CTB2200

Kansrekening & Statistiek

Tentamenbundel Civiele Techniek

Het Gezelschap "Practische Studie"

LET OP! EEN REPRODUCERENDE LEERSTIJL IS SCHADELIJK VOOR DE ACADEMISCHE VORMING

April 2017
Januari 2017
Januari 2016*

April 2015*
Januari 2015*
Oefententamen*

*LET OP! TENTAMENS VAN VOOR COLLEGEJAAR 16/17 ZIJN NIET REPRESENTATIEF!!

Vul dit schema in om een overzicht te hebben van welke onderwerpen van het vak al goed gaan en welke je nog moet oefenen

	April 2017		Januari 2017		Januari 2016		April 2015		Januari 2015		Oefententamen	
Vraag	Onderwerp	Goed?	Onderwerp	Goed?	Onderwerp	Goed?	Onderwerp	Goed?	Onderwerp	Goed?	Onderwerp	Goed?
1												
2												
3												
4												
5												

Hertentamen Kansrekening en Statistiek CTB2200
11 april 2017, 13.30 – 16.30

Bij dit tentamen is het gebruik van een **niet-grafische** rekenmachine en een onbeschreven formuleblad toegestaan.
Dit tentamen bestaat uit vijftien 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:

$$\text{Cijfer} = \frac{MC + OV}{3} + 1,$$

waarbij MC het aantal punten voor meerkeuze-deel en OV het aantal punten voor open-vragen-deel is.
Toelichting meerkeuzevragen: Maak op het bijgeleverde meerkeuze-antwoordformulier het hokje behorende bij het door u gekozen alternatief zwart of blauw. Tip: kruis uw antwoord eerst aan op het tentamenblad, herzie eventueel later uw antwoord en vul het daarna pas in op het antwoordformulier.
Ten slotte, vergeet niet de versie, de vakcode, uw naam en uw studienummer in te vullen. De laatste dient u ook aan te strepen.

Meerkeuzevragen
Versie A

1. Bekijk de gebeurtenissen (*events*) A en B met $\mathrm{P}(A \mid B) = \frac{1}{2}, \mathrm{P}(B \mid A) = \frac{1}{2}$ en $P(A \cap B) = \frac{1}{6}$. Bereken $P(A \cup B)$.

   a. $\frac{3}{4}$ **b.** $\frac{1}{6}$ **c.** $\frac{5}{6}$ **d.** $\frac{1}{2}$ **e.** $\frac{1}{3}$ **f.** $\frac{2}{3}$

2. Laat $U \sim U(0,1)$ en beschouw de volgende gebeurtenissen:

$$A = \left\{U < \frac{1}{2}\right\}, \quad B = \left\{\frac{1}{4} < U < \frac{3}{4}\right\}, \quad C = \left\{U > \frac{1}{2}\right\}$$

   Welke paar (paren) van gebeurtenissen is (zijn) onafhankelijk (*independent*)?

   a. Geen **b.** A en B
   c. A en C **d.** B en C
   e. A en B, A en C **f.** A en B, B en C
   g. B en C, A en C **h.** A en B, A en C, B en C

3. Het aantal inwoners van steden kunnen we beschrijven met de stochast (*random variable*) X, waarvan de verdelingsfunctie (*distribution function*) gegeven wordt door

$$F_X(x) = \begin{cases} 0 & x < 100, \\ 1 - \frac{100}{x} & x \geq 100. \end{cases}$$

   Hierbij is x het aantal inwoners in duizenden[1].

   Bereken de kans (in 2 decimalen) dat een stad tussen de 1 miljoen en 10 miljoen inwoners heeft.

   a. 0.9 **b.** 0.23 **c.** 0.77 **d.** 0.01 **e.** 0.09 **f.** 0.99

[1] 300 000 inwoners komt dus overeen met $x = 300$

4. Er zijn twee partijen met landmeters. Partij A bestaat uit 20 landmeters waarvan er 1 stuk is, terwijl partij B 10 landmeters bevat waarvan er 2 stuk zijn. Een vriend van je kiest een van de twee partijen volstrekt willekeurig en kiest een landmeter uit (ook volstrekt willekeurig). Hij vertelt je dat de landmeter stuk is.

   Bereken de voorwaardelijke kans (*conditional probability*) dat je vriend partij A gekozen heeft.

   a. $\frac{1}{5}$ **b.** $\frac{2}{3}$ **c.** $\frac{1}{10}$ **d.** $\frac{1}{2}$ **e.** $\frac{1}{4}$ **f.** $\frac{1}{3}$

5. Na dit tentamen wil je het geleerde in de praktijk toepassen: je gaat met vrienden naar het casino om roulette te spelen. Een van je vrienden heeft de volgende strategie. Ze begint met 1 euro in te zettten op Rood: als ze wint, maakt ze 1 euro winst en stopt ze direct. Als ze verliest, zet ze in de volgende twee rondes allebei ook 1 euro in (ze heeft in dit geval dus in totaal 3 keer ingezet). Het roulette-wiel bestaat uit 37 getallen: 18 zwart, 18 rood en 1 groen.

   Bereken de kans (in 2 decimalen) dat je vriend een positieve hoeveelheid geld wint met deze strategie.

   a. 0.36 **b.** 0.12 **c.** 0.49 **d.** 0.64 **e.** 0.61 **f.** 0.24

6. Je bent nog steeds in het casino met je vrienden. Een andere vriend van je pakt het anders aan: hij zet iedere ronde 1 euro in op 8 én 11 tegelijk[2]: als een van beide gedraaid wordt, maakt hij 17 euro winst, anders verliest hij zijn inzet van 1 euro. Het roulette-wiel bestaat nog steeds uit 37 getallen.

   Stel dat je vriend genoeg geld bij zich heeft om deze strategie 1000 keer toe te passen. Wat is dan zijn verwachte resultaat na 1000 rondes?

   a. -1 euro **b.** -486 euro **c.** -132 euro **d.** -407 euro **e.** 0 euro **f.** -27 euro

7. Een multiple choice toets heeft 20 vragen. Iedere vraag heeft vier mogelijke antwoorden, waarvan er één goed is. Een goed antwoord levert 3 punten op, maar voor ieder fout antwoord wordt een punt afgetrokken. Een student besluit alle vragen volstrekt willekeurig te beantwoorden. Laat de stochast X het aantal punten zijn dat hij gescoord heeft.

   Bereken de verwachting (*expectation*) en variantie (*variance*) van X.

   a. $\mathrm{E}[X] = 0$ and $\mathrm{Var}(X) = 60$ **b.** $\mathrm{E}[X] = 20$ and $\mathrm{Var}(X) = 30$
   c. $\mathrm{E}[X] = 0$ and $\mathrm{Var}(X) = 30$ **d.** $\mathrm{E}[X] = -20$ and $\mathrm{Var}(X) = 30$
   e. $\mathrm{E}[X] = -20$ and $\mathrm{Var}(X) = 60$ **f.** $\mathrm{E}[X] = 20$ and $\mathrm{Var}(X) = 60$

8. De gezamenlijke dichtheidsfunctie (*joint density function*) f van twee continue stochasten X en Y wordt gegeven door

$$f(x,y) = \begin{cases} xe^{-y} & 0 \leq x \leq y, \\ 0 & \text{anders.} \end{cases}$$

   Bereken de marginale dichtheidsfunctie (*marginal density function*) f_Y van Y voor $y \geq 0$.

   a. ye^{-y^2} **b.** y^2e^{-y} **c.** $\frac{1}{2}ye^{-y}$ **d.** ye^{-y} **e.** e^{-y^2} **f.** $\frac{1}{2}y^2e^{-y}$

[2] Dit heet een *cheval*: het fiche wordt op het lijntje tussen de 8 en de 11 gelegd.

9. Beschouw de volgende twee beweringen:

   (I) Als de stochasten X en Y ongecorreleerd zijn, dan zijn ze ook onafhankelijk.

   (II) Als de gebeurtenissen A en B onafhankelijk zijn, dan geldt $P(A \cup B) = P(A) + P(B)$.

   Welke van deze beweringen is/zijn waar?

   a. Beide **b.** Geen **c.** Alleen (I) **d.** Alleen (II)

10. De sterkte S van een brug heeft een normale verdeling met verwachting $\mu_S = 40$ MPa en variantie $\sigma_S^2 = 11$. De belading B van de brug heeft een normale verdeling met verwachting $\mu_B = 32$ MPa en variantie $\sigma_B^2 = 5$. De sterkte en belading van de brug zijn onafhankelijke stochasten.

    Bereken de kans dat de belading groter is dan de sterkte van de brug.

    a. 0.0005 **b.** 0.0918 **c.** 0.1587 **d.** 0.0228 **e.** 0.0749 **f.** 0.3594

11. Beschouw de volgende twee beweringen:

    (I) Als een steekproef twee keer zo groot wordt en alle overige factoren gelijk blijven, wordt een betrouwbaarheidsinterval twee keer zo klein.

    (II) Een type II fout treedt op als de nulhypothese ten onrechte niet verworpen wordt.

    Welke van deze beweringen is/zijn waar?

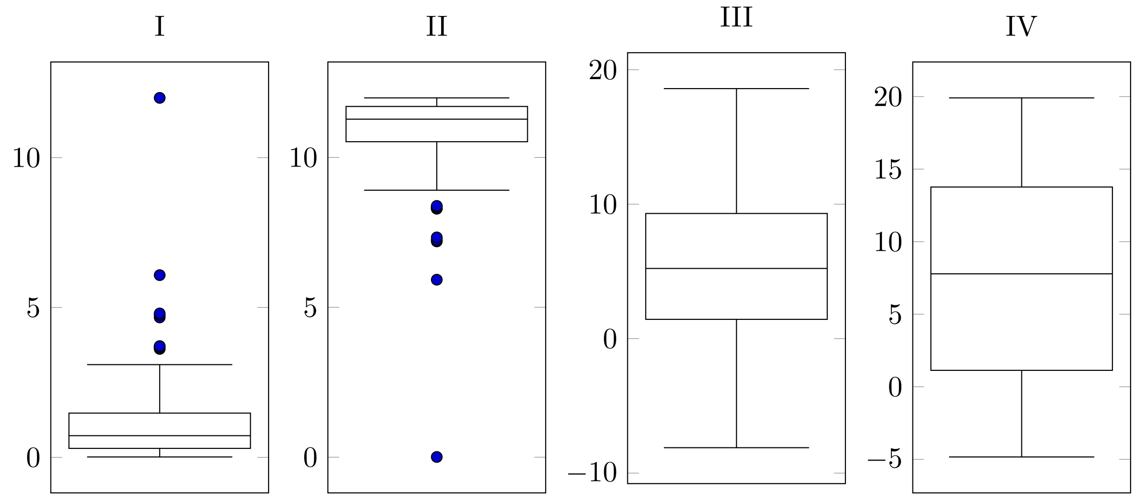
    a. Beide **b.** Geen **c.** Alleen (I) **d.** Alleen (II)

12. Je hebt 100 planken en de lengtes van de planken zijn onafhankelijk en gelijk verdeeld met verwachting $\mu = 100$ cm en variantie $\sigma^2 = 9$.

    Bereken de kans dat de totale lengte van de 100 planken minder is dan 10025 cm.

    a. 0.7967 **b.** 0.4681 **c.** 0.6103 **d.** 0.3897 **e.** 0.2033 **f.** 0.5319

13. Je ziet hieronder vier boxplots van vier verschillende datasets met 200 elementen.



    Welke dataset is afkomstig uit een normale verdeling en welke uit een exponentiële verdeling?

    a. Normaal: III; Exponentieel: I,II **b.** Normaal: IV; Exponentieel: I

    c. Normaal: III,IV; Exponentieel: I,II **d.** Normaal: IV; Exponentieel: II

    e. Normaal: III,IV; Exponentieel: II **f.** Normaal: III; Exponentieel: I

14. Van een steekproef ter grootte 100 heb je geteld hoeveel waarnemingen vallen in de cellen $[1, 1.5], (1.5, 2], (2, 2.5], (2.5, 3], (3, 4], (4, 5]$:

Cel	$[1, 1.5]$	$(1.5, 2]$	$(2, 2.5]$	$(2.5, 3]$	$(3, 4]$	$(4, 5]$
Aantal waarnemingen	19	24	33	15	7	2

Je wilt een histogram maken van de dataset met bovenstaande cellen. Wat is de hoogte van het histogram op de cel $(2.5, 3]$?

a. 15 **b.** 7.5 **c.** 30 **d.** 0.075 **e.** 0.30 **f.** 0.15

15. Laat $X \sim Exp(\lambda)$. Je wilt $H_0 : \lambda = 2$ toetsen tegen $H_1 : \lambda < 2$. Je verwerpt H_0 ten gunste van H_1 als $X \geq 1.5$.

Bereken de kans (in 2 decimalen) op een type I fout.

a. 0.02 **b.** 0.78 **c.** 0.05 **d.** 0.98 **e.** 0.22 **f.** 0.95

Hertentamen Kansrekening en Statistiek CTB2200
11 april 2017, 13.30 – 16.30

Naam:

Cijfer:

Studentnummer:

Open vragen

Motiveer altijd je antwoord en schrijf leesbaar.

1. De bekende bioloog Henk de Rijn heeft een vaardigheidstest voor chimpansees bedacht. We nemen aan dat iedere chimpansee kans p heeft om te slagen voor de test, onafhankelijk van eerdere pogingen en van andere apen. Laat X_i het aantal pogingen zijn dat chimpansee i nodig heeft om te slagen. Verder is $T = \bar{X}_n$ een schatter is voor $\frac{1}{p}$.

   (a) (2 punten) Bepaal de gemiddelde kwadratische fout (MSE) van T.

Precies 100 apen deden mee aan de test en hieronder zie je een samenvatting van het aantal pogingen dat de apen nodig hadden om te slagen voor de test:

Pogingen	1	2	3	meer dan 3
Aantal apen	49	27	15	9

   (b) (2 punten) Bepaal de meest aannemelijke schatting (*maximum likelihood estimate*) voor p.

2. In Tennessee (Verenigde Staten) is de concentratie opgeloste zuurstof gemeten in de stroming direct na een dam. Dit is bij 6 dammen gedaan en leverde als steekproefgemiddelde $\bar{x}_6 = 2.7$ (mg/L) en steekproefvariantie $s_6^2 = 3.1$ op. Neem aan dat de gemeten concentratie uit een normale verdeling met onbekende verwachting μ en variantie σ^2 komt.

   (a) (2 punten) Construeer een tweezijdig 95% betrouwbaarheidsinterval voor μ en bepaal hiermee of $H_0 : \mu = 1$ verworpen wordt ten gunste van $H_1 : \mu \neq 1$ bij een significantieniveau van $\alpha = 0.05$.

Voor het onderhoud aan de dammen is beton nodig en daarom wordt de druksterkte van 4 blokken beton, gemaakt volgens dezelfde procedure, getest in een bouwplaats. Het verkregen steekproefgemiddelde is $\bar{x}_4 = 32.1$ (MPa). Neem aan dat de druksterktes een $N(\mu, 10)$-verdeling hebben.

   (b) (2 punten) Onderzoek m.b.v. een geschikte statistische toets of de verwachte druksterkte μ minder dan 35 MPa is bij een significantieniveau van $\alpha = 0.10$.

3. Een veiligheidsconstructie bestaat uit twee systemen. De kans dat het eerste systeem niet werkt is 0.1. De kans dat het tweede systeem niet werkt als het eerste niet werkt is 0.5. Ten slotte, de kans dat het tweede systeem niet werkt als het eerste wel werkt is 0.1.

   Laat de stochast X de waarde 1 aannemen als het eerste system werkt en 0 anders. En de stochast Y is gelijk aan 1 als het tweede systeem werkt en 0 anders.

   (a) (1.5 punten) Bepaal de gezamenlijke kansmassafunctie van X en Y.

   Als je (a) niet opgelost hebt kun je deze incorrecte gezamenlijke kansmassafunctie gebruiken: $p(0,0) = 0.01$, $p(0,1) = 0.04$, $p(1,0) = 0.03$, $p(1,1) = 0.92$.

   (b) (1.5 punten) Bepaal of X en Y positief gecorreleerd, negatief gecorreleerd of ongecorreleerd zijn.

Z.O.Z. voor opgave 3(c)

Als je (a) niet opgelost hebt kun je deze incorrecte gezamenlijke kansmassafunctie gebruiken: $p(0,0) = 0.01$, $p(0,1) = 0.04$, $p(1,0) = 0.03$, $p(1,1) = 0.92$.

(c) (1 punt) Bereken $P(Y = 1|XY = 0)$.

Extra ruimte

Antwoorden:

1 d.

2 f.

3 e.

4 a.

5 e.

6 f.

7 a.

8 f.

9 b.

10 d.

11 d.

12 a.

13 f.

14 e.

15 c.

Tentamen Kansrekening en Statistiek CTB2200
24 januari 2017, 13.30 – 16.30

Bij dit tentamen is het gebruik van een **niet-grafische** rekenmachine en een onbeschreven formuleblad toegestaan.
Dit tentamen bestaat uit vijftien 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:

$$\text{Cijfer} = \frac{MC + OV}{3} + 1,$$

waarbij MC het aantal punten voor meerkeuze-deel en OV het aantal punten voor open-vragen-deel is.
Toelichting meerkeuzevragen: Maak op het bijgeleverde meerkeuze-antwoordformulier het hokje behorende bij het door u gekozen alternatief zwart of blauw. Tip: kruis uw antwoord eerst aan op het tentamenblad, herzie eventueel later uw antwoord en vul het daarna pas in op het antwoordformulier.
Ten slotte, vergeet niet de versie, de vakcode, uw naam en uw studienummer in te vullen. De laatste dient u ook aan te strepen.

Meerkeuzevragen
Versie A

1. A en B zijn disjuncte gebeurtenissen (*disjoint events*) met $P(A) = \frac{1}{3}$ en $P(B) = \frac{1}{4}$.

   Wat is de kans dat A optreedt, maar B niet?

   a. $\frac{1}{4}$ **b.** $\frac{1}{3}$ **c.** $\frac{1}{6}$ **d.** $\frac{7}{12}$ **e.** $\frac{1}{2}$ **f.** $\frac{1}{12}$

2. In de stad Technopolis wonen twee keer zo veel vrouwen als mannen. Het is bekend dat 0.25% van de vrouwen en 5% van de mannen en kleurenblind is.

   Gegeven dat een willekeurige inwoner van Technopolis kleurenblind is, wat is de voorwaardelijke kans (*conditional probability*) in 3 decimalen dat deze inwoner een vrouw is?

   a. 0.25 **b.** 0.667 **c.** 0.005 **d.** 0.91 **e.** 0.091 **f.** 0.063

3. Je gooit drie keer met een zuivere munt. Bekijk de volgende drie gebeurtenissen:

$$\begin{aligned} A &= \{\text{minimaal een keer kop en minimaal een keer munt}\} \\ B &= \{\text{maximaal een keer kop}\} \\ C &= \{\text{eerste worp is kop}\} \end{aligned}$$

   Welk(e) paar (paren) van gebeurtenissen is (zijn) onafhankelijk (*independent*)?

   a. geen **b.** B en C **c.** A en B
   d. A en C, B en C **e.** A en B, A en C **f.** A en C

4. Laat X een $Par(2)$-verdeling hebben.

   Bereken $\mathrm{E}\left[\frac{1}{X}\right]$.

   a. $\frac{1}{3}$ **b.** $\frac{1}{4}$ **c.** 1 **d.** $\frac{1}{2}$ **e.** $\frac{3}{4}$ **f.** $\frac{2}{3}$

5. Laat $X_1, \dots, X_n$ een aselecte steekproef (*random sample*) zijn uit een $Ber(p)$-verdeling. Bekijk de volgende twee beweringen:

(I) $\frac{1}{n}\sum_{i=1}^{n} X_i$ is een zuivere schatter (*unbiased estimator*) voor p.

(II) $\frac{1}{n}\sum_{i=1}^{n} X_i^2$ is een zuivere schatter voor p.

Welke van deze beweringen is/zijn waar?

a. Alleen (II) **b.** Alleen (I) **c.** Beide **d.** Geen

6. Laat de stochast X een exponentiële verdeling hebben met parameter $\lambda = 3$.

Bereken de voorwaardelijke kans $\mathrm{P}(X > 3 \mid X > 1)$.

a. e^{-9} **b.** e^{-6} **c.** e^{-1} **d.** $1 - e^{-3}$ **e.** $1 - e^{-\frac{2}{3}}$ **f.** $e^{-\frac{2}{3}}$

7. Laat $X \sim U(0, 1)$ en $Y \sim U(0, 2)$. De stochasten X en Y zijn onafhankelijk.

Bereken de kans dat X groter is dan Y.

a. $\frac{1}{4}$ **b.** 0 **c.** $\frac{1}{3}$ **d.** $\frac{3}{4}$ **e.** $\frac{1}{2}$ **f.** $\frac{2}{3}$

8. Beschouw de volgende twee beweringen:

(I) Er geldt altijd dat $\mathrm{E}[aX + bY] = a\mathrm{E}[X] + b\mathrm{E}[Y]$.

(II) Als X en Y positief gecorreleerd (*positively correlated*) zijn, dan zijn $U = 1 - X$ en $V = 2 + Y$ ook positief gecorreleerd.

