**Tentamen Statistiek KW/MBW (deel 2, eerste kans)**

Afdeling: Propedeuse KW/MBW 2019-2020

Examinator: Dr. J.B.M. Melissen

Datum: 5 juni 2020, 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. **Lever de antwoorden in op het geprinte antwoordformulier (zet je naam erop), de berekeningen en uitleg op gelinieerd papier.**

7. **De opgaven dienen na afloop van het tentamen ingeleverd te worden.**

Dit tentamen bestaat uit vier opgaven (20, 20, 20, 30 punten). Score = Puntentotaal/9

**Opgave 1 (Totaal 20 punten)**

Tijdens oefeningen wordt gebruik gemaakt van standaard NATO 24-uursrantsoenen. Deze zijn per 12 stuks verpakt in kartonnen dozen. In een depot van de logistieke dienst van waaruit deze dozen geleverd worden is van een aantal voorgaande maanden het aantal uitgeleverde dozen bijgehouden: 738, 621, 810, 389, 470, 622, 515. Neem aan dat de aantallen normaal verdeeld zijn. Uit jarenlange bedrijfsvoering is bekend dat de standaarddeviatie in goede benadering gelijk is aan dozen per maand.

**1a. [10pt]** Bereken een 90% betrouwbaarheidsinterval voor op grond van bovengenoemde steekproef, zonder daarbij gebruik te maken van de optie TESTS van de grafische rekenmachine. Rond de grenzen van dit interval af op gehele getallen en wel zodanig dat de 90% betrouwbaarheid gewaarborgd blijft.

Het betrouwbaarheidsinterval met betrouwbaarheid voor bij gegeven wordt bepaald door te schatten met het steekproefgemiddelde van meetwaarden en dat als midden te nemen van een interval met linker- en rechteroverschrijdingskansen :

Bepaal eerst de -waarde bij een rechteroverschrijdingskans van . Het gevraagde interval is dan , **2pt**

want de standaarddeviatie van het gemiddelde van waarden is .

In ons geval is en . **2pt**

Bij hoort bij een rechteroverschrijdingskans de waarde

, **2pt**

Dus het interval is . **2pt**

Afronden op gehele getallen naar buiten (zodat het interval hooguit groter wordt):

(501 , 689). **2pt**

**1b. [5pt]** De commandant van het depot vindt het zojuist bepaalde interval te ruim, hij streeft naar een interval met een lengte van hoogstens 100 bij gelijkblijvende betrouwbaarheid. Bereken voor hoeveel maanden de uitgeleverde hoeveelheid pakjes bekend moet zijn om aan deze eis te voldoen. Als het niet mogelijk is, leg dan uit waarom dit niet kan.

De breedte van het betrouwbaarheidsinterval is . **1pt**

Door de steekproefgrootte te vergroten kan de breedte van het interval willekeurig klein worden gemaakt. **1pt**

De breedte van het interval moet zijn . **1pt**

Hieruit volgt dat . Naar boven afronden voor alle zekerheid: . **2pt**

**1c. [5pt]** Toets: tegen . Bepaal de toetsuitslag via het berekenen van een kritiek gebied. Kies hierbij α = 0,05.

De toetsingsgrootheid is normaal verdeeld met gemiddelde μ = 700 (volgens de worst case situatie van de aanname ) en standaarddeviatie . **2pt**

Het kritieke gebied is met . **2pt**

Nu ligt de gemeten waarde in dit kritieke gebied, dus wordt verworpen. **1pt**

**Opgave 2 (Totaal 20 punten)**

De Rijksdienst voor het Wegverkeer doet onderzoek naar het aantal geregistreerde auto’s zonder WA verzekering.

**2a. [10pt]** Bij een steekproef van 800 auto’s bleken 69 auto’s niet WA verzekerd te zijn. Bereken een 98% betrouwbaarheidsinterval voor de fractie onverzekerde auto’s.

Auto’s zijn wel of niet verzekerd met een vaste kans, dus we hebben te maken met een binomiale verdeling. We berekenen (Clopper-Pearson, zie slides) twee extreme waarden , waarvoor

Stap 1: ofwel:

Stap 2: ofwel:

Met de GR-solver vinden we (let op dat de startwaarde tussen 0 en 1 ligt)

Het 98% betrouwbaarheidsinterval voor de fractie onverzekerde auto’s is dus:

**2b. [10pt]** Voor een steekproef van 28 onverzekerde auto’s werd de gemiddelde leeftijd bepaald: 8,9 jaar met een steekproefstandaarddeviatie van 1,6 jaar. Bereken een 98% betrouwbaarheidsinterval voor de leeftijd van onverzekerde auto’s.

We gaan voor de leeftijd van onverzekerde auto’s uit van een normale verdeling, maar omdat de standaarddeviatie geschat is m.b.v. een steekproef en moeten we de -verdeling gebruiken. Bereken de -waarde door op te lossen

Dit levert met de GR solver: . Het 98% betrouwbaarheidsinterval is dus

..

**Opgave 3 (Totaal 20 punten).**

**Opgave 3 (Totaal 20 punten).** Het aantal brandmeldingen per week in een bepaalde stad is gedurende 100 weken geregistreerd (zie tabel).

|  |  |
| --- | --- |
| **Branden**  **per week** | **Frequentie** |
| 0 | 24 |
| 1 | 40 |
| 2 | 15 |
| 3 | 8 |
| ≥ 4 | 4 |
| Totaal | 91 |

**3a. [10pt]** Toets of het aantal branden per week is te beschouwen als een kansvariabele die een Poissonverdeling volgt met , door middel van uitrekenen van een -waarde. Kies als betrouwbaarheid 98%.

