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

Afdeling: Propedeuse KW/MBW 2019-2020

Examinator: Dr. J.B.M. Melissen

Datum: 7 mei 2021, 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, 25, 30 punten). Score = Puntentotaal/10

**Opgave 1 (Totaal 25 punten)**

Tijdens een oefening wordt dagelijks bijgehouden hoeveel liter drinkwater wordt verbruikt. Dit leidt tot het volgende lijstje: 5215, 5889, 7021, 3983, 3974, 6003. Neem aan dat de aantallen normaal verdeeld zijn.

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

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

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

**1d. [5pt]** Hoeveel liter water moet er dagelijks minimaal op voorraad zijn wil met 99% zekerheid aan de dagelijkse behoefte kunnen worden voldaan?

**Opgave 2 (Totaal 20 punten).** Het aantal acute meldingen per week waarvoor de Explosieven Opruimingsdienst moest uitrukken is gedurende 40 weken geregistreerd (zie tabel).

|  |  |
| --- | --- |
| **Meldingen**  **per week** | **Frequentie** |
| 0 | 10 |
| 1 | 16 |
| 2 | 9 |
| 3 | 5 |
| ≥ 4 | 2 |
| Totaal | 42 |

**2a. [4pt]** Leg uit waarom het aannemelijk is dat aantal meldingen per week beschreven kan worden door een Poissonverdeling. Welke waarde van kies je daarbij?

**2b. [10pt]** Toets of het aantal meldingen per week is te beschouwen als een kansvariabele die een Poissonverdeling volgt. Doe deze toetsing door middel van uitrekenen van een -waarde. Kies als betrouwbaarheid 97%.

**2c. [6pt]** Bereken een 95% betrouwbaarheidsinterval voor .

**Opgave 3 (Totaal 25 punten).** Uit gegevens van het RIVM is een overzicht gemaakt van de vijftigplussers die tot nu toe zijn overleden aan corona (drie kolommen: Overleden man / vrouw / totaal), uitgesplitst naar geslacht en leeftijd.

Verder is uit gegevens van het CBS op basis van de overleden totalen per leeftijdscategorie (kolom Overleden totaal) een schatting gemaakt over hoe deze totalen normaal gesproken (buiten corona) zouden zijn verdeeld over mannen en vrouwen (zie kolommen Overleden verwacht m / v).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Leeftijd** | **Overleden**  **man** | **Overleden**  **vrouw** | **Overleden**  **totaal** | **Overleden**  **verwacht m** | **Overleden**  **verwacht v** |
| 50-59 | 246 | 124 | 370 | 217 | 153 |
| 60-69 | 820 | 463 | 1283 | 752 | 531 |
| 70-79 | 2774 | 1554 | 4326 | 2538 | 1788 |
| 80-89 | 4140 | 3389 | 7529 | 3259 | 4270 |
| 90-99 | 1335 | 2246 | 3581 | 1550 | 2031 |
| Totaal | 9315 | 7776 | 17089 | 8316 | 8773 |

**3a. 10pt]** Voer een homogeniteitsanalyse uit op de kolommen “Overleden man” en “Overleden vrouw”. Bereken daarvoor de waarde van en de -waarde.

**3b. [10pt]** Voer een aanpassingsanalyse uit op de kolommen “Overleden man” en “Overleden verwacht man”. Bereken daarvoor de waarde van en de -waarde.

**3c. [5pt]** Doe aan de hand van je berekeningen zo sterk mogelijke, maar statistisch verantwoorde uitspraken voor conclusies die je kunt trekken, **zowel 3a, als voor 3b**. Bijvoorbeeld (fictief):

* Het aantal vrouwen en mannen dat overlijdt aan corona is niet verschillend bij een significantieniveau van
* Er kan met een significantieniveau van geconcludeerd worden dat het overlijden door corona afhankelijk is van leeftijd.

**Opgave 4 (Totaal 30 punten)**

In de tabel hieronder is van zes studenten het eindexamencijfer wiskunde en het eindcijfer Statistiek vermeld.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Student** | **1** | **2** | **3** | **4** | **5** | **6** |
| **Eindexamencijfer wiskunde** | 8 | 6 | 7 | 6 | 9 | 8 |
| **Cijfer Statistiek** | 7,3 | 5,7 | 7,5 | 5,7 | 8,9 |  |

**4a [10pt]** Bereken handmatig de correlatiecoëfficiënt van Pearson. Bepaal of er sprake is van een lineaire correlatie tussen de twee cijfers.

**4b [10pt]** Bereken de regressielijn en bepaal hiermee een statistisch verantwoorde voorspelling van het verwachte Statistiekcijfer dat een student haalt als hij op zijn eindexamen een 5 zou hebben gehaald.

**4c [10pt]** Voer een homogeniteitstoets met betrouwbaarheid 95% uit om te bepalen of er afhankelijkheid bestaat tussen de twee cijfers, door berekening van de -waarde.

======== XXXXXXXX ========