Welke van deze beweringen is/zijn waar?

a. Beide **b.** Geen **c.** Alleen (I) **d.** Alleen (II)

9. Een meting van de zeegolfhoogte bij Hoek van Holland heeft verwachting (*expectation*) $\mu = 200$ (cm) en variantie (*variance*) $\sigma^2 = 3969$. Laat $\bar{X}_{81}$ het gemiddelde zijn van 81 onafhankelijke metingen van de zeegolfhoogte bij Hoek van Holland.

Benader de kans dat $\bar{X}_{81}$ kleiner is dan 210 cm.

a. 0.9236 **b.** 0.0764 **c.** 0.4920 **d.** 0.5080 **e.** 0.2420 **f.** 0.7580

10. Laat $X \sim U(-\theta, \theta)$. Je wilt $H_0 : \theta = 1$ toetsen tegen $H_0 : \theta < 1$. Je verwerpt H_0 ten gunste van H_1 als $|X| \leq 0.1$.

Bereken de kans op een type II fout als de ware waarde van θ gelijk aan $\frac{1}{2}$ is.

a. $\frac{9}{10}$ **b.** $\frac{2}{5}$ **c.** $\frac{1}{10}$ **d.** $\frac{1}{5}$ **e.** $\frac{4}{5}$ **f.** $\frac{3}{5}$

11. De sterkte S van een brug heeft een normale verdeling met verwachting $\mu_S = 35$ MPa en variantie $\sigma_S^2 = 6$. De belading B van de brug heeft een normale verdeling met verwachting $\mu_B = 29$ MPa en variantie $\sigma_B^2 = 3$. De sterkte en belading van de brug zijn onafhankelijke stochasten.

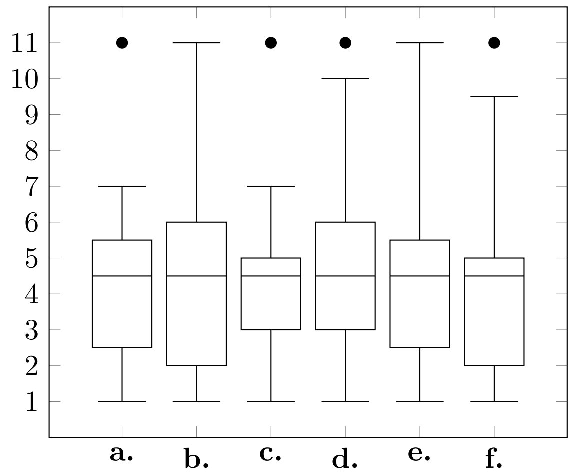
Bereken de kans dat de belading groter is dan de sterkte van de brug.

a. 0.1151 **b.** 0.0037 **c.** 0.3594 **d.** 0.0228 **e.** 0.2514 **f.** 0.0694

12. Bekijk de volgende geordende dataset:

$$1 \quad 1 \quad 3 \quad 3 \quad 4 \quad 5 \quad 5 \quad 5 \quad 7 \quad 11$$

Welke van onderstaande boxplots is de boxplot behorende bij deze dataset?



13. Laat T_1 en T_2 twee onafhankelijke zuivere schatters voor θ zijn. Er geldt: $\text{Var}(T_1) = 1$ en $\text{Var}(T_2) = 2$. Bekijk de nieuwe schatter $T_3 = 2T_1 - T_2$.

Wat is de gemiddelde kwadratische fout (*mean squared error*) van T_3?

a. 2 **b.** 0 **c.** $4 + \theta^2$ **d.** 6 **e.** 4 **f.** θ^2

14. Beschouw de volgende twee beweringen:

(I) Een $100(1-\alpha)\%$ betrouwbaarheidsinterval (*confidence interval*) wordt groter als de onbetrouwbaarheid α groter wordt (en alle overige factoren gelijk blijven)

(II) De p-waarde is de kans dat nulhypothese juist is.

Welke van deze beweringen is/zijn waar?

a. Beide **b.** Geen **c.** Alleen (I) **d.** Alleen (II)

15. Laat X en Y twee Bernoulli verdeelde discrete stochasten zijn, en laat verder gegeven zijn dat

$$\text{P}(X = 0, Y = 0) = \text{P}(X = 0, Y = 1) = \text{P}(X = 1, Y = 0) = \frac{1}{3}$$

.

Welke van de volgende uitspraken is juist?

a. X en Y zijn ongecorreleerd (*uncorrelated*)en afhankelijk (*dependent*).
b. X en Y zijn ongecorreleerd en onafhankelijk.
c. X en Y zijn gecorreleerd en onafhankelijk.
d. X en Y zijn positief gecorreleerd en afhankelijk.
e. X en Y zijn negatief gecorreleerd en afhankelijk.
f. Er zijn niet genoeg gegevens om deze opgave op te kunnen oplossen.

Lever de vragen ook weer in!

Tentamen Kansrekening en Statistiek CTB2200
24 januari 2017, 13.30 – 16.30

Naam:

Cijfer:

Studentnummer:

Open vragen

Motiveer altijd je antwoord en schrijf leesbaar.

1. Na afloop van een nieuwjaarsgala komt Nina bij haar huis. Helaas heeft ze iets te diep in de shotglaasjes gekeken: ze weet niet meer wat haar huissleutel is. Van de drie sleutels die ze heeft, kiest ze elke keer op volstrekt willekeurige wijze een sleutel uit om te proberen.

   (a) (1 punt) Bereken de kans dat Nina in maximaal 4 keer haar voordeur weet te openen.

Denise, een huisgenootje van Nina, was ook op het gala. Zij ging eerder naar huis en is er wat beter aan toe: als ze een van haar drie sleutels op volstrekt willekeurige wijze gekozen heeft en deze past niet, probeert ze deze later niet opnieuw.

   (b) (1 punt) Laat D het nummer van de poging zijn waarin Denise haar voordeursleutel vindt. Bepaal de kansmassafunctie (*probability mass function*) van D.

   (c) (2 punten) Bereken de kans dat Denise minder pogingen dan Nina nodig had om haar deur te openen.

2. Bekijk de volgende geordende dataset:

0.110	0.127	0.171	0.187	0.207	0.209	0.241	0.252	0.339	0.452
0.468	0.489	0.581	0.767	0.932	0.968	1.145	1.338	2.000	2.335

(a) (2 punten) Teken het histogram voor deze dataset met cellen (*bins*) $[0, 0.5], (0.5, 1], (1.0, 1.5], (1.5, 2], (2, 2.5]$.

Neem aan dat bovenstaande dataset een realisatie is van een steekproef uit een $Exp(\lambda)$-verdeling.

(b) (2 punten) Bereken de meest aannemelijke schatting (*maximum likelihood estimate*) voor λ.

Enkele gegevens van de dataset (slechts één is nodig voor de oplossing):
$\sum x_i = 13.318$, $\sum x_i^2 = 16.363$, $\sum \ln x_i = -15.949$, $\sum e^{-x_i} = 11.835$.

3. Als je een muntje in een frisdrankautomaat gooit, weegt de automaat het muntje om te bepalen hoeveel het waard is.

   Van 100 vijftigcentmuntjes (geproduceerd in Nederland) is het gewicht bepaald door een frisdrankautomaat: $\bar{x} = 7.49$ gram en $s_{NL}^2 = 0.011$.

   (a) (2 punten) Construeer een (benaderend) tweezijdig 95%-betrouwbaarheidsinterval voor μ en bepaal hiermee of $H_0 : \mu = 7.5$ verworpen wordt ten gunste van $H_1 : \mu \neq 7.5$ op een significantieniveau van $\alpha = 0.05$.

Bij 80 vijftigcentmuntjes die geproduceerd zijn in België is hetzelfde gedaan en dit leverde $\bar{y} = 7.52$ en $s_B^2 = 0.015$ op.

   (b) (2 punten) Onderzoek m.b.v. een geschikte statistische toets of vijftigcentmunten die geproduceerd zijn in Nederland en België even zwaar zijn, op significantieniveau $\alpha = 0.10$.

Extra ruimte

Antwoorden:

1 b.

2 e.

3 e.

4 f.

5 c.

6 b.

7 a.

8 c.

9 a.

10 e.

11 d.

12 a.

13 d.

14 b.

15 e.

a $1 - (\frac{2}{3})^4 = \frac{65}{81}$.

b $p_D(1) = \frac{1}{3} = p_D(2) = p_D(3)$.

c $\frac{2}{9} + \frac{4}{27} + \frac{8}{81} = \frac{38}{81}$.

a Hoogtes: $1.2, 0.4, 0.2, 0.1, 0.1$

b $\hat{\lambda} = 1.50$

a 95%-BI is $(7.47, 7.51)$.

b $t \approx -1.76$, $p-$waarde $\approx 0.0784 < 0.10$.

Extra ruimte

Tentamen Kansrekening en Statistiek CTB2200

24 januari 2017, 13.30 – 16.30

Naam:

Cijfer:

Studentnummer:

Open vragen

Motiveer altijd je antwoord en schrijf leesbaar.

1. Na afloop van een nieuwjaarsgala komt Nina bij haar huis. Helaas heeft ze iets te diep in de shotglaasjes gekeken: ze weet niet meer wat haar huissleutel is. Van de drie sleutels die ze heeft, kiest ze elke keer op volstrekt willekeurige wijze een sleutel uit om te proberen.

   (a) (1 punt) Bereken de kans dat Nina in maximaal 4 keer haar voordeur weet te openen.

$N = \#$pogingen $\sim Geo(\frac{1}{3})$.

$P(N \le 4) = 1 - P(N > 4) = 1 - (\frac{2}{3})^4 = \frac{65}{81} \approx 0.8025$.

Ook goed: $P(N \le 4) = \sum_{k=1}^{4} P(N=k) = \frac{1}{3} + \frac{1}{3}\cdot\frac{2}{3} + \frac{1}{3}(\frac{2}{3})^2 + \frac{1}{3}(\frac{2}{3})^3$.

Denise, een huisgenootje van Nina, was ook op het gala. Zij ging eerder naar huis en is er wat beter aan toe: als ze een van haar drie sleutels op volstrekt willekeurige wijze gekozen heeft en deze past niet, probeert ze deze later niet opnieuw.

   (b) (1 punt) Laat D het nummer van de poging zijn waarin Denise haar voordeursleutel vindt. Bepaal de kansmassafunctie (*probability mass function*) van D.

$P(D=1) = \frac{1}{3}$, $P(D=2) = \frac{2}{3}\cdot\frac{1}{2} = \frac{1}{3}$, $P(D=3) = \frac{2}{3}\cdot\frac{1}{2} = \frac{1}{3}$.

   (c) (2 punten) Bereken de kans dat Denise minder pogingen dan Nina nodig had om haar deur te openen.

$P(D<N) = P(D=1; N>1) + P(D=2; N>2) + P(D=3; N>3)$.

$P(D=1; N>1) = P(D=1)P(N>1) = \frac{1}{3}\cdot\frac{2}{3} = \frac{2}{9}$.

$P(D=2; N>2) = P(D=2)P(N>2) = \frac{1}{3}(\frac{2}{3})^2 = \frac{4}{27}$

$P(D=3; N>3) = P(D=3)P(N>3) = \frac{1}{3}(\frac{2}{3})^3 = \frac{8}{81}$

Dus: $P(D<N) = \frac{2}{9} + \frac{4}{27} + \frac{8}{81} = \frac{18+12+8}{81} = \frac{38}{81} \approx 0.47$.

Ook goed: $P(D<N) = 1 - P(D \ge N) = 1 - (P(N=1; D=1) + P(N=1; D=2) + \cdots + P(N=3; D=3))$

2. Bekijk de volgende geordende dataset:

0.110	0.127	0.171	0.187	0.207	0.209	0.241	0.252	0.339	0.452
0.468	0.489	0.581	0.767	0.932	0.968	1.145	1.338	2.000	2.335

(a) (2 punten) Teken het histogram voor deze dataset met cellen (*bins*) $[0, 0.5]$, $(0.5, 1]$, $(1.0, 1.5]$, $(1.5, 2]$, $(2, 2.5]$.

Hoogte op cel $[0,0.5] = \dfrac{\#\{x_i : 0 \le x_i \le 0.5\}}{n \times \text{lengte}([0,0.5])} = \dfrac{12}{20 \cdot 0.5} = 1.2$

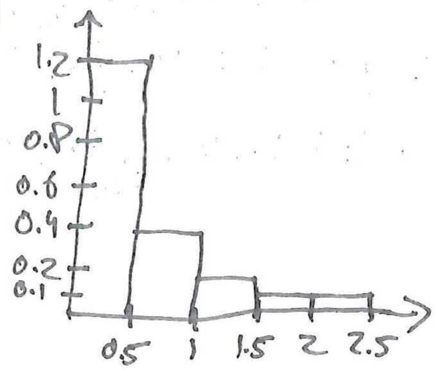
" $(0.5,1] = \frac{4}{10} = 0.4$

" $(1,1.5] = \frac{2}{10} = 0.2$

" $(1.5,2] = \frac{1}{10} = 0.1$

" $(2,2.5] = \frac{1}{10} = 0.1$

Histogram



Neem aan dat bovenstaande dataset een realisatie is van een steekproef uit een $Exp(\lambda)$-verdeling.

(b) (2 punten) Bereken de meest aannemelijke schatting (*maximum likelihood estimate*) voor λ.

Enkele gegevens van de dataset (slechts één is nodig voor de oplossing): $\sum x_i = 13.318$, $\sum x_i^2 = 16.363$, $\sum \ln x_i = -15.949$, $\sum e^{-x_i} = 11.835$.

$L(\lambda) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_i x_i}$

$\ell(\lambda) = \ln(L(\lambda)) = n \ln \lambda - \lambda \sum_i x_i$

$\ell'(\lambda) = \frac{n}{\lambda} - \sum_i x_i = 0$ als $\lambda = \frac{n}{\sum x_i}$.

$n = 20$, $\sum x_i = 13.318$, dus $\hat{\lambda} = \frac{20}{13.318} \approx 1.50$

3. Als je een muntje in een frisdrankautomaat gooit, weegt de automaat het muntje om te bepalen hoeveel het waard is.

Van 100 vijftigcentmuntjes (geproduceerd in Nederland) is het gewicht bepaald door een frisdrankautomaat: $\bar{x} = 7.49$ gram en $s^2_{NL} = 0.011$.

(a) (2 punten) Construeer een (benaderend) tweezijdig 95%-betrouwbaarheidsinterval voor μ en bepaal hiermee of $H_0 : \mu = 7.5$ verworpen wordt ten gunste van $H_1 : \mu \neq 7.5$ op een significantieniveau van $\alpha = 0.05$.

Variantie en verdeling onbekend, maar n groot, dus benaderend $100(1-\alpha)\%$ B.I. voor μ is $\bar{x}_n \pm z_{\alpha/2} \frac{s_n}{\sqrt{n}}$.

Hier $\alpha = 0.05$, dus $z_{0.025} = 1.96$.

We krijgen: $\bar{x}_n \pm z_{0.025} \frac{s_n}{\sqrt{n}} = 7.49 \pm 1.96 \frac{\sqrt{0.011}}{\sqrt{100}} = (7.47, 7.51)$

Omdat $\mu_0 = 7.5 \in (7.47, 7.51)$ verwerpen we H_0 niet.

Bij 80 vijftigcentmuntjes die geproduceerd zijn in België is hetzelfde gedaan en dit leverde $\bar{y} = 7.52$ en $s^2_B = 0.015$ op.

(b) (2 punten) Onderzoek m.b.v. een geschikte statistische toets of vijftigcentmunten die geproduceerd zijn in Nederland en België even zwaar zijn, op significantieniveau $\alpha = 0.10$.

$H_0 : \mu_{NL} = \mu_B$, $H_1 : \mu_{NL} \neq \mu_B$.

Verdelingen onbekend, maar n & m groot, dus $T = \frac{\bar{X}_n - \bar{Y}_m}{S_d} \approx N(0,1)$ met $S_d = \sqrt{\frac{S^2_{NL}}{n} + \frac{S^2_B}{m}}$.

Waargenomen waarde: $S_d = \sqrt{\frac{0.011}{100} + \frac{0.015}{80}} \approx 0.017$

$\rightarrow t = \frac{7.49 - 7.52}{0.017} = -1.76$.

H_1 is tweezijdig, dus kritieke waarden zijn $z_{0.05} = 1.645$ en $-z_{0.05} = -1.645$.

Omdat $t = -1.76 < -1.645$, verwerpen we H_0 t.g.v. H_1.

Ook goed: p-waarde $= \overset{\text{tweezijdig toetsen!}}{2} \cdot P(T \le -1.76 \mid H_0) = 2 \cdot 0.0392 = 0.0784 < 0.10$

dus verwerp H_0 t.g.v. H_1

Tentamen Kansrekening en Statistiek CTB2200
24 januari 2017, 13.30 – 16.30

Bij dit tentamen is het gebruik van een **niet-grafische** rekenmachine en een onbeschreven formuleblad toegestaan.
Dit tentamen bestaat uit vijftien 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:

$$\text{Cijfer} = \frac{MC + OV}{3} + 1,$$

waarbij MC het aantal punten voor meerkeuze-deel en OV het aantal punten voor open-vragen-deel is.
Toelichting meerkeuzevragen: Maak op het bijgeleverde meerkeuze-antwoordformulier het hokje behorende bij het door u gekozen alternatief zwart of blauw. Tip: kruis uw antwoord eerst aan op het tentamenblad, herzie eventueel later uw antwoord en vul het daarna pas in op het antwoordformulier.
Ten slotte, vergeet niet de versie, de vakcode, uw naam en uw studienummer in te vullen. De laatste dient u ook aan te strepen.

Meerkeuzevragen
Versie A

1. A en B zijn disjuncte gebeurtenissen (*disjoint events*) met $P(A) = \frac{1}{3}$ en $P(B) = \frac{1}{4}$.
   Wat is de kans dat A optreedt, maar B niet? $P(A \cap B^c) = P(A)$, want A en B disjunct.

   a. $\frac{1}{4}$ **b.** $\frac{1}{3}$ **c.** $\frac{1}{6}$ **d.** $\frac{7}{12}$ **e.** $\frac{1}{2}$ **f.** $\frac{1}{12}$

A

B

2. In de stad Technopolis wonen twee keer zo veel vrouwen als mannen. Het is bekend dat 0.25% van de vrouwen en 5% van de mannen en kleurenblind is.
   Gegeven dat een willekeurige inwoner van Technopolis kleurenblind is, wat is de voorwaardelijke kans (*conditional probability*) in 3 decimalen dat deze inwoner een vrouw is?

   a. 0.25 **b.** 0.667 **c.** 0.005 **d.** 0.91 **e.** 0.091 **f.** 0.063

$P(V|K) = \frac{P(K|V)P(V)}{P(K|V)P(V) + P(K|M)P(M)} = \frac{0.0025 \times \frac{2}{3}}{0.0025 \times \frac{2}{3} + 0.05 \times \frac{1}{3}} \approx 0.091$

3. Je gooit drie keer met een zuivere munt. Bekijk de volgende drie gebeurtenissen:

   $A = \{\text{minimaal een keer kop en minimaal een keer munt}\} = \{KMM, MKM, MMK, KKM, KMK, MKK\}$
   $B = \{\text{maximaal een keer kop}\} = \{KMM, MKM, MMK, MMM\}$
   $C = \{\text{eerste worp is kop}\} = \{KMM, KMK, KKM, KKK\}$.

   Welk(e) paar (paren) van gebeurtenissen is (zijn) onafhankelijk (*independent*)?

   a. geen **b.** B en C **c.** A en B
   d. A en C, B en C **e.** A en B, A en C **f.** A en C

$A \cap B = \{KMM, MKM, MMK\}$
$A \cap C = \{KMM, KMK, KKM\}$
$B \cap C = \{KMM\}$.
Alle uitkomsten hebben kans $\frac{1}{8}$.

4. Laat X een $Par(2)$-verdeling hebben.
   Bereken $\mathrm{E}\left[\frac{1}{X}\right]$.

   a. $\frac{1}{3}$ **b.** $\frac{1}{4}$ **c.** 1 **d.** $\frac{1}{2}$ **e.** $\frac{3}{4}$ **f.** $\frac{2}{3}$

$E\left[\frac{1}{X}\right] = \int_1^\infty \frac{1}{x} \cdot \frac{2}{x^3}\,dx = \left.\frac{-2}{3}\frac{1}{x^3}\right|_{x=1}^{\infty} = \frac{2}{3}$

$E[X_i] = 0\cdot(1-p) + 1\cdot p = p$, $E[X_i^2] = 0^2\cdot(1-p) + 1^2\cdot(p) = p$.

5. Laat $X_1, \ldots, X_n$ een aselecte steekproef (*random sample*) zijn uit een $Ber(p)$-verdeling. Bekijk de volgende twee beweringen: $E[\frac{1}{n}\sum_i X_i] = \frac{1}{n}\sum_i E[X_i] = \frac{1}{n}\cdot n\cdot p = p.$

   (I) $\frac{1}{n}\sum_{i=1}^{n} X_i$ is een zuivere schatter (*unbiased estimator*) voor p.

   (II) $\frac{1}{n}\sum_{i=1}^{n} X_i^2$ is een zuivere schatter voor p. $E[\frac{1}{n}\sum_i X_i^2] = \frac{1}{n}\sum_i E[X_i^2] = \frac{1}{n}\cdot n\cdot p = p.$

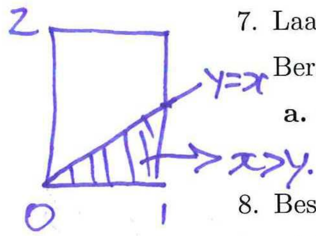
   Welke van deze beweringen is/zijn waar?

   a. Alleen (II) **b.** Alleen (I) **c.** Beide **d.** Geen

6. Laat de stochast X een exponentiële verdeling hebben met parameter $\lambda = 3$. Bereken de voorwaardelijke kans $P(X > 3 \mid X > 1)$.

   a. e^{-9} **b.** e^{-6} **c.** e^{-1} **d.** $1 - e^{-3}$ **e.** $1 - e^{-\frac{2}{3}}$ **f.** $e^{-\frac{2}{3}}$

   $P(X>3 \mid X>1) = \frac{P(X>3; X>1)}{P(X>1)} = P(X>3)/P(X>1) = e^{-3\cdot 3}/e^{-3\cdot 1} = e^{-6}$



7. Laat $X \sim U(0,1)$ en $Y \sim U(0,2)$. De stochasten X en Y zijn onafhankelijk. Bereken de kans dat X groter is dan Y.

   a. $\frac{1}{4}$ **b.** 0 **c.** $\frac{1}{3}$ **d.** $\frac{3}{4}$ **e.** $\frac{1}{2}$ **f.** $\frac{2}{3}$

   $P(X>Y) = \text{opp}(\triangle)/\text{opp}(\square) = \frac{1/2}{2} = \frac{1}{4}.$

8. Beschouw de volgende twee beweringen:

   (I) Er geldt altijd dat $E[aX + bY] = aE[X] + bE[Y]$. Waar, zie Hfst. 10.

   (II) Als X en Y positief gecorreleerd (*positively correlated*) zijn, dan zijn $U = 1 - X$ en $V = 2 + Y$ ook positief gecorreleerd. Niet waar: $Cov(U,V) = (-1)\cdot 1\cdot Cov(X,Y) = -Cov(X,Y) < 0.$

   Welke van deze beweringen is/zijn waar?

   a. Beide **b.** Geen **c.** Alleen (I) **d.** Alleen (II)

9. Een meting van de zeegolfhoogte bij Hoek van Holland heeft verwachting (*expectation*) $\mu = 200$ (cm) en variantie (*variance*) $\sigma^2 = 3969$. Laat $\bar{X}_{81}$ het gemiddelde zijn van 81 onafhankelijke metingen van de zeegolfhoogte bij Hoek van Holland. Benader de kans dat $\bar{X}_{81}$ kleiner is dan 210 cm.

   a. 0.9236 **b.** 0.0764 **c.** 0.4920 **d.** 0.5080 **e.** 0.2420 **f.** 0.7580

   $P(\bar{X}_{81} < 210) = P\left(\frac{\bar{X}_{81}-200}{\sqrt{3969/81}} < \frac{210-200}{\sqrt{\cdots}}\right) \approx P(Z<1.43) = 1 - 0.0764 = 0.9236$

10. Laat $X \sim U(-\theta, \theta)$. Je wilt $H_0 : \theta = 1$ toetsen tegen $H_1 : \theta < 1$. Je verwerpt H_0 ten gunste van H_1 als $|X| \leq 0.1$. $P(\text{type II fout} \mid \theta = \frac{1}{2}) = P(|X| > 0.1 \mid \theta = \frac{1}{2}) = \frac{0.4+0.4}{1} = \frac{4}{5}.$

    Bereken de kans op een type II fout als de ware waarde van θ gelijk aan $\frac{1}{2}$ is.

    a. $\frac{9}{10}$ **b.** $\frac{2}{5}$ **c.** $\frac{1}{10}$ **d.** $\frac{1}{5}$ **e.** $\frac{4}{5}$ **f.** $\frac{3}{5}$

11. De sterkte S van een brug heeft een normale verdeling met verwachting $\mu_S = 35$ MPa en variantie $\sigma_S^2 = 6$. De belading B van de brug heeft een normale verdeling met verwachting $\mu_B = 29$ MPa en variantie $\sigma_B^2 = 3$. De sterkte en belading van de brug zijn onafhankelijke stochasten.

    $S - B \sim N(6, 6+3) = N(6,9)$

    Bereken de kans dat de belading groter is dan de sterkte van de brug.

    a. 0.1151 **b.** 0.0037 **c.** 0.3594 **d.** 0.0228 **e.** 0.2514 **f.** 0.0694

    $P(B>S) = P(S-B<0) = P\left(\frac{S-B-6}{3} < \frac{0-6}{3}\right) = P(Z<-2) = 0.0228$

$Q_1 = q_{10}(0.25) = x_{(2)} + 0.75(x_{(3)} - x_{(2)}) = 1 + 0.75(3-1) = 2.5$

$0.25 \times (10+1) = 2.75$

mediaan $= q_{10}(0.5) = \frac{4+5}{2} = 4.5$. $Q_3 = x_{(8)} + 0.25(x_{(9)} - x_{(8)}) = 5 + 0.25(7-5)$

$= 5.5$

12. Bekijk de volgende geordende dataset:

$0.75 \times (10+1) = 8.25$ $n = 10$

$$1 \quad 1 \quad 3 \quad 3 \quad 4 \quad 5 \quad 5 \quad 5 \quad 7 \quad 11$$

$IQR = 5.5 - 2.5 = 3.$

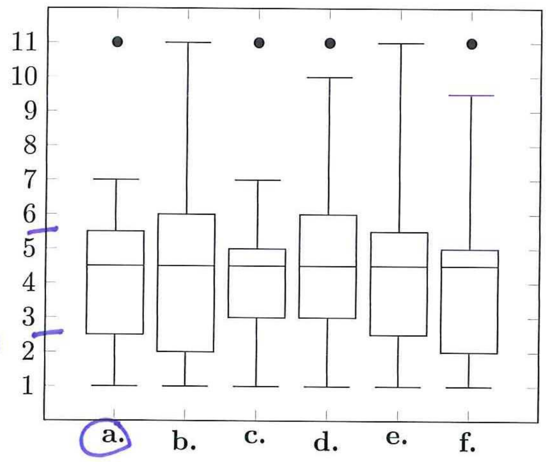
Welke van onderstaande boxplots is de boxplot behorende bij deze dataset? $1.5 \times IQR = 4.5.$

$4.5 + 5.5 = 10$,

dus a.

Q_3

Q_1



$MSE(T_3) = Var(T_3) + (E[T_3] - \theta)^2$. $E[T_3] = 2E[T_1] - E[T_2] = 2\theta - \theta = \theta$, dus

13. Laat T_1 en T_2 twee onafhankelijke zuivere schatters voor θ zijn. Er geldt: $\text{Var}(T_1) = 1$ en $\text{Var}(T_2) = 2$. Bekijk de nieuwe schatter $T_3 = 2T_1 - T_2$.

$MSE(T_3) = Var(T_3)$
$= 4Var(T_1) + Var(T_2)$
$= 4 \cdot 1 + 2 = 6.$

Wat is de gemiddelde kwadratische fout (*mean squared error*) van T_3?

a. 2 **b.** 0 **c.** $4 + \theta^2$ **d.** 6 **e.** 4 **f.** θ^2

14. Beschouw de volgende twee beweringen:

Niet waar: $z_\alpha \downarrow$ als $\alpha \uparrow$

(I) Een $100(1-\alpha)\%$ betrouwbaarheidsinterval (*confidence interval*) wordt groter als de onbetrouwbaarheid α groter wordt (en alle overige factoren gelijk blijven)

(II) De p-waarde is de kans dat nulhypothese juist is. Niet waar: Zie Hfst. 25

Welke van deze beweringen is/zijn waar?

a. Beide **b.** Geen **c.** Alleen (I) **d.** Alleen (II)

15. Laat X en Y twee Bernoulli verdeelde discrete stochasten zijn, en laat verder gegeven zijn dat

$$P(X = 0, Y = 0) = P(X = 0, Y = 1) = P(X = 1, Y = 0) = \frac{1}{3}$$

.

		X	
		0	1
Y	0	1/3	1/3
	1	1/3	0.

Welke van de volgende uitspraken is juist?

a. X en Y zijn ongecorreleerd (*uncorrelated*)en afhankelijk (*dependent*).
b. X en Y zijn ongecorreleerd en onafhankelijk.
c. X en Y zijn gecorreleerd en onafhankelijk.
d. X en Y zijn positief gecorreleerd en afhankelijk.
e. X en Y zijn negatief gecorreleerd en afhankelijk.
f. Er zijn niet genoeg gegevens om deze opgave op te kunnen oplossen.

$E[XY] = 0 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot 0 = 0$

$E[X] = 0 \cdot \frac{2}{3} + 1 \cdot \frac{1}{3} = \frac{1}{3}$.

Lever de vragen ook weer in!

Dus $Cov(X,Y) =$
$= E[XY] - E[X]E[Y]$
$= 0 - \frac{1}{3} \cdot \frac{1}{3} < 0.$
dus ook afhankelijk.

Tentamen Kansrekening en Statistiek
CTB2200 19 januari 2016 13.30 - 16.30 uur

Bij dit examen is het gebruik van een (evt. grafische) rekenmachine toegestaan. Tevens krijgt u een formuleblad uitgereikt—na afloop inleveren alstublieft.

Meerkeuzevragen

Toelichting:
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 èn aan te strepen.

---

1. Een kast heeft drie laden. In de eerste la liggen twee gouden munten, in de tweede twee zilveren en in de derde één gouden en één zilveren. Blindelings wordt een la gekozen en hieruit wordt willekeurig een munt genomen: deze blijkt van goud te zijn. Hoe groot is de kans dat ook de andere munt in die la van goud is?

a. $\frac{1}{12}$ **b.** $\frac{1}{6}$ **c.** $\frac{1}{3}$ **d.** $\frac{1}{2}$ **e.** $\frac{2}{3}$ **f.** $\frac{5}{6}$

2. Een automobilist die een aanrijding veroorzaakt, moet zich onderwerpen aan een bloedproef. De ervaring heeft geleerd dat, als zo iemand "onder invloed" verkeert, er 75% kans is dat het resultaat van de bloedproef positief is; is hij niet onder invloed dan is er maar 2% kans dat de bloedproef toch positief uitvalt. Aangenomen mag worden dat 5% van de automobilisten die een aanrijding veroorzaken onder invloed is. Hoe groot is de kans dat iemand die een aanrijding veroorzaakt onder invloed is als de bloedproef positief uitvalt?

a. 33,6% **b.** 66,4% **c.** 100% **d.** 44,1% **e.** 43,4% **f.** 56,6%

3. In 1994 was de dienstplicht nog niet afgeschaft. Bij de verplichte keuring werd 35% van de jongens afgekeurd.
In een studentenflat woonden 30 jongens. Hoe groot is de kans dat meer dan de helft van hen werd goedgekeurd?

a. $\sum_{k=0}^{14} \binom{30}{k} (0.35)^k (0.65)^{30-k}$
b. $\sum_{k=0}^{15} \binom{30}{k} (0.35)^k (0.65)^{30-k}$
c. $\sum_{k=0}^{16} \binom{30}{k} (0.35)^k (0.65)^{30-k}$
d. $\sum_{k=1}^{14} \binom{30}{k} (0.35)^k (0.65)^{30-k}$
e. $\sum_{k=1}^{15} \binom{30}{k} (0.35)^k (0.65)^{30-k}$
f. $\sum_{k=1}^{16} \binom{30}{k} (0.35)^k (0.65)^{30-k}$

4. Stel F is een discrete stochast die maar 2 waarden kan aannemen, namelijk 0 en 4. Gegeven is dat $\mathrm{E}[F] = 1$. Gevraagd $\mathrm{E}[F^3]$.

a. 0 **b.** 1 **c.** 2 **d.** 4 **e.** 12 **f.** 16

5. Gegeven is een Geometrisch verdeelde stochast G met parameter $\frac{1}{8}$. Bereken $\mathrm{Var}\left(-\frac{1}{2}G - \frac{1}{2}\right)$.

a. 8 **b.** 4 **c.** 14 **d.** 7 **e.** $13\frac{1}{2}$ **f.** $6\frac{1}{2}$

6. Stel X is een continue stochast met dichtheid $f_X(x) = \begin{cases} \frac{1}{4}xe^{-\frac{1}{2}x}, & \text{als } x > 0 \\ 0, & \text{als } x \leq 0 \end{cases}$

Bereken $\mathrm{E}\left[\frac{1}{X}\right]$.

a. $e^{\frac{1}{2}}$ **b.** 1 **c.** $\frac{1}{2}$ **d.** $\frac{1}{4}$ **e.** 0 **f.** ∞

7. X en Y zijn twee discrete stochasten met een gezamenlijke kansmassafunctie volgens de volgende tabel:

		a		
		0	1	2
	0	2/45	9/45	4/45
b	1	7/45	5/45	3/45
	2	6/45	1/45	8/45

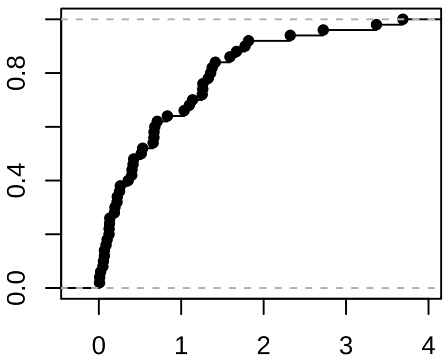
Bereken $\mathrm{Cov}(X, Y)$.

a. -2 **b.** -1 **c.** 0 **d.** 0.5 **e.** 1 **f.** 2

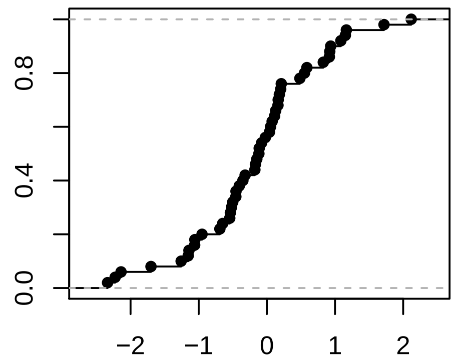
8. De politie controleert op snelheid op de A13 door op drie punten de gereden snelheid te meten. Als snelheid neemt men het gemiddelde $\bar{X}_3$ van deze 3 gemeten snelheden. De hypothese die de politie test is $H_0 : \mu = 100$ tegen $H_1 : \mu > 100$. Je mag aannemen dat de verdeling van elke afzonderlijk gemeten snelheid een $N(\mu, 4)$ verdeling is.
De beslissingsregel om een automobilist te bekeuren is: verwerp H_0 zodra $\bar{X}_3 \geq 101.7$. Bereken de kans op een type II fout als een auto met gemiddelde snelheid $\bar{X}_3 = 100.5$ passeert.

a. 88.8% **b.** 85.1% **c.** 73.6 **d.** 27.4% **e.** 14.9% **f.** 11.2%

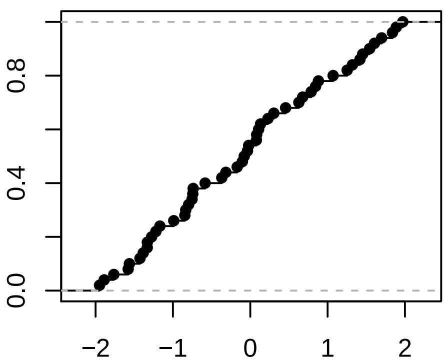
9. Van vier datasets ter grootte 50 is de empirische verdelingsfunctie weergegeven. De datasets zijn steekproeven uit vier verschillende verdelingen: een $Exp(1)$, een $Exp(1/2)$, een $N(0,1)$ en een $U(-2,2)$ verdeling. Welke dataset komt uit de $Exp(1)$ en welke dataset komt uit de $U(-2,2)$ verdeling?



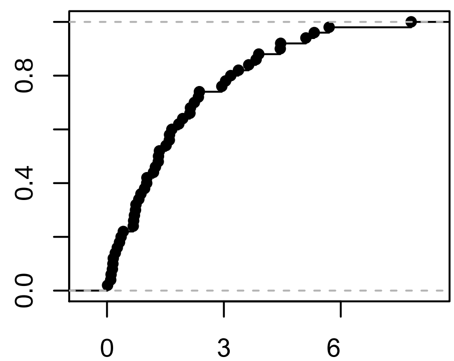
I



III



II



IV

a. $Exp(1)$: I; $U(-2,2)$: III
b. $Exp(1)$: I; $U(-2,2)$: II
c. $Exp(1)$: III; $U(-2,2)$: I
d. $Exp(1)$: IV; $U(-2,2)$: III
e. $Exp(1)$: IV; $U(-2,2)$: II
f. $Exp(1)$: II; $U(-2,2)$: IV

10. Men vult machinaal frisdrankflessen met een hoeveelheid frisdrank, die, gemeten in milliliters, $N(\mu, \sigma^2)$ verdeeld is met onbekende σ. Bij een steekproef van 25 flessen werd een gemiddelde inhoud van 748 ml gevonden. De waarde van de steekproefstandaardafwijking $s_{25} = 2,75$. Geef het 95%-betrouwbaarheidsinterval voor μ.

a. (744,472 , 751,528)
b. (746,824 , 749,135)
c. (746,865 , 749,135)
d. (747,294 , 751,528)
e. (747,765 , 751,528)
f. (747,803 , 748,197)

11. De landelijke scores op een rekentoets zijn normaal verdeeld met $\mu = 60$. Op een bepaalde school twijfelt men eraan of de landelijke norm met betrekking tot het gemiddelde wel wordt gehaald. Om dit na te gaan wordt een steekproef van 25 leerlingen getrokken om de hypothese te toetsen. In de steekproef wordt een gemiddelde score S van 58 gehaald en de steekproefstandaardafwijking is 10 punten.
Stel voor deze toets de hypotheses op en bereken de waarde t van de toetsingsgrootheid T.

a. $H_0 : \mu = 60$, $H_1 : \mu < 60$ en $t = 1$
b. $H_0 : \mu = 60$, $H_1 : \mu \neq 60$ en $t = 1$
c. $H_0 : \mu = 60$, $H_1 : \mu > 60$ en $t = 1$
d. $H_0 : \mu = 60$, $H_1 : \mu < 60$ en $t = -1$
e. $H_0 : \mu = 60$, $H_1 : \mu \neq 60$ en $t = -1$
f. $H_0 : \mu = 60$, $H_1 : \mu > 60$ en $t = -1$

12. Twee-en-veertig proefpersonen doen mee aan een experiment ter beperking van gewichtstoename. Zij worden aselect verdeeld over een experimentele en een controle groep. De controle groep krijgt een placebo (een pilletje zonder werking) terwijl de aan de experimentele groep toegewezen personen een pilletje toegediend krijgen met eetlustremmende werking. Beide groepen bestaan uit 21 personen en krijgen hetzelfde eten voorgezet en mogen zoveel eten als ze willen.
In de experimentele groep is de gemiddelde gewichtstoename na een maand $\bar{X} = 0.5$ met $S_X = 0.1$. In de controle groep geldt $\bar{Y} = 2.5$ met $S_Y = 0.1$ (meting in kilogrammen). Het geschatte effect van het pilletje is $\bar{X} - \bar{Y} = -2$ kg. Toets de hypothese dat het slikken van het eetlustremmend pilletje de gewichtstoename vermindert op significantieniveau $\alpha = 0.05$:
$H_0 : \mu_X = \mu_Y$ tegen $H_1 : \mu_X < \mu_Y$ Welke van de volgende uitspraken is de enige juiste (gebaseerd op deze data)?

a. De kritieke waarde voor deze toets is 2.042; verwerp H_0 niet
b. De kritieke waarde voor deze toets is 2.042; verwerp H_0 wel
c. De kritieke waarde voor deze toets is 1.697; verwerp H_0 niet
d. De kritieke waarde voor deze toets is 1.697; verwerp H_0 wel
e. De kritieke waarde voor deze toets is 1.684; verwerp H_0 niet
f. De kritieke waarde voor deze toets is 1.684; verwerp H_0 wel

Open vragen

1. Stel $X_1, X_2, \ldots, X_{100}$ zijn exponentieel verdeeld met parameter $\lambda = 2$.
Benader $\mathrm{P}(X_1 + X_2 + \cdots + X_{100} \geq 60)$ met behulp van de centrale limietstelling.

2. Gemiddeld worden er in de maand september 1472 baby's geboren in Cleveland. In januari 1977 was de stad geparalyseerd door een ernstige sneeuwstorm. Negen maanden later, in september 1977 was het aantal geboorten 1718. Modelleer het aantal geboortes in september met een Poisson verdeelde stochast T, met parameter μ, het gemiddeld aantal geboortes. Toets de hypothese dat de toename van het aantal geboortes toeval was met behulp van het getelde aantal van 1718.

- **a.** Formuleer de te toetsen hypotheses H_0 en H_1.
- **b.** De verdeling van T mag worden benaderd met een $N(\mu, \mu)$-verdeling. Gebruik dit feit om het kritieke gebied van de toets te bepalen op significantie niveau $\alpha = 0.05$.

3. De Pareto-verdeling is een kansverdeling waarbij de cumulatieve verdelingsfunctie wordt gegeven door $F(x) = 1 - x^{-\alpha}$ voor $x \geq 1$ en $F(x) = 0$ als $x < 1$. Hier is $\alpha > 0$ een (onbekende) parameter. Voor een dataset $x_1, x_2, \ldots, x_8$ veronderstellen we een Pareto verdeling.

