Technische Universiteit Delft — Mekelweg 4
Faculteit Elektrotechniek, Wiskunde en Informatica — 2628 CD Delft

Proeftentamen D2 Kansrekening en Statistiek EE1M31

Bij dit tentamen is het gebruik van een **niet-grafische** rekenmachine toegestaan. Ook toegestaan is een formuleblad, dat wordt uitgedeeld.
Dit tentamen bestaat uit negen meerkeuze- en drie open vragen. De drie open vragen dient u op het tentamenvel te maken.

Cijferbepaling: Iedere meerkeuzevraag telt voor 1 punt, bij de open vragen is het aantal punten per onderdeel aangegeven. Er geldt:

$$\text{Cijfer} = \frac{MC + OV}{2} + 1,$$

waarbij MC het aantal punten voor meerkeuze-deel en OV het aantal punten voor open-vragen-deel is.
Toelichting meerkeuze-vragen: Maak op het bijgeleverde meerkeuze-antwoordformulier het hokje behorende bij het door u gekozen alternatief zwart of blauw. Doorstrepen van een fout antwoord heeft geen zin: u moet het òf uitgummen, òf verwijderen met correctievloeistof òf een nieuw formulier invullen. Tenslotte, vergeet niet de versie, de vakcode, uw naam en uw studienummer in te vullen. De laatste dient u ook aan te strepen. Uw tentamen wordt anders niet gecorrigeerd.

Meerkeuze vragen

1. $X_1, X_2, \ldots, X_{300}$ zijn stochasten met onafhankelijke uniforme verdelingen op $\{1, 2, 3\}$, dus $\mathrm{P}(X_i = k) = \frac{1}{3}$, voor $k = 1, 2, 3$.
   De som $S_{300} = \sum_{i=1}^{300} X_i$ heeft dan bij benadering de normale verdeling

   a. $N(200, 300)$ **b.** $N(300, 300)$ **c.** $N(600, 300)$
   d. $N(200, 200)$ **e.** $N(300, 200)$ **f.** $N(600, 200)$

2. We beschikken over een dataset van omvang 20 die we opvatten als een realisatie van steekproef $X_1, X_2, \ldots, X_{20}$ uit de kansdichtheid

$$f(x) = \begin{cases} \dfrac{2x}{\theta^2} & \text{, als } 0 \le x \le \theta, \\ 0 & \text{, elders,} \end{cases}$$

   waarbij $\theta > 0$ een onbekende parameter is. Definieer de volgende schatter voor θ:

$$T = a(X_1 + X_2 + \cdots + X_{20}).$$

   Voor welke a is T een zuivere schatter voor θ?

   a. $\frac{1}{40}$ **b.** $\frac{1}{20}$ **c.** $\frac{3}{20}$ **d.** $\frac{1}{10}$ **e.** $\frac{5}{40}$ **f.** $\frac{3}{40}$

3. Men beschikt over een dataverzameling die een realisatie vormt van een steekproef $X_1, \ldots, X_{16}$ uit een $N(\mu, 4)$ verdeling. Men toetst $H_0 : \mu = 1$ tegen $H_1 : \mu > 1$ met toetsingsgrootheid $T = \bar{X}_{16}$. Men verwerpt H_0 ten gunste van H_1 als voor de data $\bar{x}_{16} \ge 2$. In dit geval is de kans op een type I fout

   a. 0.1151 **b.** 0.2734 **c.** 0.0556 **d.** 0.1587 **e.** 0.0228 **f.** 0.2302

4. Beschouw betrouwbaarheidsintervallen voor het gemiddelde van normale data met bekende variantie σ^2.
   Welk van de beweringen is/zijn waar: het betrouwbaarheidsinterval wordt groter als
   (A) de steekproefomvang groter wordt (bij gelijkblijvende betrouwbaarheid en variantie);
   (B) de betrouwbaarheid groter wordt (bij gelijkblijvende steekproefomvang en variantie);
   (C) de variantie groter wordt (bij gelijkblijvende steekproefomvang en betrouwbaarheid).

   a. alle drie onwaar **b.** alleen (A) waar **c.** alleen (B) waar **d.** alleen (C) waar
   e. (A) en (B) waar **f.** (A) en (C) waar **g.** (B) en (C) waar **h.** alle drie waar

5. Van een dataset van grootte 120 zijn vier waarden van de empirische verdelingsfunctie gegeven:
   $$F_{120}(5) = 0.10, \quad F_{120}(7) = 0.20, \; F_{120}(8) = 0.40 \quad \text{en} \quad F_{120}(11) = 0.70.$$
   Als we een histogram maken wat is dan de sterkste uitspraak die je kun doen over de hoogte h op de cel $(6, 10]$?

