

U krijgt de beschikking over een tabellenboekje.

Gebruik van een niet-programmeerbare rekenmachine is toegestaan

1. Uw docent algemene statistiek wil onderzoeken of de drie groepen studenten (econometrie, wiskunde en BWI) gelijk presteren bij de eerste deelttoets. Hij gebruikt daartoe de uitslagen van de toets van oktober 1999. Laat $X_1, \dots, X_{26}$ de score bij econometrie zijn, $Y_1, \dots, Y_{21}$ de score bij BWI en $Z_1, \dots, Z_5$ de score bij wiskunde. Op één decimaal afgerond zijn de volgende scores behaald:

econometrie: 3.1 3.2 3.5 3.9 4.3 4.5 4.6 4.6 4.6 4.8 4.9 5.0 5.0 5.1 5.5 5.7 5.7 5.9 6.8 7.0 7.0 7.0 7.3 7.6 7.9 8.1

BWI: 2.6 3.8 3.9 4.3 4.5 4.6 4.6 4.7 4.8 4.9 4.9 5.1 5.5 5.6 5.7 5.8 6.0 6.3 6.3 6.4 7.1

wiskunde: 5.2 7.7 7.7 8.7 9.3

Verder geldt

$$\bar{x}_{26} = 5.48, \quad \bar{y}_{21} = 5.11, \quad \bar{z}_5 = 7.72$$

en

$$\sum_{k=1}^{26} (x_k - \bar{x}_{26})^2 = 52.83, \quad \sum_{k=1}^{21} (y_k - \bar{y}_{21})^2 = 21.85, \quad \sum_{k=1}^5 (z_k - \bar{z}_5)^2 = 9.81$$

We mogen aannemen: $X_i \sim \mathcal{N}(\mu_1, \sigma^2)$, $i = 1, \dots, 26$, $Y_j \sim \mathcal{N}(\mu_2, \sigma^2)$, $j = 1, \dots, 21$ en $Z_k \sim \mathcal{N}(\mu_3, \sigma^2)$, $k = 1, \dots, 5$.

- Toets de nulhypothese dat studenten wiskunde en BWI gelijk presteren tegen het alternatief dat beide groepen verschillend presteren bij onbetrouwbaarheidsdrempel $\alpha = 0.05$. Formuleer nauwkeurig het statistisch model, de gehanteerde hypothesen en de gebruikte toetsingsgrootheid. Wat is uw conclusie?
 - Bepaal de overschrijdingskans op basis van de bij (a) berekende waarde van de toetsingsgrootheid. Als deze niet uit de tabellen in het tabellenboekje gehaald kan worden, kunt u volstaan met een zo nauwkeurig mogelijke bovengrens op basis van de gegevens in het tabellenboekje.
 - Geef een 90% betrouwbaarheidsinterval voor het verschil in score tussen studenten econometrie en BWI. Wat is uw conclusie m.b.t. de nulhypothese dat beide groepen gelijk presteren?
2. De stochastische grootheden $X_1, \dots, X_n$ zijn onderling onafhankelijk en identiek verdeeld met kansdichtheid

$$f_\theta(x) = \theta^2 x e^{-\theta x} 1_{\{x>0\}}.$$

Hierbij is $\theta > 0$ een onbekende parameter. Er geldt $E_\theta X_i = \frac{2}{\theta}$ en $E_\theta X_i^2 = \frac{6}{\theta^2}$.

- Bepaal de meest aannemelijke schatter voor θ op basis van $X_1, \dots, X_n$.
- Bepaal de "plug-in" schatter voor de Fisher informatie i_θ en bepaal de waargenomen Fisher informatie voor θ .
- Bepaal een benaderend 95% betrouwbaarheidsinterval voor θ gebaseerd op de meest aannemelijke schatter.

3. Zij $X_1, \dots, X_n$ een steekproef uit een $\text{hom}[0, \theta]$ -verdeling. We willen $\theta > 0$ schatten.

(a) Laat zien dat

$$f_{X_{(n)}}(x) = \begin{cases} \frac{n}{\theta^n} x^{n-1} & , 0 \leq x \leq \theta, \\ 0 & , \text{elders.} \end{cases}$$

en

$$f_{X_{(1)}}(x) = \begin{cases} \frac{n}{\theta^n} (\theta - x)^{n-1} & , 0 \leq x \leq \theta, \\ 0 & , \text{elders.} \end{cases}$$

kansdichtheden van $X_{(n)} = \max(X_1, \dots, X_n)$ en $X_{(1)} = \min(X_1, \dots, X_n)$ zijn.

(b) Laat zien dat $T_1 = X_{(n)} + X_{(1)}$ een zuivere schatter voor θ is.

(c) Is $W = (X_{(n)}, X_{(1)})$ een voldoende statistiek voor de familie van $\text{hom}[0, \theta]$ -verdelingen? Motiveer uw antwoord.

(d) In het collegedictaat is aangetoond dat $X_{(n)}$ volledig is voor de familie van $\text{hom}[0, \theta]$ -verdelingen. Toon aan, via de definitie van volledigheid, dat $W = (X_{(n)}, X_{(1)})$ niet volledig is voor de familie van $\text{hom}[0, \theta]$ -verdelingen.

(e) Andere zuivere schatters voor θ zijn $T_2 = 2\bar{X}_n$ en $T_3 = \frac{n+1}{n}X_{(n)}$. Is er één van de drie schatters T_1 , T_2 en T_3 aan welke u de voorkeur geeft? Motiveer uw antwoord.

4. Zij $X_1, \dots, X_n$ een steekproef uit een verdeling met kansdichtheid

$$f_\theta(x) = \begin{cases} \theta 2^\theta x^{-(\theta+1)} & , x \geq 2 \\ 0 & , x < 2 \end{cases}$$

waarbij $\theta > 0$.

(a) Bepaal de momentenschatter voor θ op basis van $X_1, \dots, X_n$ (neem hier aan $\theta > 1$).

(b) De meest aannemelijke schatter voor θ is

$$\hat{\theta} = \frac{n}{\sum_{i=1}^n \log \frac{X_i}{2}}.$$

Bepaal de meest aannemelijke schatter voor $1/\theta$.

(c) Bepaal een volledige en voldoende statistiek voor deze familie van verdelingen.

(d) Bepaal een ondergrens voor de variantie van zuivere schatters van $1/\theta$ op basis van de steekproef, gebruik makend van de stelling van Cramér-Rao. Wordt deze ondergrens bereikt door de UMVZ schatter voor $1/\theta$?

U mag hierbij het volgende gebruiken: de dichtheid van $Y_i = \log X_i$ wordt gegeven door

$$g_\theta(y) = \begin{cases} \theta 2^\theta e^{-\theta y} & , y \geq \log 2 \\ 0 & , y < \log 2 \end{cases}$$

en $E_\theta \log X_i = \log 2 + \frac{1}{\theta}$ en $E_\theta (\log X_i)^2 = (\log 2)^2 + \frac{2}{\theta} \log 2 + \frac{2}{\theta^2}$

Normering:

1(a): 9	2(a): 8	3(a): 5	4(a): 7
1(b): 4	2(b): 9	3(b): 6	4(b): 3
1(c): 8	2(c): 4	3(c): 4	4(c): 6
		3(d): 5	4(d): 6
		3(e): 6	

Eindcijfer = (totaal + 10)/10