

Kansrekening en statistiek
WI2211TI / WI2105IN–deel 2
2 februari 2012, 14.00–16.00 uur

VOOR WI2211TI: Bij dit examen is het gebruik van een (evt. grafische) rekenmachine toegestaan. Een formuleblad is niet toegestaan.

VOOR WI2105IN: Bij dit examen is het gebruik van een (evt. grafische) rekenmachine toegestaan. Een formuleblad wordt uitgereikt.

Meerkeuzevragen

Toelichting: In het algemeen zijn niet altijd twee van de drie alternatieven 100% fout, het juiste antwoord is het meest volledige antwoord. Maak op het bijgeleverde 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. Vergeet niet uw studienummer in te vullen en aan te strepen.

1. Gegeven is een dataverzameling die een realisatie is van een steekproef $X_1, X_2, \dots, X_n$ uit een geometrische verdeling met parameter p . We toetsen de hypothese $H_0 : p = \frac{1}{2}$ tegen $H_1 : p < \frac{1}{2}$ met toetsingsgrootte $T = \bar{X}_n$. We verwerpen H_0 ten gunste van H_1
 - a. voor waarden van T in de buurt van 2
 - b. voor hele grote waarden van T
 - c. voor waarden van T dichtbij 0
2. Men beschikt over een dataverzameling die een realisatie vormt van een steekproef $X_1, X_2, \dots, X_5$ uit een $N(\mu, 1)$ verdeling. Iemand toetst $H_0 : \mu = 0$ tegen $H_1 : \mu < 0$ met toetsingsgrootte $\bar{X}_n$. Stel hij besluit H_0 te verwerpen ten gunste van H_1 als voor de data geldt $\bar{x}_5 \leq -0.5$. Bereken in dat geval de kans op een type I fout.
 - a. 0.26
 - b. 0.31
 - c. 0.13
3. Vervolg op de voorgaande opgave. Bereken de type II fout als in werkelijkheid $\mu = -0.8$.
 - a. 0.07
 - b. 0.90
 - c. 0.25
4. Tien monsters van een nieuw product leveren de volgende waarden (in grammen)

15.02	14.78	15.11	15.16	14.88
15.35	15.35	15.10	15.18	15.14

Het gemiddelde hiervan is 15.11 en de (steekproef)standaardafwijking 0.181. De producent vraagt zich af of er sprake is van een significante afwijking ten opzichte van de gewenste waarde 15.0 gram. Toets hiervoor met de daarvoor meest geschikte toetsingsvariabele T de hypothese $H_0 : \mu = 15.0$ tegen $H_1 : \mu > 15.0$. Er geldt: T heeft een Student-verdeling met

- a. 9 vrijheidsgraden en H_0 wordt verworpen bij significantieniveau 0.05
 - b. 10 vrijheidsgraden en H_0 wordt niet verworpen bij significantieniveau 0.05
 - c. 9 vrijheidsgraden en H_0 wordt niet verworpen bij significantieniveau 0.05
5. Men vult machinaal wijnflessen met een hoeveelheid wijn die, gemeten in milliliters, $N(\mu, \sigma^2)$ verdeeld is met $\sigma = 3$ ml. We zijn voornemens voor een aantal wijnflessen de hoeveelheid wijn te bepalen. Stel dat we μ schatten met het steekproefgemiddelde van deze metingen. Als we met 95% zeker willen zijn dat de echte waarde van μ niet meer dan 1 ml afwijkt van onze schatting, hoeveel wijnflessen moeten we dan meten om dit te bereiken?
 - a. 138
 - b. 35
 - c. 75

6. De Pareto-verdeling is een kansverdeling waarbij de cumulatieve verdelingsfunctie wordt gegeven door

$$F(x) = \begin{cases} 0 & \text{als } x < 1 \\ 1 - \frac{1}{x^\alpha} & \text{als } x \geq 1 \end{cases}.$$

Hier is $\alpha > 0$ een (onbekende) parameter. Voor een dataset $x_1, x_2, \dots, x_n$ veronderstellen we een Pareto verdeling. De likelihoodfunctie voor α wordt gegeven door

a. $L(\alpha) = \left(1 - \frac{1}{x_1^\alpha}\right) \left(1 - \frac{1}{x_2^\alpha}\right) \cdots \left(1 - \frac{1}{x_n^\alpha}\right)$

b. $L(\alpha) = \frac{\alpha^n}{x_1^{\alpha+1} x_2^{\alpha+1} \cdots x_n^{\alpha+1}}$

c. $L(\alpha) = \frac{\alpha^n}{x_1^\alpha x_2^\alpha \cdots x_n^\alpha}$

7. Veertig metingen van het kookpunt van een nieuw ontwikkelde legering leveren het 98 % betrouwbaarheidsinterval (924°C , 932°C). Chemiprac besluit nog een serie van veertig bepalingen te doen. Wat valt er te zeggen over het betrouwbaarheidsinterval voor de gehele serie van tachtig metingen.

