

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

Afdeling: Propedeuse MBW/KW 2020-2021

Examinator: Dr. J.B.M. Melissen

Datum: 23 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, 20, 30 punten) plus een bonusopgave met 10 bonuspunten). Score = Puntentotaal/10. De totale score kan de 10 niet overschrijden.

Opgave 1 (Totaal 30 punten)

In het kader van een NAVO missie zijn vier F16's van de Koninklijke Luchtmacht gestationeerd op de RAF luchtmachtbasis van Akrotiri op Cyprus om van daaruit dagelijks missies te vliegen boven brandhaarden in het Midden-Oosten. Een week lang wordt dagelijks de verbruikte hoeveelheid kerosine bijgehouden. Dit leidt tot de volgende waarden: 26.500, 22.300, 31.800, 19.900, 25.500, 29.200, 27.400 (lbs = Amerikaanse pound). Neem aan dat de hoeveelheden x (kansvariabele die het kerosineverbruik per dag in lbs voorstelt) normaal verdeeld zijn, elke dag met dezelfde verwachtingswaarde en standaarddeviatie.

1a. [4pt] Bereken van de gemeten waarden het steekproefgemiddelde.

Laat met behulp van een berekening zien dat de steekproefstandaarddeviatie gelijk is aan 4024,6857.

1b. [2pt] In welke eenheid moet de steekproefstandaarddeviatie worden uitgedrukt?

Opties:

- Geen, het is gewoon een getal, dimensieloos
- lbs net als μ
- lbs² want in de berekening wordt gekwadraterd
- liters, want het is een volume
- Kan niet worden bepaald op grond van de gegevens

1c. [6pt] Bereken een 90% betrouwbaarheidsinterval voor het verwachte dagelijkse kerosineverbruik μ , 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 veelvouden van 100 lbs en wel zodanig dat de 90% betrouwbaarheid behouden blijft.

1d. [8pt] Toets: $H_0: \mu \leq 23.800$ tegen $H_1: \mu > 23.800$. Bepaal de toetsuitslag door het berekenen van een kritiek gebied op basis van de gegeven steekproef van zeven dagen kerosineverbruik. Kies als onbetrouwbaarheid $\alpha = 0,10$.

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

1e. [5pt] Hoeveel kerosine moet er **dagelijks** minimaal op voorraad zijn voor het Nederlands contingent, wil met 99% zekerheid aan de dagelijkse behoefte kunnen worden voldaan? (Antwoord afgerond op honderden lbs).

1f. [5pt] Hoeveel kerosine moet er **op weekbasis** minimaal op voorraad zijn wil met 99% zekerheid aan de dagelijkse behoefte van de F16's kunnen worden voldaan? (Antwoord afgerond op honderden lbs).

Leg uit waarom deze hoeveelheid minder is dan zevenmaal de hoeveelheid die in 1e is berekend.

Opgave 2 (Totaal 20 punten). Het aantal ontplofte bermbommen per week in en om de door de terreurbeweging Talibom belaagde stad Al-â-Kabum is gedurende 100 weken geregistreerd (zie tabel).

IED's per week	Frequentie
0	35
1	33
2	18
3	3
≥ 4	2
Totaal	91



2a. [8pt] Toets of het aantal bermbomexplosies per week voldoet aan een Poissonverdeling met $\mu = 0,91$. Voer deze toets uit door het uitrekenen van een p -waarde. Kies als betrouwbaarheid 97%.

2b. [5pt] Voer de toets ook uit door berekening van het kritieke gebied.

2c. [2pt] Kan het zijn dat de aanpakken van 2a en 2b leiden tot verschillende conclusies? Leg uit!

2d. [5pt] Leg uit waarom de waarde $\mu = 0,91$ de meest geschikte is om te kiezen voor de Poissonverdeling.

Opgave 3 (Totaal 20 punten)

In 2017 werden 400 personen in een online testpanel bevraagd over hun politieke voorkeur en hun begrip voor maatregelen die toenmalig president Trump van de VS per twitter afkondigde. Voor het gemak wordt hierbij onderscheid gemaakt naar de politieke groeperingen PVV, VVD en Overig. De resultaten – gerangschikt naar politieke voorkeur en het wel/niet begrip tonen voor de decreten van Trump – staan hieronder vermeld:

	Wel begrip	Geen begrip	totaal
PVV	58	1	59
VVD	41	40	81
Overig	61	199	260
Totaal	160	240	400



3a. [10pt] Ga met een homogeniteitstoets na of er sprake is van samenhang tussen *politieke voorkeur* en *het wel/geen begrip hebben voor Trump*. Formuleer eerst een nulhypothese en een alternatieve hypothese. Bepaal de uitslag van de toets via een kritiek gebied. Kies $\alpha = 0,05$.

Volgens de Peilingwijzer kon de PVV rekenen op 19% aanhang, de VVD op 16%, terwijl de overige politieke groeperingen 65% voor hun rekening namen.

3b. [10pt] Is deze steekproef van 400 respondenten een afspiegeling van de verdeling die door de Peilingwijzer werd gemeten? Beantwoord de vraag door het uitvoeren van een aanpassingstoets. Maak gebruik een berekening van de p -waarde. Kies $\alpha = 0,05$.

