

# 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 10 november 2023, 8:30 – 10:30. Duur tentamen: 2 uur

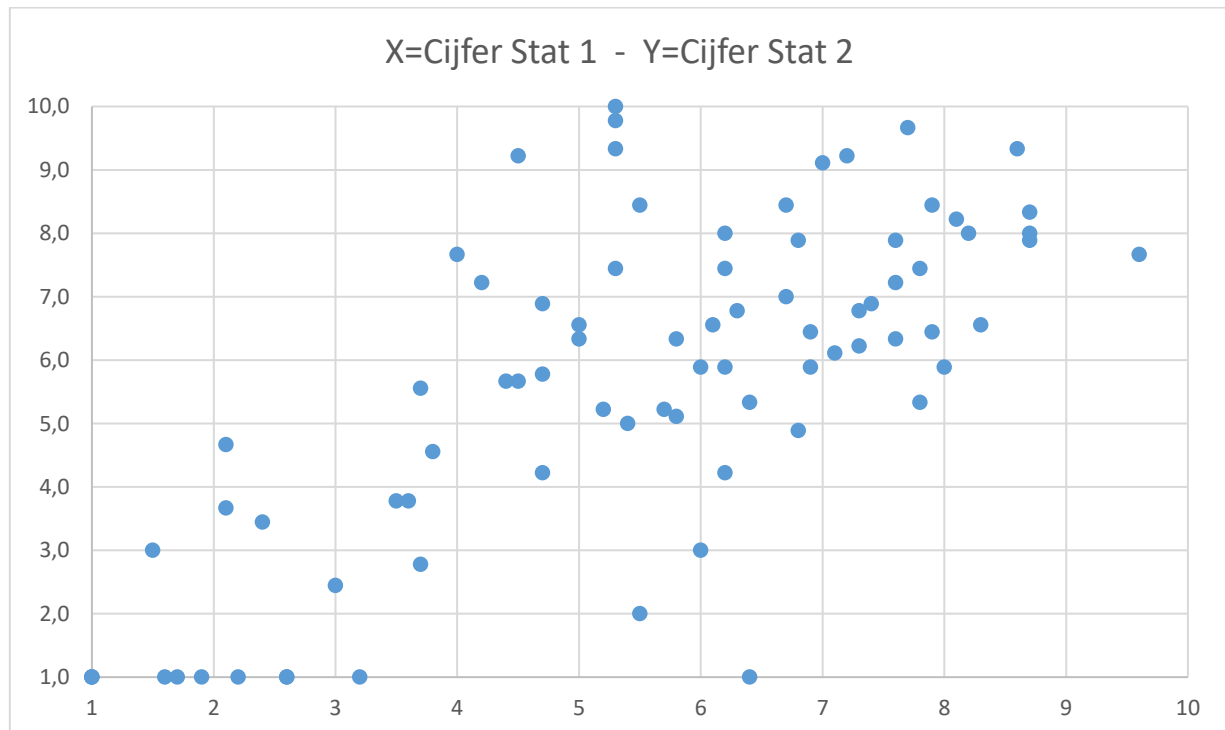
## 1. Alle antwoorden moeten gemotiveerd worden!

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). Score = Puntentotaal/10.

## Opgave 1 (Totaal 30 punten)

Van 82 cadetten en adelborsten zijn de cijfers van Statistiek deel 1 ( $X$ ) en Statistiek deel 1 ( $Y$ ) uit 2023 uitgezet in een spreidingsdiagram:



De vraag is of er een lineair verband is tussen de cijfers.

De volgende waarden zijn uitgerekend:

$$\bar{X} = 5,445, \quad \bar{Y} = 5,680, \quad \overline{X^2} = 34,629, \quad \overline{Y^2} = 39,250, \quad \overline{XY} = 35,234$$

**1a [6pt]** Bereken met behulp van de bovenstaande waarden de correlatiecoëfficiënt van Pearson.

Bepaal of er sprake is van een lineaire correlatie tussen de scores van de twee cijfers. 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.

**1b [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  $X$ .

**1c [2pt]** Bereken met de regressielijn een voorspelling voor de Statistiek 2 score (in vier decimalen) van een cadet die voor Statistiek 1 een 5,5 scoorde.

