

Multiple choice vragen

1. De NS meet dagelijks de vertrektijd van de eerste intercity naar Leiden vanuit Amsterdam. Wat voor type data betreft dit?
 - A. Interval data
 - B. Nominale data
 - C. Ordinale data
 - D. Ratio data

2. Een grote stationsketen start twee projecten. Project A: de ontwikkeling van een applicatie om mobiel te betalen. Project B: het vervangen van de traditionele spaarkaart door een mobiele app. De kans op een succesvolle introductie van project A is 20%. De kans op een succesvolle introductie van project B is 30%. De kans dat er tenminste een van beide projecten slaagt is 40%. Wat is de kans dat beide projecten slagen?
 - A. 0,06
 - B. 0,10
 - C. 0,66
 - D. 0,90

3. Als barman moet je de juiste hoeveelheid bier in een glas kunnen tappen. De juiste hoeveelheid bier in een glas betreft 0,3 liter. Een brouwerij voert een kwaliteitscontrole uit in een bar in Groningen. Zij nemen een steekproef van 41 glazen bier en berekenen een steekproef gemiddelde van 0,28 liter en een steekproef standaarddeviatie van 0,05 liter. Zij berekenen een p-waarde om te testen of de gemiddelde hoeveelheid getapt bier in een glas verschilt van 0,3 liter. Welke p-waarde is is juist?
 - A. Minder dan 0,5%.
 - B. Tussen 0,5% en 1%.
 - C. Tussen 1% en 2%.
 - D. Meer dan 2%.

4. Een parfumerie verkoopt per maand gemiddeld vijftig flessen parfum van het merk Hermés. De maandelijkse verkopen zijn normaal verdeeld. Om de benodigde voorraad parfum flessen te bepalen willen zij een schatting van de populatie standaarddeviatie, σ , van de maandelijkse verkopen. De manager verkrijgt de maandelijkse verkopen van dit item over de afgelopen 10 maanden. De steekproef variantie, s^2 , is 17. Wat is het 95% betrouwbaarheidsinterval van de populatie standaarddeviatie van het aantal per maand verkochte Hermés parfum flessen?
 - A. [1,951; 13,744]
 - B. [2,836; 7,528]
 - C. [8,043; 56,667]
 - D. [9,043; 46,015]

5. Welke van de volgende twee beweringen over berekende steekproef statistieken is juist?
 - Bewering 1: Het verschil tussen de mediaan en het eerste kwartiel is net zo groot als de interkwartielafstand.
 - Bewering 2: het eerste kwartiel is groter dan de modus.
 - A. Beide beweringen zijn juist.
 - B. Beide beweringen zijn onjuist.
 - C. Alleen de eerste bewering is juist.
 - D. Alleen de tweede bewering is juist.

6. Een bedrijf koopt Smart-TVs die binnen de garantietermijn ter reparatie worden aangeboden. Hiervan kan 85% worden gerepareerd, terwijl de overige 15% wordt vervangen door een nieuw exemplaar. Als een bedrijf 15 van deze Smart-TVs opkoopt, wat is de kans dat er ten minste 4 worden vervangen door een nieuw exemplaar?
 - A. 0,062
 - B. 0,177
 - C. 0,823
 - D. 0,938

7. Je schrijft een business plan om een nieuwe internet provider te starten. Als input heb jij een schatting nodig van het gemiddelde aantal minuten internet gebruik per maand van een huishouden. Omdat het kostbaar is om deze informatie te verkrijgen besluit jij om het benodigde aantal observaties te bepalen. De informatie is bruikbaar ter onderbouwing van het plan als jij een 95%-betrouwbaarheidsinterval van het populatie gemiddelde hebt met een maximale breedte van 3 minuten. Neem aan dat het wekelijkse internet gebruik normaal verdeeld is met een populatie variatie van 40 minuten². Hoeveel huishoudens moeten minimaal worden ondervraagd?

- A. 17
- B. 18
- C. 68
- D. 69

8. De AIVD evalueert een nieuw systeem om leugens te detecteren in de context van een veiligheidsonderzoek. Testen laten zien dat het systeem in 90% van de gevallen waarin een persoon liegt dit juist detecteert. Verder geeft het systeem in 95% van de gevallen waarin een persoon de waarheid spreekt dit correct aan. Neem aan dat 5% van alle geteste personen liegt.

Hoe groot is de kans dat een persoon die door het systeem wordt aangemerkt als een leugenaar daadwerkelijk liegt (afgerond op 3 decimalen)?

