

Tentamen Statistiek MBW/KW deel 2 26-07-2024

Afdeling: Propedeuse MBW/KW 2023-2024

Examinatoren: Dr. J.B.M. Melissen, Dr. Ir. D.A.M.P. Blom

Datum: vrijdag 26 juli 2022 09:00 – 11:00, 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, 40, 15 punten). Score = Puntentotaal/10

Opgave 1 (Totaal 25 punten)

Op een luchtmachtbasis is met behulp van een steekproef van tien dagen bepaald dat het dagelijks brandstofverbruik op de basis gemiddeld 82.000 kg bedraagt met een standaarddeviatie van 10.000 kg. Er wordt aangenomen dat het dagelijks brandstofverbruik normaal verdeeld is.

1a [5pt] Bereken op grond van deze steekproef een 95% betrouwbaarheidsinterval voor het gemiddelde dagelijkse brandstofverbruik. Rond de grenzen van dit interval af op duizenden kilo's en wel zodanig dat de betrouwbaarheid gewaarborgd blijft.

1b [5pt] Bereken op grond van deze steekproef een 95% betrouwbaarheidsinterval voor standaarddeviatie in het dagelijkse brandstofverbruik. Maak hiervoor gebruik van de χ^2 -verdeling. Rond de grenzen van dit interval af op duizenden kilo's en wel zodanig dat de betrouwbaarheid gewaarborgd blijft.

1c [5pt] Toets: $H_0: \mu \leq 90.000$ tegen $H_1: \mu > 90.000$ (μ is het gemiddeld dagelijks brandstofverbruik). Bepaal de toetsuitslag door het berekenen van een kritiek gebied op basis van de gegeven steekproef van tien dagen brandstofverbruik. Kies als onbetrouwbaarheid $\alpha = 0,1$.

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

1d [5pt] Ga er vanuit dat de voorraad dagelijks wordt aangevuld. Hoeveel kg brandstof moet er elke dag minimaal op voorraad zijn, wil met 99% zekerheid aan de dagelijkse behoefte kunnen worden voldaan? (Antwoord in veelvoud van 1000 kg).

1e [5pt] Stel nu dat men een maandvoorraad (30 dagen) op de basis in voorraad wil hebben.

Hoeveel kg brandstof moet er dan minimaal op voorraad zijn zodat met 99% zekerheid aan de dagelijkse behoefte kan worden voldaan? (Antwoord in veelvoud van 1000 kg).

Opgave 2 (Totaal 20 punten). Van een klas studenten is tijdens een tentamen bijgehouden hoeveel blaadjes papier ze gebruiken:

Aantal blaadjes	Frequentie
0	0
1	3
2	4
3	8
4	7
5	3
6	2
7	1

2a [5pt] Bereken met behulp van de gegevens uit de tabel het aantal studenten en het aantal blaadjes tentamenpapier waarop de tabel is gebaseerd. Bereken ook het gemiddelde aantal blaadjes dat een student nodig had.

2b [10pt] Toets of het aantal blaadjes papier dat een student nodig heeft is te beschouwen als een kansvariabele die aan een Poissonverdeling voldoet, door middel van uitrekenen van een p -waarde. Kies als betrouwbaarheid 98% en gebruik in je berekening de verwachte frequenties in één decimaal nauwkeurig.

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

Opgave 3 (Totaal 40 punten)

In het artikel “*Using Linear Regression in the Context of Military Power Enhancement*” van Ivan Okromtchedlishvili (*Defence and Science* 2, 2023) wordt met lineaire regressie onderzocht of de militaire sterkte van een land afhangt van de volgende drie variabelen:

1. Defensieuitgaven in 2020 (*Defense expenditure*),
2. Defensieuitgaven in 2020 als percentage van het bruto binnenlands product (*Defense expenditure as percentage of real GDP*), en
3. De hoeveelheid militair personeel in 2020 (*Military personnel*).

Hiervoor worden de gegevens van 31 voornamelijk NAVO landen gebruikt. Hieronder zie je het eerste deel van de tabel die hierbij werd gebruikt.

Variables	Dependent		Independent		
Country	Military Strength Ranking for 2022	Military Strength Ranking for 2022 (reversed)	Defense expenditure in US dollars at current prices and exchange rates for 2020	Defense expenditure as a percentage of real GDP, for 2020	Military Personnel, for 2020
United States	0.0453	35.8958	784952000000	3.72	1346000
France	0.1283	5.7275	52727000000	2.03	208000
United Kingdom	0.1382	5.2184	61925000000	2.29	156200
Italy	0.1801	4.7871	26071000000	1.38	175500
Turkey	0.1961	4.4351	13396000000	1.86	437200
Germany	0.2322	4.3067	58902000000	1.55	186900
Spain	0.2901	3.7337	12828000000	1.00	122500

3a [4pt] In het artikel wordt de onderzoeksvraag geformuleerd als een hypothesetoets:

H0: *The independent variables - Defense expenditure, Defense expenditure as a percentage of real GDP, and Military personnel - do not exert an impact on the dependent variable - the Military Power Index.*

H1: *The independent variables - Defense expenditure, Defense expenditure as a percentage of real GDP, and Military personnel - do indeed have an impact on the dependent variable - the Military Power Index.*

Stel dat het in deze formulering lukt om H0 te verwerpen is dan de conclusie dat de genoemde variabelen **alle drie** invloed op de *Military Power Index* (= *Military Strength Ranking*) hebben, of kan het ook zijn dat dit maar **voor één of twee van de variabelen** geldt?

3b [5pt] In het artikel worden de drie onafhankelijke variabelen tegelijk behandeld, maar wij bekijken nu specifiek naar het lineaire verband tussen X = de totale nationale militaire uitgaven in 2020 in miljarden dollars en Y = de *Military Strength Ranking 2020 (reversed)*, dit is een getal waarvan de grootte een indicatie is voor de militaire sterkte van een land. Hiervoor kunnen met de tabel de volgende waarden worden berekend:

$$\bar{X} = 35,8772, \bar{Y} = 3,3378, \overline{XY} = 953,3821, \overline{X^2} = 20271,09, \overline{Y^2} = 48,8061$$

Gebruik deze waarden om de correlatiecoëfficiënt van Pearson te berekenen.

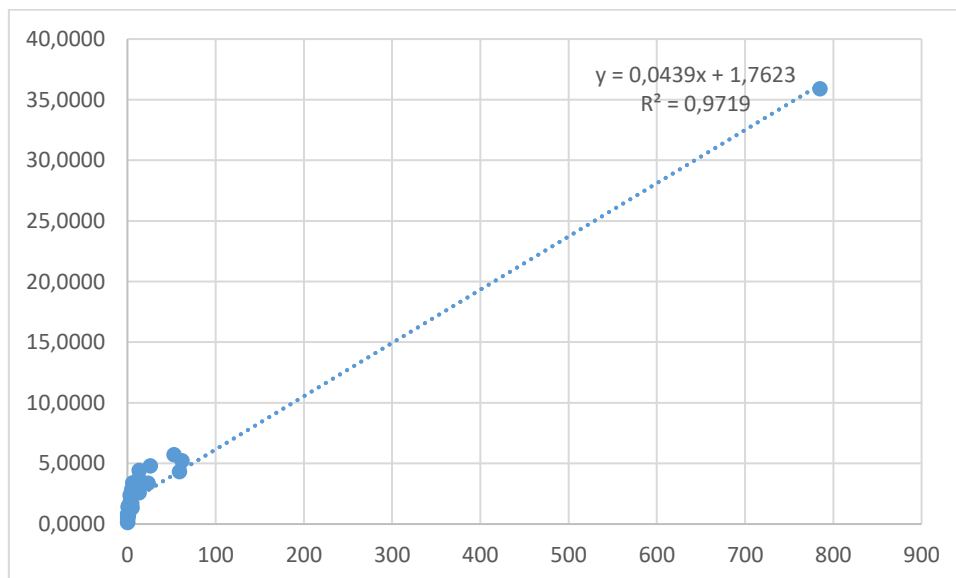
Bepaal of er sprake is van een lineaire correlatie tussen de defensieuitgaven en de militaire sterkte. Leg uit hoe daarbij het teken en de grootte van de berekende coëfficiënt een rol spelen.

3c [8pt] Bereken de regressielijn tussen de variabelen X en Y .

3d [2pt] In de tabel staan voor Nederland de waarde $X = 13,125$. Bereken met de regressielijn een statistisch verantwoorde voorspelling voor de militaire sterkte van Nederland. In de tabel staan voor Nederland de waarde $Y = 2,5771$. Wat kun je hieruit concluderen?

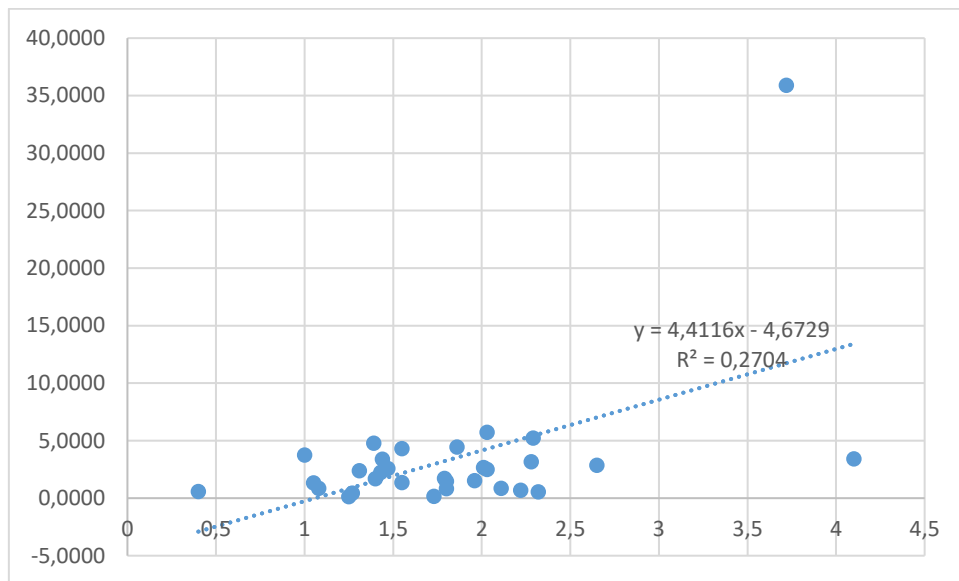
3e [10pt] Bereken een 98% voorspellingsinterval voor de waarde die in 3d is berekend.

3f [6pt] De regressielijn tussen de variabelen X en Y ziet er als volgt uit:



Welke conclusie(s) kun je hieruit trekken?

3g [5pt] Als je de militaire sterkte uitzet tegen het percentage defensieuitgaven van het bruto binnenlands product krijg je de volgende regressielijn:



De twee meest rechtse punten zijn van de Verenigde Staten (rechtsboven) en Oekraïne (rechtsonder). Kun je verklaren waarom de correlatie zo klein is en wat hier aan de hand is?

Opgave 4 (Totaal 15 punten). In het wetenschappelijk artikel Nasioudis, D., Palaiodimos, L., Dagiasis, M., Katsarou, A., & Ntouros, E. (2015). Depression in military medicine cadets: a cross-sectional study. *Military Medical Research*, 2, 1-5. werd onderzoek gedaan naar depressie onder 55 vrouwelijke en 91 mannelijke cadetten van de bacheloropleiding Geneeskundige Dienst van de Griekse Hellenic Military School of Combat Support Officers. Depressie werd gemeten met behulp van de *Zung score* die kan worden bepaald met een vierpunts scorelijst van 20 items die door de cadetten werd ingevuld. Dit levert een Zung depressiescore op. Een score onder 50 is normaal, tussen 50 en 59 is er sprake van een milde depressie, tussen 60 en 69 een behoorlijke depressie en boven de 70 is de depressie extreem. De cadetten werden onder andere ingedeeld op *Low risk* (depressiescore <45) en *High risk* (score ≥45) en vervolgens ingedeeld op geslacht, academische resultaten (*Average* betekent dat de gemiddelde score over alle vakken binnen één standaarddeviatie van het gemiddelde ligt) en studiejaar.

Table 1 Risk factors for presence of depressive symptoms

Risk factor	Low risk ^a [n(% ^b)]	High risk ^a [n(% ^b)]	P value from univariate analysis
Gender			0.87*
Male	55 (61.8 %)	36 (63.2 %)	
Female	34 (38.2 %)	21 (36.8 %)	
Academic performance			
High	17(19.1 %)	10 (18.5 %)	0.67*
Average	57 (64 %)	34 (62.3 %)	
Low	15 (16.9 %)	13 (19.2 %)	
Year of training			0.45*
1st	21 (23.6 %)	15 (26.3 %)	
2nd	34 (38.2 %)	16 (28.1 %)	
3rd	34 (38.2 %)	26 (45.6 %)	

*P value from Chi-square test

^ahigh and low risk for the presence of depression according to Zung self-reporting depression scale

^bpercentage within group

4a [10pt] In de tabel wordt steeds een *p* gegeven op grond van een Chi-Square waarde. Reken de gegeven waarde voor de *Year of training* tabel na in minstens drie decimalen nauwkeurig, d.w.z. maak de tabel met *Expected* waarden, bereken de waarde van χ^2 en vervolgens de overschrijdingskans *p*.

4b [5pt] Welke conclusies kun je trekken uit de *p*-waarden van Table 1 uit het artikel?