**Tentamen Statistiek MBW/KW deel 2, 2e kans 28-11-2022**

Afdeling: Propedeuse MBW/KW 2021-2022

Examinator: Dr. J.B.M. Melissen, Dr. R.J. Nijboer

Datum: maandag 28 november 2022 09:00 – 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 (25, 20, 30, 25 punten). Score = Puntentotaal/10

**Opgave 1 (Totaal 25 punten)**

De Koninklijke Landmacht heeft 3000 DAF YA-4442 viertonners in gebruik. Van een steekproef van 12 viertonners is in een week het aantal gereden kilometers geregistreerd, dit leverde de volgende aantallen kilometers: 250, 135, 540, 35, 125, 785, 0, 520, 255, 185, 1240, 470.

**1a [5pt]** Bereken van de gemeten waarden het steekproefgemiddelde en de steekproefstandaarddeviatie van het gereden aantal kilometers van een viertonner.

km **3pt**

km **2pt. (Gedeeld door i.p.v. , 1372,702 i.p.v. s: 1pt)**

**1b [5pt]** Bereken op grond van bovengenoemde steekproef een 90% betrouwbaarheidsinterval voor het verwachte aantal wekelijkse gereden kilometers van een viertonner, zonder daarbij gebruik te maken van de optie TESTS/Interval van de grafische rekenmachine. Rond de grenzen van dit interval af op veelvouden van 5 kilometers en wel zodanig dat de betrouwbaarheid gewaarborgd blijft.

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 wekelijks gereden kilometers van een truck is normaal verdeeld met een geschat gemiddelde van en een standaarddeviatie van .

**3pt**

Afronden: .

**1c [5pt]** Bereken op grond van bovengenoemde steekproef en met behulp van de -verdeling een 90% betrouwbaarheidsinterval voor de standaarddeviatie van het verwachte wekelijkse aantal gereden kilometers van een viertonner, zonder daarbij gebruik te maken van de optie TESTS/Interval van de grafische rekenmachine. Rond de grenzen van dit interval af op gehele kilometers en wel zodanig dat de betrouwbaarheid gewaarborgd blijft.

met betrouwbaarheid . ,

Los op met de GR: en vind =4,5748.

Los op en vind .

Schattingsinterval is

**Afgerond: [268 , 557]**

**1d [5pt]** Bereken een 90% betrouwbaarheidsinterval voor het aantal banden dat jaarlijks door de KL vervangen moet worden voor de 3000 viertonners. Neem daarbij aan dat er 50 weken in een jaar zijn, dat een viertonner op zes banden rijdt en dat banden na 20.000 gereden kilometers worden vervangen.

Neem aan dat het aantal kilometers dat elke vrachtwagen in een week rijdt, normaal verdeeld is met een vast (onbekend) gemiddelde en standaarddeviatie (dit is een aanname, omdat het de mogelijkheid uitsluit dat er vrachtwagens zijn die gedurende een jaar intensief worden gebruikt en vrachtwagens die nauwelijks worden gebruikt). Deze aanname leidt tot de conclusie dat het aantal kilometers dat 3000 vrachtwagens per jaar rijden normaal verdeeld is met een gemiddelde van km met een standaarddeviatie van km. Het aantal banden dat bij 3000 trucks jaarlijks vervangen moet worden is dus normaal verdeeld met een gemiddelde van banden en een standaarddeviatie van banden.

Bij 90% betrouwbaarheid hoort een tweezijdige -waarde van . Het 90% betrouwbaarheidsinterval voor het jaarlijks aantal banden is dus

**1e [5pt]** Hoeveel banden moeten er **wekelijks** minimaal op voorraad zijn, wil met 99% zekerheid aan de wekelijkse behoefte kunnen worden voldaan? (Antwoord in gehele banden).  
het aantal kilometers dat 3000 vrachtwagens per week rijden is normaal verdeeld is met een gemiddelde van km met een standaarddeviatie van km. Het aantal banden dat bij 3000 trucks wekelijks vervangen moet worden is dus normaal verdeeld met een gemiddelde van banden en een standaarddeviatie van banden.

Bij 99% betrouwbaarheid hoort een enkelzijdige z-waarde van . Een 99% betrouwbaarheidsondergrens voor het wekelijks aantal banden is dus

banden, ofwel 3419 banden.

