**Tentamen Statistiek MBW/KW deel 2, 1e kans 28-07-2023**

Afdeling: Propedeuse MBW/KW 2022-2023

Examinator: Dr. J.B.M. Melissen, T. Zijlstra MSc.

Datum: vrijdag 28 juli 2023, 9:00 – 11:00. **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. **De opgaven dienen na afloop van het tentamen ingeleverd te worden.**

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

# Opgave 1 (Totaal 30 punten)

Van 20 MTW cadetten staan in de tabel hieronder de scores (op een schaal van 0 – 100) van de vakken Analyse 1 () en Lineaire Algebra 1 ():

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Student** | **=Analyse** | **=Algebra** | **2** | **2** |  |
| **1** | 40 | 53 | 1600 | 2809 | 2120 |
| **2** | 47 | 56 | 2209 | 3136 | 2632 |
| **3** | 51 | 91 | 2601 | 8281 | 4641 |
| **4** | 51 | 59 | 2601 | 3481 | 3009 |
| **5** | 53 | 58 | 2809 | 3364 | 3074 |
| **6** | 54 | 70 | 2916 | 4900 | 3780 |
| **7** | 54 | 42 | 2916 | 1764 | 2268 |
| **8** | 55 | 62 | 3025 | 3844 | 3410 |
| **9** | 56 | 77 | 3136 | 5929 | 4312 |
| **10** | 56 | 55 | 3136 | 3025 | 3080 |
| **11** | 58 | 49 | 3364 | 2401 | 2842 |
| **12** | 65 | 87 | 4225 | 7569 | 5655 |
| **13** | 69 | 68 | 4761 | 4624 | 4692 |
| **14** | 70 | 58 | 4900 | 3364 | 4060 |
| **15** | 77 | 75 | 5929 | 5625 | 5775 |
| **16** | 86 | 90 | 7396 | 8100 | 7740 |
| **17** | 92 | 80 | 8464 | 6400 | 7360 |
| **18** | 21 | 35 | 441 | 1225 | 735 |
| **19** | 56 | 82 | 3136 | 6724 | 4592 |
| **20** | 86 | 89 | 7396 | 7921 | 7654 |
| **Gemiddeld:** | **59,85** | **66,80** | **3848,05** | **4724,30** |  |

**1a [2pt]** Maak de tabel compleet door het uitrekenen van de waarde van (Narekenen!)

. **2pt**

**1b [6pt]** Bereken met behulp van de tabel de correlatiecoëfficiënt van Pearson.

Bepaal of er sprake is van een lineaire correlatie tussen de scores van de twee vakken.

Leg in woorden uit wat de betekenis is van de grootte en het teken van de correlatiecoëfficiënt voor een mogelijke relatie tussen de scores voor de twee vakken.

Is er sprake van een causaal verband?

De score voor Analyse is de verklarende variabele, die kies je als X, de score voor Lineaire Algebra is Y.

De correlatiecoëfficiënt van Pearson is een getal tussen -1 en +1 dat aangeeft hoe goed twee variabelen aan een lineair verband voldoen.

De correlatiecoëfficiënt is

**3pt**

Hoe dichter bij 1 (of -1), hoe beter de correlatie. In dit geval dus een behoorlijk goede correlatie. Dat betekent dat er een behoorlijk goed lineair verband zal zijn tussen X en Y, dus het is verantwoord om lineaire regressie toe te passen. **1pt**

De correlatiecoëfficiënt is positief, dus er is een positieve correlatie (d.w.z. bij een langere oefening horen meer kosten), het lineaire verband tussen duur oefening en kosten is een rechte lijn die stijgend is. **1pt**

Het is onwaarschijnlijk dat er een causaal verband is, want dat zou betekenen dat de score voor Lineaire algebra het gevolg is van de score voor Analyse. Het is waarschijnlijker dat studenten die goed of minder goed zijn in wiskunde daardoor voor beide vakken goed of minder goed scoren.  **1pt**

**1c [7pt]** Bereken de regressielijn door berekening van de coëfficiënten en met behulp van de tabel (narekenen!). Neem hierbij de score van Analyse als de verklarende variabele .

De regressielijn is met

**3pt**

**3pt**

**1pt**

**1d [2pt]** Bereken met de regressielijn een voorspelling voor de Lineaire Algebra score van een cadet die voor Analyse 74 punten scoorde. Rond de score af op een geheel getal.

