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

Afdeling: Propedeuse KW/MBW 2019-2020

Examinator: Dr. J.B.M. Melissen

Datum: 15 juli 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 (30, 20, 25, 25 punten). Score = Puntentotaal/10

**Opgave 1 (Totaal 30 punten)**

Tijdens een landmachtoefening wordt een week lang dagelijks bijgehouden hoeveel liter drinkwater wordt verbruikt. Dit leidt tot de volgende waarden: 5215, 5879, 7021, 3983, 3974, 6003, 7972. Neem aan dat de hoeveelheden (kansvariabele die het waterverbruik per dag in liters voorstelt) normaal verdeeld zijn, elke dag met dezelfde verwachtingswaarde en standaarddeviatie.

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

**1b. [7pt]** Bereken een 96% betrouwbaarheidsinterval voor het verwachte dagelijkse waterverbruik van een dergelijke oefening, 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 96% betrouwbaarheid gewaarborgd blijft.

Omdat de standaarddeviatie niet bekend is (en bovendien de steekproefgrootte kleiner dan 30) moet met de -verdeling worden gewerkt. De -waarde bij 96% betrouwbaarheid voor een tweezijdig interval is (de linkeroverschrijdingskans van is dan 0,96 in het betrouwbaarheidsinterval + 0,02 in het interval links daarvan).

Het gemiddelde van 7 dagwaarden (steekproefgemiddelde) is normaal verdeeld met gemiddelde en standaarddeviatie . We gebruiken hiervoor de schattingen en en de -verdeling.

Het betrouwbaarheidsinterval van is dan

Afronden mag het interval niet kleiner maken, dus afronden naar buiten: .

**1c. [8pt]** Toets: tegen . Bepaal de toetsuitslag door het berekenen van een kritiek gebied op basis van de gegeven steekproef van zeven dagen waterverbruik. Kies nu als onbetrouwbaarheid α = 0,05.

Leg in simpele bewoordingen uit wat de uitslag van deze toets betekent voor het dagelijks waterverbruik.

We zoeken de (kleinste) grens van het kritieke gebied , zodanig dat de kans op en fout van de eerste soort (je verwerpt H0, terwijl H0 toch waar is) kleiner is dan α = 0,05:

is hierin het waterverbruik van één dag, dat is normaal verdeeld met gemiddelde (worst case situatie wanneer H0 geldt) en standaarddeviatie . Voor gebruiken de schatting en de -verdeling.

Er geldt dus , dus .

(Let op dat de berekening van de -waarde anders gaat dan in 1b., omdat daar het interval tweezijdig was en hier enkelzijdig).

De steekproefwaarde is groter dan deze waarde, ligt dus in het kritieke gebied en H0 wordt verworpen.

Dit betekent dat met 95% betrouwbaarheid kan worden gesteld dat het dagelijks waterverbruik minimaal 2830 liter is.

**1d. [5pt]** Hoeveel liter water moet er **dagelijks** minimaal op voorraad zijn wil met 98% zekerheid aan de dagelijkse behoefte kunnen worden voldaan? (Antwoord in gehele liters).  
Noem de gezochte minimale dagelijkse hoeveelheid , dan is de kans dat het werkelijke waterverbruik deze hoeveelheid voldoende is minstens 98%, dus

, dus, worst case is .

Nu is , het waterverbruik van één dag, normaal verdeeld met gemiddelde en standaarddeviatie . We gebruiken hiervoor de schattingen en en de -verdeling.

We zoeken een linkszijdig interval met kans 0,98. Dat correspondeert met een -waarde van , dus .

**1e. [5pt]** Hoeveel liter water moet er **op weekbasis** minimaal op voorraad zijn wil met 98% zekerheid aan de dagelijkse behoefte kunnen worden voldaan? (Antwoord in gehele liters).

Leg uit waarom deze hoeveelheid niet gelijk is aan zevenmaal de hoeveelheid die in 1d is berekend.

We noemen het waterverbruik van één week, dat is normaal verdeeld met gemiddelde en standaarddeviatie . We gebruiken hiervoor de schattingen en en de -verdeling.

Noem de gezochte minimale wekelijkse hoeveelheid , dan is de kans dat het werkelijke waterverbruik deze hoeveelheid voldoende is minstens 98%, dus

, dus, worst case is .

We zoeken een linkszijdig interval met kans 0,98. Dat correspondeert weer met een -waarde van , dus .

Zevenmaal de waarde van 1d zou veel teveel zijn omdat je nu alleen maar aan het eind van de week goed uit moet komen, i.p.v. elke dag.

**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 kritiek gebied. Kies als betrouwbaarheid 98%.

**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 en met mei 2021 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 en precies mogelijke, maar statistisch verantwoorde uitspraken voor conclusies die je kunt trekken, **zowel 3a, als voor 3b**. Bijvoorbeeld (fictief, dit zijn mogelijk niet de juiste formuleringen hier):

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

**Opgave 4 (Totaal 25 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 | 5,8 | 8,8 | 7,2 |

**4a [5pt]** 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 [5pt]** Voer een homogeniteitstoets met betrouwbaarheid 95% uit om te bepalen of er afhankelijkheid bestaat tussen de twee cijfers.

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