- **a.** Geef de likelihoodfunctie voor α.
- **b.** Bepaal de maximum likelihood schatter voor α.

Antwoorden multiple choice:

1 **e.**

2 **b.**

3 **a.**

4 **f.**

5 **c.**

6 **c.**

7 **c.**

8 **b.**

9 **b.**

10 **c.**

11 **d.**

12 **f.**

Antwoorden open vragen:

1 0.0228.

2a $H_0 : \mu = 1472$ tegen $H_1 : \mu > 1472$.

2b $K = \{1536, 1537, \cdots\}$

3a $L(\alpha) = \frac{\alpha^8}{x_1^{\alpha+1} x_2^{\alpha+1} \cdots x_8^{\alpha+1}}$

3b $\alpha = \frac{8}{ln(x_1)+\cdots+ln(x_8)}$

Uitwerkingen multiple choice:

1 Laat G de gebeurtenis zijn dat het een gouden munt is. Laat L de gebeurtenis zijn dat er twee gouden munten in de la zijn. Dan is $\mathrm{P}(L) = 1/3$ en $\mathrm{P}(G) = 1/2$. De gevraagde kans is:

$$\mathrm{P}(L|G) = \frac{\mathrm{P}(L \cap G)}{\mathrm{P}(G)} = \frac{1/3}{1/2} = 2/3$$

2 Laat I de gebeurtenis zijn dat de automobilist onder invloed is, T de gebeurtenis dat de bloedtest positief is, dan is gegeven: $\mathrm{P}(T|I) = 0.75, \mathrm{P}(I) = 0.05, \mathrm{P}(T|I^c) = 0.02$. De gevraagde kans is

$$\begin{aligned} \mathrm{P}(I|T) &= \frac{\mathrm{P}(T|I)\,\mathrm{P}(I)}{\mathrm{P}(T|I)\,\mathrm{P}(I) + \mathrm{P}(T|I^c)\,\mathrm{P}(I^c)} \\ &= \frac{0.75 \cdot 0.05}{0.75 \cdot 0.05 + 0.02 \cdot 0.95} \\ &= 0.664 \end{aligned}$$

3 Het aantal afgekeurde jongens X is $Bin(30, 0.35)$ verdeeld. Gevraagd wordt de kans dat meer dan de helft (16 of meer) wordt goedgekeurd, dat betekent dat er dus 14 of minder studenten worden afgekeurd.

$$\mathrm{P}(X \leq 14) = \sum_{k=0}^{14} \binom{30}{k} (0.35)^k (0.65)^{30-k}.$$

4 Aangezien $\mathrm{E}[F] = 0 \cdot \mathrm{P}(F = 0) + 4 \cdot \mathrm{P}(F = 4) = 1$, vinden we dat $\mathrm{P}(F = 4) = \frac{1}{4}$. Dus de kansmassatabel van F is

f	0	4
$\mathrm{P}(F = f)$	3/4	1/4

Dus $\mathrm{E}[F^3] = 0^3 \cdot \frac{3}{4} + 4^3 \cdot \frac{1}{4} = 16$

5 Volgens het formuleblad is de variantie van G gelijk aan $\frac{1-\frac{1}{8}}{(\frac{1}{8})^2} = 56$ dus

$$\mathrm{Var}\left(-\frac{1}{2}G - \frac{1}{2}\right) = \frac{1}{4}\mathrm{Var}(G) = \frac{56}{4} = 14.$$

6 Om de verwachting van een continue stochast uit te rekenen moeten we integreren met betrekking tot de dichtheid, deze is al gegeven dus:

$$\begin{aligned} \mathrm{E}\left[\frac{1}{X}\right] &= \int_0^\infty \frac{1}{x} f_X(x)\,dx \\ &= \int_0^\infty \frac{1}{x}\frac{1}{4} x e^{-\frac{1}{2}x}\,dx \\ &= \int_0^\infty \frac{1}{4} e^{-\frac{1}{2}x}\,dx \\ &= \left[-\frac{1}{2}e^{-\frac{1}{2}x}\right]_0^\infty \\ &= \frac{1}{2} \end{aligned}$$

7 De tabel compleet met de marginale verdelingen:

		a			
		0	1	2	$\mathrm{P}(Y=b)$
	0	2/45	9/45	4/45	15/45
b	1	7/45	5/45	3/45	15/45
	2	6/45	1/45	8/45	15/45
$\mathrm{P}(X=a)$		15/45	15/45	15/45	1

$$\mathrm{E}[X] = 2 \cdot 15/45 + 1 \cdot 15/45 = 1 = \mathrm{E}[Y]$$

en

$$\mathrm{E}[XY] = 0 \cdot 28/45 + 1 \cdot 5/45 + 2 \cdot 4/45 + 4 \cdot 8/45 = 1$$

dus

$$\mathrm{Cov}(X,Y) = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] = 0$$

8 Bereken $\mathrm{P}\left(\bar{X}_3 < 101.7 | \mu = 100.5\right)$. De stochast $\bar{X}_3 \sim N(100.5, \frac{4}{3})$, dus

$$\frac{\bar{X}_3 - 100.5}{2/\sqrt{3}} = Z \sim N(0,1).$$

De gevraagde kans is dus gelijk aan $\mathrm{P}(Z < 1.04) = 1 - 0.1492 = 0.8508$.

9 Uit de vorm van de grafieken is het duidelijk dat I en IV de twee exponentiele verdelingen zijn. Aangezien $Exp(1)$ een kleinere verwachting heeft dan de $Exp(1/2)$, moet IV bij $Exp(1/2)$ horen. (Bovendien is de mediaan af te lezen.) Verder kan III niet bij $U(-2,2)$ horen, aangezien er waarnemingen zijn kleiner dan -2. Dus II moet bij $U(-2,2)$ horen.

10 Uitrekenen met de formule voor het betrouwbaarheidsinterval $(\bar{x}_n \pm t_{24,0.025} \cdot \frac{s_{25}}{5})$ geeft **c.**

11 De hypothese die het vermoeden beschrijft is dat de score lager is dan 60, dus $H_1 : S < 60$.

$$T = \frac{S - \mu_0}{s_n/\sqrt{n}} = \frac{58 - 60}{10/\sqrt{25}} = -1$$

.

12 De gepoolede variantie is

$$S_p^2 = \frac{20 \cdot (0.1)^2 + 20 \cdot (0.1)^2}{40}\left(\frac{2}{21}\right) = 0.00095$$

Dus

$$T_p = \frac{\mu_X - \mu_Y}{S_p} = \frac{-2}{\sqrt{0.00095}} = -64.81$$

T_p volgt een $t(40)$ verdeling. We toetsen eenzijdig, dus de kritieke waarde is $t_{40,0.05} = -1.684$. De gevonden waarde van T_p is kleiner dan deze kritieke waarde, dus we moeten H_0 verwerpen.

Uitwerkingen open vragen:

1 Volgens de centrale limietstelling heeft

$$Z = \frac{\bar{X}_{100} - \mu}{\sigma/\sqrt{100}}$$

bij benadering een $N(0,1)$ verdeling. Nu is $\mu = \mathrm{E}[X_1] = 1/\lambda = 1/2$ en $\sigma = \sqrt{\mathrm{Var}(X_1)} = 1/\lambda = 1/2$ Dus,

$$\begin{aligned} \mathrm{P}(X_1 + X_2 + \cdots + X_{100} \geq 60) &= \mathrm{P}\left(\bar{X}_{100} \geq 0.6\right) \\ &= \mathrm{P}\left(\frac{\bar{X}_{100} - \mu}{\sigma\sqrt{n}} \geq \frac{0.6 - 1/2}{1/2/\sqrt{100}}\right) \\ &= \mathrm{P}(Z \geq 2) \approx 0.0228. \end{aligned}$$

2a $H_0 : \mu = 1472$ tegen $H_1 : \mu > 1472$.

2b Omdat de verdeling van T te benaderen is met de normale $N(1472, 1472)$ geldt dat

$$\frac{T - 1472}{\sqrt{1472}}$$

standaard normaal verdeeld is. Om het kritieke gebied te bepalen zoek je c waarvoor $\mathrm{P}(T \geq c) = 0.05$. Dus

$$\mathrm{P}(T \geq c) = \mathrm{P}\left(\frac{T - 1472}{\sqrt{1472}} \geq \frac{c - 1472}{\sqrt{1472}}\right) \approx \mathrm{P}\left(Z \geq \frac{c - 1472}{\sqrt{1472}}\right).$$

Je vindt c dus door op te lossen:

$$\frac{c - 1472}{\sqrt{1472}} = 1.645$$

Dus $c = 1535.1$. Aangezien c alleen maar gehele waarden kan aannemen is het kritieke gebied dus de verzameling $K = \{1536, 1537, \cdots\}$.

3a De kansdichtheid wordt gegeven door $f(x) = \alpha/x^{\alpha+1}$. Gebruik verder de definitie van de likelihoodfunctie. $L(\alpha) = \frac{\alpha^8}{x_1^{\alpha+1} x_2^{\alpha+1} \cdots x_8^{\alpha+1}}$

3b De loglikelihood is

$$l(\alpha) = 8ln(\alpha) - (\alpha + 1)\left(ln(x_1) + \cdots ln(x_8)\right)$$

De afgeleide hiervan is

$$l'(\alpha) = \frac{8}{\alpha} - \left(ln(x_1) + \cdots + ln(x_8)\right)$$

Dit is gelijk aan nul als $\alpha = \frac{8}{ln(x_1) + \cdots ln(x_8)}$

Hertentamen Kansrekening en Statistiek
CTB2200
13 april 2015

Dit tentamen bestaat uit 15 meerkeuzevragen en 3 open vragen. Graag één vel per open vraag inleveren.
Bij dit examen is het gebruik van een (evt. grafische) rekenmachine toegestaan. Tevens krijgt u een formuleblad uitgereikt.

Meerkeuzevragen

Toelichting:
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 èn aan te strepen.

---

1. Van de studenten aan de TU Delft is 97% in het bezit van een smartphone. Er staat een groepje van 8 studenten bij elkaar. Hoe groot is de kans dat hoogstens 2 van deze studenten geen smartphone heeft?

a. 0.1% **b.** 99.9% **c.** 97.0% **d.** 3.0% **e.** 0.0% **f.** 25%

2. Gegeven zijn twee gebeurtenissen A en B waarvan gegeven is $\mathrm{P}(A) = 3/4$, $\mathrm{P}(B) = 1/4$ en $\mathrm{P}(A \cup B) = 1/2$. Bereken $\mathrm{P}(A^c|B^c)$.

a. $\frac{2}{3}$ **b.** $\frac{1}{3}$ **c.** $\frac{1}{4}$ **d.** $\frac{3}{4}$ **e.** $\frac{1}{8}$ **f.** $\frac{3}{8}$

3. Bij een inbraak gaat een allarminstallatie af met kans 0.9. Als er geen inbraak plaatsvindt is de kans op een vals alarm per nacht 0.02. De kans op een inbraak is 0.001 voor een willekeurige nacht. Hoe groot is de kans op een inbraak als op zekere nacht het alarm afgaat?

a. 0.955 **b.** 0.045 **c.** 0.980 **d.** 0.020 **e.** 0.957 **f.** 0.043

4. Van een discrete stochast X is de kansmassafunctie gegeven in onderstaande tabel. Bereken $\mathrm{E}\left[X^3\right]$.

a	-2	0	2
$p_X(a)$	$1/3$	$1/6$	$1/2$

a. $\frac{1}{27}$ **b.** $\frac{4}{3}$ **c.** $\frac{20}{3}$ **d.** $\frac{1}{9}$ **e.** $\frac{5}{3}$ **f.** $\frac{125}{27}$

5. Van een continue stochast X is de verdelingsfunctie $F_X(x)$ gegeven door

- $F_X(x) = 0$, als $x \leq 0$
- $F_X(x) = \sqrt{x}$, als $0 \leq x \leq 1$
- $F_X(x) = 1$,als $x \geq 1$

Bereken $\mathrm{E}\left[X^2\right]$

a. $\frac{1}{3}$ **b.** $\frac{2}{3}$ **c.** $\frac{1}{2}$ **d.** 1 **e.** $\frac{1}{5}$ **f.** $\frac{2}{5}$

6. Stel $X \sim$ N(-2,4). Laat $Y = -\frac{1}{2}X - 3$.
Bepaal $\mathrm{E}[Y]$ en $\mathrm{Var}(Y)$.

a. $\mathrm{E}[Y] = -2$ en $\mathrm{Var}(Y) = 1$
b. $\mathrm{E}[Y] = -2$ en $\mathrm{Var}(Y) = 2$
c. $\mathrm{E}[Y] = -2$ en $\mathrm{Var}(Y) = -2$
d. $\mathrm{E}[Y] = 1$ en $\mathrm{Var}(Y) = 1$
e. $\mathrm{E}[Y] = 1$ en $\mathrm{Var}(Y) = 2$
f. $\mathrm{E}[Y] = 1$ en $\mathrm{Var}(Y) = -2$

7. Laat X en Y twee stochasten zijn met onderstaande gezamenlijke kansmassatabel. Hierbij kan X dus de waarden $0, 1$ en 2 aannemen.
Wat is $\mathrm{P}(X = 1|Y = 1)$?

	a		
b	0	1	2
-1	1/4	1/4	1/4
1	0	1/8	1/8

a. $\frac{1}{8}$ **b.** $\frac{3}{8}$ **c.** $\frac{1}{2}$ **d.** $\frac{3}{4}$ **e.** $\frac{1}{4}$ **f.** 1

8. Stel je gooit k keer met een zuivere dobbelsteen. Je telt het aantal keren dat je 6 gooit.
Bepaal de kans op hoogstens één keer 6 in k worpen.

a. $1 - \left(k(\frac{1}{6})^k + (\frac{5}{6})^k\right)$
b. $1 - k(\frac{1}{6})^k$
c. $1 - (\frac{1}{6})^k$
d. $(\frac{5+k}{6})\left(\frac{5}{6}\right)^{k-1}$
e. $(k+1)(\frac{5}{6})^k$
f. $k(\frac{1}{6})^k\frac{5}{6}$

9. Je hebt 12 flessen rum besteld, die elk 0,75 l horen te bevatten. De gemiddelde inhoud van de 12 flessen is 0,73 l. Neem aan dat de 12 inhouden beschouwd mogen worden als een realisatie van een Normale verdeling met onbekende parameters. Construeer een 98% betrouwbaarheidsinterval voor de verwachte inhoud van een fles rum. De steekproefvariantie is 0.0009.

a. (0.7145 , 0.7456)
b. (0.7109 , 0.7491)
c. (0.7065 , 0.7535)
d. (0.7065 , 0.7456)
e. (0.7109 ,0.7535)
f. (0.7145 , 0.7491)

10. Stel $X \sim N(3, 9)$ en $Y \sim N(-1, 16)$, en X en Y zijn onafhankelijk. Bereken $\mathrm{P}(X + Y < 1)$.

a. 0.4129 **b.** 0.4207 **c.** 0.4335 **d.** 0.5793 **e.** 0.5160 **f.** 0.4840

11. Laat $X_1, X_2, \cdots, X_{100}$ onafhankelijk, gelijkverdeelde stochasten zijn uit een Pareto-verdeling met parameter 3. Bereken met behulp van de Centrale Limietstelling $\mathrm{P}(X_1 + X_2 + \cdots + X_{100} \leq 175)$.

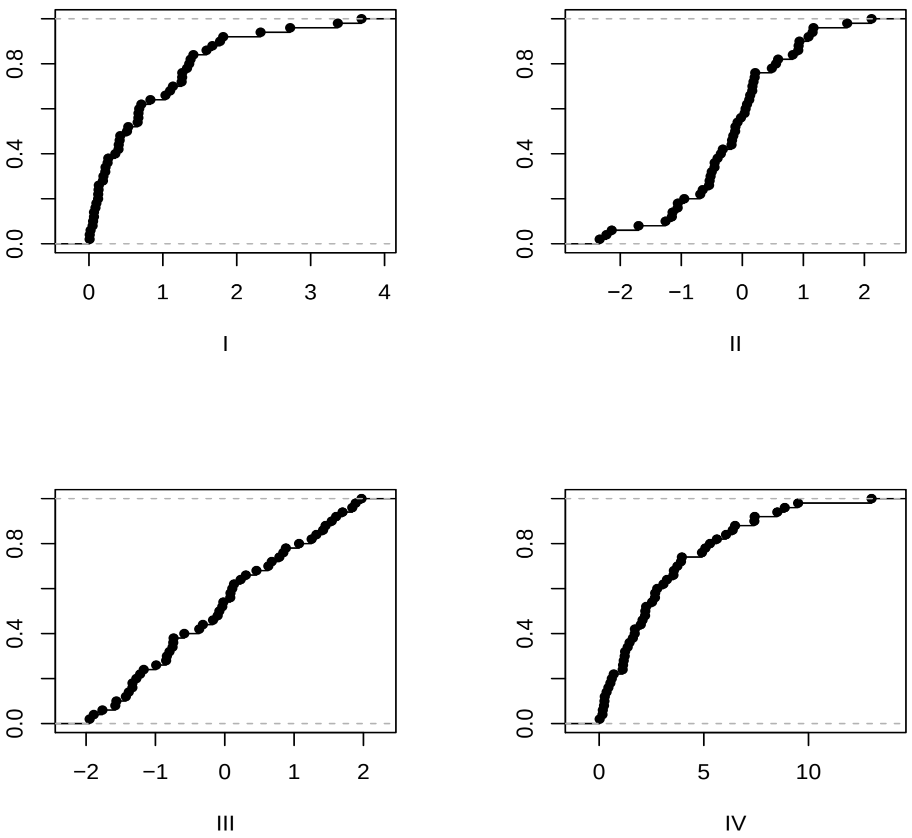
a. 0.9974 **b.** 0.0026 **c.** 0.9996 **d.** 0.0004 **e.** 0.0019 **f.** 0.9981

12. We werpen met een munt tot voor de tweede keer KOP optreedt. Laat X het aantal keren zijn dat we moeten gooien om de tweede keer KOP te krijgen. Stel de kans op KOP bij dit munststuk is $\frac{1}{3}$. Neem aan dat de uitkomst van elke worp onafhankelijk is van de eerdere uitkomsten.

Bereken $\mathrm{P}(X = 3)$.

a. $\frac{1}{3}$ **b.** $\frac{2}{3}$ **c.** $\frac{4}{9}$ **d.** $\frac{2}{9}$ **e.** $\frac{4}{27}$ **f.** $\frac{2}{27}$

13. Van vier datasets ter grootte 50 is de empirische verdelingsfunctie weergegeven. De datasets zijn steekproeven uit vier verschillende verdelingen: een $Exp(1)$, een $Exp(1/3)$, een $N(0,1)$ en een $U(-2,2)$ verdeling. Welke dataset komt uit de $Exp(1/3)$ en welke dataset komt uit de $U(-2,2)$ verdeling?



a. $Exp(1/3)$: I; $U(-2,2)$: III
b. $Exp(1/3)$: IV; $U(-2,2)$: III
c. $Exp(1/3)$: III; $U(-2,2)$: II
d. $Exp(1/3)$: III; $U(-2,2)$: I
e. $Exp(1/3)$: IV; $U(-2,2)$: II
f. $Exp(1/3)$: II; $U(-2,2)$: IV

14. Gegeven is een getal t, de realisatie van een stochastische variabele T met een Normale verdeling met parameters μ en 9. Men toetst met behulp van de toetsingsgrootheid T:
$H_0 : \mu = 0$ tegen $H_1 : \mu \neq 0$.
Men besluit om H_0 te verwerpen als de uitkomst van de toetsingsgrootheid $|T| > 1.5$. Bereken de kans op een type I fout.

a. 0.6170 **b.** 0.3830 **c.** 0.8831 **d.** 0.1169 **e.** 0.4801 **f.** 0.3632

15. Twintig proefpersonen doen mee aan een experiment ter bevordering van examenresultaten. Zij worden aselect verdeeld over een experimentele en een controle groep. De controle groep krijgt geen behandeling terwijl de aan de experimentele groep toegewezen personen voor het tentamen een week lang meditatie-oefeningen doen. Beide groepen bestaan uit 10 personen. In de experimentele groep is het gemiddelde examenresultaat $\bar{X} = 6.8$ met $S_X = 0.25$. In de controle groep geldt $\bar{Y} = 6.5$ met $S_Y = 0.25$. Het geschatte effect van het doen van meditatie-oefeningen is $\bar{X} - \bar{Y} = 0.3$ punt. Toets de hypothese dat het doen van meditatie-oefeningen het examenresultaat verhoogt op significantieniveau $\alpha = 0.05$:
$H_0 : \mu_X = \mu_Y$ tegen $H_1 : \mu_X > \mu_Y$.
Welke van de volgende uitspraken is de enige juiste (gebaseerd op deze data)?

a. De kritieke waarde voor deze toets is 2.042; verwerp H_0 niet
b. De kritieke waarde voor deze toets is 2.042; verwerp H_0 wel
c. De kritieke waarde voor deze toets is 1.697; verwerp H_0 niet
d. De kritieke waarde voor deze toets is 1.697; verwerp H_0 wel
e. De kritieke waarde voor deze toets is 1.734; verwerp H_0 niet
f. De kritieke waarde voor deze toets is 1.734; verwerp H_0 wel

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. Graag één vel per open vraag inleveren.