**Opgave 2 (Totaal 20 punten).** Bij een groot bedrijf zijn in een jaar (50 weken) 432 sollicitatiegesprekken gevoerd. In de onderstaande tabel is te zien hoeveel gesprekken er per week plaatsvonden.

|  |  |
| --- | --- |
| **Gesprekken**  **per week** | **Frequentie** |
| 0 - 4 | 4 |
| 5 - 9 | 26 |
| 10 - 14 | 18 |
| 15 - 19 | 2 |

**2a [5pt]** Laat door berekenen **met behulp van de gegevens uit de tabel** zien dat het gemiddelde aantal gesprekken per week waarop de tabel is gebaseerd gelijk is aan 8,80 (neem daarvoor de gemiddelde waarde per categorie). Laat door berekening zien dat de standaarddeviatie daarin gelijk is aan 3,43 gesprekken per week.

De som van de frequenties is het aantal weken: 4 + 26 + 15 + 18 + 2 = 50 weken.

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

Het gemiddeld aantal gesprekken per week is dus 440/50 = 8,80 **1pt**

3,46 **2pt**

Of

|  |  |
| --- | --- |
| **Gesprekken**  **per week** | **Frequentie** |
| 0 - 4 | 4 |
| 5 - 9 | 26 |
| 10 - 14 | 18 |
| 15 - 19 | 2 |

**2b [10pt]** Toets of het aantal gesprekken per week is te beschouwen als een kansvariabele die normaal verdeeld is met gemiddelde waarde en standaarddeviatie zoals uitgerekend in 2a. Voer de toets uit door middel van uitrekenen van een -waarde. Kies als betrouwbaarheid 92% en gebruik in je berekening de verwachte frequenties in één decimaal nauwkeurig.

|  |  |  |
| --- | --- | --- |
| **Gesprekken**  **per week** | **Frequentie** | **Verwachte**  **frequentie** |
| 0 - 4 | 4 | 50\*normalcdf(0 , 4.5 , 8.8 , 3.43) = 5,0818 |
| 5 - 9 | 26 | 50\*normalcdf(4.5 , 9.5 , 8.8 , 3.43) = 23.7935 |
| 10 - 14 | 18 | 50\*normalcdf(9.5 , 14.5 , 8.8 , 3.43) = 18,5435 |
| 15 - 19 | 2 | 50\*normalcdf(14.5 , 19.5 , 8.8 , 3.43) = 2,3685 |

De laatste verwachte frequentie is te klein, voeg de laatste twee rijen samen:

|  |  |  |
| --- | --- | --- |
| **Gesprekken**  **per week** | **Frequentie** | **Verwachte**  **frequentie** |
| 0 - 4 | 4 | 5,0818 |
| 5 - 9 | 26 | 23,7935 |
| 10 - 19 | 20 | 20,9110 |

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 normale verdeling met en .

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 tabel kan vanuit het oogpunt van deze toets worden verklaard met een normale verdeling.

**2c [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

Dat geeft De waarde ligt niet in het kritieke gebied , dus H0 wordt niet verworpen.

**Opgave 3 (Totaal 30 punten).** Onlangs namen 40 eerstejaars MBW en KW cadetten en adelborsten deel aan een herkansingstentamen Statistiek. Van de 32 voldoendes werden er 13 door een vrouw gehaald en er waren 6 mannen met een onvoldoende.

**3a [5pt]** Maak een kruistabel waarinde aantallen studenten zijn uitgesplitst naar geslacht en het wel of niet gehaald hebben van dit tentamen.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **voldoende** | **onvoldoende** | **Totaal** |
| **Vrouw** | **13** | **2** | **15** |
| **Man** | **19** | **6** | **25** |
| **totaal** | **32** | **8** | **40** |

**3b [10pt]** Zoek uit of je met behulp van een homogeniteitstoets de volgende hypothese met een betrouwbaarheid van 95% kunt verwerpen:

**H0: Het wel of niet slagen voor dit tentamen is onafhankelijk van het geslacht van de studenten.**

Houd hierbij geen rekening met het feit dat verwachte aantallen geslaagden mogelijk te laag zijn om deze toets toe te passen.

Als het wel of niet slagen onafhankelijk zou zijn van het geslacht, dan zou de tabel er als volgt uit zien:

|  |  |  |  |
| --- | --- | --- | --- |
|  | voldoende | onvoldoende | Totaal |
| Vrouw | 12 | 3 | 15 |
| Man | 20 | 5 | 25 |
| totaal | 32 | 8 | 40 |

De toetsingsgrootheid is67

Het aantal vrijheidsgraden is (2-1)\*(2-1) = 1, en de overschrijdingskans is

Deze kans is groter dan 0,05. Dit betekent dat de nulhypothese (er is geen afhankelijkheid) niet kan worden verworpen.