Vul in en je krijgt een bijbehorende voorspelling van de score: =76 **2pt**

**1e [7pt]** Bereken een **92%** voorspellingsinterval voor de voorspelling uit 1d. Rond de grenzen af op gehele getallen.

Het voorspellingsinterval is is de -waarde die hoort bij de betrouwbaarheid van 92% met vrijheidsgraden. Bij een betrouwbaarheid van 92% is de linker overschrijdingskans 0,92 + 0,08/2 = 0,96 en

**2pt**

**1pt**

**1pt**

**1pt**

**2pt**

**1f [3pt]** Bereken de Analyse score waarvoor de regressieformule een Lineaire Algebra score voorspelt die gelijk is aan de Analyse score. Rond deze score af op een geheel getal.

Uit volgt **3pt**

**1g [3pt]** Stel dat een student als Analyse score toevallig de gemiddelde Analyse score van alle cadetten heeft. Laat door berekening zien dat de regressielijn voor deze student een Lineaire Algebra score voorspelt die gelijk is aan de gemiddelde Lineaire Algebra score van alle studenten.  
Toon vervolgens aan dat dit geen toeval is, namelijk, dat altijd geldt: .

(Hint: gebruik de formules voor en/of).  
  
 **2pt**  
De formule voor: is omgeschreven: . **1pt**

**Opgave 2 (Totaal 25 punten)**

Vliegtuigbanden slijten bij het landen (10% van de slijtage), het remmen na de landing (ca. 30%) en het taxiën en starten (60%). Daarom moeten vliegtuigbanden na een bepaald aantal vluchten worden vervangen. Voor tien vliegtuigbanden werd het aantal vluchten genoteerd waarna ze moesten worden vervangen: 250, 171, 302, 119, 234, 272, 112, 262, 203, 189.

**2a [6pt]** Bereken van de gemeten waarden het steekproefgemiddelde en de steekproefstandaarddeviatie.

vluchten **3pt**

vluchten **3pt**

**(Gedeeld door i.p.v. , 60,8801 i.p.v. s: 1pt**

**Formule met wortel verkeerd: max 1 pt)**

**2b [6pt]** Bereken op grond van de bovengenoemde steekproef een 90% betrouwbaarheids-interval voor het verwachte aantal operationele vluchten dat een band meegaat, zonder

daarbij gebruik te maken van de optie TESTS/Interval van de grafische rekenmachine. Rond de grenzen van dit interval af op gehele aantallen en wel zodanig dat de betrouwbaarheid gewaarborgd blijft.

De standaarddeviatie is niet bekend, (en steekproefgrootte is kleiner dan 30) dus de **-verdeling** moet worden gebruikt. De -waarde voor 90% tweezijdige betrouwbaarheid is **2pt**

Het aantal vluchten per band is normaal verdeeld met een geschat gemiddelde van vluchten en een standaarddeviatie van vluchten.

Het 90% schattingsinterval voor , de verwachtingswaarde van het aantal vluchten dat een band meegaat, is dan

**3pt**

Afronden: . **1pt**

**2c [7pt]** Bereken op grond van bovengenoemde steekproef en met behulp van de -verdeling een 90% betrouwbaarheidsinterval (zie formuleblad) **voor de standaarddeviatie** van het aantal vluchten dat per band kan worden gevlogen, zonder daarbij gebruik te maken van de optie TESTS/Interval van de grafische rekenmachine. Rond de grenzen van dit interval af één decimaal.

met betrouwbaarheid . ,

Los op met de GR: en vind . **2pt**

Los op en vind . **2pt**

Schattingsinterval is **2pt**

**Afgerond: [46,8 , 105,6] 1pt**

**2d [2pt]** Als de kans dat één band vluchten kan uitvoeren zonder vervangen te worden gelijk is aan 0,98305, toon dan met een berekening aan dat de kans dat drie banden minstens vluchten meegaan gelijk is aan 0,9500.  
  
De kans dat drie banden tegelijk gedurende vluchten meegaan is **2pt**

**2e [4pt]** Een F35-Lightning II (Joint Strike Fighter) wordt voorzien van 3 nieuwe banden. Bereken een aantal vluchten waarvan je met 95% zekerheid kunt aantonen dat deze F35 zoveel vluchten kan maken zonder dat een band vervangen hoeft te worden. (Hint: Denk aan een rechteroverschrijdingskans van 0,98305.)  
  