**3a.** We berekenen eerst de frequenties zoals ze uit de Poissonverdeling zouden volgen:

,

,

,

Omdat de verwachte frequenties en te klein zijn voor toepassing van Chikwadraat (< 5) nemen we de laatste twee categorieën samen tot , waardoor de verwachting daarvan groter dan 5 wordt (zie tabel).

|  |  |  |
| --- | --- | --- |
| **Branden**  **per week** | **Frequentie**  **Observed** | **Frequentie**  **Expected** |
| 0 | 24 | 36,6 |
| 1 | 40 | 33,3 |
| 2 | 15 | 15,2 |
| ≥ 3 | 12 | 5,9 |
| Totaal | 91 | 91 |

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

We toetsen hiermee

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

H1: De waargenomen frequenties kunnen niet zo worden verklaard.

Dat kan het snelst door de -waarde uit te rekenen (met vrijheidsgraden):

Dit is kleiner dan dus H0 wordt verworpen, dus de tabel kan worden niet verklaard met de Poissonverdeling met met een betrouwbaarheid van 98%. Er zijn met name veel te veel weken met drie of meer branden.

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

3b. Je kunt ook met een kritiek gebied en grenswaarde werken, dan moet je met de GR oplossen

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

**3c. [5pt]** Leg uit waarom je zou toetsen of de Poissonverdeling met de metingen verklaart.

**3c.** Een Poissonverdeling beschrijft verschijnselen die niet vaak, maar wel geregeld en onafhankelijk van elkaar optreden per vaste periode. Dat geldt waarschijnlijk/meestal wel voor branden, als er geen pyromanen actief zijn of seizoenseffecten zijn (bos-, schoorsteen- en barbecue-branden), maar dat is een stad minder relevant of kan uitmiddelen. In 100 weken waren er 91 branden, dat is gemiddeld per week, dat verklaart .

# Opgave 4 (Totaal 30 punten)

In het kader van het verminderen van milieubelasting en het verbeteren van efficiëntie werd binnen Defensie de actie Paper Tiger gehouden. Tijdens deze actie werd gewerkt aan bewustwording bij het personeel en werden maatregelen genomen om het gebruik van papier in de bedrijfsvoering terug te dringen. Om het effect van de actie te evalueren werd bij zes depots de uitgifte van papier vlak voor en direct na de actie gemeten. In de tabel hieronder is het aantal pallets papier weergegeven dat in een maand door elk depot werd verstrekt.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Depot** | **1** | **2** | **3** | **4** | **5** | **6** |
| **Uitgifte voor Paper Tiger** | 20 | 26 | 18 | 29 | 28 | 22 |
| **Uitgifte na Paper Tiger** | 13 | 22 | 6 | 20 | 18 | 12 |

**4a [10pt]** Bereken met behulp van de tabel op het antwoordformulier de correlatiecoëfficiënt van Pearson. Bepaal of er sprake is van een lineaire correlatie tussen aantallen pallets papier dat door deze depots per maand vóór de actie werd verstrekt en het aantal daarna. Leg uit hoe daarbij het teken en de grootte van de berekende coëfficiënt een rol spelen.

4a.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Depot** |  |  |  |  |  |
| **1** | **20** | **13** | 260 | 400 | 169 |
| **2** | **26** | **22** | 572 | 676 | 484 |
| **3** | **18** | **6** | 108 | 324 | 36 |
| **4** | **29** | **20** | 580 | 841 | 400 |
| **5** | **28** | **17** | 476 | 784 | 289 |
| **6** | **23** | **12** | 276 | 529 | 144 |
| **Gem.** | 24 | 15 | 378,6667 | 592,3333 | 253,6667 |

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 = uitgifte van papier vóór het PT en = de uitgifte erna.

De correlatiecoëfficiënt is

want

De correlatiecoëfficiënt is positief, dus er is een positieve correlatie (d.w.z. bij een grotere uitgifte vóór PT hoort een grotere uitgifte erna, het lineaire verband tussen uitgifte voor en na is een rechte lijn die stijgend is.

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.

**4b [10pt]** In een zevende depot zijn in de maand vóór de actie 26 pallets papier uitgegeven. Bereken de regressielijn en bepaal hiermee een statistisch verantwoorde voorspelling van het aantal verstrekte pallets papier in dit depot in de maand na Paper Tiger. Rond je antwoord af op gehele pallets.

**4b.** De regressielijn is met

Vul verkoop vóór WK in: in en je krijgt een bijbehorende voorspelling na WK van 2857.

**4c [10pt]** Voer een homogeniteitstoets met betrouwbaarheid 95% uit om te bepalen of er onafhankelijkheid bestaat tussen de variabelen uitgifte voor/na Paper Tiger en het depot, door berekening van de -waarde.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Expected** | **1** | **2** | **3** | **4** | **5** | **6** | **Tot** |
| **Voor PT** | 20,1667 | 29,3333 | 14,6667 | 29,9444 | 28,1111 | 20,7778 | 143 |
| **Na PT** | 12,8333 | 18,6667 | 9,3333 | 19,0556 | 17,8889 | 13,2222 | 91 |
| **Totaal** | 33 | 48 | 24 | 49 | 46 | 34 | 234 |

0,00138 + 0,37878 + 0,75756 + 0,02978 + 0,00044 + 0,07189 +

0,00217 + 0,59523 + 1,19046 + 0,04680 + 0,00069 + 0,11297 = 3,1882

De p-waarde is niet kleiner dan 0,05, dus de H0 wordt niet verworpen, dus de variabelen zijn met 95% betrouwbaarheid onafhankelijk.