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

Afdeling: Propedeuse MBW/KW 2020-2021

Examinator: Dr. J.B.M. Melissen

Datum: 19 november 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!

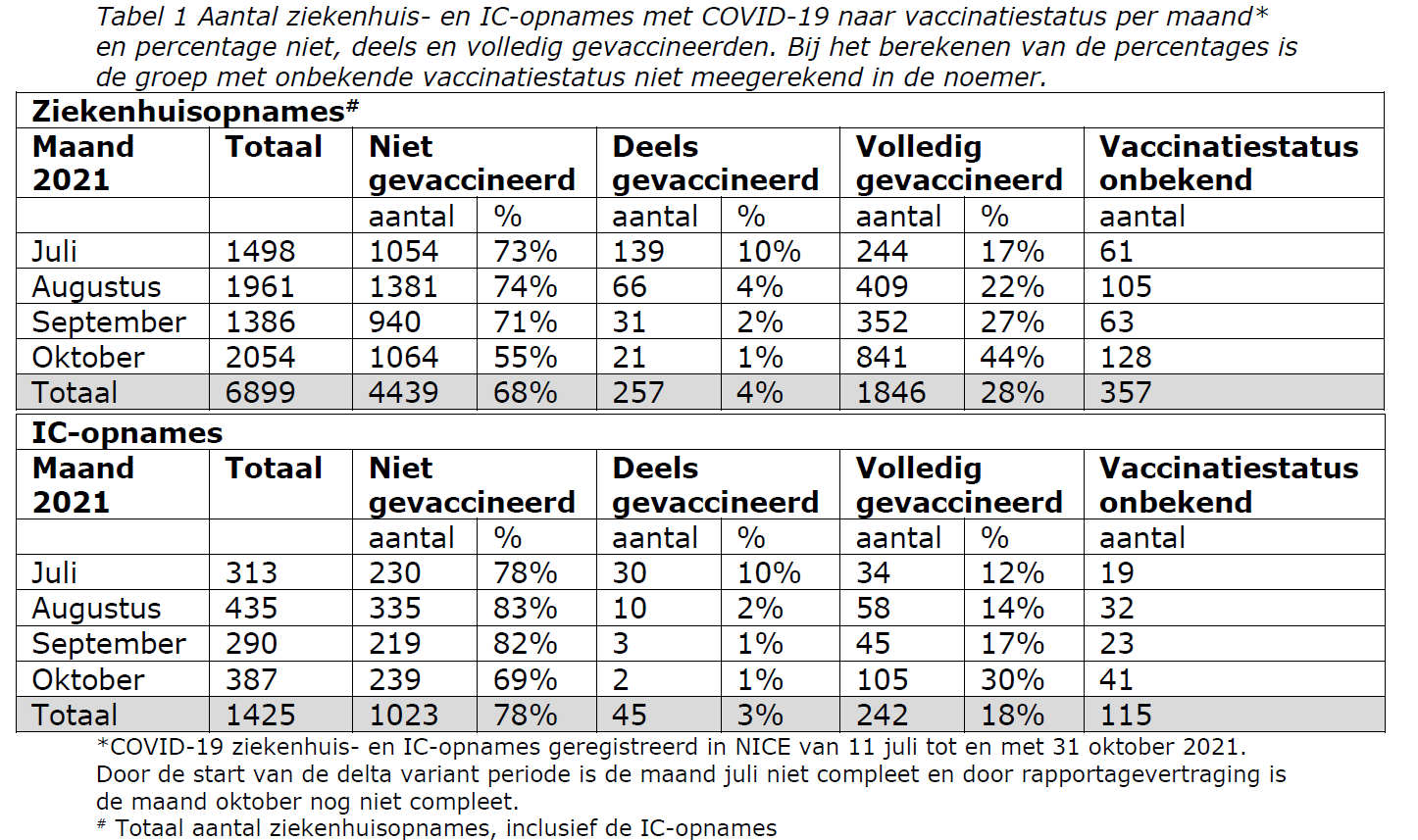
6. **De opgaven dienen na afloop van het tentamen ingeleverd te worden.**

Dit tentamen bestaat uit drie opgaven (55, 25, 20 punten). Score = Puntentotaal/10.