**1d [10pt]** Bereken een **90%** voorspellingsinterval voor de voorspelling uit 1c. Doe de berekeningen in 4 decimalen. Rond de antwoorden af op 2 decimalen

**1e [3pt]** Bereken de Statistiek 1 score waarvoor de regressieformule een Statistiek 2 score voorspelt die 0,2 lager is dan de Statistiek 1 score. Rond deze score af op een decimaal.

**1f [2pt]** Stel dat een student als Statistiek 1 score toevallig de gemiddelde Statistiek 1 score van alle cadetten heeft. Laat door berekening zien dat de regressielijn voor deze student een Statistiek 2 score voorspelt die gelijk is aan de gemiddelde Statistiek 2 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  $b$ ).

**1g [1pt]** Verbind de punten in het spreidingsdiagram zodanig dat een ludieke figuur ontstaat.

## Opgave 2 (Totaal 20 punten)

Een genie-eenheid heeft in het kader van een vredesmissie de opdracht om voormalige oorlogsterreinen te zuiveren van mijnen, boobytraps en niet ontplofte granaten. Tien dagen lang is bijgehouden hoeveel devices onklaar zijn gemaakt: 12, 21, 20, 15, 17, 9, 18, 19, 16, 13.

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

**2b [6pt]** Bereken op grond van de bovengenoemde steekproef een 95% betrouwbaarheidsinterval voor het verwachte aantal opgeruimde devices per dag ( $\mu$ ), 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 [8pt]** Bereken op grond van bovengenoemde steekproef en met behulp van de  $\chi^2$ -verdeling een 95% betrouwbaarheidsinterval (zie formuleblad) **voor de standaarddeviatie** van het aantal opgeruimde devices per dag 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.

## Opgave 3 (Totaal 20 punten).

**3a [7pt]** In 2003 waren er volgens het CBS 191.032 gezinnen met één zoon en 180.974 gezinnen met één dochter. Er wordt vaak gedacht dat de kans op een jongen of een meisje bij geboorte gelijk is. Voer een  $\chi^2$  aanpassingstoets uit om aan de hand van deze steekproef de volgende hypothese te toetsen met een betrouwbaarheidsniveau van 99,9% en leg in woorden uit at je resultaat betekent:

**H<sub>0</sub>:** Bij geboorte is de kans op een jongen of een meisje gelijk.

**3b [6pt]** Voer de toets ook uit door berekening van het kritieke gebied.

**3c [7pt]** Bereken op grond van de gegevens van het CBS een schatting voor de kans op een jongen.

Bereken vervolgens met behulp van Clopper-Pearson een 99% betrouwbaarheidsinterval voor deze schatting.

**Opgave 4 (Totaal 30 punten).** In het wetenschappelijk artikel Nasioudis, D., Palaiodimos, L., Dagiasis, M., Katsarou, A., & Ntouro, E. (2015). Depression in military medicine cadets: a cross-sectional study. *Military Medical Research*, 2, 1-5. werd onderzoek gedaan onder 55 vrouwelijke en 91 mannelijke cadetten in de bacheloropleiding Geneeskundige Dienst van de Griekse Hellenic Military School of Combat Support Officers gedaan naar depressie. 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 <sup>a</sup> [n(% <sup>b</sup> )]	High risk <sup>a</sup> [n(% <sup>b</sup> )]	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

<sup>a</sup>high and low risk for the presence of depression according to Zung self-reporting depression scale

<sup>b</sup>percentage within group

**4a [10pt]** In de tabel wordt steeds een *p* gegeven op grond van een Chi-Square waarde en een *p* gegeven. 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  $\chi^2$  en vervolgens de overschrijdingskans *p*.

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

**4c [5pt]** In het artikel wordt een ander artikel aangehaald waarin bij niet-militaire medische studenten een high risk percentage werd geconstateerd van 30,4%. Kun je nu met een 99% betrouwbaarheid stellen dat militaire medische studenten een hoger risico op depressie hebben?

**4d [10pt]** Bereken de kans dat een cadet een *Average* studiescore heeft. Hoeveel van de cadetten zouden er eigenlijk *Average* hebben moeten scoren?  
Bereken de kans dat van de 146 cadetten er hoogstens 91 een gemiddelde studiescore hebben. Wat kun je hieruit concluderen als je een betrouwbaarheid van 95% aanhoudt?