Opgave 4 (Totaal 25 punten)

In een stad is de gemeente zich bewust van de toenemende vereenzaming bij ouderen. Daarom is men een actie begonnen, waarbij gemeentemedewerkers de wijken in trekken en daar bewoners aanspreken en voorlichten in de hoop dat zich vrijwilligers aanmelden die af en toe willen gaan buurten bij een contactbeperkte buurtgenoot. De actie heeft gelopen in acht wijken en de wethouder wil laten onderzoeken of er een verband is tussen het aantal uren dat zijn medewerkers in een wijk besteden en het aantal vrijwilligers dat dit oplevert. De resultaten tot nu toe staan in de volgende tabel. Hierbij gaat het om vrijwilligers die zich niet alleen hebben aangemeld, maar ook daadwerkelijk actief zijn met begeleiden.

Wijk	uren	vrijwilligers
1	38	8
2	40	8
3	44	10
4	47	12
5	52	15
6	35	7
7	46	12
8	50	14

4a [5pt] Bereken handmatig de correlatiecoëfficiënt van Pearson. Bepaal of er sprake is van een lineaire correlatie tussen de uren die door de gemeente werden besteed en het aantal vrijwilligers dat dit opleverde.

4b [7pt] Bereken de regressielijn en maak een schets van de puntenwolk en de regressielijn.

4c [3pt] De regressielijn voorspelt dat nul uren van de gemeente een negatief aantal vrijwilligers oplevert, m.a.w. niets doen kost vrijwilligers! Hoe valt dit uit te leggen? Probeer minstens twee verklaringen te vinden.

4d [3pt] Bereken een statistisch verantwoorde voorspelling van het verwachte aantal vrijwilligers dat zich aanmeldt als de gemeente 60 uur in een wijk aan voorlichting besteedt.

4e [7pt] Bereken een 95% voorspellingsinterval voor aantal vrijwilligers dat 60 uur werk van de gemeente waarschijnlijk gaat opleveren.

Bonusopgave 5 (Totaal 10 extra punten]

De cadet Wagesnich is erg geïnteresseerd in de handel met bitcoins. Hij wil kijken of hij kan voorspellen of de koers morgen gaat stijgen of dalen aan de hand van het gedrag van vandaag. Hij heeft daarvoor een lijst gemaakt met slotkoersen van de afgelopen 151 dagen (16 februari 2021 t/m 16 juli 2021). Hij gebruikt de slotkoersen om de procentuele stijging per dag te berekenen. Het blijkt dat in die periode de stijging per dag ligt tussen -14,4% en +11,8%.

Om verbanden te onderzoeken maakt hij drie groepen:

Daling: De dagen waarop koersdaling sterker is dan 2,5%. Dit geldt voor 37 dagen. De gemiddelde koersstijging op deze dagen is -5,76% per dag (een flinke daling dus).

Gelijk: De dagen waarop de waarde hooguit 2,5% verandert. Dit geldt voor 78 dagen. De gemiddelde koersstijging op deze dagen is -0,14% per dag (nauwelijks veranderd dus).

Stijging: De dagen waarop de koersstijging meer dan 2,5% is. Dit geldt voor 36 dagen. De gemiddelde koersstijging op deze dagen is +5,45% per dag (een flinke stijging dus).

De cadet Wagesnich heeft vervolgens de onderstaande tabel gemaakt, waarin te zien is hoe vaak de koers de volgende dag ging dalen/gelijk bleef/ging stijgen, tegenover dalen/gelijk blijven/stijgen op de huidige dag.

Het kwam bijvoorbeeld op **23** dagen voor dat een koersdaling tijdens een bepaalde dag werd gevolgd door het gelijk blijven op de dag erna.

Gemeten 16-02-2021 t/m 16-07-2021				
Vandaag↓\Morgen→	Daling	Gelijk	Stijging	Totaal
Daling	5	24	8	37
Gelijk	26	31	20	77
Stijging	6	23	7	36
Totaal	37	78	35	150

Wagesnich's peristaltiek (gut feeling) zegt hem dat er een verband zou moeten zijn tussen de koers van vandaag en die van morgen. Hij kan dat alleen niet aantonen, omdat hij tijdens de lessen Statistiek vaak onderhevig was aan heftige aanvallen van ongebalanceerde prioriteringen.

Verwachte waarden op basis van onafhankelijkheid				
Vandaag↓\Morgen→	Daling	Gelijk	Stijging	Totaal
Daling	9,1267	19,2400	8,6333	37
Gelijk	18,9933	40,0400	17,9667	77
Stijging	8,8800	18,7200	8,4000	36
Totaal	37	78	35	150

5a. Bepaal of met een betrouwbaarheid van 99% of er kan worden gesproken over een afhankelijkheid tussen verandering van de bitcoin koers op een dag en de daaropvolgende dag.

5b. Bereken voor daling, gelijk blijven en voor stijgen van de koers op een willekeurige dag telkens de verwachtingswaarde van procentuele stijging op de dag erna.

Op grond van de resultaten kiest Wagesnich voor de volgende taktiek:

1. Als hij bitcoins heeft verkoopt hij ze als de koers op een dag gelijk gebleven is.
2. Als hij geen bitcoins heeft koopt hij bitcoins als de koers op een dag gedaald is.

Hij berekent dat een investering van 1000 euro zonder handelen over de 151 dagen 648 euro opgeleverd had, terwijl het algoritme (dat hij voor zichzelf alvast het Wagenich-algoritme noemt) over de zelfde periode weliswaar eerst tot 848 euro zou zijn gedaald, om dan weer te stijgen tot een eindwaarde van 1169 euro. Wel dertig keer bitcoins kopen en verkopen, maar dan heb je ook wat.

===== XXXXXXXX =====