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

Afdeling: Propedeuse KW/MBW 2022-2023

Examinator: Dr. J.B.M. Melissen

Datum: 3 mei 2024, 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 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, 20, 30, 30 punten). Score = Puntentotaal/10

**Opgave 1 (Totaal 20 punten)**

In de PSU shop bedraagt het aantal uitgegeven T-shirts in zes opeenvolgende maanden 245, 178, 314, 223, 205, 188. Neem aan dat de maandelijks uitgeleverde aantallen normaal verdeeld zijn.

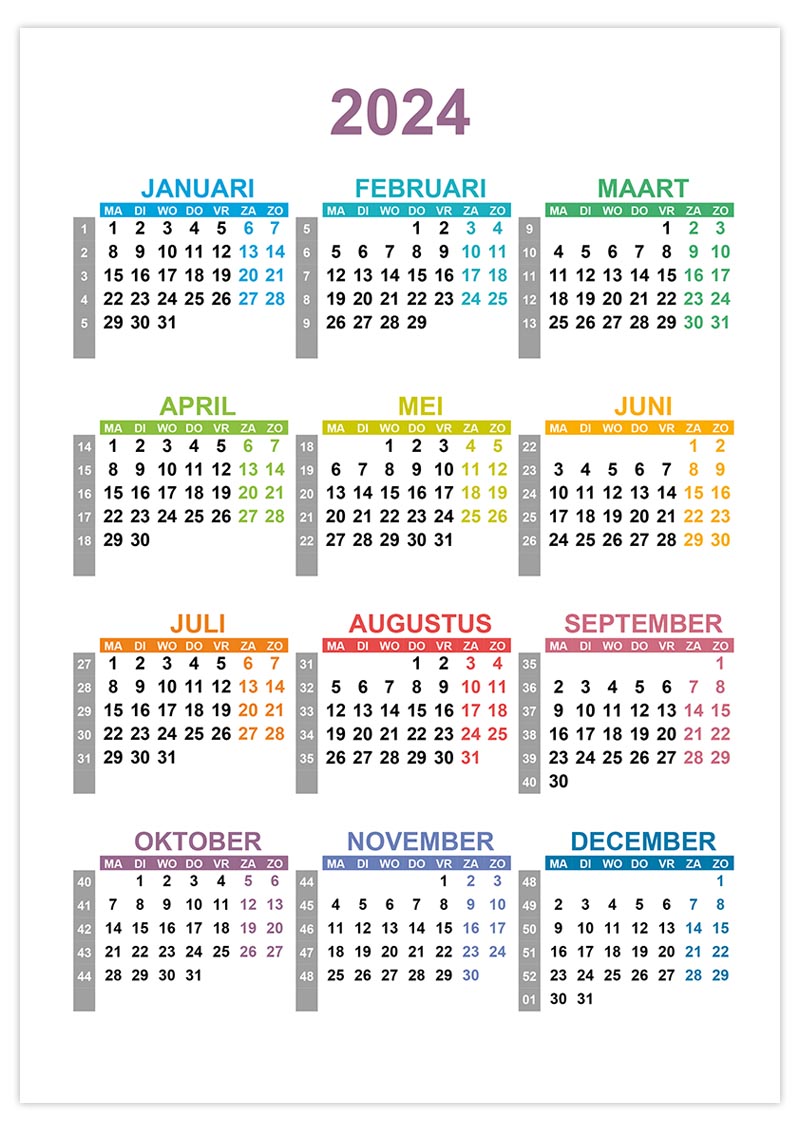
**1a. [6pt]** Bereken een 97% schattingsinterval voor de gemiddelde maandelijks uitgifte op grond van bovengenoemde steekproef, zonder daarbij gebruik te maken van de optie TESTS van de grafische rekenmachine. Rond de grenzen van dit interval af op gehele getallen en wel zodanig dat de 97% betrouwbaarheid gewaarborgd blijft.

**1b. [6pt]** Bereken voor de bovengrens uit 1a hoeveel maandvoorraad minimaal aanwezig moet zijn bij die waarde van om met 98% zekerheid aan de vraag te kunnen voldoen.

**1c. [8pt]** Toets: tegen . Bepaal de toetsuitslag via het berekenen van een kritiek gebied. Kies hierbij α = 0,05.

De steekproef zit niet in het kritieke gebied dus H0 kan niet worden verworpen.

**Opgave 2 (Totaal 20 punten)**



**2a. [3pt]** Bereken de kans dat een vrijdag in 2024 de dertiende van de maand blijkt te zijn.

Er zijn 52 vrijdagen, waarvan er 2 op de 13e vallen, dus kans 2/52 = 0,03846.

**2b.** **[3pt]** Bereken de kans dat een dertiende van een maand in 2024 op een vrijdag valt.

Er zijn 12 dertienden van een maand, waarvan er 2 op vrijdag vallen, dus kans 2/12 = 0,1667

**2c.** **[4pt]** Bereken de kans dat een dag in 2024 een vrijdag de dertiende is (Let op).

Er zijn 366 dagen (schrikkeljaar), dus kans 2/366=0,005464.

We nemen een steekproef van zeven **verschillende** dagen. Van deze zeven dagen blijkt er één op vrijdag de 13e te vallen.

**2d.** **[10pt]** Bereken op grond van deze steekproef een 95% schattingsinterval voor de kans dat een dag in 2024 op een vrijdag de dertiende valt.

Los op (Clopper-Pearson):

Levert:

**Opgave 3 (Totaal 30 punten)**

Een militaire eenheid voert patrouilles uit in een bepaald gebied om vijandelijke activiteiten te monitoren. Er wordt aangenomen dan het aantal vijandelijke ontmoetingen per dag een Poissonverdeling volgt. De eenheid heeft historische gegevens verzameld over het aantal ontmoetingen per dag gedurende een periode van 100 dagen. Het kwam 13 keer voor dat er geen ontmoeting op een dag was, 19 keer was er één ontmoeting, 26 keer twee ontmoetingen, 20 keer 3 ontmoetingen, 11 keer vier ontmoetingen en ook 11 keer vijf ontmoetingen.