1. Bekijk de data punten

$$(0,0),\ (1,0),\ (2,1),\ (2,2).$$

We gaan op zoek naar de rechte lijn die het beste past bij de data in de zin van de kleinste kwadratenoplossing. Het model is

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

De matrixformulering voor dit regressiemodel is

$$\mathbf{y} = X\beta + \epsilon,$$

waarbij $\epsilon_i \sim N(0, \sigma^2)$.

a. Geef de designmatrix X en de observatievector $\mathbf{y}$.

b. Geef de regressielijn door deze datapunten, door kleinste kwadratenschattingen voor het startgetal en de richtingscoëfficiënt te berekenen met behulp van de matrixformulering.

2. De landelijke scores op een rekentoets zijn normaal verdeeld met $\mu = 60$. Op een bepaalde school twijfelt men eraan of de landelijke norm met betrekking tot het gemiddelde wel wordt gehaald. Om dit na te gaan wordt een steekproef van 25 leerlingen getrokken om de hypothese te toetsen. In de steekproef wordt een gemiddelde score S van 58 gehaald en de steekproefstandaardafwijking is 10 punten.
Toets het vermoeden op significantieniveau $\alpha = 0.05$:

a. Stel de relevante hypotheses op.

b. Bereken de waarde van de toetsingsgrootheid.

c. Geef de conclusie van de toets.

3. De bevolkingsdichtheid voor een vierkant gebied van 3 km bij 3 km kan worden beschreven met de functie

$$f_{X,Y}(x, y) = \frac{x + y}{27}, \qquad \text{voor } 0 < x, y < 3.$$

a. Bereken de kans dat een willekeurig persoon die je daar tegenkomt uit het gebied komt met $x < 2$ en $y < 2$.

b. Bereken de marginale dichtheid van X.

Antwoorden multiple choice:

1 **b.**

2 **a.**

3 **f.**

4 **b.**

5 **e.**

6 **a.**

7 **c.**

8 **d.**

9 **c.**

10 **b.**

11 **f.**

12 **e.**

13 **b.**

14 **a.**

15 **f.**

Antwoorden open vragen:

1a De vector $\mathbf{y} = [0, 0, 1, 2]^T$ en de matrix X is:

$$\mathbf{X} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 2 \end{pmatrix}.$$

1b $y = -\frac{3}{11} + \frac{9}{11}x$.

2a $H_0 : S = 60$, $H_1 : S < 60$.

2b $t = -1$

2c H_0 niet verwerpen.

3a $\frac{8}{27}$

3b $\frac{1}{9}x + \frac{1}{6}$.

Uitwerkingen multiple choice:

1 Laat X het aantal studenten in het groepje zijn emph zonder smartphone, dan is $X \sim Bin(8, 0.03)$, en de gevraagde kans is

$$\begin{aligned}
\mathrm{P}(X \leq 2) &= \mathrm{P}(X=0) + \mathrm{P}(X=1) + \mathrm{P}(X=2) \\
&= \binom{8}{0}(0.03)^0(0.97)^8 + \binom{8}{1}(0.03)^1(0.97)^7 + \binom{8}{2}(0.03)^2(0.97)^6 \\
&= 0.999
\end{aligned}$$

2

$$\begin{aligned}
\mathrm{P}(A^c|B^c) &= \frac{\mathrm{P}(A^c \cap B^c)}{\mathrm{P}(B^c)} \\
&= \frac{\mathrm{P}((A \cup B)^c)}{1 - \mathrm{P}(B)} \\
&= \frac{1 - \mathrm{P}(A \cup B)}{1 - \mathrm{P}(B)} \\
&= \frac{2}{3}
\end{aligned}$$

3 Laat A de gebeurtenis zijn dat het alarm afgaat, I de gebeurtenis dat er een inbraak is, dan is gegeven: $\mathrm{P}(A|I) = 0.9, \mathrm{P}(I) = 0.001, \mathrm{P}(A|I^c) = 0.02$. De gevraagde kans is

$$\begin{aligned}
\mathrm{P}(I|A) &= \frac{\mathrm{P}(A|I)\,\mathrm{P}(I)}{\mathrm{P}(A|I)\,\mathrm{P}(I) + \mathrm{P}(A|I^c)\,\mathrm{P}(I^c)} \\
&= \frac{0.9 \cdot 0.001}{0.9 \cdot 0.001 + 0.02 \cdot 0.999} \\
&= 0.043
\end{aligned}$$

4 $\mathrm{E}\left[X^3\right] = (-2)^3 \cdot \frac{1}{3} + 2^3 \cdot \frac{1}{2} = \frac{4}{3}$

5 De formule voor de verwachting is $\mathrm{E}\left[X^2\right] = \int x^2 f_X(x)\,dx$, met f_X de dichtheid van X. Deze is te berekenen als de afgeleide van de verdelingsfunctie, dus $f_X(x) = F'(x) = \frac{1}{2\sqrt{x}}$. Dus

$$\mathrm{E}\left[X^2\right] = \int_0^1 x^2 \frac{1}{2\sqrt{x}}\,dx = \int_0^1 \frac{1}{2}(x)^{3/2}\,dx = \frac{1}{5}$$

6 $X \sim$ N(-2,4) betekent $\mathrm{E}[X] = -2$ en $\mathrm{Var}(X) = 4$. Volgens de rekenregels voor verwachting en variantie krijgen we dus $\mathrm{E}[Y] = -\frac{1}{2}\mathrm{E}[X] - 3 = -2$ en $\mathrm{Var}(Y) = (-\frac{1}{2})^2\mathrm{Var}(X) = 1$.

7

$$\begin{aligned}
\mathrm{P}(X=1|Y=1) &= \frac{\mathrm{P}(X=1, Y=1)}{\mathrm{P}(Y=1)} \\
&= \frac{1/8}{1/8 + 1/8} = \frac{1}{2}
\end{aligned}$$

8 Laat X het aantal keren 6 zijn in k worpen. Dan is X dus binomiaal verdeeld met parameters k en $\frac{1}{6}$. Dus de gevraagde kans is

$$\begin{aligned}
\mathrm{P}(X \leq 1) &= \mathrm{P}(X = 0) + \mathrm{P}(X = 1) \\
&= \binom{k}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^k + \binom{k}{1}\frac{1}{6}\left(\frac{5}{6}\right)^{k-1} = \\
&= \left(\frac{5}{6}\right)^k + \frac{1}{6}k\left(\frac{5}{6}\right)^{k-1} \\
&= \frac{5+k}{6}\left(\frac{5}{6}\right)^{k-1}
\end{aligned}$$

9 Het betrouwbaarheidsinterval wordt gegeven door

$$\left(\bar{x}_n \pm t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}}\right).$$

$\bar{x}_n = 0.73$, $s_n = \sqrt{0.0009} = 0.03$, $n = 12$, $t_{11,0.01} = 2.718$ invullen geeft het antwoord.

10 $X + Y \sim N(2, 25)$, dus

$$\mathrm{P}(X + Y < 1) = \mathrm{P}\left(\frac{X+Y-2}{5} < \frac{1-2}{5}\right) = \mathrm{P}(Z < -0.2) = \mathrm{P}(Z > 0.2) = 0.4207$$

11 Aangezien $X \sim \mathrm{Par}(3)$, geldt $\mathrm{E}[X] = 1.5$ en $\mathrm{Var}(X) = 0.75$. Dus

$$\begin{aligned}
\mathrm{P}(X_1 + X_2 + \cdots + X_{100} \leq 175) &= \mathrm{P}\left(\frac{X_1 + X_2 + \cdots + X_{100}}{100} \leq \frac{175}{100}\right) \\
&= \mathrm{P}\left(\frac{\bar{X}_{100} - \mu}{\sigma/\sqrt{n}} \leq \frac{\frac{175}{100} - 1.5}{\sqrt{0.75}/10}\right) \\
&\approx \mathrm{P}(Z \leq 2.89) = 1 - \mathrm{P}(Z > 2.89) \\
&= 1 - 0.0019 = 0.9981
\end{aligned}$$

12 Als $X = 3$, dan hebben we óf KMK gegooid óf MKK. De kans hierop is $2 \cdot (\frac{1}{3})(\frac{2}{3})^2$.

13 Uit de vorm van de grafieken is het duidelijk dat I en IV de twee exponentiele verdelingen zijn. Aangezien $Exp(1)$ een kleinere verwachting heeft dan de $Exp(1/3)$, moet IV bij $Exp(1/3)$ horen. Verder kan II niet bij $U(-2, 2)$ horen, aangezien er waarnemingen zijn kleiner dan -2. Dus III moet bij $U(-2, 2)$ horen.

14 Kans op een type I fout is

$$\begin{aligned}
\mathrm{P}(H_0 \text{ verwerpen} \,|\, \mu = 0) &= \mathrm{P}(|T| > 1.5 | \mu = 0) \\
&= 2\mathrm{P}(T > 1.5) \\
&= 2\mathrm{P}\left(\frac{T-\mu}{\sigma} > \frac{1.5-0}{3}\right) \\
&= 2\mathrm{P}(Z \geq 0.5) \\
&= 2 \cdot 0.3085 = 0.6170
\end{aligned}$$

15 De gepoolede variantie is

$$S_p^2 = \frac{9 \cdot (0.25)^2 + 9 \cdot (0.25)^2}{18}(\frac{1}{10}) = 0.00625$$

Dus

$$T_p = \frac{\mu_X - \mu_Y}{S_p} = \frac{0.3}{\sqrt{0.00625}} = 3.795$$

T_p volgt een $t(18)$ verdeling. We toetsen eenzijdig, dus de kritieke waarde is $t_{18,0.05} = 1.734$. De gevonden waarde van T_p is groter dan deze kritieke waarde, dus we moeten H_0 verwerpen.

Uitwerkingen open vragen:

1a De vector $\mathbf{y} = [0, 0, 1, 2]^T$ en de matrix X is:

$$\mathbf{X} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 2 \end{pmatrix}.$$

1b Er geldt

$$\mathbf{X^T X} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 5 \\ 5 & 9 \end{pmatrix}$$

$$\mathbf{X^T y} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ 6 \end{pmatrix}.$$

De normaalvergelijkingen zijn

$$\begin{pmatrix} 4 & 5 \\ 5 & 9 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 3 \\ 6 \end{pmatrix}.$$

De oplossing van de normaal vergelijkingen is

$$\begin{aligned} \mathbf{b} = (\mathbf{X^T X})^{-1}\mathbf{X^T y} &= \begin{pmatrix} 4 & 5 \\ 5 & 9 \end{pmatrix}^{-1} \begin{pmatrix} 3 \\ 6 \end{pmatrix} \\ &= \frac{1}{11} \begin{pmatrix} 9 & -5 \\ -5 & 4 \end{pmatrix} \begin{pmatrix} 3 \\ 6 \end{pmatrix} \\ &= \frac{1}{11} \begin{pmatrix} -3 \\ 9 \end{pmatrix} = \begin{pmatrix} -3/11 \\ 9/11 \end{pmatrix} \end{aligned}$$

Dus de lijn is

$$y = -\frac{3}{11} + \frac{9}{11}x.$$

2a $H_0 : S = 60$, $H_1 : S < 60$.

2b

$$T = \frac{S - \mu_0}{s_n/\sqrt{n}} = \frac{58 - 60}{10/\sqrt{25}} = -1$$

2c De kritieke waarde waarmee we de uitkomst van (b) moeten vergelijken is $t_{n-1,\alpha} = -t_{24,0.05} = -1.711$. De gevonden waarde van T ligt dus niet in het kritieke gebied, dus we verwerpen H_0 niet en kunnen dus niet concluderen dat de school niet voldoet aan de norm.

3a De gevraagde kans is

$$\begin{aligned}
\mathrm{P}(X < 2, Y < 2) &= \int_0^2 \int_0^2 \frac{x+y}{27}\, dydx \\
&= \int_0^2 \left[\frac{xy}{27} + \frac{1}{54}y^2\right]_{y=0}^{y=2} dx \\
&= \int_0^2 \frac{2}{27}x + \frac{2}{27}\, dx \\
&= \left[\frac{1}{27}x^2 + \frac{2}{27}x\right]_0^2 \\
&= \frac{8}{27}
\end{aligned}$$

3b

$$\begin{aligned}
f_X(x) &= \int_0^3 f_{X,Y}(x,y)\, dy \\
&= \int_0^3 \frac{x+y}{27}\, dx \\
&= \left[\frac{xy}{27} + \frac{y^2}{54}\right]_0^3 \\
&= \frac{1}{9}x + \frac{1}{6}.
\end{aligned}$$

Tentamen Kansrekening en Statistiek
CTB2200
19 januari 2015

Dit tentamen bestaat uit 15 meerkeuzevragen en 3 open vragen. Bij dit examen is het gebruik van een (evt. grafische) rekenmachine toegestaan. Tevens krijgt u een formuleblad uitgereikt.

Meerkeuzevragen

Toelichting:
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 èn aan te strepen.

---

1. Van de groep studenten die dit tentamen maakt, is bekend dat 90% bij de faculteit CiTG studeert, en 10% aan een andere faculteit.
Laat A de gebeurtenis zijn dat een student het tentamen haalt, en B de gebeurtenis dat een student bij CiTG studeert. Bekend is dat de kans dat een student het tentamen haalt, gegeven dat hij aan een andere faculteit dan CiTG studeert, 45% is. Bereken $\mathrm{P}(A \cap B^c)$

a. 0.1000 **b.** 0.045 **c.** 0.4500 **d.** 0.9000 **e.** 0.0550 **f.** 0.5500

2. Van de Nederlandse bevolking lijdt 1 op de 100 mensen aan reumatoïde artritis. Er bestaat een test, de "reumatest", die bij reumapatiënten meestal positief is en bij niet-reumapatiënten meestal negatief. De test is echter niet 100% waterdicht en heeft een specificiteit (dat wil zeggen de kans op een negatieve test als de ziekte afwezig is) van 0,8 en een sensitiviteit (kans op een positieve test bij aanwezigheid van de ziekte) van 0,7.

Met Z geven we aan dat de testpersoon aan de ziekte lijdt en met T dat de uitslag van de test positief is. Bereken de kans $\mathrm{P}(Z|T)$ dat een persoon de ziekte heeft, gegeven dat de persoon positief test.

a. 0.198 **b.** 0.070 **c.** 0.034 **d.** 0.160 **e.** 0.140 **f.** 0.560

3. Van een discrete stochast X is de kansmassafunctie gegeven in onderstaande tabel. Bereken $\mathrm{E}\left[X^2\right]$.

a	-2	0	2
$p_X(a)$	$1/3$	$1/6$	$1/2$

a. $\frac{5}{3}$ **b.** $\frac{1}{3}$ **c.** $\frac{2}{3}$ **d.** $\frac{1}{6}$ **e.** $\frac{5}{6}$ **f.** $\frac{10}{3}$

4. Van een continue stochast X is de verdelingsfunctie $F_X(x)$ gegeven door

- $F_X(x) = 0$, als $x \leq 0$
- $F_X(x) = \sqrt{x}$, als $0 \leq x \leq 1$
- $F_X(x) = 1$,als $x \geq 1$

Bereken $\mathrm{E}[X]$

a. $\frac{1}{3}$ **b.** $\frac{2}{3}$ **c.** $\frac{1}{2}$ **d.** 1 **e.** $\frac{1}{5}$ **f.** $\frac{2}{5}$

5. Laat $X \sim \mathrm{Par}(\frac{1}{2})$. Bekijk de functie $g : [1, \infty) \to \mathbb{R}$ gedefinieerd door $g(x) = \frac{1}{x^2}$, voor $x \geq 1$. Bereken $\mathrm{E}[g(X)]$.

a. $\frac{1}{5}$ **b.** $\frac{1}{2}$ **c.** ∞ **d.** $\frac{2}{7}$ **e.** -1 **f.** $\frac{5}{2}$

6. Stel $X \sim \mathrm{Exp}(0.2)$. Laat $Y = -\frac{1}{5}X - 3$.
Bepaal $\mathrm{E}[Y]$ en $\mathrm{Var}(Y)$.

a. $\mathrm{E}[Y] = -1$ en $\mathrm{Var}(Y) = 1$
b. $\mathrm{E}[Y] = -1$ en $\mathrm{Var}(Y) = 2$
c. $\mathrm{E}[Y] = -1$ en $\mathrm{Var}(Y) = 25$
d. $\mathrm{E}[Y] = -4$ en $\mathrm{Var}(Y) = 1$
e. $\mathrm{E}[Y] = -4$ en $\mathrm{Var}(Y) = -2$
f. $\mathrm{E}[Y] = -4$ en $\mathrm{Var}(Y) = 2$

7. Laat X en Y twee stochasten zijn met gezamenlijke verdeling gegeven door onderstaande tabel. Hierbij kan X dus de waarden $0, 1$ en 2 aannemen.
Wat is $\mathrm{P}(X = 2)$?

		a	
b	0	1	2
1	1/4	1/4	1/4
-1	0	1/8	1/8

a. $\frac{1}{8}$ **b.** $\frac{3}{8}$ **c.** $\frac{1}{2}$ **d.** $\frac{3}{4}$ **e.** $\frac{1}{4}$ **f.** 1

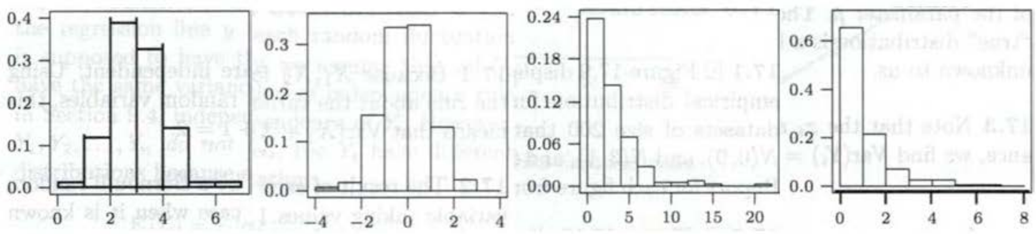
8. Stel je gooit k keer met een zuivere munt. Je telt het aantal keren dat je KOP gooit. Bepaal de kans op hoogstens één keer KOP in k worpen.

a. $1 - \left(k(\frac{1}{2})^k + (\frac{1}{2})^k\right)$
b. $1 - k(\frac{1}{2})^k$
c. $1 - (\frac{1}{2})^k$
d. $(\frac{1}{2})^k$
e. $(k+1)(\frac{1}{2})^k$
f. $k(\frac{1}{2})^k$

9. Je hebt 10 zakken cement besteld, die elk 94 kg horen te wegen. Het gemiddelde gewicht van de 10 zakken is 93.5 kg. Neem aan dat de 10 gewichten beschouwd mogen worden als een realisatie van een Normale verdeling met onbekende parameters. Construeer een 95% betrouwbaarheidsinterval voor het verwachte gewicht van een zak cement. De steekproefstandaarddeviatie is 0.75.

a. (93.109 , 93.890)
b. (93.035 , 93.965)
c. (92.972 , 94.028)
d. (93.065 , 93.935)
e. (92.964 , 94.037)
f. (92.072 , 94.228)

10. Bekijk de onderstaande histogrammen.



Dataset I Dataset II Data set III Data set IV

Welke bewering is de enige juiste?

a. Dataset I komt uit een N(0,1)-verdeling en dataset III uit een Exp(1/3)-verdeling
b. Dataset II komt uit een N(0,1)-verdeling en dataset III uit een Exp(1)-verdeling
c. Dataset I komt uit een N(0,1)-verdeling en dataset IV uit een Exp(1/3)-verdeling
d. Dataset I komt uit een N(3,1)-verdeling en dataset II uit een Norm(2,4)-verdeling
e. Dataset I komt uit een N(3,1)-verdeling en dataset III uit een Exp(1/3)-verdeling
f. Dataset I komt uit een N(3,1)-verdeling en dataset IV uit een Exp(1/3)-verdeling

11. Laat $X_1, X_2, \cdots, X_{625}$ onafhankelijk, gelijkverdeelde stochasten zijn uit een Exponentiële verdeling met parameter 1/10. Bereken met behulp van de Centrale Limietstelling $P(X_1 + X_2 + \cdots + X_{625} \leq 6450)$.

a. 0.4681 **b.** 0.2148 **c.** 0.2112 **d.** 0.2109 **e.** 0.7762 **f.** 0.7881

12. Gegeven is een getal t, de realisatie van een stochastische variabele T met een Normale verdeling met parameters μ en 4. Men toetst met behulp van de toetsingsgrootheid T:
$H_0 : \mu = 0$ tegen $H_1 : \mu \neq 0$. Men besluit om H_0 te verwerpen als de uitkomst van de toetsingsgrootheid $|T| > 0.3$. Bereken de kans op een type II fout als μ in werkelijkheid 0.4 is.

a. 0.5120 **b.** 0.4880 **c.** 0.8831 **d.** 0.1169 **e.** 0.4801 **f.** 0.3632

13. Dit tentamen bestaat uit 15 zeskeuze vragen en drie open vragen. De normering is als volgt:
cijfer = 0.4 punt voor elke meerkeuzevraag en 1 punt voor elke open vraag + 1.
Stel je weet op zes meerkeuzevragen het antwoord zeker en van de open vragen heb je er 2 goed en eentje helemaal fout. De rest van de meerkeuzevragen moet je gokken. Laat X de stochast zijn die het aantal goed gegokte vragen weergeeft. Dan is de verdeling van X gegeven door:

a. $Bin(6, \frac{1}{6})$ **b.** $Bin(15, \frac{1}{6})$ **c.** $Bin(9, \frac{1}{6})$ **d.** $Bin(18, \frac{1}{6})$ **e.** $Ber(\frac{1}{6})$ **f.** $Ber(\frac{5}{6})$

14. Stel $X \sim N(3, 25)$. Bereken $\mathrm{P}(X < 4)$.

a. 0.4207 **b.** 0.4129 **c.** 0.4335 **d.** 0.5793 **e.** 0.5160 **f.** 0.4840

15. Tweeëndertig proefpersonen doen mee aan een experiment ter bevordering van de nachtrust. Zij worden aselect verdeeld over een experimentele en een controle groep. De controle groep krijgt geen behandeling terwijl de aan de experimentele groep toegewezen personen voor het slapen gaan een beker warme melk met honing drinken. Beide groepen bestaan uit 16 personen. In de experimentele groep is de gemiddelde duur van de slaap $\bar{X} = 7$ met $S_X = 0.25$. In de controle groep geldt $\bar{Y} = 6.75$ met $S_Y = 0.25$ (meting in uren). Het geschatte effect van het drinken van een beker melk is $\bar{X} - \bar{Y} = 0.25$ uur (een kwartier langer slaap). Toets de hypothese dat het drinken van een beker melk de nachtrust verlengt op significantieniveau $\alpha = 0.05$:
$H_0 : \mu_X = \mu_Y$ tegen $H_1 : \mu_X > \mu_Y$ Welke van de volgende uitspraken is de enige juiste (gebaseerd op deze data)?

a. De kritieke waarde voor deze toets is 2.042; verwerp H_0 niet
b. De kritieke waarde voor deze toets is 2.042; verwerp H_0 wel
c. De kritieke waarde voor deze toets is 1.697; verwerp H_0 niet
d. De kritieke waarde voor deze toets is 1.697; verwerp H_0 wel
e. De kritieke waarde voor deze toets is 2.048; verwerp H_0 niet
f. De kritieke waarde voor deze toets is 2.048; verwerp H_0 wel

Zie volgende bladzijde voor de Open Vragen

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. Bekijk de data punten

$$(1,0),\ (2,1),\ (2,2),\ (3,2).$$

We gaan op zoek naar de rechte lijn die het beste past bij de data in de zin van de kleinste kwadratenoplossing. Het model is

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

De matrixformulering voor dit regressiemodel is

$$\mathbf{y} = X\beta + \epsilon.$$

a. Geef de matrix X en de observatievector $\mathbf{y}$.

b. Bepaal het startgetal β_0 en de richtingscoëfficiënt β_1 van de regressielijn door deze datapunten, door gebruik te maken van de matrixformulering.