Als de levensduur van een band is, dan weten we dat normaal verdeeld is met vluchten en vluchten. Als we zo bepalen dat , dan is de kans dat drie zulke banden minstens vluchten meegaan minstens 0,95.

Als , dan is , en de -waarde voor een linkeroverschrijdingskans is . **2pt**

De bijbehorende waarde is vluchten. **2pt**

Dus één band gaat met kans 0,98305 minstens 51 vluchten mee en drie van zulke banden gaan minstens 51 vluchten mee met kans 0,95.

**Opgave 3 (Totaal 20 punten).** Bij Defensie solliciteerden in 2021 14.500 personen naar 3.600 beschikbare functies als militair. In de onderstaande tabel is voor een aantal functies te zien hoeveel sollicitanten zich per functie hadden aangemeld.

|  |  |
| --- | --- |
| **Sollicitanten per functie** | **Frequentie** |
| 0-1 | 10 |
| 2-3 | 32 |
| 4-5 | 36 |
| 6-7 | 15 |
| 8-9 | 5 |
|  | 2 |

**3a [5pt]** Gebruik **de gegevens uit de tabel** om het aantal sollicitanten en het aantal functies te berekenen/schatten waarop deze tabel is gebaseerd. Bereken ook het gemiddeld aantal sollicitanten per functie en de steekproefstandaarddeviatie.

De som van de frequenties is het aantal functies: 10 + 32 + 36 + 15 + 5 + 2 = **100** functies.

Het aantal gesprekken kun je niet exact bepalen, maar is ongeveer sollicitanten, we hebben daarbij de gemiddelde waarden van de gegeven ranges gebruikt. **2pt**

Het gemiddeld aantal sollicitanten per functie is dus 407/100 = **4,07** **1pt**

**2pt**

**3b [10pt]** Toets aan de hand van de steekproef in de tabel of het aantal sollicitanten per functie is te beschouwen als een kansvariabele die Poisson verdeeld is met gemiddelde waarde van 4 sollicitanten per functie. Voer de toets uit door middel van uitrekenen van een -waarde. Kies als betrouwbaarheid 95% en gebruik in je berekening de verwachte frequenties in één decimaal nauwkeurig.

|  |  |  |
| --- | --- | --- |
| **Gesprekken**  **per week** | **Frequentie** | **Verwachte**  **frequentie** |
| 0-1 | 10 | 100\*(Poissonpdf(4 , 0)+ Poissonpdf(4 , 1)) = 9,2  Of: 100\* Poisson**cdf**(4 , 1)= 9,2 |
| 2-3 | 32 | 100\*(Poissonpdf(4 , 2)+ Poissonpdf(4 , 3)) = 34,2  100\*(Poisson**cdf**(4 , 3) - Poisson**cdf**(4 , 1)) = 34,2 |
| 4-5 | 36 | 100\*(Poissonpdf(4 , 4)+ Poissonpdf(4 , 5)) = 35,2 |
| 6-7 | 15 | 100\*(Poissonpdf(4 , 6)+ Poissonpdf(4 , 7)) = 16,4 |
| 8-9 | 5 | 100\*(Poissonpdf(4 , 8)+ Poissonpdf(4 , 9)) = 4,3 |
|  | 2 | 100\*(1- Poisson**cdf**(4 , 9)) = 0,8 |

**4pt**

De laatste verwachte frequenties zijn te klein, dus voeg de laatste twee rijen samen:

|  |  |  |
| --- | --- | --- |
| **Gesprekken**  **per week** | **Geobserveerde**  **frequentie** | **Verwachte**  **frequentie** |
| 0-1 | 10 | 9,2 |
| 2-3 | 32 | 34,2 |
| 4-5 | 36 | 35,2 |
| 6-7 | 15 | 16,4 |
|  | 7 | 5,1 |

**2pt**

Kijken of Ei en Oi voldoende op elkaar lijken doen we met een aanpassingstoets. De toetsingsgrootheid is

**2pt**

We toetsen hiermee

H0: De waargenomen frequenties kunnen worden verklaard met een Poissonverdeling met .

H1: De waargenomen frequenties kunnen niet zo worden verklaard.