**3a [4pt]** Bereken het gemiddeld aantal ontmoetingen per dag.

Het aantal ontmoetingen in 100 dagen is , dus het gemiddeld aantal per dag is

**3b [6pt]** Bereken een 95% betrouwbaarheidsinterval voor het gemiddeld aantal ontmoetingen per dag (zie formuleblad bij Poissonverdeling).

Los op: ()

Interval is:

**3c [12pt]** Voer een ()toets met een betrouwbaarheid van 90% uit om te verifiëren of de gemeten data overeenkomen met een Poissonverdeling met een gemiddelde waarde van μ= 2,2 ontmoetingen per dag.

|  |  |  |
| --- | --- | --- |
| Aantal | Observed | Expected |
| 0 | 13 | 11,08 |
| 1 | 19 | 24,37 |
| 2 | 26 | 26,81 |
| 3 | 20 | 19,66 |
| 4 | 11 | 10,81 |
| ≥5 | 11 | 7,25 |

Bereken

De overschrijdingskans is

Dit betekent dat H0 niet kan worden verworpen want p > 0,1.

**3d [8pt]** Laat door berekening zien dat de steekproefstandaarddeviatie van het aantal ontmoetingen per dag gelijk is aan 1,508. Voer een toets met een betrouwbaarheid van 90% uit om te verifiëren of de gemeten data overeenkomen met een normale verdeling met een gemiddelde waarde van μ = 2,3 ontmoetingen per dag en een standaarddeviatie van 1,5. Moet je hiervoor de normale verdeling of de t-verdeling gebruiken?

**Opgave 4 (Totaal 30 punten)**

Veilig Verkeer Nederland heeft onderzoek gedaan naar de invloed van alcohol op het bloedalcoholpromillage. Bij 10 mannen heeft men het promillage gemeten twee uur na het drinken van vijf glazen bier. Ook is het lichaamsgewicht gemeten. De resultaten staan in de onderstaande tabel.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Persoon** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| **Gewicht (kg)** | 80 | 70 | 61 | 75 | 73 | 82 | 90 | 86 | 67 | 96 |
| **Promillage (o/oo)** | 0,83 | 0,95 | 1,06 | 0,84 | 0,96 | 0,77 | 0,73 | 0,72 | 0,99 | 0,65 |

De volgende waarden zijn uitgerekend: . Hierin is het gewicht en het gemeten promillage.

Men wil weten of er een lineair verband is tussen gewicht en promillage.

**4a [2pt]** Wat is in dit geval de verklarende variabele en wat de te verklaren variabele?

Verklarende variabele (X): gewicht.  
Te verklaren variabele (Y): promillage

**4b [4pt]** Bereken de correlatiecoëfficiënt van Pearson. Hierbij mag je gebruik maken van de bovenstaande waarden.

De correlatiecoëfficiënt van Pearson is een getal tussen -1 en +1 dat aangeeft hoe goed twee variabelen aan een lineair verband voldoen.

De correlatiecoëfficiënt is in dit geval

**4pt**

**4c [2pt]** Bepaal of er sprake is van een lineaire correlatie tussen gewicht en promillage. Leg daarbij in woorden uit wat de betekenis is van de grootte en het teken van de correlatiecoëfficiënt.

Hoe dichter bij 1 (of -1), hoe beter de correlatie. In dit geval dus een redelijk sterke 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**

De correlatiecoëfficiënt is negatief, dus er is een negatieve correlatie (d.w.z. bij een hoger gewicht wordt een lager promillage voorspeld), het lineaire verband tussen gewicht en alcoholpromillage is een rechte lijn die dalend is. **1pt**

**4d [6pt]** Bereken de regressielijn door berekening van de coëfficiënten en met behulp van de tabel en de gegeven waarden.

De regressielijn is met

**3pt**

**2pt**

**1pt**

**4e [2pt]** Bereken met de regressielijn een voorspelling voor het promillage van een man van 70 kilo.

Vul in en je krijgt een bijbehorende voorspelling van de score: **2pt**

**4f [6pt]** Bereken een 90% voorspellingsinterval voor de voorspelling uit 4e. Doe de berekeningen in 4 decimalen. Rond de antwoorden af op 2 decimalen.

Het voorspellingsinterval is is de -waarde die hoort bij de betrouwbaarheid van 90% met vrijheidsgraden. Bij een betrouwbaarheid van 90% is de linker overschrijdingskans

en

**2pt**

**2pt**

**1pt**

**1pt**

**2pt**

(correct afronden, links naar beneden, recht naar boven) **2pt**

**4g [4pt]** Is het in 4e voorspelde promillage gelijk aan het promillage van de proefpersoon van 70kg in de tabel? Leg dit uit.   
Ligt het promillage van de proefpersoon van 70kg uit de tabel in het 90% voorspellingsinterval uit 4f? Licht dit toe.

Het voorspelde promillage is niet gelijk aan dat van de proefpersoon van 70kg, maar dat hoef ook niet, het is een voorspelling op grond van alle steekproefwaarden.

De waarde ligt wel in het voorspellingsinterval (dat zou in 90% van de gevallen moeten gelden, in dit geval is het zo.

**4h [4pt]** Uit onderzoek is gebleken dat het alcoholpromillage na het drinken van zes glazen bier bij mannen 18% hoger ligt dan na het drinken van vijf glazen bier. Stel de regressielijn voor dit geval op.

Vermenigvuldig de waarden met 1,18: