

# Tentamen Statistiek MBW/KW deel 2, 1<sup>e</sup> kans 28-07-2023

Afdeling: Propedeuse MBW/KW 2022-2023

Examinator: Dr. J.B.M. Melissen, T. Zijlstra MSc.

Datum: vrijdag 28 juli 2023, 9:00 – 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 (30, 25, 20, 25 punten). Score = Puntentotaal/10.

### Opgave 1 (Totaal 30 punten)

Van 20 MTW cadetten staan in de tabel hieronder de scores (op een schaal van 0 – 100) van de vakken Analyse 1 ( $X$ ) en Lineaire Algebra 1 ( $Y$ ):

Student	$X$ =Analyse	$Y$ =Algebra	$X^2$	$Y^2$	$XY$
1	40	53	1600	2809	2120
2	47	56	2209	3136	2632
3	51	91	2601	8281	4641
4	51	59	2601	3481	3009
5	53	58	2809	3364	3074
6	54	70	2916	4900	3780
7	54	42	2916	1764	2268
8	55	62	3025	3844	3410
9	56	77	3136	5929	4312
10	56	55	3136	3025	3080
11	58	49	3364	2401	2842
12	65	87	4225	7569	5655
13	69	68	4761	4624	4692
14	70	58	4900	3364	4060
15	77	75	5929	5625	5775
16	86	90	7396	8100	7740
17	92	80	8464	6400	7360
18	21	35	441	1225	735
19	56	82	3136	6724	4592
20	86	89	7396	7921	7654
Gemiddeld:	59,85	66,80	3848,05	4724,30	

**1a [2pt]** Maak de tabel compleet door het uitrekenen van de waarde van  $\overline{XY}$  (Narekenen!)

**1b [6pt]** Bereken met behulp van de tabel de correlatiecoëfficiënt van Pearson. Bepaal of er sprake is van een lineaire correlatie tussen de scores van de twee vakken. Leg in woorden uit wat de betekenis is van de grootte en het teken van de correlatiecoëfficiënt voor een mogelijke relatie tussen de scores voor de twee vakken. Is er sprake van een causaal verband?

**1c [7pt]** Bereken de regressielijn  $Y = aX + b$  door berekening van de coëfficiënten  $a$  en  $b$  met behulp van de tabel (narekenen!). Neem hierbij de score van Analyse als de verklarende variabele  $X$ .

**1d [2pt]** Bereken met de regressielijn een voorspelling voor de Lineaire Algebra score van een cadet die voor Analyse 74 punten scoorde. Rond de score af op een geheel getal.

**1e [7pt]** Bereken een **92%** voorspellingsinterval voor de voorspelling uit 1d. Rond de grenzen af op gehele getallen.

**1f [3pt]** Bereken de Analyse score waarvoor de regressieformule een Lineaire Algebra score voorspelt die gelijk is aan de Analyse score. Rond deze score af op een geheel getal.

**1g [3pt]** Stel dat een student als Analyse score toevallig de gemiddelde Analyse score van alle cadetten heeft. Laat door berekening zien dat de regressielijn voor deze student een Lineaire Algebra score voorspelt die gelijk is aan de gemiddelde Lineaire Algebra score van alle studenten.

Toon vervolgens aan dat dit geen toeval is, namelijk, dat altijd geldt:  $\bar{Y} = a\bar{X} + b$ .  
(Hint: gebruik de formules voor  $a$  en/of  $b$ ).

## Opgave 2 (Totaal 25 punten)



Vliegtuigbanden slijten bij het landen (10% van de slijtage), het remmen na de landing (ca. 30%) en het taxiën en starten (60%). Daarom moeten deze banden na een bepaald aantal vluchten worden vervangen. Voor tien vliegtuigbanden werd het aantal vluchten genoteerd waarna ze moesten worden vervangen: 250, 171, 302, 119, 234, 272, 112, 262, 203, 189.

**2a [6pt]** Bereken van de gemeten waarden het steekproefgemiddelde en de steekproefstandaarddeviatie.

**2b [6pt]** Bereken op grond van de bovengenoemde steekproef een 90% betrouwbaarheidsinterval voor het verwachte aantal operationele vluchten ( $\mu$ ) dat een band meegaat, zonder daarbij gebruik te maken van de optie TESTS/Interval van de grafische rekenmachine. Rond de grenzen van dit interval af op gehele aantallen en wel zodanig dat de betrouwbaarheid gewaarborgd blijft.

**2c [7pt]** Bereken op grond van bovengenoemde steekproef en met behulp van de  $\chi^2$ -verdeling een 90% betrouwbaarheidsinterval (zie formuleblad) **voor de standaarddeviatie** van het aantal vluchten dat per band kan worden gevlogen, zonder daarbij gebruik te maken van de optie TESTS/Interval van de grafische rekenmachine. Rond de grenzen van dit interval af één decimaal.

**2d [2pt]** Als de kans dat één band  $n$  vluchten kan uitvoeren zonder vervangen te worden gelijk is aan 0,98305, toon dan met een berekening aan dat de kans dat drie banden minstens  $n$  vluchten meegaan gelijk is aan 0,9500.

**2e [4pt]** Een F35-Lightning II (Joint Strike Fighter) wordt voorzien van 3 nieuwe banden. Bereken een aantal vluchten waarvan je met 95% zekerheid kunt aantonen dat deze F35 zoveel vluchten kan maken zonder dat een band vervangen hoeft te worden.  
(Hint: Denk aan een rechteroverschrijdskans van 0,98305.)

### Opgave 3 (Totaal 20 punten).

Bij Defensie solliciteerden in 2021 14.500 personen naar 3.600 beschikbare functies als militair. In de onderstaande tabel is voor een aantal functies te zien hoeveel sollicitanten zich per functie hadden aangemeld.

Sollicitanten per functie	Frequentie
0-1	10
2-3	32
4-5	36
6-7	15
8-9	5
$\geq 10$	2

**3a [5pt]** Gebruik **de gegevens uit de tabel** om het aantal sollicitanten en het aantal functies te berekenen/schatten waarop deze tabel is gebaseerd. Bereken ook het gemiddeld aantal sollicitanten per functie en de steekproefstandaarddeviatie.

**3b [10pt]** Toets aan de hand van de steekproef in de tabel of het aantal sollicitanten per functie is te beschouwen als een kansvariabele die Poisson verdeeld is met gemiddelde waarde van 4 sollicitanten per functie. Voer de toets uit door middel van uitrekenen van een  $p$ -waarde. Kies als betrouwbaarheid 95% en gebruik in je berekening de verwachte frequenties in één decimaal nauwkeurig.

**3c [5pt]** Voer de toets ook uit door berekening van het kritieke gebied.

**Opgave 4 (Totaal 25 punten).** In een recent wetenschappelijk artikel (*Education as a Pathway to Help Job Searching Military Veterans' Transition to Civilian Life*) worden tabellen van 56 US combat veterans en 43 non-combat veterans gegeven die zijn uitgesplitst naar of ze na hun militaire carrière wel of niet actief op zoek zijn naar een baan en of ze momenteel wel of niet een opleiding volgen:

<http://journal.julypress.com/index.php/jed>

Vol. 5, No. 1; April, 2021

Table 4. Chi-Square tests comparing combat and non-combat veterans who are actively job searching versus not and who are currently going to school to further their education versus not

	Combat Veterans		Non-combat Veterans	
	Going to School		Going to School	
	Yes	No	Yes	No
Yes	19	1	15	2
Actively				
Job Searching				
No	21	15	20	6
Chi-Square (1, n = 56) = 6.82, p = .01		Chi-Square (1, n = 43) = 1.02, p = .33		

**4a [10pt]** Onder de linker tabel wordt een Chi-Square waarde en een  $p$  gegeven. Reken deze twee waarden na in minstens drie decimalen nauwkeurig, d.w.z. maak de tabel met *expected* waarden, bereken de waarde van  $\chi^2$  en vervolgens de overschrijdingskans  $p$ .

**4b [5pt]** In beide tabellen in Table 4 komen waarden voor die kleiner zijn dan 5. Leg met behulp van de *expected* tabellen uit waarom de  $\chi^2$ -test wel toepasbaar is in de linker tabel (*combat veterans*) maar niet in de rechter (*non-combat veterans*).

**4c [3pt]** De tabel hoort bij de volgende onderzoeksvraag:

RQ2 stated that combat veterans and non-combat veterans who are actively searching for a new job are more likely to be currently going to school to further their education than those who are not actively searching. The results are shown in Table 4. The Chi-square values show that there is statistical support for combat veterans, but not non-combat veterans. Thus RQ2 was only partially supported.

Leg uit of en waarom je met de  $\chi^2$ -toets de beweringen die ze doen kunt toetsen. Wat test deze  $\chi^2$ -toets eigenlijk?

**4d [2pt]** voor het testen van 2x2 kruistabellen bestaat er Fischer's exacte toets die niet de beperking heeft voor kleine verwachte waarden zoals de  $\chi^2$ -toets. Fischer's toets geeft voor de twee tabellen respectievelijk de waarden  $p = 0,006936$  en  $p = 0,4462$ . Wat betekent dit voor de conclusies die de auteurs onder 4c trekken?

**4e [5pt]** Als je de twee tabellen samenvoegt tot één tabel ontstaat een tabel voor alle veteranen. Voor deze tabel blijkt  $p = 0,0036$  te gelden. Wat is nu de conclusie die je hieruit kunt trekken? Is deze conclusie sterker of zwakker dan die van het artikel?