De -waarde (met vrijheidsgraden) is:

**1pt**

Deze kans is niet kleiner dan dus H0 wordt niet verworpen, dus de toets geeft geen reden waarom de tabel niet zou kunnen worden verklaard met een Poissonverdeling met . **1pt**

**3c [5pt]** Voer de toets ook uit door berekening van het kritieke gebied.

Je kunt ook met een kritiek gebied en grenswaarde werken, dan moet je met de GR oplossen

**3pt**

Dat geeft **1pt**

De waarde ligt niet in het kritieke gebied , dus H0 wordt niet verworpen. **1pt**

**Opgave 4 (Totaal 25 punten).** In een recent wetenschappelijk artikel (*Education as a Pathway to Help Job Searching Military Veterans’ Transition to Civilian Life*) worden tabellen van 56 US *combat veterans* en 43 *non-combat veterans* gegeven die zijn uitgesplitst naar of ze na hun militaire carrière wel of niet actief op zoek zijn naar een baan en of ze momenteel wel of niet een opleiding volgen:

# 

**4a [10pt]** Onder de linker tabel wordt een Chi-Square waarde en een gegeven. Reken deze twee waarden na in minstens drie decimalen nauwkeurig, d.w.z. maak de tabel met *expected* waarden, bereken de waarde van en vervolgens de overschrijdingskans .

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Observed | | | |  | Expected | | | |
|  | School | No school | Totaal |  |  | School | No school | Totaal |
| Search | 19 | 1 | 20 |  | Search | **14,2857** | **5,7143** | 20 |
| No search | 21 | 15 | 36 |  | No search | **25,7143** | **10,2857** | 36 |
| Totaal | 40 | 16 | 56 |  | Totaal | 40 | 16 | 56 |

**4pt**

**3pt**

**3pt**

**4b [5pt]** In beide tabellen in Table 4 komen waarden voor die kleiner zijn dan 5. Leg uit waarom de -test wel toepasbaar is in de linker tabel (*combat veterans*) maar niet in de rechter (*non-combat veterans*).

In de linker tabel is de expected waarde 20\*16/56 = 5,7143 ≥ 5 dus -test wel toepasbaar. **3pt**

In de rechter tabel is de expected waarde 17\*8/43 = 3,1628 < 5 dus -test niet toepasbaar.

**2pt**

**4c [3pt]** De tabel hoort bij de volgende onderzoeksvraag:

RQ2 stated that combat veterans and non-combat veterans who are actively searching for a new job are more

likely to be currently going to school to further their education than those who are not actively searching. The

results are shown in Table 4. The Chi-square values show that there is statistical support for combat veterans, but

not non-combat veterans. Thus RQ2 was only partially supported.

Leg uit of en waarom je met de -toets de beweringen die ze doen kunt toetsen. Wat test deze -toets eigenlijk?

De -toets test of de twee variabelen onafhankelijk zijn. Bij verwerpen van H0 is de conclusie dat er een afhankelijkheid is tussen wel of niet naar school gaan en wel of niet actief op zoek zijn van een baan. Uit de tabel zie je vervolgens dat onder degenen die actief naar een baan zoeken (20) er meer (19) dan verwacht (14,2857) naar school gaan. De conclusie is dus terecht voor de linker tabel. Voor de rechter tabel kunnen ze eigenlijk geen conclusie trekken (zie c). **3pt**

**4d [2pt]** voor het testen van 2x2 kruistabellen bestaat er Fischer’s exacte toets die niet de beperking heeft voor kleine verwachte waarden zoals de -toets. Fischer’s toets geeft voor de twee tabellen respectievelijk de waarden en . Wat betekent dit voor de conclusies die de auteurs onder 4c trekken?

Dit bevestigt de conclusie voor de linker tabel, en nu ook voor de rechter tabel. **2pt**

**4e [5pt]** Als je de twee tabellen samenvoegt tot één tabel ontstaat er een tabel voor alle veteranen. Voor deze tabel blijkt te gelden. Wat is nu de conclusie die je hieruit kunt trekken? Is deze conclusie sterker of zwakker dan die van het artikel?

Je kunt daaruit concluderen dat de conclusie voor alle veteranen geldt. **3pt**

Deze conclusie is zwakker dan de oorspronkelijke omdat die zegt dat de conclusie wel voor de combatants geldt, maar niet voor de non-combatants. **2pt**