

Tentamen Kansrekening en statistiek
wi2105IN
25 juni 2007, 14.00–17.00 uur

Bij dit examen is het gebruik van een (evt. grafische) rekenmachine toegestaan. Tevens krijgt u een formuleblad uitgereikt—na afloop inleveren alstublieft. Normering: De meerkeuze vragen en de open vragen tellen elk voor de helft van het cijfer. Bij de open vragen telt elk (vraag)onderdeel even zwaar.

Meerkeuzevragen

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 en aan te strepen.

1. Uit een groep van tien LR-, acht BK- en vier IN-studenten wordt door loting een commissie van vier samengesteld om de kwaliteit van het kansrekening- en statistiekonderwijs aan de TU Delft te onderzoeken. Hoe groot is de kans dat in deze commissie alle drie de faculteiten vertegenwoordigd zijn?
a. 0.0173 b. 0.0586 c. 0.1013 d. 0.2078 e. 0.4156 f. 0.6117
2. Uit dezelfde groep studenten wordt een commissie van *negen* leden samengesteld. Alle faculteiten moeten vertegenwoordigd zijn. Hoeveel toegelaten samenstellingen zijn er als we alleen letten op het *aantal* vertegenwoordigers per faculteit?
a. 12 b. 17 c. 18 d. 22 e. 28 f. 98
3. Gegeven zijn twee gebeurtenissen A en B waarvoor $P(B) = 0.5$, $P(A|B) = 0.8$ en $P(B|A) = 0.6$. Dan is $P(A)$:
a. 0.33 b. 0.4 c. 0.5 d. 0.6 e. 0.66 f. er zijn niet genoeg gegevens
4. De regel van Bayes luidt:
a. $P(B|A) = P(A|B)P(B)$ b. $P(B|A) = \frac{P(A \cap B)}{P(B)}$
c. $P(B|A) = \frac{P(A|B)P(B)}{P(A)}$ d. $P(B|A) = \frac{P(A \cap B)}{P(A)}$
e. $P(B|A) = \frac{P(A|B)P(A)}{P(B)}$ f. $P(A \cap B) = P(A)P(B|A)$
5. Stel $X_1, X_2, \dots, X_{15}$ zijn onafhankelijke $N(1, 1)$ variabelen.
 $P(X_1 + X_2 + X_3 + \dots + X_{15} \leq 10)$ is gelijk aan
a. 0.412 b. 0.369 c. 0.126 d. 0.098 e. 0.040 f. 0.012
6. Het aantal ongelukken per week op een bepaalde kruising is een Poisson variabele met parameter $\mu = 3$. Bereken de kans op 3 of meer ongelukken in een bepaalde week als gegeven is dat er ten minste een ongeluk in die week is gebeurd.
a. 0.177 b. 0.298 c. 0.325 d. 0.577 e. 0.607 f. 0.950
7. U is uniform verdeeld over het interval $(0, 1)$. F_Y is de verdelingsfunctie van $Y = U^2$. Dan geldt dat $F_Y(y)$, voor $0 \leq y \leq 1$, gelijk is aan:
a. $\frac{1}{2\sqrt{y}}$ b. $\sqrt{y}$ c. y^2 d. $\frac{1}{y}$ e. $\frac{1}{y^2}$ f. $2\sqrt{y}$

