7 |
| 7 | 3 |

4a [5pt] Bereken met behulp van de gegevens uit de tabel het aantal dagen waarop er data is verzameld en het gemiddelde aantal gedetecteerde vijandelijke onderzeeboten per dag.

Uitwerking

Het aantal dagen n waarop er data is verzameld is gelijk aan

$$n = 40 + 80 + 100 + 85 + 35 + 15 + 7 + 3 = 365$$

(1pt)

Het totaal aantal detecties is gelijk aan

$$\# \text{detecties} = 0 \cdot 40 + 1 \cdot 80 + 2 \cdot 100 + 3 \cdot 85 + 4 \cdot 35 + 5 \cdot 15 + 6 \cdot 7 + 7 \cdot 3 = 813$$

(2pt)

Het gemiddelde aantal detecties is dus gelijk aan

$$\text{gemiddelde} = \frac{\# \text{detecties}}{n} = \frac{813}{365} \approx 2.2274.$$

(2pt)

4b [12pt] Toets of het aantal gedetecteerde onderzeeboten per dag X is te beschouwen als een Poisson-verdeelde kansvariabele. Schrijf hierbij specifiek de nul- en alternatieve hypothese uit en bepaal de toetsuitslag aan de hand van het berekenen van een p -waarde. Kies als betrouwbaarheid 95% en gebruik in je berekening je antwoord op vraag 4a.

Uitwerking

Om deze bewering te toetsen, moeten we werken met de volgende hypothesen:

H_0 : de kansvariabele X volgt een Poissonverdeling.

H_1 : de kansvariabele X volgt NIET een Poissonverdeling.

(2pt)

Omdat we een toets willen uitvoeren om te testen of een kansvariabele een bepaalde kansverdeling volgt, moeten we een chi-kwadraattoets voor aanpassing gebruiken. Om de toetsingsgrootte χ^2 te berekenen, hebben we allereerst de verwachte frequenties nodig als we zouden werken met een Poissonverdeling met parameter $\lambda = \frac{813}{365}$.

(1pt)

| Aantal gedetecteerde onderzeeboten | Aantal dagen (observed) | Aantal dagen (expected) |
|------------------------------------|-------------------------|--|
| 0 | 40 | $365 \cdot \text{poissonpdf}(\lambda = \frac{813}{365}; k = 0) = 39.3502$ |
| 1 | 80 | $365 \cdot \text{poissonpdf}(\lambda = \frac{813}{365}; k = 1) = 87.6484$ |
| 2 | 100 | $365 \cdot \text{poissonpdf}(\lambda = \frac{813}{365}; k = 2) = 97.6140$ |
| 3 | 85 | $365 \cdot \text{poissonpdf}(\lambda = \frac{813}{365}; k = 3) = 72.4750$ |
| 4 | 35 | $365 \cdot \text{poissonpdf}(\lambda = \frac{813}{365}; k = 4) = 40.3577$ |
| 5 | 15 | $365 \cdot \text{poissonpdf}(\lambda = \frac{813}{365}; k = 5) = 17.9785$ |
| 6 | 7 | $365 \cdot \text{poissonpdf}(\lambda = \frac{813}{365}; k = 6) = 6.6742$ |
| 7 | 3 | $365 \cdot (1 - \text{poissoncdf}(\lambda = \frac{813}{365}; k = 6)) = 2.9020$ |

(3pt)

Omdat er verwachte frequenties zijn die kleiner dan 5 zijn (vuistregel), moeten we rijen gaan samenvoegen, namelijk de laatste twee rijen:

| Aantal gedetecteerde onderzeeboten | Aantal dagen (observed) | Aantal dagen (expected) |
|---------------------------------------|----------------------------|----------------------------|
| 0 | 40 | 39.3502 |
| 1 | 80 | 87.6484 |
| 2 | 100 | 97.6140 |
| 3 | 85 | 72.4750 |
| 4 | 35 | 40.3577 |
| 5 | 15 | 17.9785 |
| ≥ 6 | 10 | 9.5763 |

(1pt)

De toetsingsgrootte kunnen we nu bepalen als volgt:

$$\begin{aligned}
 \chi^2 &= \frac{(O_0 - E_0)}{E_0} + \frac{(O_1 - E_1)}{E_1} + \dots + \frac{(O_{\geq 6} - E_{\geq 6})}{E_{\geq 6}} \\
 &= \frac{(40 - 39.3502)}{39.3502} + \frac{(80 - 87.6484)}{87.6484} + \dots + \frac{(10 - 9.5763)}{9.5763} \\
 &\approx 4.1245
 \end{aligned}$$

(2pt)

Aangenomen dat de nulhypothese waar zou zijn, is deze toetsingsgrootte getrokken uit een chi-kwadraatverdeling met aantal vrijheidsgraden

$$df = \# \text{categorieën} - 1 - \# \text{geschatte parameters} = 7 - 1 - 1 = 5.$$

De p -waarde die hoort bij deze uitkomst is

$$p = P(\chi^2 > 4.1245) = \chi^2 \text{cdf}(4.1245; 10^{99}; 5) \approx 0.5316.$$

(2pt)

Tevens is er voor een betrouwbaarheidsniveau van 95% ($\alpha = 0.05$) gekozen. Omdat $p > \alpha$, wordt H_0 niet verworpen. Er is onvoldoende bewijs om de bewering te verwerpen dat de kansvariabele een Poissonverdeling zou volgen.

(1pt)

4c [8pt] De marine definieert een dag met hoge vijandelijke onderzeese activiteit als een

dag waarop minstens vier onderzeeboten worden gedetecteerd. Bereken met de methode van Clopper-Pearson een 95%-betrouwbaarheidsinterval voor de kans p dat er op een willekeurige dag hoge vijandelijke activiteit is.

Uitwerking

Volgens de gegevens van de marine zijn op $35 + 15 + 7 + 3 = 60$ van de 365 dagen minstens vier onderzeeboten gedetecteerd. Een puntschatting voor de

(1pt)

kans dat er op een willekeurige dag hoge vijandelijke onderzeese activiteit is, is gelijk aan $\hat{p} = \frac{60}{365} = 0.1644$.

(1pt)

Voor het 95%-betrouwbaarheidsinterval (dus $\alpha = 0.05$) berekenen we de grenzen aan de hand van

1. We berekenen p_1 door de vergelijking $P(X \leq 60) = \frac{\alpha}{2} = 0.025$ op te lossen. Hieruit volgt:

$$\text{binomcdf}(n = 365; p_1 = ?; k = 60) = 0.025 \rightarrow p_1 = 0.2065$$

(2pt)

2. We berekenen p_2 door de vergelijking $P(X \geq 60) = 1 - P(X \leq 59) = \frac{\alpha}{2} = 0.025$ op te lossen. Hieruit volgt:

$$1 - \text{binomcdf}(n = 365; p_2 = ?; k = 59) = 0.025 \rightarrow p_2 = 0.1278$$

(3pt)

Dit levert het 95% Clopper-Pearson betrouwbaarheidsinterval $[0.1278; 0.2065]$ voor de kans p dat er op een willekeurige dag hoge vijandelijke onderzeese activiteit is.

(1pt)