**Onderdelen 1e, 1f, 3c niet meegerekend Totaal 80pt**

**Opgave 1 (Totaal 55 punten)**

Op 3 november 2021 kwam het RIVM met een rapportage over de effectiviteit van COVID-19 vaccins over de periode 11 juli en 31 oktober 2021. De gegevens waren gebaseerd op de volgende gegevens:



In het rapport werden de volgende beweringen gedaan:

1. De COVID-19 vaccin-effectiviteit (VE) is in Nederland nog steeds hoog, namelijk 94% (95% BI 94-94) tegen ziekenhuisopname en 97% (95% BI 96-97) tegen intensive care (IC)-opname over de hele periode juli-oktober.
2. Op dit moment is 84 procent van iedereen van 18 jaar en ouder gevaccineerd.
3. De werkzaamheid van COVID-19 vaccinatie tegen ziekenhuisopname (94%) en IC-opname (97%) over de periode juli-oktober is hoog.
4. Dit betekent dat de kans voor volledig gevaccineerden om met COVID-19 in het ziekenhuis opgenomen te worden 17 keer kleiner is dan voor ongevaccineerden. De kans om op de IC opgenomen te worden is 33 keer kleiner.

De VE is de kans dat een gevaccineerde persoon in een bepaalde periode niet hoeft te worden opgenomen.

**1a. [10pt]** Bereken op grond van de gegevens uit Tabel 1 de vergelijking van de regressielijn voor de totale ziekenhuisopnamen door COVID per maand als functie van de maanden juli (1), augustus (2), september (3) en oktober (4) (gebruik deze nummers voor de maanden in je berekeningen).

Bereken ook de correlatiecoëfficiënt van deze lijn. Geef de coëfficiënten van de lijn en de correlatiecoëfficiënt in drie cijfers nauwkeurig.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| X | Y | X\*X | Y\*Y | X\*Y |
| 1 | 1498 | 1 | 2244004 | 1498 |
| 2 | 1961 | 4 | 3845521 | 3922 |
| 3 | 1386 | 9 | 1920996 | 4158 |
| 4 | 2054 | 16 | 4218916 | 8216 |
|  |  |  |  |  |
| 2,5 | 1724,75 | 7,5 | 3057359,25 | 4448,5 |
|  |  |  |  |  |

**3+3+4pt**

**1b. [10pt]** Gebruik de lijn (met het maandnummer zoals eerder genoemd, en de VE in procenten), om een 95% voorspellingsinterval te berekenen voor de VE in de maand november.

November levert

is de -waarde die hoort bij de opgegeven betrouwbaarheid met vrijheidsgraden.

**1+2+2+2+2+1pt**

**1c. [10pt]** Toets de bewering

**H0:** De vergelijking die werd genoemd in 1b geeft een goede weerspiegeling van de totale ziekenhuisopnamen over de maanden juli tot en met oktober.

Doe hiervoor een aanpassingstoets met een betrouwbaarheidsniveau van 97% en werk in de berekening met een kritiek gebied.

Formuleer een begrijpelijke conclusie van je berekening.

De kritieke waarde vind je door op te lossen

Dit levert . De waarde is

|  |  |  |  |
| --- | --- | --- | --- |
| X | Y=Oi | Ei | (Oi-Ei)^2/Ei |
| 1 | 1498 | 1561 | 2,542601 |
| 2 | 1961 | 1670 | 50,70719 |
| 3 | 1386 | 1779 | 86,81788 |
| 4 | 2054 | 1888 | 14,59534 |
|  |  |  |  |
|  |  |  | 154,663 |

. Die is veel groter dan , dus moet H0 worden verworpen. De lijn geeft geen goede weerspiegeling van de ziekenhuisopnamen. Dit bleek ook al uit de kleine waarde van de correlatiecoëfficiënt. **4+4+2pt**

**1d. [10pt]** Onderzoek met behulp van een homogeniteitstoets of er een afhankelijkheid bestaat tussen de volledig gevaccineerden met IC opname en de deels gevaccineerden met IC opname. Gebruik een betrouwbaarheidsniveau van 99% en doe de berekeningen met behulp van een -waarde.

Formuleer een begrijpelijke conclusie van je berekening.