   a. $0 \leq h \leq 1$ **b.** $0.05 \leq h \leq 0.10$ **c.** $0.1 \leq h \leq 0.75$
   d. $0.025 \leq h \leq 0.6$ **e.** $0 \leq h \leq 0.10$ **f.** $0.05 \leq h \leq 0.15$

6. Voor een bivariate dataverzameling $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, beschouwen we een lineair regressie model zonder intercept: $Y_i = \beta x_i + U_i$ voor $i = 1, 2, \ldots, n$, waarbij $U_1, U_2, \ldots, U_n$ onafhankelijk zijn met verwachting 0 en variantie σ^2. De kleinste-kwadraten-schatting voor β wordt gegeven door

   a. $\dfrac{\sum y_i}{\sum x_i}$ **b.** $\dfrac{1}{n}\sum \dfrac{y_i}{x_i}$ **c.** $\dfrac{\sum x_i y_i}{\sum x_i^2}$ **d.** $\dfrac{\sum x_i^2 y_i^2}{\sum x_i^4}$ **e.** $\dfrac{\sqrt{\sum y_i^2}}{\sqrt{\sum x_i^2}}$ **f.** $\sqrt{\dfrac{1}{n}\sum \dfrac{y_i^2}{x_i^2}}$

7. In een zevendaagse studie naar het effect van ozon is een groep van 14 ratten in een ozon-vrije omgeving, en een groep van 13 ratten is een ozon-rijke omgeving geplaatst. Voor iedere rat is de gewichtstoename (in grammen) gemeten. We vragen ons af of ozon een negatieve invloed heeft op gewichtstoename. We onderzoeken dit door de hypothese $H_0 : \mu_1 = \mu_2$ te toetsen tegen de alternatieve hypothese $H_1 : \mu_1 > \mu_2$. Hierbij zijn μ_1 en μ_2 de gewichtstoename voor een rat in de ozon-vrije en ozon-rijke omgeving respectievelijk. In de ozon-vrije omgeving is de gemiddelde gewichtstoename $\bar{x}_{14} = 22.40$; in de ozon-rijke omgeving $\bar{y}_{13} = 11.01$. De gepoolde standaard-deviatie is 4.58.

   Veronderstel dat de data normaal verdeeld zijn, met gelijke variantie in beide groepen. De p-waarde van de t-toets is ongeveer gelijk aan: (kies het antwoord dat het beste past)

   a. 0.005 **b.** 0.05 **c.** 0.02 **d.** 0.01 **e.** 0.006 **f.** 0.001

8. Zij $\bar{X}_n$ and $\bar{Y}_m$ de steekproefgemiddelden van twee onafhankelijke steekproeven van omvang n respektievelijk m uit dezelfde kansverdeling met verwachting μ en variantie σ^2. We combineren de twee schatters tot een nieuwe schatter

$$T = r\bar{X}_n + (1-r)\bar{Y}_m,$$

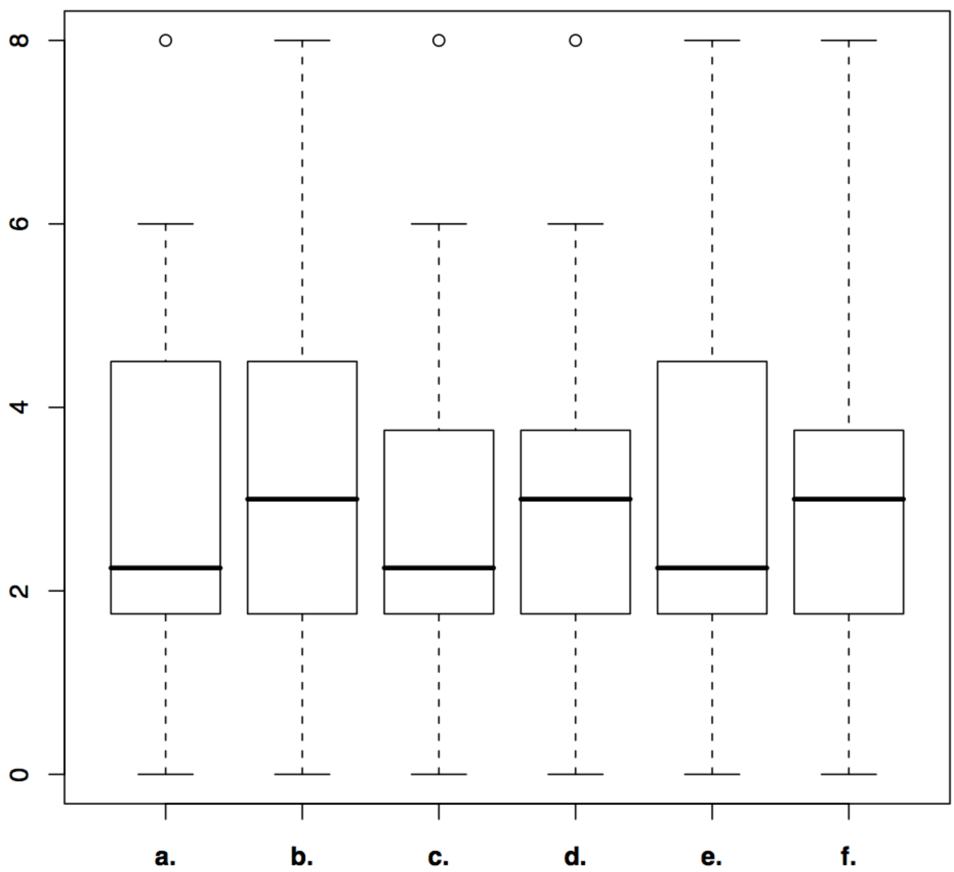
waarbij r is een getal is tussen 0 en 1. Dan geldt