U mag aannemen dat de ‘meetfouten’ normaal verdeeld zijn.

- a. Het blijft gecentreerd rond 928°C , en wordt ongeveer gehalveerd.
b. Het midden zal iets verschuiven en het interval wordt ongeveer half zo lang.
c. Het midden zal iets verschuiven en de lengte wordt ongeveer 6.

8. Zij $X_1, X_2, \dots, X_n$ een steekproef van positieve getallen uit een verdeling met verwachting λ en variantie λ , waarbij $\lambda > 0$ een onbekende parameter is. Een schatter voor λ wordt gegeven door $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$. De MSE (mean squared error) van deze schatter worden gegeven door:

a. λ^2/n^2

b. λ/n

c. λ/n^2

9. Om de slaagkans van een experiment te onderzoeken wordt door elk van honderd personen een serie experimenten uitgevoerd tot ze voor het eerst succes hebben. Een model om dit te beschrijven is: het aantal pogingen benodigde N heeft een $Geo(p)$ -verdeling. We willen p schatten met de maximum likelihood methode. De gegevens:

aantal pogingen	1	2	3	≥ 4
frequentie	40	30	15	15

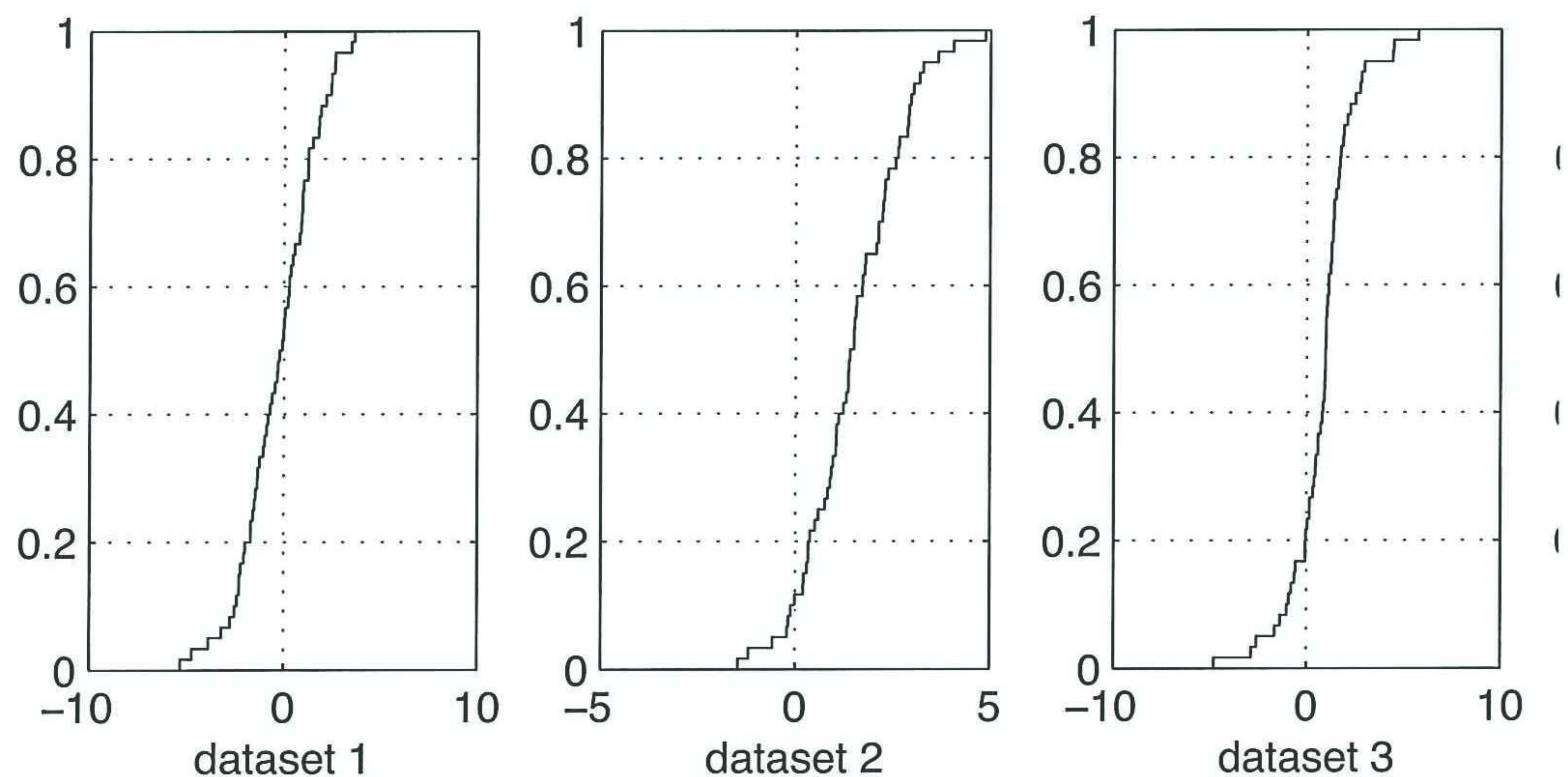
De likelihoodfunctie wordt dan

a. $p^{85}(1-p)^{15}$

b. $p^{85}(1-p)^{105}$

c. $p^{85}(1-p)^{60}$

10. Voor WI2105IN: Onderstaand plaatje toont de empirische verdelingsfuncties van drie datasets van 60 elementen. Welk van deze datasets heeft de kleinste mediaan?



a. dataset 1

b. dataset 2

c. dataset 3

10. Voor WI2211TI: Beschouw het tweesteeekproevenprobleem, waarbij we aannemen dat de data in beide groepen realisaties zijn van onafhankelijke normaal verdeelde stochastische variabelen. Welke van de onderstaande beweringen zijn waar?

A De gepoolde en ongepoolde variantieschatters geven altijd een verschillende waarde.

B De toetsingsgrootheid heeft een t -verdeling onder de nulhypothese, indien we de gepoolde variantieschatter gebruiken.

a. alleen [A]

b. alleen [B]

c. geen van beide

