

Vraag:	1	2	3	4	5	Totaal
Punten:	2	3	6	2	1	14
Score:						

Examen Statistiek I

18 januari 2019

Naam :

Richting :

Lees volgende aanwijzingen alvorens aan het examen te beginnen

- Wie de vragen aanneemt en bekijkt, moet minstens 1 uur blijven zitten.
- Schrijf op 1ste blad duidelijk je volledige naam en richting (en op elk blad je naam).
- Je mag gebruik maken van niet-grafisch rekenmachine, formularium en statistische tabellen. Op het formularium en de tabellen mag niets geschreven staan! Berekeningen moeten altijd schriftelijk uitgevoerd worden tot het moment dat je de waarde zou kunnen opzoeken in een statistische tabel. Bijvoorbeeld: het uitrekenen van een kans onder een normale verdeling moet herleid worden tot een kans onder een standaardnormale verdeling. Wanneer het nodige aantal vrijheidsgraden niet in de tabel staat, mag je gaan kijken bij het dichtstbijzijnde aantal dat wel in de tabel staat. Werk met 3 cijfers na de komma!
- Alle communicatie-apparatuur is strikt verboden.
- Gebruik de voorziene ruimte om te antwoorden op de vragen (voor- en achterkant).
- Bij het indienen van je examen, geef je ook kladpapier af (maar daar wordt geen rekening mee gehouden tijdens verbetering).
- Let op
 - correct (numeriek) antwoord zonder uitleg (of foute uitleg) is weinig/niets waard!
 - fout (numeriek) antwoord zonder uitleg is niets waard.
 - fout numeriek antwoord (bvb ten gevolge van een rekenfout) met juiste afleiding is veel waard.

Toon dus **DUIDELIJK** aan hoe je tot ieder numeriek resultaat komt (telegramstijl is toegelaten). Vermeld de gebruikte formules en ook je berekening. Gebruik zoveel mogelijk de wiskundige notatie zoals die in de leerstof is aangebracht. Verklaar nieuwe symbolen.

- Je hebt **3 uur** tijd om het examen op te lossen.

VEEL SUCCES !

1. (2 punten) In een boxplot worden observaties die voorbij de *whiskers* liggen aangeduid als uitschieters. Stel dat X exponentieel verdeeld is met verwachtingswaarde $E[X] = 2$.
 - (a) Hoeveel observaties verwacht je dat er als uitschieters worden aangeduid bij de boxplot van een steekproef uit X bij een steekproefomvang van $n = 1000$?
 - (b) Verklaar dit resultaat aan de hand van de vorm van de verdeling van X als je weet dat voor een normale verdeling ongeveer 0.37% van de observaties boven de bovenste whisker liggen en 0.37% van de observaties onder de onderste whisker.

2. (3 punten) Zij X een discrete toevalsvariabele met de volgende kansdichtheid, die afhangt van een onbekende parameter θ (met $0 \leq \theta \leq 1$):

x	0	1	2
$P(X = x)$	θ	$\frac{2(1-\theta)}{3}$	$\frac{1-\theta}{3}$

Op basis van een o.o.i.v. steekproef X_1, X_2, \dots, X_n kan men de volgende schatters voor θ bepalen:

$$T_1 = 1 - \frac{3}{4}\bar{X} \quad \text{en} \quad T_2 = \frac{N_0}{n}$$

met \bar{X} het steekproefgemiddelde en N_0 het aantal X_i 's gelijk aan 0.

- (a) Bereken de MSE van T_1

Naam:

4

(b) Bereken de MSE van T_2

(c) Welk van beide schatters zou je verkiezen? Beantwoord deze vraag door de maximale MSE in functie van θ voor beide schatters met elkaar te vergelijken.

3. (6 punten) In *De Morgen* kon je op 30/11/2018 de volgende krantenkop lezen:

Minstens één op de zes Vlaamse specialisten verdient meer dan de premier

In de onderstaande tabel vind je lonen van 30 willekeurig uitgekozen Vlaamse specialisten (in 1000 euro).

mannen	248	319	292	298	314	278	266	298	269	235	208	255	340	278	389
gemiddelde	285.80														
variantie	1973.46														
<hr/>															
vrouwen	282	245	212	323	294	126	258	252	352	401	214	158	247	198	339
gemiddelde	260.07														
variantie	5520.07														