2. Bij een groep van 25 proefpersonen wordt de reactietijd in seconden op een bepaalde taak vastgesteld, zowel onmiddelijk vóór als een half uur na de consumptie van één liter bier.
Per persoon wordt het verschil D in reactietijd voor en na het drinken bepaald. Het vermoeden is dat de alcoholconsumptie de reactietijd doet toenemen, met andere woorden, men vermoedt dat D positief is.
De uitkomsten van de 25 proefpersonen geven een gemiddeld verschil $\bar{D}_{25}$ van 1.32 sec met steekproefstandaarddeviatie $S_{25} = 0.7056$.
Toets het vermoeden op significantieniveau $\alpha = 0.1$:

a. Stel de relevante hypotheses op.

b. Bereken de waarde van de toetsingsgrootheid.

c. Geef de conclusie van de toets.

3. X en Y hebben een gezamenlijke verdeling volgens de volgende tabel:

$P(X=a, Y=b)$		b				
		0	1	2	3	$P(X=a)$
a	-1	0.05	0.15	0.05	0.10	0.35
	0	0.10	0.05	0.05	0.05	0.25
	1	0.10	0.20	0.05	0.05	0.40
$P(Y=b)$		0.25	0.40	0.15	0.20	1

a. Bereken $E[XY]$

b. Bereken $\mathrm{Cov}(X, Y)$.

Einde

Tentamen Kansrekening en Statistiek
CTB2200
19 januari 2015

Dit tentamen bestaat uit 15 meerkeuzevragen en 3 open vragen. Bij dit examen is het gebruik van een (evt. grafische) rekenmachine toegestaan. Tevens krijgt u een formuleblad uitgereikt.

Meerkeuzevragen

Toelichting:
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 èn aan te strepen.

---

1. Van de groep studenten die dit tentamen maakt, is bekend dat 90% bij de faculteit CiTG studeert, en 10% aan een andere faculteit.
Laat A de gebeurtenis zijn dat een student het tentamen haalt, en B de gebeurtenis dat een student bij CiTG studeert. Bekend is dat de kans dat een student het tentamen haalt, gegeven dat hij aan een andere faculteit dan CiTG studeert, 45% is. Bereken $\mathrm{P}(A \cap B^c)$

a. 0.1000 **b.** 0.045 **c.** 0.4500 **d.** 0.9000 **e.** 0.0550 **f.** 0.5500

2. Van de Nederlandse bevolking lijdt 1 op de 100 mensen aan reumatoïde artritis. Er bestaat een test, de "reumatest", die bij reumapatiënten meestal positief is en bij niet-reumapatiënten meestal negatief. De test is echter niet 100% waterdicht en heeft een specificiteit (dat wil zeggen de kans op een negatieve test als de ziekte afwezig is) van 0,8 en een sensitiviteit (kans op een positieve test bij aanwezigheid van de ziekte) van 0,7.

Met Z geven we aan dat de testpersoon aan de ziekte lijdt en met T dat de uitslag van de test positief is. Bereken de kans $\mathrm{P}(Z|T)$ dat een persoon de ziekte heeft, gegeven dat de persoon positief test.

a. 0.198 **b.** 0.070 **c.** 0.034 **d.** 0.160 **e.** 0.140 **f.** 0.560

3. Van een discrete stochast X is de kansmassafunctie gegeven in onderstaande tabel. Bereken $\mathrm{E}\left[X^2\right]$.

a	-2	0	2
$p_X(a)$	$1/3$	$1/6$	$1/2$

a. $\frac{5}{3}$ **b.** $\frac{1}{3}$ **c.** $\frac{2}{3}$ **d.** $\frac{1}{6}$ **e.** $\frac{5}{6}$ **f.** $\frac{10}{3}$

4. Van een continue stochast X is de verdelingsfunctie $F_X(x)$ gegeven door

- $F_X(x) = 0$, als $x \leq 0$
- $F_X(x) = \sqrt{x}$, als $0 \leq x \leq 1$
- $F_X(x) = 1$,als $x \geq 1$

Bereken $\mathrm{E}[X]$

a. $\frac{1}{3}$ **b.** $\frac{2}{3}$ **c.** $\frac{1}{2}$ **d.** 1 **e.** $\frac{1}{5}$ **f.** $\frac{2}{5}$

5. Laat $X \sim \mathrm{Par}(\frac{1}{2})$. Bekijk de functie $g : [1, \infty) \to \mathbb{R}$ gedefinieerd door $g(x) = \frac{1}{x^2}$, voor $x \geq 1$. Bereken $\mathrm{E}[g(X)]$.

a. $\frac{1}{5}$ **b.** $\frac{1}{2}$ **c.** ∞ **d.** $\frac{2}{7}$ **e.** -1 **f.** $\frac{5}{2}$

6. Stel $X \sim \mathrm{Exp}(0.2)$. Laat $Y = -\frac{1}{5}X - 3$.
Bepaal $\mathrm{E}[Y]$ en $\mathrm{Var}(Y)$.

a. $\mathrm{E}[Y] = -1$ en $\mathrm{Var}(Y) = 1$
b. $\mathrm{E}[Y] = -1$ en $\mathrm{Var}(Y) = 2$
c. $\mathrm{E}[Y] = -1$ en $\mathrm{Var}(Y) = 25$
d. $\mathrm{E}[Y] = -4$ en $\mathrm{Var}(Y) = 1$
e. $\mathrm{E}[Y] = -4$ en $\mathrm{Var}(Y) = -2$
f. $\mathrm{E}[Y] = -4$ en $\mathrm{Var}(Y) = 2$

7. Laat X en Y twee stochasten zijn met gezamenlijke verdeling gegeven door onderstaande tabel. Hierbij kan X dus de waarden $0, 1$ en 2 aannemen.
Wat is $\mathrm{P}(X = 2)$?

		a	
b	0	1	2
1	1/4	1/4	1/4
-1	0	1/8	1/8

a. $\frac{1}{8}$ **b.** $\frac{3}{8}$ **c.** $\frac{1}{2}$ **d.** $\frac{3}{4}$ **e.** $\frac{1}{4}$ **f.** 1

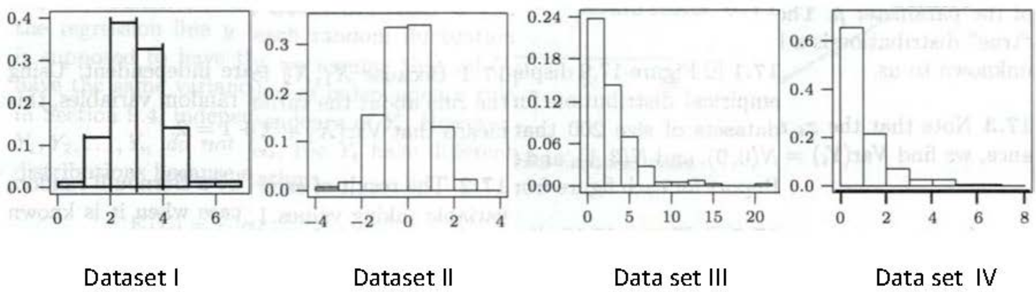
8. Stel je gooit k keer met een zuivere munt. Je telt het aantal keren dat je KOP gooit. Bepaal de kans op hoogstens één keer KOP in k worpen.

a. $1 - \left(k(\frac{1}{2})^k + (\frac{1}{2})^k\right)$
b. $1 - k(\frac{1}{2})^k$
c. $1 - (\frac{1}{2})^k$
d. $(\frac{1}{2})^k$
e. $(k+1)(\frac{1}{2})^k$
f. $k(\frac{1}{2})^k$

9. Je hebt 10 zakken cement besteld, die elk 94 kg horen te wegen. Het gemiddelde gewicht van de 10 zakken is 93.5 kg. Neem aan dat de 10 gewichten beschouwd mogen worden als een realisatie van een Normale verdeling met onbekende parameters. Construeer een 95% betrouwbaarheidsinterval voor het verwachte gewicht van een zak cement. De steekproefstandaarddeviatie is 0.75.

a. (93.109 , 93.890)
b. (93.035 , 93.965)
c. (92.972 , 94.028)
d. (93.065 , 93.935)
e. (92.964 , 94.037)
f. (92.072 , 94.228)

10. Bekijk de onderstaande histogrammen.



Welke bewering is de enige juiste?

a. Dataset I komt uit een N(0,1)-verdeling en dataset III uit een Exp(1/3)-verdeling
b. Dataset II komt uit een N(0,1)-verdeling en dataset III uit een Exp(1)-verdeling
c. Dataset I komt uit een N(0,1)-verdeling en dataset IV uit een Exp(1/3)-verdeling
d. Dataset I komt uit een N(3,1)-verdeling en dataset II uit een Norm(2,4)-verdeling
e. Dataset I komt uit een N(3,1)-verdeling en dataset III uit een Exp(1/3)-verdeling
f. Dataset I komt uit een N(3,1)-verdeling en dataset IV uit een Exp(1/3)-verdeling

11. Laat $X_1, X_2, \cdots, X_{625}$ onafhankelijk, gelijkverdeelde stochasten zijn uit een Exponentiële verdeling met parameter 1/10. Bereken met behulp van de Centrale Limietstelling $P(X_1 + X_2 + \cdots + X_{625} \leq 6450)$.

a. 0.4681 **b.** 0.2148 **c.** 0.2112 **d.** 0.2109 **e.** 0.7762 **f.** 0.7881

12. Gegeven is een getal t, de realisatie van een stochastische variabele T met een Normale verdeling met parameters μ en 4. Men toetst met behulp van de toetsingsgrootheid T:
$H_0 : \mu = 0$ tegen $H_1 : \mu \neq 0$. Men besluit om H_0 te verwerpen als de uitkomst van de toetsingsgrootheid $|T| > 0.3$. Bereken de kans op een type II fout als μ in werkelijkheid 0.4 is.

a. 0.5120 **b.** 0.4880 **c.** 0.8831 **d.** 0.1169 **e.** 0.4801 **f.** 0.3632

13. Dit tentamen bestaat uit 15 zeskeuze vragen en drie open vragen. De normering is als volgt:
cijfer = 0.4 punt voor elke meerkeuzevraag en 1 punt voor elke open vraag + 1.
Stel je weet op zes meerkeuzevragen het antwoord zeker en van de open vragen heb je er 2 goed en eentje helemaal fout. De rest van de meerkeuzevragen moet je gokken. Laat X de stochast zijn die het aantal goed gegokte vragen weergeeft. Dan is de verdeling van X gegeven door:

a. $Bin(6, \frac{1}{6})$ **b.** $Bin(15, \frac{1}{6})$ **c.** $Bin(9, \frac{1}{6})$ **d.** $Bin(18, \frac{1}{6})$ **e.** $Ber(\frac{1}{6})$ **f.** $Ber(\frac{5}{6})$

14. Stel $X \sim N(3, 25)$. Bereken $\mathrm{P}(X < 4)$.

a. 0.4207 **b.** 0.4129 **c.** 0.4335 **d.** 0.5793 **e.** 0.5160 **f.** 0.4840

15. Tweeëndertig proefpersonen doen mee aan een experiment ter bevordering van de nachtrust. Zij worden aselect verdeeld over een experimentele en een controle groep. De controle groep krijgt geen behandeling terwijl de aan de experimentele groep toegewezen personen voor het slapen gaan een beker warme melk met honing drinken. Beide groepen bestaan uit 16 personen. In de experimentele groep is de gemiddelde duur van de slaap $\bar{X} = 7$ met $S_X = 0.25$. In de controle groep geldt $\bar{Y} = 6.75$ met $S_Y = 0.25$ (meting in uren). Het geschatte effect van het drinken van een beker melk is $\bar{X} - \bar{Y} = 0.25$ uur (een kwartier langer slaap). Toets de hypothese dat het drinken van een beker melk de nachtrust verlengt op significantieniveau $\alpha = 0.05$:
$H_0 : \mu_X = \mu_Y$ tegen $H_1 : \mu_X > \mu_Y$ Welke van de volgende uitspraken is de enige juiste (gebaseerd op deze data)?

a. De kritieke waarde voor deze toets is 2.042; verwerp H_0 niet
b. De kritieke waarde voor deze toets is 2.042; verwerp H_0 wel
c. De kritieke waarde voor deze toets is 1.697; verwerp H_0 niet
d. De kritieke waarde voor deze toets is 1.697; verwerp H_0 wel
e. De kritieke waarde voor deze toets is 2.048; verwerp H_0 niet
f. De kritieke waarde voor deze toets is 2.048; verwerp H_0 wel

Zie volgende bladzijde voor de Open Vragen

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. Bekijk de data punten

$$(1,0),\ (2,1),\ (2,2),\ (3,2).$$

We gaan op zoek naar de rechte lijn die het beste past bij de data in de zin van de kleinste kwadratenoplossing. Het model is

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

De matrixformulering voor dit regressiemodel is

$$\mathbf{y} = X\beta + \epsilon.$$

a. Geef de matrix X en de observatievector $\mathbf{y}$.

b. Bepaal het startgetal β_0 en de richtingscoëfficiënt β_1 van de regressielijn door deze datapunten, door gebruik te maken van de matrixformulering.

2. Bij een groep van 25 proefpersonen wordt de reactietijd in seconden op een bepaalde taak vastgesteld, zowel onmiddelijk vóór als een half uur na de consumptie van één liter bier.
Per persoon wordt het verschil D in reactietijd voor en na het drinken bepaald. Het vermoeden is dat de alcoholconsumptie de reactietijd doet toenemen, met andere woorden, men vermoedt dat D positief is.
De uitkomsten van de 25 proefpersonen geven een gemiddeld verschil $\bar{D}_{25}$ van 1.32 sec met steekproefstandaarddeviatie $S_{25} = 0.7056$.
Toets het vermoeden op significantieniveau $\alpha = 0.1$:

a. Stel de relevante hypotheses op.

b. Bereken de waarde van de toetsingsgrootheid.

c. Geef de conclusie van de toets.

3. X en Y hebben een gezamenlijke verdeling volgens de volgende tabel:

$P(X = a, Y = b)$		$b = 0$	1	2	3	$P(X = a)$
a	-1	0.05	0.15	0.05	0.10	0.35
	0	0.10	0.05	0.05	0.05	0.25
	1	0.10	0.20	0.05	0.05	0.40
$P(Y = b)$		0.25	0.40	0.15	0.20	1

a. Bereken $E[XY]$

b. Bereken $\mathrm{Cov}(X, Y)$.

Einde

Uitwerkingen

1 b. Gegeven is dat $\mathrm{P}(A|B^c) = 0.45$ en $\mathrm{P}(B^c) = 0.1$. Dus

$$\mathrm{P}(A \cap B^c) = \mathrm{P}(A|B^c)\,\mathrm{P}(B^c) = 0.45 \cdot 0.1 = 0.045$$

2 c. De gegevens kunnen we vertalen als: $\mathrm{P}(T|Z) = 0.7$, $\mathrm{P}(Z) = 0.01$ en $\mathrm{P}(T^c|Z^c) = 0.8$ Met de regel van Bayes kunnen we nu berekenen:

$$\mathrm{P}(Z|T) = \frac{\mathrm{P}(T|Z)\,\mathrm{P}(Z)}{\mathrm{P}(T|Z)\,\mathrm{P}(Z) + \mathrm{P}(T|Z^c)\,\mathrm{P}(Z^c)} = \frac{0,70 \times 0,01}{0,70 \times 0,01 + 0,20 \times 0,99} = 0,034$$

3 f. De kansverdeling van X^2 is gegeven door:

a	0	4
$p_X(a)$	1/6	5/6

Dus $\mathrm{E}\left[X^2\right] = 0 \cdot \frac{1}{6} + 4 \cdot \frac{5}{6} = \frac{10}{3}$

4 a. De formule voor de verwachting is $\mathrm{E}[X] = \int x f_X(x)\,dx$, met f_X de dichtheid van X. Deze is te berekenen als de afgeleide van de verdelingsfunctie, dus $f_X(x) = F'(x) = \frac{1}{2\sqrt{x}}$. Dus

$$\mathrm{E}[X] = \int_0^1 x \frac{1}{2\sqrt{x}}\,dx = \int_0^1 \frac{1}{2}\sqrt{x}\,dx = \frac{1}{3}$$

5 a.

$$\begin{aligned} \mathrm{E}[g(X)] &= \int_1^\infty \frac{1}{x^2}\frac{1}{2x^{3/2}}\,dx \\ &= \frac{1}{2}\int_1^\infty \frac{1}{x^{7/2}}\,dx \\ &= \frac{1}{2}[-\frac{2}{5}x^{-5/2}]_1^\infty \\ &= \frac{1}{5} \end{aligned}$$

6 d. $X \sim \mathrm{Exp}(0.2)$ betekent $\mathrm{E}[X] = 5$ en $\mathrm{Var}(X) = 25$. Volgens de rekenregels voor verwachting en variantie krijgen we dus $\mathrm{E}[Y] = -\frac{1}{5}\mathrm{E}[X] - 3 = -4$ en $\mathrm{Var}(Y) = (-\frac{1}{5})^2\mathrm{Var}(X) = 1$.

7 b. Om die kans te berekenen, moeten we de kansen in de kolom bij $X = 2$ bij elkaar optellen. Dit geeft $\frac{1}{4} + \frac{1}{8} = \frac{3}{8}$.

8 e. Laat X het aantal keren KOP zijn in k worpen. Dan is X dus binomiaal verdeeld met parameters k en $\frac{1}{2}$. Dus de gevraagde kans is

$$\begin{aligned}
P(X \leq 1) &= P(X=0) + P(X=1) \\
&= \binom{k}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^k + \binom{k}{1}\frac{1}{2}\left(\frac{1}{2}\right)^{k-1} = \\
&= \left(\frac{1}{2}\right)^k + k\left(\frac{1}{2}\right)^{k-1} \\
&= (k+1)\left(\frac{1}{2}\right)^k
\end{aligned}$$

9 e. Het betrouwbaarheidsinterval wordt gegeven door

$$\left(\bar{x}_n \pm t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}}\right).$$

$\bar{x}_n = 93.5$, $s_n = 0.75$, $n = 10$, $t_{9,0.025} = 2.262$ invullen geeft het antwoord.

10 e. Dit is een deel van opgave 17.1

11 f. Aangezien $X \sim \text{Exp}(0.1)$, geldt $\text{E}[X] = 10$ en $\text{Var}(X) = 100$. Dus

$$\begin{aligned}
P(X_1 + X_2 + \cdots + X_{625} \leq 6450) &= P\left(\frac{X_1 + X_2 + \cdots + X_{625}}{625} \leq \frac{6450}{625}\right) \\
&= P\left(\frac{\bar{X}_{625} - \mu}{\sigma/\sqrt{n}} \leq \frac{\frac{6450}{625} - 10}{10/25}\right) \\
&= P(Z \leq 0.8) = 1 - P(Z > 0.8) \\
&= 1 - 0.2119 = 0.7881
\end{aligned}$$

12 d. Kans op een type II fout is

$$\begin{aligned}
P(H_0 \text{ onterecht niet verwerpen}\,|\mu = 0.4) &= P(|T| \leq 0.3|\mu = 0.4) \\
&= P(-0.3 \leq T \leq 0.3) \\
&= P\left(\frac{-0.3 - 0.4}{2} \leq \frac{T - \mu}{\sigma} \leq \frac{0.3 - 0.4}{2}\right) \\
&= P(-0.35 \leq Z \leq -0.05) \\
&= P(0.05 \leq Z \leq 0.35) \\
&= 0.4801 - 0.3632 = 0.1169
\end{aligned}$$

13 c. De rest van de meerkeuze vragen is $Bin(9, \frac{1}{6})$ verdeeld.