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. Een bepaalde soort heeft drie mogelijke typen bloed: AA, OO of OA, met kansen: θ^2 , $(1 - \theta)^2$ en $2\theta(1 - \theta)$ respectievelijk. Een steekproef van 150 exemplaren wordt onderzocht en de aantallen met bloedgroep AA, OO en OA blijken te zijn 15, 70 en 65 respectievelijk. Leid op grond hiervan de maximum-likelihoodschatting voor θ af.
2. 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. Bereken de kans op het begaan van een type I fout.
 - b. Bereken de kans op het begaan van een type II fout als de echte waarde van θ gelijk is aan 3.5.
 - c. Bepaal het kritieke gebied van de toets (met als toetsingsgrootte de waarneming X), als we eisen dat de kans op een type I fout precies 0.1 moet zijn.
3. Alleen voor WI2105IN Gegeven zijn metingen $x_1, x_2, \dots, x_{10}$.
 - a. Geef aan hoe je een bootstrap steekproef kunt genereren volgens de empirische bootstrap.
 - b. We nemen nu aan dat we de metingen kunnen opvatten als realisaties van stochastische variabelen $X_1, X_2, \dots, X_n$ met een exponentiële verdeling met onbekende parameter λ . Geef aan hoe je een bootstrap steekproef kunt genereren volgens de parametrische bootstrap.
3. Alleen voor WI2211TI Gegeven is een deel van een ANOVA tabel van een studie die is uitgevoerd volgens randomised block design. Het totaal aantal metingen in de studie bedraagt 12. Op vier plaatsen is de waarde weggelaten. Dit zijn de plaatsen waar A, B, C en D staan.

	Df	Sum sq	Mean Sq
block	4	0.00073	A
factor	3	B	0.35431
Residuals	C	D	0.00142

- a. Hoeveel niveaus (levels) heeft de factor?
- b. Geef de waarden van A, B en C in de tabel,

Puntenverdeling bij open vragen:

Opgave:	1	2a	2b	2c	3a	3b
Punten:	3	2	2	2	1	2

Bepaling van het eindcijfer: Bij de berekening van het eindcijfer wordt een apart cijfer berekend voor het meerkeuze- en open-vragen-deel. Bij het meerkeuzedeel zal een correctie voor gokken worden toegepast. Het eindcijfer is een gewogen gemiddelde van de twee verkregen cijfers, waarbij het meerkeuze-deel voor 2/3 en het open-vragen-deel voor 1/3 meetelt.