8. De stochastische variabelen $X_1, X_2, \dots, X_7$ zijn onafhankelijk en voor elke i , $1 \leq i \leq 7$, geldt $E[X_i] = 1$ en $\text{Var}(X_i) = 3$. $\bar{X}_7$ is het gemiddelde van $X_1, X_2, \dots, X_7$, dus $\bar{X}_7 = (X_1 + X_2 + \dots + X_7)/7$. Bereken $E[\bar{X}_7]$ en $\text{Var}(\bar{X}_7)$.
- a. $E[\bar{X}_7] = 1$, $\text{Var}(\bar{X}_7) = \frac{3}{7}$. b. $E[\bar{X}_7] = 7$, $\text{Var}(\bar{X}_7) = \frac{3}{49}$.
c. $E[\bar{X}_7] = 1$, $\text{Var}(\bar{X}_7) = \frac{3}{49}$. d. $E[\bar{X}_7] = 7$, $\text{Var}(\bar{X}_7) = 21$.
e. $E[\bar{X}_7] = 1$, $\text{Var}(\bar{X}_7) = 21$. f. $E[\bar{X}_7] = 7$, $\text{Var}(\bar{X}_7) = \frac{3}{7}$.
9. Bij een schattingsprobleem voor een parameter p beschikken we over X_1 die binomiaal verdeeld met parameters $n = 100$ en $p/2$, en X_2 die $\text{Bin}(150, 2p/3)$ verdeeld is. De schatters $T_1 = X_1/50$ en $T_2 = X_2/100$ zijn beide zuiver voor p . De relatieve efficiëntie (relative efficiency) van T_2 ten opzichte van T_1 is
- a. $\frac{(2-p)}{(9-6p)}$ b. $\frac{(9-6p)}{(2-p)}$ c. $\frac{(2-p)}{(2-6p)}$ d. $\frac{(6-3p)}{(3-2p)}$ e. 2 f. 3
10. De likelihood in een aaneenschakeling van 3 experimenten met onbekende parameter p wordt gegeven door:
- $$L(p) = p^2(1-p)p^7(1-p)^{13}p^{11}$$
- Dan is de maximum likelihood schatting voor p :
- a. $p = \frac{3}{10}$ b. $p = \frac{7}{17}$ c. $p = \frac{6}{14}$ d. $p = \frac{7}{10}$ e. $p = \frac{10}{17}$ f. $p = \frac{11}{20}$
11. Voor het toetsen van een nulhypothese H_0 beschikken we over een toetsingsgrootheid T , die een continue verdeling heeft. We willen toetsen met $\alpha = 0.05$ en we weten dat H_0 verworpen moet worden als de realisatie t te ver in de rechterstaart van de verdeling van T ligt. Welke conclusies kunnen bij de volgende twee scenario's worden getrokken?
- scenario I: $t = 2.31$ en $P(T \leq 2.14) = 0.92$
scenario II: $t = 2.15$ en $P(T \geq 2.025) = 0.016$.
- a. in beide scenario's wordt H_0 verworpen
b. in scenario I wordt H_0 niet verworpen, in scenario II wordt H_0 wel verworpen
c. in beide scenario's is de p -waarde kleiner dan 0.05
d. in scenario I is de p -waarde groter dan 0.05, in scenario II kleiner dan 0.05
e. in scenario I is de p -waarde groter dan 0.05, bij scenario II wordt H_0 verworpen
f. in scenario I is de p -waarde kleiner dan 0.10, in scenario II kleiner dan 0.025
12. Voor het uitvoeren van een parametrische bootstrap is het nodig te simuleren aan de hand van de volgende verdelingsfunctie: $F(x) = 1 - e^{-0.05x^2}$ voor $x > 0$ (en $F(x) = 0$ voor $x < 0$). Als U een $U(0, 1)$ verdeelde stochast is dan heeft X verdelingsfunctie F indien
- a. $X = -20 \ln U$ b. $X = -0.05 \ln U$ c. $X = \sqrt{-20 \ln U}$
d. $X = \sqrt{-\ln 20U}$ e. $X = 1 - e^{-0.05U^2}$ f. $X = e^{-0.05U^2}$
13. Voor een bivariate dataverzameling $(x_1, y_1), (x_2, y_2), \dots, (x_{20}, y_{20})$, beschouwen we een lineair regressie model: $Y_i = \alpha + \beta x_i + U_i$ voor $i = 1, 2, \dots, 20$, waarbij $U_1, U_2, \dots, U_{20}$ onafhankelijk zijn met verwachting 0 en variantie σ^2 . Voor de dataverzameling vinden we

$$\hat{\alpha} = 3, \quad \hat{\beta} = 1.5, \quad s_a = 0.5, \quad s_b = 0.25.$$

We toetsen $H_0 : \beta = 1$ tegen $H_1 : \beta > 1$ door middel van een t -toets bij significantieniveau 0.05. Noteer de waarde van de toetsingsgrootheid door t en de rechter kritieke waarde door c_r . Dan geldt

- a. $t = 2$ en $c_r = 1.734$ b. $t = 2$ en $c_r = 2.101$ c. $t = 4$ en $c_r = 1.645$
d. $t = 4$ en $c_r = 2.101$ e. $t = 6$ en $c_r = 1.645$ f. $t = 6$ en $c_r = 1.734$

14. We bekijken de volgende dataset van 50 getallen:

11.4	14.5	16.8	17.8	18.8	19.6	20.8	22.3	22.8	23.0
24.0	25.2	26.0	26.2	26.7	27.4	27.9	28.3	28.4	28.5
28.8	30.4	31.8	33.2	34.2	36.0	36.6	36.6	37.0	37.5
37.6	39.2	40.8	42.9	45.2	48.2	49.8	52.2	52.3	52.5
52.6	53.2	54.0	55.2	59.4	63.5	64.1	64.5	64.8	66.7

Bereken de waarden van de empirische verdelingsfunctie in 28.7 en 72.0.

- a. $F_{50}(28.7) = 20$, $F_{50}(72.0) = 50$
b. $F_{50}(28.7) = 20$, $F_{50}(72.0)$ is niet gedefinieerd
c. $F_{50}(28.7) = 0.2$, $F_{50}(72.0) = 1.0$
d. $F_{50}(28.7) = 0.2$, $F_{50}(72.0)$ is niet gedefinieerd
e. $F_{50}(28.7) = 0.4$, $F_{50}(72.0) = 1.0$
f. $F_{50}(28.7) = 0.4$, $F_{50}(72.0)$ is niet gedefinieerd
15. Een steekproef van omvang 5 uit een normale verdeling met onbekende μ en onbekende σ resulteert in:

129.0 129.4 130.8 131.1 132.2

Het gemiddelde is 130.50 en de standaarddeviatie is 1.3038. Een symmetrisch tweezijdig 90%-betrouwbaarheidsinterval voor μ wordt gegeven door:

- a. $129.11 < \mu < 131.89$ b. $129.26 < \mu < 131.74$ c. $129.33 < \mu < 131.67$
d. $129.54 < \mu < 131.46$ e. $129.61 < \mu < 131.39$ f. $129.75 < \mu < 131.25$

Open vragen

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

1. Een darter werpt een pijl op een cirkelvormig bord met een straal van 1 meter. Het is niet zo'n beste speler: hij raakt het op een uniform over het bord verdeelde lokatie. Noem de afstand tot de roos X . Voor $0 \leq t \leq 1$ geldt $P(X \leq t) = t^2$.
- a. Bepaal $E[X]$.
- b. Hij gooit in totaal drie pijlen. Neem aan dat de worpen onafhankelijk zijn. Beschouw de pijl die het verst van de roos terechtkomt; noem de afstand tot de roos Y . Bepaal de verwachting van Y . Hint: bepaal eerst de verdelingsfunctie.
2. Bedelaars Jan en Piet staan op straat en ontvangen van elke voorbijganger: niets, 20 ct of 50 ct. De simultane verdeling is gegeven door:

P($X = a, Y = b$)		a			
		0	20	50	
b	0	0	0.2	0.1	0.3
	20	0.2	0.1	0.1	0.4
	50	0.1	0.1	0.1	0.3
		0.3	0.4	0.3	1

Verder is gegeven: $E[X] = 23$ en $E[X^2] = 910$.

- a. Bepaal $\rho(X, Y)$, de correlatie tussen X en Y .
- b. Bepaal $\text{Var}\left(\frac{X+Y}{2}\right)$.

- c. Er komen 10 voorbijgangers langs. Bepaal de kans dat Jan in totaal exact 200 ct ontvangt. Je mag aannemen dat de bedragen van verschillende voorbijgangers onafhankelijk zijn.
3. Bij de periodieke keuring van een paraglidingscherm wordt van het doek op een aantal plaatsen de porositeit (luchtdoorlaatbaarheid) gemeten. Cirkelvormige stukjes doek worden met een speciaal apparaatje getest op doorlaatbaarheid, door een zekere overdruk aan te brengen en te meten hoe lang het duurt om een bepaalde hoeveelheid lucht door te laten. Als dit te kort duurt is de porositeit te hoog en het scherm niet veilig meer. Bij nieuw doek duurt het ongeveer 2 minuten; de keuringsnorm is: langer dan 80 seconden. Omdat schermstijtage niet gelijkmatig is zijn 23 metingen gedaan die we opvatten als een steekproef aan een onbekende verdeling, waarvan de verwachting μ aan de eis zou moeten voldoen om het scherm goed te keuren. Hieronder de 23 metingen in seconden, naar grootte geordend:

10	47	66	72	80	96	97	98
105	107	108	109	112	113	113	116
116	116	117	117	117	117	117	119

Het gemiddelde is 98.6; de standaardafwijking 27.2. We vatten de keuring op als een toetsingsprobleem met significantieniveau 0.01, en kiezen: $H_0 : \mu = 80$ (het scherm is net niet goed genoeg) en $H_1 : \mu > 80$.

- a. Ga even uit van de veronderstelling dat de metingen afkomstig zijn van een normale verdeling. Voer de toepasselijke t -toets uit en bepaal ook de p -waarde. Wat is uw (keurings)conclusie?
- b. Omdat we niet op de normaliteitsaannname willen vertrouwen hebben we een gestudentiseerde bootstrap uitgevoerd met 10 000 bootstrapsteekproeven. Een deel van de gesorteerde t^* -waarden is hieronder gegeven. Bepaal de bootstrapschatting van de p -waarde. Wat is nu uw (keurings)conclusie?

i	$t_{(i)}^*$	i	$t_{(i)}^*$	i	$t_{(i)}^*$	i	$t_{(i)}^*$
10	-2.55	200	-1.71	8000	1.09	9900	4.93
20	-2.31	400	-1.48	8200	1.22	9910	5.09
30	-2.27	600	-1.31	8400	1.36	9920	5.34
40	-2.22	800	-1.19	8600	1.52	9930	5.49
50	-2.15	1000	-1.10	8800	1.69	9940	5.66
60	-2.11	1200	-1.00	9000	1.91	9950	6.11
70	-2.04	1400	-0.93	9200	2.21	9960	6.41
80	-2.00	1600	-0.86	9400	2.57	9970	7.23
90	-1.96	1800	-0.79	9600	3.04	9980	8.89
100	-1.94	2000	-0.73	9800	3.96	9990	10.08

- c. Welke van de twee methoden zou gebruikt moeten worden? Motiveer uw antwoord.

Antwoorden multiple choice:

1 e. Een van de drie faculteiten kan met 2 studenten vertegenwoordigd worden, de andere twee met één. Het aantal commissiesamenstellingen met 2 LR, 1 BK en 1 IN student is $\binom{10}{2} \cdot 8 \cdot 4 = 1440$; met 2 BK is het $10 \cdot \binom{8}{2} \cdot 4 = 1120$ en met 2 IN $10 \cdot 8 \cdot \binom{4}{2} = 480$. Het totaal hiervan is 3040. Er zijn dus 3040 toegelaten samenstellingen van in totaal $\binom{22}{4} = 7315$ mogelijkheden, ofwel een kans van $32/77 \approx 0.4156$.

2 d. Dit lossen we op met genererende functies, het feit dat het aantal BK en IN studenten beperkt is (kleiner dan 9) maakt andere oplossingen ingewikkeld(er). Het aantal komt overeen met de coefficient van x^9 in de volgende uitdrukking

$$(x + x^2 + \dots + x^{10})(x + x^2 + \dots + x^8)(x + x^2 + \dots + x^4)$$

ofwel die van x^6 uit

$$(1 + x + \dots + x^9)(1 + x + x^2 + \dots + x^7)(1 + x + \dots + x^3).$$

Herschrijven met behulp van $1 + x + \dots + x^n = (1 - x^{n+1})/(1 - x)$ geeft

$$(1 - x^{10})(1 - x^8)(1 - x^4)(1 - x)^{-3} = (1 - x^4 - x^8 - x^{10} - \dots)(1 - x)^{-3}$$

Nu geldt $(1 - x)^{-3} = \sum_{i=0}^{\infty} \binom{i+2}{i} x^i$, en het is duidelijk dat alleen de termen $i = 6$ en $i = 2$ van belang zijn:

$$(1 - x^4 - x^8 - x^{10} - \dots)(1 + \binom{3}{1}x + \binom{4}{2}x^2 + \dots + \binom{8}{6}x^6 + \dots$$

want geen enkele andere term kan uitkomen op x^6 . We vinden dus dat het gevraagde aantal gelijk is aan $\binom{8}{6} - \binom{4}{2} = 22$.

3 e. Schrijf de twee conditionele kansen uit en vul $P(B)$ in. Uit de resulterende vergelijking is $P(A)$ makkelijk op te lossen.

4 c. Zie boek, Chapter 3.

5 d. $P(X_1 + X_2 + X_3 + \dots + X_{15} \leq 10) = P(Z < 10 - 15/\sqrt{15}) = P(Z < -1.29) \approx 0.098$.

6 e. Noem de Poisson variabele X ; gevraagd wordt $P(X \geq 3 | X \geq 1)$. Deze kans is met de definitie:

$$P(X \geq 3 | X \geq 1) = \frac{P(X \geq 3)}{P(X \geq 1)} = \frac{0.5768}{0.9502} = 0.607.$$

Immers $P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-3} = 0.9502$ en $P(X \geq 3) = 1 - P(X = 0) - P(X = 1) - P(X = 2) = 0.5768$.

7 b. Het juiste antwoord is $\sqrt{y}$, immers $P(Y \leq y) = P(U^2 \leq y) = P(0 \leq U \leq \sqrt{y}) = \sqrt{y}$, voor $0 \leq y \leq 1$.

8 a. $E[\bar{X}_7] = E[X_1] = 1$ en $\text{Var}(\bar{X}_7) = \text{Var}(X_1)/7 = 3/7$. Zie Chapter 13.

9 d.

$$\text{Var}(T_1) = \frac{\text{Var}(X_1)}{2500} = \frac{100 \cdot \frac{1}{2}p \cdot (1 - \frac{1}{2}p)}{2500} = \dots = \frac{p(2-p)}{100}.$$

Analoog:

$$\text{Var}(T_2) = \frac{\text{Var}(X_2)}{10000} = \frac{150 \cdot \frac{2}{3}p \cdot (1 - \frac{2}{3}p)}{10000} = \frac{p(3-2p)}{300}$$

Dus

$$\frac{\text{Var}(T_1)}{\text{Var}(T_2)} = \frac{p(2-p)/100}{p(3-2p)/300} = \frac{6-3p}{3-2p}.$$

10 e. $L(p) = p^{20}(1-p)^{14}$, zodat de loglikelihood gegeven wordt door:

$$l(p) = \ln(L(p)) = 20 \ln p + 14 \ln(1-p).$$

Afgeleiden naar p nemen levert

$$l'(p) = \frac{20}{p} - \frac{14}{1-p}, \quad l''(p) = -\frac{20}{p^2} - \frac{14}{(1-p)^2} < 0,$$

We kunnen dus $l'(p) = 0$ stellen en vinden zo $p = \frac{10}{17}$ als ML schatting.

11 f. Er wordt verworpen als t te ver in de rechterstaart ligt. De p -waarde is dus $P(T \geq t)$. In het eerste scenario geldt $P(T \geq 2.14) = 0.08$, dus $P(T \geq t) = P(T \geq 2.31) \leq 0.08$. Van de genoemde conclusies is alleen de bewering dat de p -waarde kleiner is dan 0.10 zeker. Bij het tweede scenario kan geconcludeerd worden dat $P(T \leq t) \leq 0.016$, en dat is in ieder geval < 0.025 (ook wordt de nulhypothese verworpen, maar dat is voor het antwoord niet meer relevant).

12 c. Los op naar x : $F(x) = u$ voor $0 \leq u \leq 1$, dan vind je $x = \sqrt{-20 \ln u}$. Zie verder paragraaf 6.2.

13 a. De waarde van de toetsingsgrootte is $t_b = (1.5 - 1)/0.25 = 2$. Onder de nulhypothese heeft de toetsingsgrootte een $t(18)$ verdeling. We verwerpen alleen voor grote waarden van T_b , dus de rechter kritieke waarde is $t_{18,0.05} = 1.734$.

$$\mathbf{14 \ e.} \quad F_{50}(28.7) = \frac{\text{aantal datapunten} \leq 28.7}{50} = \frac{20}{50} = 0.4.$$

$$\text{Evenzo } F_{50}(72) = \frac{50}{50} = 1.$$

15 b. De grenzen van het betrouwbaarheidsinterval worden gegeven door $\bar{x}_n \pm t_{n-1, \alpha/2} s_n / \sqrt{n}$, met $\alpha = 0.1$, en $n = 5$. Dit geeft $130.5 \pm 2.132 \cdot 1.3038 / \sqrt{5}$.