Nulhypothese is dat de twee grootheden onafhankelijk zijn

|  |  |  |
| --- | --- | --- |
| 30 | 34 | **64** |
| 10 | 58 | **68** |
| 3 | 45 | **48** |
| 2 | 105 | **107** |
| **45** | **242** | **287** |

Expected tabel op grond van onafhankelijkheid:

|  |  |  |
| --- | --- | --- |
| 10,0348 | 53,9652 | **64** |
| 10,6620 | 57,3380 | **68** |
| 7,5261 | 40,4739 | **48** |
| 16,7770 | 90,2230 | **107** |
| **45** | **242** | **287** |

Alle verwachte frequenties zijn minstens 5, dus samenvoegen hoeft niet.

Berekening van :

|  |  |
| --- | --- |
| 39,72234 | 7,386386 |
| 0,041106 | 0,007644 |
| 2,721966 | 0,506151 |
| 13,01543 | 2,420224 |

Deze overschrijdingskans is kleiner dan 0,01, dus nulhypothese verwerpen, dus geen onafhankelijkheid, dus afhankelijkheid. Welke afhankelijkheid kun je niet zeggen. **1+3+2+2+2pt**

**1e. [10pt]** In Bewering 1 wordt een 95% betrouwbaarheidsinterval (BI) voor de VE voor ziekenhuisopname genoemd van (94, 94). Dit is merkwaardig, want dit is een interval met lengte 0, dat zou dus betekenen dat de werkelijke waarde van VE **gelijk is** aan 94% met een betrouwbaarheid van 95%. Bereken dit interval (gebruik de binomiale verdeling) en geef het resultaat in twee decimalen nauwkeurig.

**1f. [5pt]** In bewering 4 wordt gesteld dat de kans om op de IC te worden opgenomen voor volledig gevaccineerden 33 keer kleiner in dan voor ongevaccineerden. Dit zou betekenen dat

Ofwel:

Ga met behulp van de gegevens uit Tabel 1 of dit juist is of niet. (Je hebt tijdens je berekening het totale aantal inwoners van Nederland nodig, maar dat valt uiteindelijk weer weg, dus het maakt niet uit wat je daar doet). Doe hiervoor drie verschillende berekeningen:

1. Bekijk alleen de volledig gevaccineerden en de niet gevaccineerden.
2. Neem de Deels gevaccineerden en de Vaccinatiestatus onbekend wel mee en neem aan dat die allemaal niet gevaccineerd zijn.
3. Neem de Deels gevaccineerden en de Vaccinatiestatus onbekend wel mee en neem aan dat die allemaal wel gevaccineerd zijn.

Maak eventueel gebruik van Bewering 2 en dat de verdeling over de groepen Volledig gevaccineerd, Niet gevaccineerd, Deels gevaccineerd en Vaccinatiestatus is geconstateerd.

Trek een conclusie uit de drie getallen die je hebt berekend.

**Opgave 2 (Totaal 25 punten)**

De bottelier op een mijnenveger van de Koninklijke Marine is verantwoordelijk voor de bevoorrading van proviand. De mijnenveger heeft een bemanning van 40 opvarenden en de laatste acht maanden is het maandelijks verbruik van aardappels als volgt geweest:

400, 340, 390, 370, 370, 400, 320, 410 kg aardappels.

**2a. [5pt]** Bereken van de bekende waarden het steekproefgemiddelde en de steekproefstandaarddeviatie.

**2b. [2pt]** In welke eenheid zou de steekproef**variantie** moeten worden uitgedrukt?

Opties:

* Geen, gewoon een getal, dimensieloos
* In aardappels
* In kilo aardappels
* In kilo vierkante aardappels
* In vierkante kilo aardappels (kg2)

De variantie is het kwadraat van de standaarddeviatie, die heeft dezelfde eenheid als gemiddelde en meting (allebei kg) dus variantie is in vierkante kilo aardappels (kg2).

**2c. [8pt]** Bereken een 95% betrouwbaarheidsinterval voor het verwachte maandelijkse aardappelverbruik op grond van bovengenoemde steekproef, zonder daarbij gebruik te maken van de optie TESTS/Interval van de grafische rekenmachine. De aardappels worden geleverd in zakken van 10 kg, dus geef de grenzen van het interval in aantal zakken. Leg uit hoe en waarom je de grenzen hebt afgerond.

