**Tentamen Statistiek MBW/KW deel 2, finale kans 12-05-2023**

Afdeling: Propedeuse MBW/KW 2021-2022

Examinator: Dr. J.B.M. Melissen

Datum: vrijdag 12 mei 2023 13:00 – 15:00. 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 (20, 25, 30, 25 punten). Score = Puntentotaal/10

**Opgave 1 (Totaal 20 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 aantal vluchten worden vervangen. Hier volgt een lijstje van aantal vluchten waarna 10 banden moesten worden vervangen: 250, 171, 302, 119, 234, 272, 112, 262, 203, 189 vluchten.

**1a [5pt]** Bereken van de gemeten waarden het steekproefgemiddelde en de steekproefstandaarddeviatie.

vluchten **2pt**

vluchten **3pt**

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

**1b [5pt]** Bereken op grond van de bovengenoemde steekproef een 90% betrouwbaarheidsinterval voor het verwachte aantal operationele vluchten van een band, 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.

Standaard deviatie 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 .

Het 90% schattingsinterval voor is dan

**3pt**

Afronden: .

**1c [5pt]** Bereken op grond van bovengenoemde steekproef en met behulp van de -verdeling een 90% betrouwbaarheidsinterval (zie formuleblad) voor de standaarddeviatie van het verwachte 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 =4,5748.

Los op en vind .

Schattingsinterval is

**Afgerond: [268 , 557]**

**1d [5pt]** Een F35-Lightning II (Joint Strike Fighter) wordt voorzien van 3 nieuwe banden. Bereken het maximale aantal vluchten dat deze F35 met 95% zekerheid kan vliegen zonder dat er een band vervangen hoeft te worden. (Hint: als één band met kans 0,98305 vluchten kan uitvoeren zonder vervangen te worden, wat is dan de kans dat drie banden vluchten kunnen uitvoeren zonder vervangen te worden?)

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

**Opgave 2 (Totaal 25 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 |
|  | 7 |

**2a [10pt]** Gebruik **de gegevens uit de tabel** om het aantal sollicitanten en het aantal functies te berekenen of te schatten waarop deze tabel is gebaseerd. Bereken ook het gemiddeld aantal sollicitanten per functie en de steekproefstandaarddeviatie daarin.

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

**2b [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 - 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).** Vorig jaar namen 51 eerstejaars MBW en KW cadetten en adelborsten deel aan het herkansingstentamen Statistiek deel 1. In de volgende tabel zijn de aantallen behaalde cijfers uitgesplitst naar beginletter van de achternaam. Gebruik in deze opgave een significantieniveau van 95%.

|  |  |  |
| --- | --- | --- |
|  | **Achternaam A-M** | **Achternaam N-Z** |
| **Cijfer <5** | **8** | **4** |
| **Cijfer van 5 t/m 6** | **6** | **8** |
| **Cijfer >6** | **13** | **12** |

**3a [10pt]** Pas een homogeniteitstoets toe om vast te stellen of er een afhankelijkheid is tussen de hoogte van het tentamencijfer en de plaats van de achternaam in het alfabet.

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.

**3b [5pt]** Een onderwijskundige beweert dat een optimale verdeling van de cijfers van een tentamen zou zijn een normale verdeling met en . Bereken voor elk van de drie categorieën (cijfer in [0,5], [5,6] of [6,10]) wat de kans is om met het behaalde cijfer in deze categorie te eindigen als deze optimale verdeling zou gelden.

**3c [5pt]** Test met behulp van een aanpassingstoets en het antwoord van 3b of de verdeling van de cijfers in de tabel voldoet aan de aanname van de onderwijskundige.

**3d [10pt]** Bereken met de gegevens uit de tabel een 95% betrouwbaarheidsinterval voor het percentage van eerstejaars studenten waarvan de beginletter van de achternaam begint met A t/m M.

# 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)

Het Oekraïens Ministerie van Defensie geeft dagelijks een schatting van het aantal gedode Russische militairen. De opgegeven aantallen op de eerste van de maand staan in de volgende tabel:

|  |  |  |
| --- | --- | --- |
| **Datum** | **X** | **Y=Aantal gedode Russen** |
| 1-09-2022 | 1 | 350 |
| 1-10-2022 | 2 | 530 |
| 1-11-2022 | 3 | 650 |
| 1-12-2022 | 4 | 560 |
| 1-01-2023 | 5 | 760 |
| 1-02-2023 | 6 | 920 |
| 1-03-2023 | 7 | 550 |

**4a [10pt]** Bereken met behulp van een tabel de correlatiecoëfficiënt van Pearson. Bepaal of er sprake is van een lineaire correlatie tussen de maand en het aantal opgegeven doden op de eerste van de maand.

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 het opgegeven aantal gedode Russische militairen op 1 juni 2023.

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

**4d [5pt]** Bereken een 90% voorspellingsinterval voor de voorspelling 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**