Antwoorden open vragen:

1a De kansdichtheid van X is $f(t) = 2t$ voor $0 \leq t \leq 1$, en $f(t) = 0$ elders. Er volgt $E[X] = \int_0^1 t \cdot 2t dt = \dots = \frac{2}{3}$.

1b Er geldt $Y = \max\{X_1, X_2, X_3\}$ waarbij X_1, X_2, X_3 onafhankelijk zijn, en verdeeld als de gegeven X . Er volgt (zie paragraaf 8.4): $F_Y(t) = t^6$ voor $0 \leq t \leq 1$. De kansdichtheid van Y is $f(t) = 6t^5$ voor $0 \leq t \leq 1$, en $f(t) = 0$ elders. Er volgt $E[Y] = \int_0^1 t \cdot 6t^5 dt = \dots = 6/7$.

2a Merk op dat de verdeling symmetrisch is. Uit de gegevens volgt $\text{Var}(X) = \text{Var}(Y) = 910 - (23)^2 = 381$. Voorts

$$E[XY] = (20 \cdot 20) \cdot 0.1 + (20 \cdot 50) \cdot 0.1 + (50 \cdot 20) \cdot 0.1 + (50 \cdot 50) \cdot 0.1 = 490,$$

en dus

$$\text{Cov}(X, Y) = E[XY] - E[X] \cdot E[Y] = 490 - (23)^2 = -39,$$

waaruit volgt: $\rho(X, Y) = -39/381 = -0.102$.

2b We hadden gevonden $\text{Cov}(X, Y) = -39$. Dus geldt

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) = 381 + 381 - 2 \cdot 39 = 684$$

en

$$\text{Var}\left(\frac{X + Y}{2}\right) = \frac{1}{4} \cdot \text{Var}(X + Y) = \frac{1}{4} \cdot 684 = 171.$$

2c De uitkomst 200 kan op drie manieren verkregen worden: $200 = 4 \times 50 + 0 \times 20 + 6 \times 0$, $200 = 2 \times 50 + 5 \times 20 + 3 \times 0$ en $200 = 0 \times 50 + 10 \times 20 + 0 \times 0$. Voor de eerste combinatie vinden we als kans

$$\binom{10}{4, 0, 6} 0.3^4 \cdot 0.4^0 + 0.3^6 = \binom{10}{4} 0.3^{10} = 0.00124;$$

de andere twee kansen komen uit op 0.06271 en 0.000104. De som is 0.06405.

3a We voeren de t -toets uit. De toetsingsgrootte is $t = (98.6 - 80)/(27.2/\sqrt{23}) = 3.28$. De p -waarde volgt nu uit de $t(22)$ -verdeling. We vinden in de tabel: $t_{22, 0.0025} = 3.119$ en $t_{22, 0.001} = 3.505$. De p -waarde ligt tussen 0.001 en 0.0025. Conclusie: verwerp H_0 , het scherm is goed.

3b Nu zoeken we in deze tabel en vinden $t_{(9600)}^* = 3.04 < 3.28 < 3.96 = t_{(9800)}^*$, er zijn dus tussen de 200 en de 400 waarden groter dan de waargenomen t -waarde. Vertaald in p -waarde: deze moet tussen 0.02 en 0.04 inliggen (interpolerend in de tabel zou ik 0.035 schatten; de werkelijke waarde is 0.033). Conclusie: H_0 niet verwerpen, vertrouw het doek maar niet.

3c De bootstrapmethode. Om de p -waarde te bepalen onder H_0 moeten de verdeling van het gestudentiseerde gemiddelde kennen. Wanneer de data afkomstig is van een normale verdeling vinden we de $t(n - 1)$ verdeling. Als n groot is geldt op grond van de centrale limietstelling dat de verdeling ongeveer standaard normaal is. Blijft over de categorie: niet al te grote n , geen normale verdeling. Hier is de bootstrap de juiste aanpak. Uit de boxplot zien we dat de verdeling erg scheef is. Die scheefheid maakt dat het gestudentiseerde gemiddelde een flinke staart naar rechts heeft, waardoor de p -waarde hoger uitkomt dan die volgt uit de $t(22)$ -verdeling.