**3c [2pt]** Welke groep studenten levert de grootste bijdrage aan de conclusie in 3b?

De grootste bijdrage aan komt van onvoldoende vrouwen (meer dan gemiddeld), en daarna van de onvoldoende mannen (minder dan gemiddeld).

**3d [3pt]** L eg uit waarom het resultaat van de homogeniteitstoets uit 3b al dan niet valide is.

Niet valide omdat het verwachte aantal vrouwen met onvoldoende kleiner is dan 5.

**3e [5pt]** Stel dat in 3b de conclusie was dat H0 niet kan worden verworpen en dat dit op valide gronden was gebeurd, kun je dan met 95% zekerheid stellen dat het wel of niet slagen onafhankelijk is van het geslacht van de studenten? Leg uit.

Je kunt op grond van deze toets niet met 95% betrouwbaarheid zeggen dat er afhankelijkheid is, maar niet dat er onafhankelijkheid is.

**3f [5pt]** Bereken met de gegevens uit de tabel een 90% betrouwbaarheidsinterval voor het slaagpercentage van de vrouwelijke studenten.

# De vrouwen hebben een slaagfractie van

Het betrouwbaarheidsinterval is uit te rekenen met Clopper-Pearson, door op te lossen

en **2pt**

**2pt**

Dit levert als 90% betrouwbaarheidsinterval [0,6366 ; 0,9758] voor de slaagfractie, ofwel voor het slaagpercentage: [64% ; 98%]. **1pt**

# Opgave 4 (Totaal 25 punten)

Met een inzet van de Koninklijke Marechaussee in het hogere geweldsspectrum zijn vaak aanzienlijke kosten gemoeid. In de onderstaande tabel zijn voor vijf eerder inzetten de kosten in euro’s te zien.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **1** | **2** | **3** | **4** | **5** | **6** |
| **Duur inzet (uren)** | 1 | 3 | 6 | 3 | 5 | 5 |
| **Kosten (1000 euro's)** | 23 | 41 | 48 | 30 | 51 | 42 |

**4a [10pt]** Bereken met behulp van een tabel de correlatiecoëfficiënt van Pearson. Bepaal of er sprake is van een lineaire correlatie tussen de duur van de inzet en de bijbehorende kosten. Leg uit hoe daarbij het teken en de grootte van de berekende coëfficiënt een rol spelen.

De duur van de inzet is de verklarende variabele, die kies je als X, de kosten zijn het effect dat is Y.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| **1** | 1 | 23 | 23 | 1 | 529 |
| **2** | 3 | 41 | 123 | 9 | 1681 |
| **3** | 6 | 48 | 288 | 36 | 2304 |
| **4** | 3 | 30 | 90 | 9 | 900 |
| **5** | 5 | 51 | 255 | 25 | 2601 |
| **6** | 5 | 42 | 210 | 25 | 1764 |
| **Gem.** | **3,8333** | **39,1667** | **164,8333** | **17,50** | **1629,8333** |

**5pt**

De correlatiecoëfficiënt van Pearson is een getal tussen -1 en +1 dat aangeeft hoe goed twee variabelen aan een lineair verband voldoen. In dit geval is dat = duur van de inzet in uren en = de kosten in 1000 euro.

De correlatiecoëfficiënt is

**3pt**

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**

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**

**4b [8pt]** Bereken de regressielijn door berekening van de coëfficiënten en met behulp van een tabel.

De regressielijn is met

**4pt**

**3pt**

**1pt**

**4c [2pt]** Bereken met de regressielijn een statistisch verantwoorde voorspelling voor de te verwachten kosten voor een inzet die twee uur gaat duren.

Vul in en je krijgt een bijbehorende voorspelling van de kosten: k€. **2pt**

**4d [5pt]** Bereken een 90% voorspellingsinterval voor de waarde die in 4c is berekend.

Het voorspellingsinterval is

is de -waarde die hoort bij de betrouwbaarheid van 90% met vrijheidsgraden. Bij een betrouwbaarheid van 95% is de linker overschrijdingskans 0,90 + 0,10/2 = 0,95 en

**1pt**

**1pt**

**1pt**

**1pt**

**1pt**