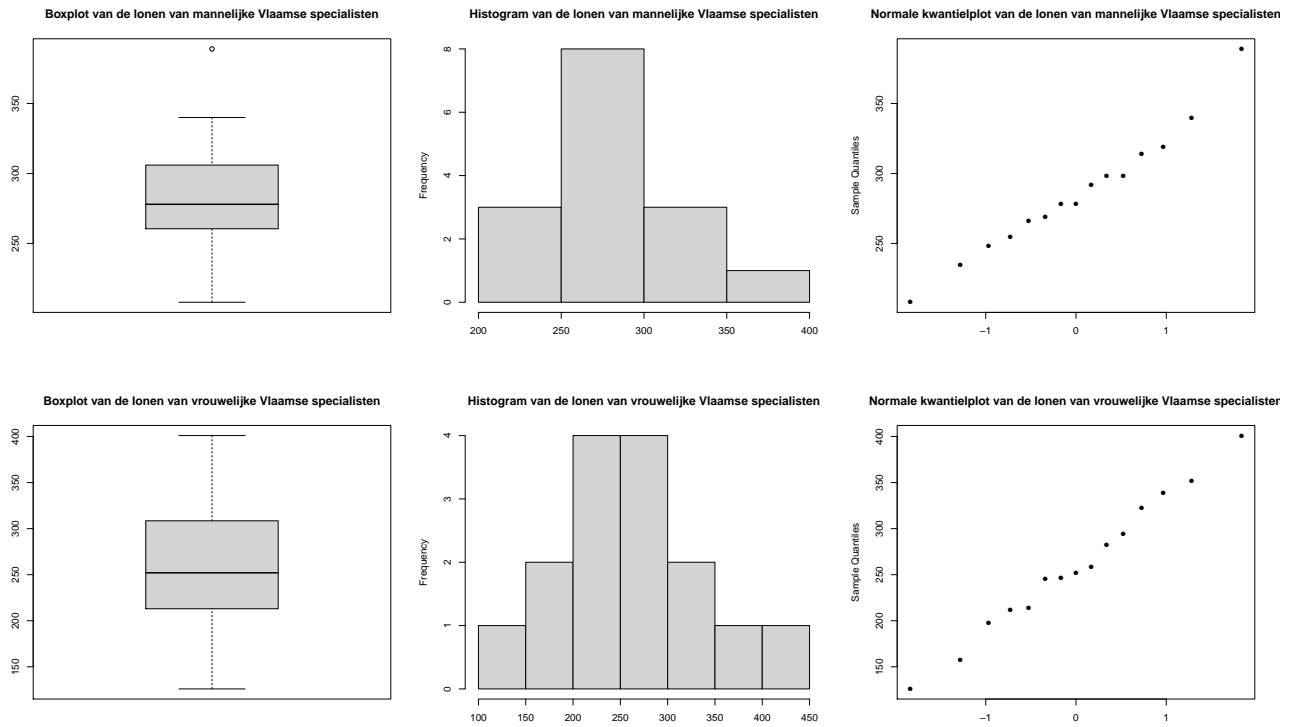
- (a) Onderzoek of je op basis van deze gegevens kan bewijzen dan meer dan één op zes Vlaamse specialisten meer verdient dan de premier, als je weet dat het loon van de premier 290 000 euro bedraagt. Doe dit aan de hand van een geschikte hypothesetest. Vermeld zeker H_0 en H_1 en de gebruikte teststatistiek met de verdeling ervan onder H_0 , de **p-waarde** van de test en formuleer je (genuanceerd) besluit op significantieniveau 0.05. Ga ook na of de voorwaarden die nodig zijn om deze test te mogen uitvoeren, voldaan zijn.

(b) Maak een schets waarop je de testwaarde en de p -waarde uit de hypothesetest in (a) aanduidt.

(c) Bereken de kans op een type II fout indien de kans dat een Vlaamse specialist meer verdient dan de premier, gelijk is aan 0.3.

- (d) Kan je op basis van deze gegevens besluiten dat een mannelijke Vlaamse specialist gemiddeld 10% meer verdient dan een vrouwelijke Vlaamse specialist? Onderzoek aan de hand van een geschikte hypothesetest. Vermeld zeker H_0 en H_1 en de gebruikte teststatistiek met de verdeling ervan onder H_0 , het **aanvaardingsgebied** van de test en formuleer je (genuanceerd) besluit op significantieniveau $\alpha = 0.01$. Ga de voorwaarden om de test te mogen uitvoeren na aan de hand van de grafieken in Figuur 1 en de R-uitvoer op pagina 9. Vermeld bij het beantwoorden van de vragen steeds welke output en/of grafieken je gebruikt om je conclusie(s) te trekken.

- (e) Leg uit hoe je de hypothesen in (d) kan testen op significantieniveau α aan de hand van een éézijdig betrouwbaarheidsinterval (je moet het betrouwbaarheidsinterval **niet** berekenen!).



Figuur 1: grafische voorstellingen van de gegevens per groep

```
> shapiro.test(loon_m)
```

Shapiro-Wilk normality test

```
data: m
W = 0.97607, p-value = 0.9356
```

```
> shapiro.test(loon_v)
```

Shapiro-Wilk normality test

```
data: v
W = 0.98822, p-value = 0.9983
```

4. (2 punten)

- (a) Beschrijf het enkelvoudige lineaire regressiemodel waarmee je de waarde van een kwantitatieve variabele Y op een lineaire manier kan voorspellen aan de hand de waarde van een kwantitatieve variabele X op basis van een steekproef van grootte n . Leg uit hoe je de parameters in dit model zou schatten. Het is **niet** nodig om de (afleiding van) de formules voor schatters te geven.

- (b) Leg uit hoe je de modelveronderstellingen kan nagaan.

- (c) Men vindt voor deze steekproef dat $R^2 = 0.35$. Interpreteer deze waarde indien aan alle modelveronderstellingen voldaan is. [Enkele zinnen volstaan !]

- (d) Welke hypothesen ga je testen om na te gaan of het zinvol is om lineaire regressie uit te voeren?

5. (1 punt) Op de figuur hieronder zie je twee verschillende OC-curves. Naar welk van beide curves zou de voorkeur van de producent uitgaan en welke heeft de voorkeur van de consument? Verklaar je antwoord.