a. T is zuiver voor μ met de kleinste MSE als $r = n/(m+n)$
b. T is zuiver voor μ met de kleinste MSE als $r = m/(m+n)$
c. T is zuiver voor μ met de kleinste MSE als $r = 1/2$
d. T is onzuiver voor μ met de kleinste MSE als $r = n/(m+n)$
e. T is onzuiver voor μ met de kleinste MSE als $r = m/(m+n)$
f. T is onzuiver voor μ met de kleinste MSE als $r = 1/2$

9. Gegeven is de dataset

0.00 1.5 1.75 1.75 2.00 2.25 3.25 4.5 4.5 6.00 8.00

Welke van de zes onderstaande boxplots hoort bij deze dataset?



Proeftentamen Deel 2 Kansrekening en Statistiek EE1M31

Naam:

Groep:

Cijfer:

Studentnummer:

Open vragen

Toelichting: Een antwoord alleen is *niet* voldoende: er dient een berekening, toelichting en/of motivatie aanwezig te zijn. Dit alles goed leesbaar en in goed Nederlands.

1. We genereren een getal x uit een uniforme verdeling op het interval $[0, \theta]$ ($\theta > 0$). We toetsen $H_0 : \theta = 3$ tegen $H_1 : \theta \neq 3$ door H_0 te verwerpen als $x \leq 0.5$ of $x \geq 2.5$.

   (a) (2 punten) Bepaal het kritieke gebied van de toets (met als toetsingsgrootheid de waarneming X), als we eisen dat het significantieniveau gelijk aan 0.1 moet zijn.

   (b) (1 punt) Bereken de kans op het begaan van een fype II fout als de echte waarde van θ gelijk is aan 3.5.

2. Volgens de wet van Hardy-Weinberg zijn gen-frequenties in evenwicht als de genotypen AA, Aa en aa voorkomen met kansen respectievelijk $(1-\theta)^2$, $2\theta(1-\theta)$ en θ^2. Hierbij is $0 < \theta < 1$ een onbekende parameter. We willen de maximum likelihood schatting berekenen voor θ. Bij een dataverzameling omtrent de Chinese bevolking van Hong Kong in 1937 kwamen de genotypen voor met de volgende aantallen:

Type	Aantal
1: AA	342
2: Aa	500
3: aa	187
Totaal	1029

(a) (1 punt) Leid af dat de likelihoodfunctie wordt gegeven door

$$L(\theta) = C \cdot 2^{500}(1-\theta)^{1184}\theta^{874},$$

waarbij C een constante is die het aantal volgordes aangeeft met 342 keer genotype 1, 500 keer genotype 2, en 187 keer genotype 3.

(b) (1.5 punten) Bepaal tevens de maximum likelihood schatting op grond van de bovenstaande dataverzameling.

3. Een steekproef van omvang 5 uit een normale verdeling met onbekende μ en onbekende σ resulteert in:

$$129.0 \quad 129.4 \quad 130.8 \quad 131.1 \quad 132.2$$

(a) (2 punten) Bereken een (symmetrisch) 90% betrouwbaarheidsinterval voor μ.

(b) (1.5 punten) Als we toetsen H_0: $\mu = 130$ tegen H_1: $\mu > 130$ bij een significantieniveau $\alpha = 0.1$, en daarvoor de t-toets gebruiken, wat is dan de conclusie? (M.a.w., verwerpen of niet verwerpen, en vergeet niet een **argument** te geven.)

Extra ruimte