14 d.

$$P(X < 4) = P\left(\frac{X-3}{5} < \frac{4-3}{5}\right) = P(Z < 0.2) = 1 - P(X > 0.2) = 1 - 0.4207 = 0.5793$$

15 d. De gepoolede variantie is

$$S_p^2 = \frac{15 \cdot (0.25)^2 + 15 \cdot (0.25)^2}{30} (\frac{1}{8}) = 0.0078125$$

Dus

$$T_p = \frac{\mu_X - \mu_Y}{S_p} = \frac{7 - 6.75}{\sqrt{0.0078125}} = 2.8284$$

T_p volgt een $t(30)$ verdeling. We toetsen eenzijdig, dus de kritieke waarde is $t_{30,0.05} = 1.697$. De gevonden waarde van T_p is groter dan deze kritieke waarde, dus we moeten H_0 verwerpen.

1a De vector $\mathbf{y} = [0, 1, 2, 2]^T$ en de matrix X is:

$$\mathbf{X} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}.$$

1b Er geldt

$$\mathbf{X^T X} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 4 & 8 \\ 8 & 18 \end{pmatrix}$$

$$\mathbf{X^T y} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 3 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 5 \\ 12 \end{pmatrix}.$$

De normaalvergelijkingen zijn

$$\begin{pmatrix} 4 & 8 \\ 8 & 18 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 5 \\ 12 \end{pmatrix}.$$

De oplossing van de normaal vergelijkingen is

$$\begin{aligned} \mathbf{b} = (\mathbf{X^T X})^{-1} \mathbf{X^T y} &= \begin{pmatrix} 4 & 8 \\ 8 & 18 \end{pmatrix}^{-1} \begin{pmatrix} 5 \\ 12 \end{pmatrix} \\ &= \frac{1}{8} \begin{pmatrix} 18 & -8 \\ -8 & 4 \end{pmatrix} \begin{pmatrix} 5 \\ 12 \end{pmatrix} \\ &= \frac{1}{8} \begin{pmatrix} -6 \\ 8 \end{pmatrix} = \begin{pmatrix} -3/4 \\ 1 \end{pmatrix} \end{aligned}$$

2a $H_0 : D = 0$, $H_1 : D > 0$.

2b

$$T = \frac{\bar{D} - \mu_0}{s_n/\sqrt{n}} = \frac{1.32 - 0}{0.7056/\sqrt{25}} = 9.354$$

2c De kritieke waarde waarmee we de uitkomst van (b) moeten vergelijken is $t_{n-1,\alpha} = t_{24,0.1} = 1.318$. De gevonden waarde van T ligt dus in het kritieke gebied, dus we verwerpen H_0 en concluderen dat er inderdaad een effect is van het drinken van alcohol op de reactietijd.

3a De kansverdeling van XY halen we uit de tabel van de gezamenlijke verdeling van X en Y:

XY	-3	-2	-1	0	1	2	3
$\mathrm{P}(XY = xy)$	0.1	0.05	0.15	0.4	0.2	0.05	0.05

Dus $\mathrm{E}[XY] = -3 \cdot 0.1 - 2 \cdot 0.05 - 1 \cdot 0.15 + 0 \cdot 0.4 + 1 \cdot 0.2 + 2 \cdot 0.05 + 3 \cdot 0.05 = -0.1$

3b

$$\mathrm{Cov}(X,Y) = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y]$$

$\mathrm{E}[X] = -1 \cdot 0.35 + 1 \cdot 0.4 = 0.05$, $\mathrm{E}[Y] = 1 \cdot 0.4 + 2 \cdot 0.15 + 3 \cdot 0.2 = 1.3$, $\mathrm{E}[XY] = -0.1$, dus $\mathrm{Cov}(X,Y) = -0.1 - (1.3 \times 0.05) = -0.165$

Oefententamen Statistiek
wi1102CT

Bij dit examen is het gebruik van een (evt. grafische) rekenmachine toegestaan. Tevens krijgt u een formuleblad uitgereikt—na afloop inleveren alstublieft.

Toelichting: In het algemeen zijn niet altijd vijf van de zes 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 èn aan te strepen.

1. IKEA heeft 1000 Billy boekenkasten in het magazijn. De kans dat er te weinig schroeven zitten in het bouwpakket van zo'n boekenkast is 0.1%. Hoe groot is de kans dat er in het magazijn precies twee boekenkasten staan met te weinig schroeven?

a. 0.184 **b.** 0.368 **c.** 0.512 **d.** 0.772 **e.** 0.801 **f.** 0.920

2. We werpen twee keer met een zuivere dobbelsteen, waarbij we aannemen dat de uitkomst van de eerste worp geen invloed heeft op de tweede worp. Definieer de stochastische variabele M als het maximum van de twee worpen. De kansverdeling van M wordt gegeven door

a.

a	1	2	3	4	5	6
$\mathrm{P}(M=a)$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{6}{36}$

b.

a	1	2	3	4	5	6
$\mathrm{P}(M=a)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

c.

a	1	2	3	4	5	6
$\mathrm{P}(M=a)$	$\frac{12}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

d.

a	1	2	3	4	5	6
$\mathrm{P}(M=a)$	$\frac{11}{36}$	$\frac{9}{36}$	$\frac{7}{36}$	$\frac{5}{36}$	$\frac{3}{36}$	$\frac{1}{36}$

e.

a	1	2	3	4	5	6
$\mathrm{P}(M=a)$	$\frac{11}{36}$	$\frac{6}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{6}{36}$	$\frac{11}{36}$

f.

a	1	2	3	4	5	6
$\mathrm{P}(M=a)$	$\frac{1}{36}$	$\frac{6}{36}$	$\frac{11}{36}$	$\frac{11}{36}$	$\frac{6}{36}$	$\frac{1}{36}$

3. Bereken de verwachting van een stochast X met een kansverdeling gegeven door

a	0	3	4	6
$\mathrm{P}(X=a)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{8}$

a. $1\frac{1}{2}$ **b.** $1\frac{3}{4}$ **c.** 2 **d.** $2\frac{1}{4}$ **e.** $2\frac{3}{4}$ **f.** 3

4. Gegeven is een stochastische variabele met kansdichtheid

$$f(x) = \begin{cases} x & 0 \le x \le 1 \\ 2-x & 1 \le x \le 2 \\ 0 & \text{anders.} \end{cases}$$

en verdelingsfunctie F. Dan is $F(0.5)$ gelijk aan

a. 0.125 **b.** 0.150 **c.** 0.167 **d.** 0.250 **e.** 0.333 **f.** 0.500

5. Drie vrienden gaan een avondje naar het casino. Ze spelen daar een spel met kans $1/4$ om te winnen. Nadat ze alle drie dit spel drie keer hebben gespeeld, wat is de kans dat precies twee van de drie vrienden geen enkele keer het spel hebben gewonnen?

a. 0.052 **b.** 0.103 **c.** 0.206 **d.** 0.309 **e.** 0.412 **f.** 0.515

6. En kansspel heeft kans 0.25 op succes. Geert speelt het spel net zo lang tot hij een keer succes heeft gehad. Wat is de verwachting van het aantal spellen dat hij moet spelen?

a. 2 **b.** 4 **c.** 5 **d.** 8 **e.** 10 **f.** 12

7. Een hotelbedrijf heeft bij aardewerk dat in gebruik is geconstateerd dat de kans, dat een bord langer dan 2 jaar heel blijft, gelijk is aan $\frac{1}{2}$. We definiëren X als de levensduur van deze borden, en we veronderstellen dat X een exponentiële verdeling heeft. Bereken $\mathrm{E}[X]$ in jaren.

a. 0.35 **b.** 0.48 **c.** 2.00 **d.** 2.73 **e.** 2.89 **f.** 3.00

8. Zij $X_1, X_2, \ldots, X_n$ 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 onzuiverheid (bias), en MSE (mean square error) van deze schatter worden gegeven door:

a. bias $= 0$, MSE $= \lambda^2/n^2$ **b.** bias $= 1/(n\lambda)$, MSE $= 2/(n\lambda^2)$
c. bias $= \lambda$, MSE $= 2/(n\lambda) + \lambda^2$ **d.** bias $= 0$, MSE $= \lambda/n$
e. bias $= 1/(n\lambda)$, MSE $= \lambda/n$ **f.** bias $= \lambda$, MSE $= \lambda^2/n^2$

9. De mediaan van een $Par(1)$ verdeling is gelijk aan

a. -3 **b.** 0 **c.** $\frac{1}{2}$ **d.** 1 **e.** 2 **f.** 4

10. De stochast B is de belasting op een betonnen ligger van een brug. Stel B heeft een normale verdeling met verwachting $\mu_B = 50$ kN en een standaardafwijking $\sigma_B = 8$ kN. De stochast S is de sterkte van de betonnen ligger. Stel S heeft een normale verdeling met verwachting $\mu_S = 60$ kN en een standaardafwijking $\sigma_S = 6$ kN. Gegeven is dat B en S ongecorreleerd zijn. Bereken de kans dat S kleiner is dan B, dus $\mathrm{P}(S < B)$. U mag gebruiken dat $S - B$ een normale verdeling heeft.

a. 0.08 **b.** 0.92 **c.** 0.24 **d.** 0.76 **e.** 0.84 **f.** 0.16

11. Gegeven is een stochastische variabele X met $\mathrm{E}[X] = 2$, $\mathrm{Var}(X) = 4$. Dan worden de verwachting en variantie van $Y = 3 - 2X$ gegeven door

a. $\mathrm{E}[Y] = -1$ en $\mathrm{Var}(Y) = 5$ **b.** $\mathrm{E}[Y] = -1$ en $\mathrm{Var}(Y) = 8$
c. $\mathrm{E}[Y] = -1$ en $\mathrm{Var}(Y) = 16$ **d.** $\mathrm{E}[Y] = -1$ en $\mathrm{Var}(Y) = 19$
e. $\mathrm{E}[Y] = 8$ en $\mathrm{Var}(Y) = 5$ **f.** $\mathrm{E}[Y] = 8$ en $\mathrm{Var}(Y) = 8$

12. Stel we hebben een steekproef X_1, X_2 uit een verdeling met onbekende verwachting μ en onbekende variantie σ^2. Vervolgens doen we een andere meting Y, met verwachting 2μ (let op de factor 2!) en variantie σ^2, onafhankelijk van de eerste steekproef. We beschouwen een zuivere schatter T voor μ van de volgende vorm:

$$T = aX_1 + aX_2 + (\frac{1}{2} - a)Y,$$

met $a \in \mathbb{R}$. Voor welke waarde van a is de variantie van T minimaal?

a. $a = \frac{1}{4}$ **b.** $a = 0$ **c.** $a = \frac{1}{6}$ **d.** $a = \frac{1}{3}$ **e.** $a = \frac{1}{2}$ **f.** $a = \frac{3}{8}$

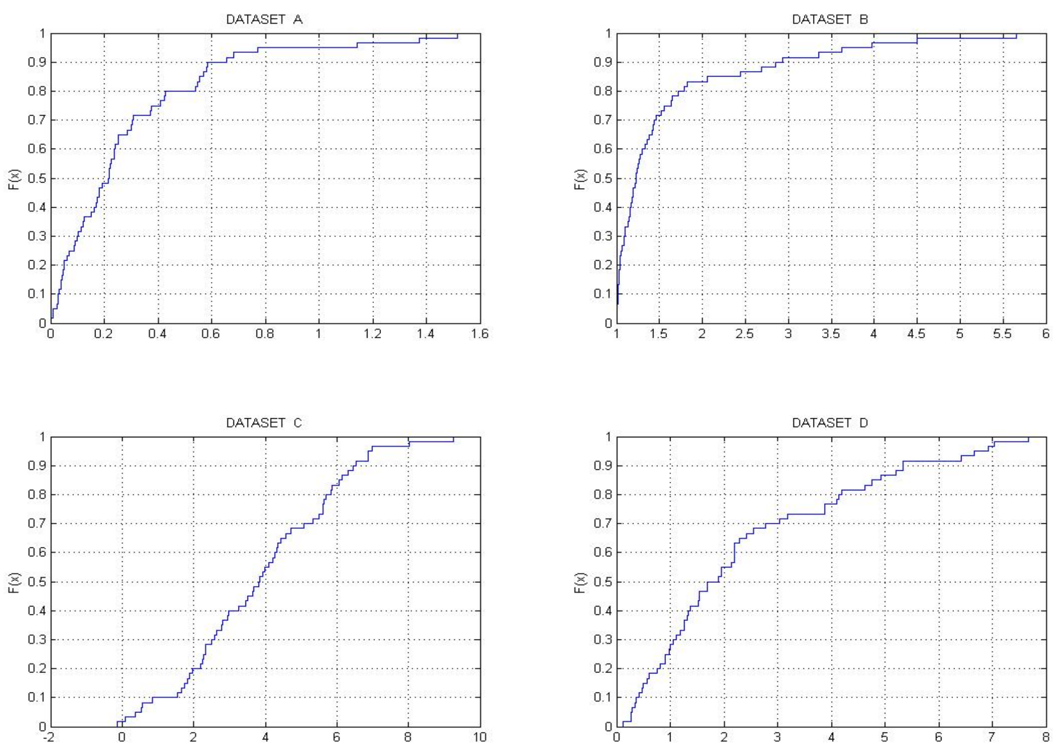
13. Gegeven is de volgende geordende dataset:

0	0	0	2	4	6	8	9	10	10
10	12	15	15	16	21	22	24	26	30
30	31	33	36	44	50	55	58	65	68
75	77	79	81	88	91	97	100	108	108
112	113	114	115	120	122	129	134	138	143
148	160	176	180	193	193	197	227	232	233
236	242	245	255	261	263	281	290	296	300
300	325	330	357	365	369	371	379	386	422
445	446	447	452	457	482	529	529	543	600
648	670	700	707	724	729	748	790	810	816
828	843	860	865	868	875	943	948	983	990
1011	1045	1064	1071	1082	1146	1160	1222	1247	1351
1435	1461	1755	1783	1800	1864	1897	2323	2930	3110
3321	4116	5485	5509	6150					

De hoogte van het histogram van deze data bij de cel (500, 1000] is gelijk aan

a. 0.00036 **b.** 0.00163 **c.** 0.04800 **d.** 0.17778 **e.** 0.27000 **f.** 0.81481

14. Van vier datasets zijn de empirische verdelingfuncties weergegeven. De datasets zijn afkomstig uit vier verschillende verdelingen: een $Exp(3)$, een $Exp(1/3)$, een $Par(3)$ en een $N(4,4)$ verdeling. Welke dataset komt uit de $Exp(3)$ en welke dataset komt uit de $Exp(1/3)$ verdeling?



a. $Exp(3)$: A; $Exp(1/3)$: C **b.** $Exp(3)$: A; $Exp(1/3)$: D
c. $Exp(3)$: B; $Exp(1/3)$: A **d.** $Exp(3)$: B; $Exp(1/3)$: D
e. $Exp(3)$: D; $Exp(1/3)$: A **f.** $Exp(3)$: D; $Exp(1/3)$: C

15. We construeren een histogram met cellen [0,1], (1,3], (3,5], (5,8], (8,11], (11,14], en (14,18]. Gegeven zijn de waarden van de empirische verdelingsfunctie in de randen van de cellen:

t	0	1	3	5	8	11	14	18
$F_n(t)$	0	0.225	0.445	0.615	0.735	0.805	0.910	1.000

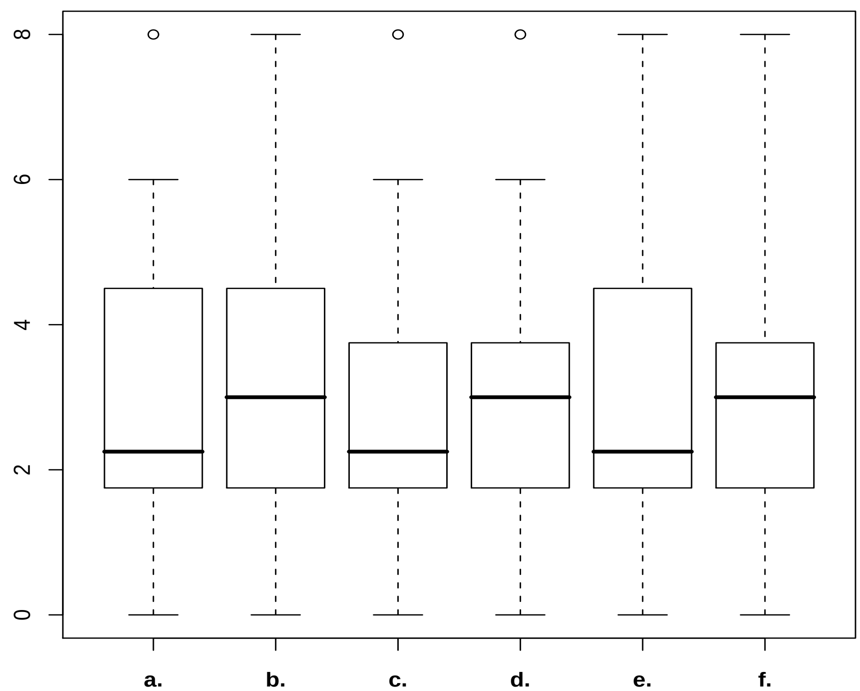
Dan is de hoogte van het histogram op de cel $(8, 11]$ gelijk aan

a. 0.0233 **b.** 0.0035 **c.** 0.0700 **d.** 0.2100 **e.** 0.2450 **f.** 0.2683

16. Gegeven is de dataset

0.00 1.75 1.75 1.75 2.00 2.50 4.50 4.50 6.00 8.00

Welke van de zes onderstaande boxplots hoort bij deze dataset?



17. Aan een tentamen met 20 meerkeuze vragen, elk met 6 mogelijkheden, doen 30 studenten mee. Alle deelnemende studenten beantwoorden de vragen onafhankelijk van elkaar en loten bij elke vraag uit de 6 mogelijkheden met gelijke kans. Zij X het aantal studenten dat minstens 1 vraag goed beantwoord heeft. We modelleren X als een stochastische variabele met een binomiale verdeling met n en p gegeven door

a. $n = 20$, $p = 1 - (\frac{5}{6})^{30}$ **b.** $n = 20$, $p = (\frac{1}{6})^{30}$

c. $n = 20$, $p = (\frac{5}{6})^{30}$ **d.** $n = 30$, $p = 1 - (\frac{5}{6})^{20}$

e. $n = 30$, $p = (\frac{1}{6})^{20}$ **f.** $n = 30$, $p = (\frac{5}{6})^{20}$

18. Gegeven is een bivariate dataset $(x_1, y_1), (x_2, y_2), \ldots, (x_{50}, y_{50})$, waarvan bekend is dat

$$\sum_{i=1}^{50} x_i = 150 \quad \text{en} \quad \sum_{i=1}^{50} y_i = 400.$$

De kleinste kwadraten schatting voor de helling van de regressielijn is $\hat{\beta} = 2$. De kleinste kwadraten schatting $\hat{\alpha}$ voor het intercept is

a. 0 **b.** 2 **c.** 14 **d.** 100 **e.** 700 **f.** niet te berekenen

19. Voor het aantal klanten X_1 dat een winkel binnen loopt tussen 14:00 en 15:00 geldt een Poisson verdeling met parameter $\lambda > 0$. Dat wil zeggen:

$$\mathrm{P}(X_1 = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{voor } k = 0, 1, 2, \ldots.$$

Het aantal klanten dat binnen loopt tussen 15:00 en 16:00, noem dit X_2, heeft dezelfde verdeling, en is onafhankelijk van X_1. De winkelbediende weet dat er tussen 14:00 en 16:00 uur 1 klant is binnengekomen. Wat is de maximum likelihood schatter voor λ, gebaseerd op dit gegeven?

a. $\hat{\lambda} = 0.25$ **b.** $\hat{\lambda} = 0.50$ **c.** $\hat{\lambda} = 1.0$ **d.** $\hat{\lambda} = 1.5$ **e.** $\hat{\lambda} = 2.0$ **f.** $\hat{\lambda} = 4.0$

20. Piet en Marijke spelen een spel op de computer. Zij hebben beiden kans p om het spel te winnen van de computer. Piet speelt drie keer, en wint 1 keer. Marijke speelt net zolang totdat ze een keer wint; dit gebeurt de vierde keer dat ze speelt. Geef de maximum likelihoodschatter voor p, gebaseerd op deze gegevens.

a. $\hat{p} = \frac{2}{7}$ **b.** $\hat{p} = \frac{3}{7}$ **c.** $\hat{p} = \frac{1}{3}$ **d.** $\hat{p} = \frac{2}{3}$ **e.** $\hat{p} = \frac{1}{5}$ **f.** $\hat{p} = \frac{2}{5}$

21. Een fabrikant van verf wenst de gemiddelde droogtijd van een nieuwe type muurverf te bepalen. Voor 12 testmuren van dezelfde oppervlakte vond hij een gemiddelde droogtijd van 66.3 minuten en een steekproefvariantie van 8.4. De 12 gemeten droogtijden vat men op als een realisatie van een steekproef uit een $N(\mu, \sigma^2)$ verdeling. Het 90% betrouwbaarheidsinterval voor de verwachte droogtijd μ wordt gegeven door

a. (61.944 , 70.655) **b.** (61.978 , 70.621) **c.** (64.730 , 67.869)
d. (64.742 , 67.857) **e.** (64.797 , 67.803) **f.** (64.809 , 67.790)

22. Serrano ham wordt in een winkel verkocht in verpakkingen van 100 gram. De manager vreest echter dat het personeel systematisch iets te veel ham in de verpakkingen stopt. Hij meet de hoeveelheid Serrano ham in $n = 12$ verpakkingen. Noem de uitkomsten hiervan $x_1, \ldots, x_{12}$. Gegeven is dat

$$\bar{x}_n = 101.03\, g \quad \text{en} \quad s_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 = 4.00\, g^2.$$

De manager gaat er vanuit dat het gewicht van de ham in een verpakking verdeeld is volgens $N(\mu, \sigma^2)$. We toetsen m.b.v. de t-toets de nul-hypothese $H_0 : \mu = 100$ tegen het alternatief $H_1 : \mu > 100$, bij een significantieniveau van 5%. Noem p de p-waarde van deze toets. Welke van de volgende zes conclusies is de juiste?

a. $0.025 < p < 0.05$, verwerp H_0. **b.** $0.025 < p < 0.05$, verwerp H_0 niet.
c. $0.05 < p < 0.10$, verwerp H_0. **d.** $0.05 < p < 0.10$, verwerp H_0 niet.
e. $p > 0.10$, verwerp H_0. **f.** $p > 0.10$, verwerp H_0 niet.