Er zijn 8 waarden gegeven, dus t-waarde gebruiken

Betrouwbaarheidsinterval is

in kg,

Of in zakken van 10 kg:

**2d. [5pt]** Toets: Elke maand zijn minstens 31 zakken van 10 kg nodig. Kies daarbij als onbetrouwbaarheid α = 0,05.

Leg in simpele bewoordingen uit wat de uitslag van deze toets betekent voor het maandelijks aardappelverbruik.

De overschrijdingskans is

Dit is kleiner dan 0.05, dus H0 wordt verworpen.

**2e. [5pt]** Hoeveel 10 kg zakken moeten er op een missie van twee maanden worden meegenomen om met 99% zekerheid te garanderen dat er zonder tussentijdse bevoorrading voldoende aardappels zijn?

Het aantal zakken voor twee maanden voldoet aan een t-verdeling met

Minimaal meenemen dus

Afronden op 89 10-kilo zakken

**Opgave 3 (Totaal 20 punten)**

Het wekelijks aantal meldingen van inbreuk op persoonlijke integriteit bij Defensie is gedurende een jaar (52 weken) geregistreerd (zie tabel, de gegevens zijn volstrekt fictief. Het werkelijk aantal integriteitsschendingen van het type “Ongewenste omgangsvormen” was volgens het Jaarverslag Integriteit Defensie 2020: 82).

|  |  |
| --- | --- |
| **Meldingen**  **per week** | **Frequentie** |
| 0 | 18 |
| 1 | 21 |
| 2 | 9 |
| 3 | 4 |
| ≥ 4 | 0 |

**3a. [5pt]** Hoeveel meldingen hebben in dit jaar plaatsgevonden?

**3b. [10pt]** Toets of het aantal meldingen per week is te beschouwen als een kansvariabele die een Poissonverdeling volgt met . Gebruik hierbij als betrouwbaarheid 97%.

BASEREN OP 52 WEKEN!!!!!!!

|  |  |  |  |
| --- | --- | --- | --- |
| **Meldingen**  **per week** | **Frequentie observed** | **Expected** |  |
| 0 | 18 | 51\*poissonpdf(1 , 0) = 18,7619 |  |
| 1 | 21 | 51\*poissonpdf(1 , 1) = 18,7619 |  |
| 2 | 9 | 51\*poissonpdf(1 , 2) = 9,3809 |  |
| 3 | 4 | 51\*poissonpdf(1 , 3) = 3,1269 |  |
| ≥ 4 | 0 | 51\*poissoncdf(1 , 0) = 0,9684 |  |

Als we met Chi-kwadraat willen toetsen of expected en observed dicht bij elkaar liggen of juist ver van elkaar, dan moeten de verwachte frequenties minimaal 5 zijn. We moeten dus de laatste drie kolommen samenvoegen (of 2 en 3 per week mag ook)

|  |  |  |  |
| --- | --- | --- | --- |
| **Meldingen**  **per week** | **Frequentie observed** | **Expected** |  |
| 0 | 18 | 18,7619 | 0,0309 |
| 1 | 21 | 18,7619 | 0,2670 |
| ≥ 2 | 13 | 13,4762 | 0,01683 |
|  |  |  | 0,3147 |

H0: De frequenties waarmee meldingen per week worden waargenomen kunnen worden verklaard met een Poissonverdeling met μ=1.

Overschrijdingskans

Deze kans is niet kleiner dan 0,03, dus de nulhypothese kan (op deze manier) niet worden verworpen, dus de conclusie is dat met 97% betrouwbaarheid er geen reden is om aan te nemen dat de frequenties niet kunnen worden verklaard door een Poissonverdeling met μ=1.

**3c. [5pt]** Leg uit waarop het niet onlogisch is te verwachten dat hier de Poissonverdeling met een rol zou kunnen spelen.

Het aantal meldingen is 51, het aantal weken waarin geobserveerd is, is 52. Als de meldingen onafhankelijk zijn dan is te verwachten dat dit beschreven kan worden door een Poissonverdeling met μ = 51/52 = 0,98. Dit getal ligt zo dicht bij 1, dat je verwacht dat μ=1 ook wel zal werken.

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