23. Iemand voert een toets uit, gebruikmakend van een toetsingsgrootheid T, die waarden heeft tussen 0 en 10. Stel dat hij besluit H_0 te verwerpen ten gunste van H_1 als $T \leq c$. Verder weten we dat onder de nulhypothese geldt: $\mathrm{P}(T \leq c) = (c/10)^2$.
Hoe groot (tot op drie decimalen) moet hij c nemen als voor het significantieniveau de waarde $\alpha = 0.10$ gekozen wordt?

a. 1.414 **b.** 2.173 **c.** 3.162 **d.** 4.225 **e.** 6.666 **f.** 8.100

24. Vijf maanden achter elkaar wordt de gemiddelde verkoopprijs van huizen bepaald in Delft. Een makelaar vat deze data op als een bivariate dataset $(x_1, y_1), \ldots, (x_5, y_5)$, waarbij $x_1, \ldots, x_5$ de vijf tijdstippen zijn van meten (hij neemt simpelweg $x_1 = 1, x_2 = 2, \ldots, x_5 = 5$), en $y_1, \ldots, y_5$ de gemiddelde verkoopprijs van huizen in die repectievelijke maanden. Hij gaat uit van een lineair regressie model voor deze data:

$$Y_i = \alpha + \beta x_i + U_i, \quad \text{voor } i = 1, 2, \ldots, 5,$$

waarbij $U_1, U_2, \ldots, U_5$ een steekproef is uit een $N(0, \sigma^2)$ verdeling. Hij vindt als kleinste kwadratenschatter

$$\hat{\beta} = -1500 \quad \text{en} \quad S_b^2 = 4.65 \cdot 10^5.$$

Hij wil laten zien dat de huizenprijs daalt door de nul-hypothese $H_0 : \beta = 0$ te toetsen tegen het alternatief $H_1 : \beta < 0$ met behulp van een t-toets, bij een significantieniveau van 5%. Geef de toetsingsgrootheid T en uw conclusie aangaande de nul-hypothese.

a. $T = -3.5$, verwerp H_0.
b. $T = -4.9$, verwerp H_0.
c. $T = -2.2$, verwerp H_0.
d. $T = -3.5$, verwerp H_0 niet.
e. $T = -4.9$, verwerp H_0 niet.
f. $T = -2.2$, verwerp H_0 niet.

25. Van een mannenkoor in New York wordt elk lid gevraagd zijn lengte in inches op te geven. Het koor bestaat uit $n = 39$ bassen en $m = 20$ tenoren. De lengtes $X_1, X_2, \ldots, X_{39}$ van de bassen worden gemodelleerd met een normale verdeling $N(\mu_1, \sigma_X^2)$. De lengtes $Y_1, Y_2, \ldots, Y_{20}$ van de tenoren worden gemodelleerd met een normale verdeling $N(\mu_2, \sigma_Y^2)$. We willen aantonen dat bassen in verwachting langer zijn dan tenoren.
Als $x_1, \ldots, x_{39}$ de opgegeven lengtes van de bassen zijn, en $y_1, \ldots, y_{20}$ de opgegeven lengtes van de tenoren, dan geldt

$$\bar{x}_n = 70.72 \text{ inch}, S_X^2 = 5.58 \text{ inch}^2, \quad \bar{y}_m = 69.15 \text{ inch}, S_Y^2 = 10.3 \text{ inch}^2.$$

Als nul-hypothese nemen we $H_0 : \mu_1 = \mu_2$. We baseren onze toets op de t-toets toetsingsgrootheid T met niet-gepoolde variantie:

$$T = \frac{\bar{x}_n - \bar{y}_m}{S_d}, \quad \text{met } S_d^2 = \frac{1}{n} S_X^2 + \frac{1}{m} S_Y^2.$$

Geef de alternatieve hypothese H_1 die van toepassing is op dit probleem, en geef uw conclusie aangaande de nulhypothese, als we een significantieniveau van 5% aanhouden.

a. $H_1 : \mu_1 \neq \mu_2$, verwerp H_0.
b. $H_1 : \mu_1 > \mu_2$, verwerp H_0.
c. $H_1 : \mu_1 < \mu_2$, verwerp H_0.
d. $H_1 : \mu_1 \neq \mu_2$, verwerp H_0 niet.
e. $H_1 : \mu_1 > \mu_2$, verwerp H_0 niet.
f. $H_1 : \mu_1 < \mu_2$, verwerp H_0 niet.

Antwoorden multiple choice:

1 a.

2 b.

3 e.

4 a.

5 d.

6 b.

7 e.

8 d.

9 e.

10 f.

11 c.

12 c.

13 a.

14 b.

15 a.

16 e.

17 d.

18 b.

19 b.

20 a.

21 e.

22 d.

23 c.

24 f.

25 b.

Uitwerkingen multiple choice:

1 a. Het aantal boekenkasten met te weinig schroeven kan gemodelleerd worden door een stochast X met een $Bin(n, p)$ met $n = 1000$ en $p = 0.001$. Zodoende is de gevraagde kans gelijk aan

$$\mathrm{P}(X = 2) = \binom{1000}{2} \times (0.001)^2 \times (0.999)^{998} = 0.184.$$

2 b. Noteer X_1 en X_2 als de uitkomsten van dobbelsteen 1 en 2. Dan is

$$\mathrm{P}(M = 6) = \mathrm{P}(\text{minstens één } X_i = 6) = 1 - \mathrm{P}(X_1 \neq 6, X_2 \neq 6) = 1 - \frac{5}{6} \cdot \frac{5}{6} = \frac{11}{36}$$

Dan zijn alleen nog **b** en **e** over als mogelijkheden. Veder is

$$\mathrm{P}(M = 1) = \mathrm{P}(X_1 = 1, X_2 = 1) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36},$$

zodat **b** het juiste antwoord is. Voor de volledigheid:

$$\begin{aligned}
\mathrm{P}(M = 5) &= \mathrm{P}((5,5),(4,5),(3,5),(2,5),(1,5),(5,4),(5,3),(5,2),(5,1)) = \frac{9}{36} \\
\mathrm{P}(M = 4) &= \mathrm{P}((4,4),(3,4),(2,4),(1,4),(4,3),(4,2),(4,1)) = \frac{7}{36} \\
\mathrm{P}(M = 3) &= \mathrm{P}((3,3),(2,3),(3,2),(1,3),(3,1)) = \frac{5}{36} \\
\mathrm{P}(M = 2) &= \mathrm{P}((1,2),(2,1),(2,2)) = \frac{3}{36}.
\end{aligned}$$

3 e.

$$\mathrm{E}[X] = 0 \cdot \frac{1}{4} + 3 \cdot \frac{1}{2} + 4 \cdot \frac{1}{8} + 6 \cdot \frac{1}{8} = 2\frac{3}{4}.$$

4 a. $F(0.5)$ is de oppervlakte onder f op het interval $(-\infty, 0.5]$. Omdat de onderhavige dichtheid een driehoek vormt, is eenvoudig in te zien dat deze oppervlakte gelijk is aan $\frac{1}{2} \times 0.5 \times 0.5 = 0.125$. Een formele berekening levert uiteraard hetzelfde

$$F(0.5) = \int_{-\infty}^{0.5} f(x)\, dx = \int_0^{0.5} x\, dx = \left[\frac{1}{2}x^2\right]_0^{0.5} = 0.125.$$

5 d. De kans dat vriend nummer 1 geen een keer wint, is gelijk aan $p = (1 - \frac{1}{4})^3 = 0.422$. De kans dat precies twee van de drie vrienden geen een keer wint, is dus de binomiale kans

$$\mathrm{P}(Bin(3, p) = 2) = \binom{3}{2} p^2 (1 - p)^1 = 0.309.$$

6 b. Het aantal rondes heeft een Geo(0.25)-verdeling met verwachting $1/0.25 = 4$.

7 e. *Zie opgave 5.11.* De mediaan $q_{0.50}$ van een $Exp(\lambda)$ verdeling is de oplossing van de vergelijking

$$F(q) = 0.5 \Leftrightarrow 1 - \mathrm{e}^{-\lambda q} = 0.5 \Leftrightarrow \mathrm{e}^{-\lambda q} = 0.5 \Leftrightarrow q = \frac{-\ln(0.5)}{\lambda} = \frac{\ln(2)}{\lambda}.$$

Hier geldt dat $q = 2$, dus $\lambda = \ln(2)/q = 0.3466$. De verwachting van een $Exp(\lambda)$ verdeling is gelijk aan $1/\lambda = 2.89$.

8 d. Bekend is dat $\mathrm{E}\left[\bar{X}_n\right] = \mathrm{E}\left[X_1\right] = \lambda$, en $\mathrm{Var}\left(\bar{X}_n\right) = \frac{\mathrm{Var}(X_1)}{n} = \lambda/n$. De schatter is zuiver zodat de bias gelijk is aan 0. Omdat de schatter zuiver is, is de MSE gelijk aan de variantie van de schatter. Het goede antwoord is dus

$$\text{bias} = 0, \quad \text{MSE} = \lambda/n.$$

9 e. *Zie huiswerk opgave 5.12.* De verdelingsfunctie van de $Par(1)$ is $F(x) = 1 - x^{-1}$. Om de mediaan te vinden moeten we oplossen

$$F(x) = \frac{1}{2} \quad \Leftrightarrow \quad 1 - \frac{1}{x} = \frac{1}{2} \quad \Leftrightarrow \quad x = 2.$$

10 f. Definieer $W = S - B$. Dan geldt dat $\mathrm{E}[W] = \mathrm{E}[S] - \mathrm{E}[B] = 10$ kN, en $\mathrm{Var}(W) = \mathrm{Var}(S) + \mathrm{Var}(B) = 6^2 + 8^2 = 100$ (kN)^2, dus W heeft een $N(10, 10^2)$ verdeling. Er geldt dus

$$\mathrm{P}(S < B) = \mathrm{P}(W < 0) = \mathrm{P}\left(\frac{W - 10}{10} < \frac{0 - 10}{10}\right) = \mathrm{P}(Z < -1) = \mathrm{P}(Z > 1) = 0.1587.$$

Hier is Z een standaard normaal verdeelde stochast.

11 c. *Zie huiswerk opgave 7.4.* Volgens de lineariteit van verwachting is

$$\mathrm{E}[3 - 2X] = 3 - 2 \times \mathrm{E}[X] = 3 - 2 \times 2 = -1.$$

Volgens de regel voor de variantie

$$\mathrm{Var}(3 - 2X) = (-2)^2 \times \mathrm{Var}(X) = 4 \times 4 = 16.$$

12 c. Een schatter van de vorm $T = aX_1 + aX_2 + (\frac{1}{2} - a)Y_1$ is zuiver, d.w.z.

$$\mathrm{E}[T] = a\mathrm{E}[X_1] + a\mathrm{E}[X_2] + (\frac{1}{2} - a)\mathrm{E}[Y] = (2a + 2(\frac{1}{2} - a))\mu = \mu \text{ voor elke } \mu \in \mathbb{R}.$$

Verder geldt:

$$\mathrm{Var}(T) = \mathrm{Var}(aX_1) + \mathrm{Var}(aX_2) + \mathrm{Var}\left((\frac{1}{2} - a)Y_1\right) = (2a^2 + (\frac{1}{2} - a)^2)\sigma^2.$$

Dit is minimaal als

$$4a - 2(\frac{1}{2} - a) = 0 \Leftrightarrow a = \frac{1}{6}.$$

Dit laat zien dat het antwoord **c.** moet zijn. U kunt natuurlijk ook alle zes de varianties uitrekenen:

a. $\frac{3}{16}\sigma^2$ **b.** $\frac{1}{4}\sigma^2$ **c.** $\frac{3}{18}\sigma^2$ **d.** $\frac{9}{36}\sigma^2$ **e.** $\frac{1}{2}\sigma^2$ **f.** $\frac{19}{64}\sigma^2$

13 a. *Zie huiswerkopgave 15.4.* De hoogte van het histogram op cel (500, 1000] is

$$\frac{\text{aantal elementen in } (500,\ 1000]}{135 \cdot 500} = 0.00036$$

14 b. De Paretovariabele neemt waarden ≥ 1 aan; dat hoort bij plaatje B.
Plaatje C valt af door de negatieve waarde die daar optreedt.
De Exp(1/3) verdeling neemt in het algemeen grotere waarden aan dan de Exp(3)-verdeling (de verwachting van een Exp(λ)-verdeling is $1/\lambda$), dus het eerste plaatje zal bij de Exp(3) dataset horen. Een andere mogelijkheid is om te kijken naar de empirische mediaan van de verdelingen: deze is te vinden door bij $F(x) = 0.5$ op de verticale as naar rechts te gaan tot de grafiek, en dan naar beneden tot aan de horizontale as.

15 a. *Zie huiswerk opgave 15.5.* De hoogte op de cel $(8, 11]$ is gelijk aan

$$\begin{aligned}\frac{\text{aantal } x_i \text{ in } (8,11]}{(11-8)n} &= \frac{\text{aantal } x_i \leq 11 - \text{aantal } x_i \leq 8}{3n} \\ &= \frac{\text{aantal } x_i \leq 11}{3n} - \frac{\text{aantal } x_i \leq 8}{3n} \\ &= \frac{1}{3}\left(\frac{\text{aantal } x_i \leq 11}{n} - \frac{\text{aantal } x_i \leq 8}{n}\right) \\ &= \frac{1}{3}\left(F_n(11) - F_n(8)\right) = \frac{1}{3}(0.805 - 0.735) = 0.0233.\end{aligned}$$

16 e. We bereken eerst de mediaan: het 5de en 6de getal (van de geordende dataset met 10 punten) is 2 en 2.5, dus de mediaan is 2.25. Dan zijn alleen **a**, **c** en **e** nog over. Vervolgens berekenen we het derde kwartiel: $\frac{3}{4} \cdot 10 = 7.5$ (volgens het boek: $\frac{3}{4} \cdot (10-1) + 1 = 7.75$), en het 7de en 8ste getal zijn beide 4.5, dus het derde kwartiel is 4.5. Dan blijven alleen **a** en **e** over. Tot slot berekenen we de Inter Quartile Range: het onderste kwartiel is 1.75 (de waarde van het tweede, derde en vierde getal), dus de IQR = 2.75. De omhooggaande "whisker" heeft maximale lengte $1.5 \cdot \text{IQR} = 4.125$, en $4.5 + 4.125 > 8$. Aangezien 8 het grootste getal is in de dataset, gaat de whisker tot aan 8. Dit leidt tot antwoord **e**.

17 d. X is het aantal studenten met 'succes', dus $n = 30$. Succes betekent minstens 1 vraag goed, dus is $p = 1 - \text{P}(\text{alle 20 vragen fout}) = 1 - (\frac{5}{6})^{20}$.

18 b. *Zie ook Quick Exercise 22.2.* Er geldt dat $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$. We hebben $\bar{x} = 150/50 = 3$ en $\bar{y} = 400/50 = 8$, dus $\hat{\alpha} = 8 - 2 \cdot 3 = 2$.

19 b. We weten dat X_1 en X_2 een Poisson verdeling hebben met parameter λ. Gegeven is dat $X_1 + X_2 = 1$, dus ofwel $X_1 = 0$ en $X_2 = 1$, ofwel $X_1 = 1$ en $X_2 = 0$. De kans $L(\lambda)$ op de data, gegeven de parameter λ, is dus:

$$L(\lambda) = \text{P}(X_1 = 0) \cdot \text{P}(X_2 = 1) + \text{P}(X_1 = 1) \cdot \text{P}(X_2 = 0) = e^{-\lambda} \cdot \lambda e^{-\lambda} + \lambda e^{-\lambda} \cdot e^{-\lambda} = 2\lambda e^{-2\lambda}.$$

We zoeken de lokatie van het maximum van $L(\lambda)$, en kunnen dit doen door de lokatie van het maximum van

$$l(\lambda) = \log(L(\lambda)) = \log(2) + \log(\lambda) - 2\lambda$$

te vinden. Aangezien

$$l'(\lambda) = 0 \Leftrightarrow \frac{1}{\lambda} - 2 = 0 \Leftrightarrow \lambda = 0.5,$$

vinden we als maximum likelihoodschatter $\lambda = 0.5$.

20 a. De kans $L(p)$ op de data is gegeven door

$$L(p) = \text{P}(\text{Piet wint 1 van de 3 keer}) \cdot \text{P}(\text{Marijke wint de 4de keer}) = 3p(1-p)^2(1-p)^3 p = 3p^2(1-p)^5.$$

Definieer $l(p) = \log(L(p)) = \log(3) + 2\log(p) + 5\log(1-p)$. Vind het maximum van l:

$$l'(p) = 0 \Leftrightarrow \frac{2}{p} - \frac{5}{1-p} = 0 \Leftrightarrow 2(1-p) = 5p \Leftrightarrow p = \frac{2}{7}.$$

Dus $\hat{p} = 2/7$.

21 e. Het 90% betrouwbaarheidsinterval voor de verwachte droogtijd μ wordt gegeven door

$$\left(\bar{x}_n - t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}},\ \bar{x}_n + t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}}\right).$$

Hierbij is $t_{n-1,\alpha/2} = t_{11,0.05} = 1.796$ en $s_n = \sqrt{8.4} = 2.898$, zodat het 90% betrouwbaarheidsinterval wordt gegeven door

$$\left(66.3 - 1.796\frac{2.898}{\sqrt{12}},\ 66.3 + 1.796\frac{2.898}{\sqrt{12}}\right) = (64.797,\ 67.803).$$

22 d. De toetsingsgrootheid wordt gegeven door

$$T = \frac{\bar{x}_n - 100}{s_n/\sqrt{n}} = 1.784.$$

Er geldt dat $t_{11,0.10} < T < t_{11,0.05}$ (merk op dat T een t-verdeling heeft met $n-1 = 11$ vrijheidsgraden!). Aangezien alleen grote waarden van T wijzen op de alternatieve hypothese H_1, geldt dus dat $0.05 < p < 0.10$. Aangezien p groter is dan het significantieniveau, verwerpen we H_0 niet.

23 c. Onder H_0 moet gelden: $\mathrm{P}(T \leq c) \leq 0.1$. De grootste c waarvoor dit het geval is voldoet aan $\frac{c}{10} = \sqrt{0.1}$, en dat is $c = 10\sqrt{0.1} \approx 3.162$.

24 f. De toetsingsgrootheid T wordt gegeven door

$$T = \frac{\hat{\beta} - 0}{S_b} = -2.2.$$

Onder de nul-hypothese geldt dat T een t-verdeling heeft met $n - 2 = 3$ vrijheidsgraden! Aangezien alleen sterk negatieve waarden van T wijzen op H_1, en aangezien $-t_{3,0.05} < -2.2$, geldt dat de p-waarde groter is dan 0.05, en dus verwerpen we H_0 niet.

25 b. We merken op dat de twee varianties behoorlijk verschillen, dus we werken niet met de pooled variance. De toetsingsgrootheid wordt

$$T = \frac{\bar{x}_n - \bar{y}_m}{S_d}, \quad \text{met} \quad S_d^2 = \frac{S_x^2}{n} + \frac{S_y^2}{m}.$$

Uitrekenen geeft $T = 1.93$. Merk op dat alleen grote waarden van T wijzen in de richting van de alternatieve hypothese, dus we verwerpen als $T > t_{n+m-2,0.05}$. Er geldt dat

$$t_{57,0.05} \leq t_{50,0.05} = 1.676,$$

dus $1.93 = T > t_{57,0.05}$, en we verwerpen H_0.