- A. 0,045
- B. 0,050
- C. 0,486
- D. 0,900

9. De gezamenlijke kansverdeling van de prijzen van twee aandelen van bedrijf X en bedrijf Y zijn weergegeven in de onderstaande tabel.

Gezamenlijke kansverdeling van de prijzen van Aandeel X en Aandeel Y

Prijs aandeel X	Prijs aandeel Y:	
	40	50
30	0,35	0,2
60	0,15	0,3

Een investeerder besluit twee aandelen van bedrijf X en 3 aandelen van bedrijf Y te kopen. Bereken de covariantie van X en Y (σ_{xy}). Bereken daarnaast de verwachte waarde van het portfolio $W = 2X + 3Y$ (μ_w).

- A. $\sigma_{xy} = 22,5$ and $\mu_w = 220,5$
- B. $\sigma_{xy} = 22,5$ and $\mu_w = 222$
- C. $\sigma_{xy} = 36$ and $\mu_w = 220,5$
- D. $\sigma_{xy} = 36$ and $\mu_w = 222$

10. We hebben de volgende data: 3, 5, 6, 9, 0, 4, 1, 3, 7, 10, 8. Bereken de volgende beschrijvende statistieken:

- Modus (Mode)
- Interkwartielafstand (Interquartile range IQR)

- A. De modus is 3 en de IQR is 5.
- B. De modus is 3 en de IQR is 10.
- C. De modus is 4 en de IQR is 10.
- D. De modus is 5 en de IQR is 5.

11. Onderzoekers willen het oorzakelijke verband bepalen van een gematigde consumptie van alcohol op de gezondheid van mensen. Na het verzamelen van de gegevens hebben zij informatie over de gezondheid, dagelijks alcohol gebruik, leeftijd en het opleidingsniveau van mensen. Zij hebben geen informatie over andere variabelen zoals de leefstijl van mensen. Welke van de onderstaande stellingen is juist?

- A. Een matched pairs design corrigeert voor de confounders (verstorende factoren) leeftijd en scholing.
- B. Een randomized experimental design corrigeert niet voor de confounder (verstorende factor) leefstijl.
- C. In een matched pairs design krijgt ieder deelnemend paar dezelfde treatment.
- D. Een observational design corrigeert voor de confounder (verstorende factor) leefstijl.

Open vragen

1. Jan verkoopt hamburgers in Amsterdam. Hij heeft de prijs van de hamburgers vanaf de start van zijn populaire restaurant niet veranderd. Hij ondervindt vaak problemen met de bevoorrading waardoor er een tekort aan hamburgers ontstaat. Hij vermoedt dat de vraag op een dag afhangt van de maximum temperatuur op dezelfde dag. Tabel 2 toont een willekeurige steekproef van het aantal verkochte hamburgers per dag en de maximum temperatuur op dezelfde dag.

Vraag	Temperatuur °C
110	25
92	22
81	23
45	11
73	17
103	28
60	13
53	14
139	27

- a. Bepaal de afhankelijke en onafhankelijke variabele en licht dit kort toe.
- b. Schat met behulp van de kleinste kwadratenmethode (least squares method) een lineaire regressielijn van de temperatuur en dagelijkse vraag naar hotdogs. Rapporteer de geschatte regressielijn en toon aan dat de hellingscoëfficiënt gelijk is aan 4,331. Om tijd te besparen, is de covariantie van de twee variabelen al berekent: $s_{xy} = 176,5$.
- c. Interpreteer de geschatte hellingcoëfficiënt b_1
- d. Bereken de geschatte variantie $s_{b_1}^2$ van de geschatte hellingscoëfficiënt b_1 . Maak gebruik van het feit dat de som van gekwadraterde fouten, SSE, gelijk is aan 1198,22.
- e. Test op het 5%-significantieniveau, tegen het tweezijdige alternatief, de null hypothese dat de hellingscoëfficiënt van de populatie regressielijn gelijk is aan 0. Doorloop alle relevante stappen bij het testen van een hypothese.
- f. Jan's werknemers hebben een 95% betrouwbaarheidsinterval van de hellingcoëfficiënt opgesteld waarbij zij uitgaan van de geschatte hellingcoëfficiënt van 4,331. Slechts een van de onderstaande intervallen is correct berekend. Leg uit zonder het interval te berekenen waarom de overige drie intervallen onjuist zijn.

Arend: [-1.424; 5.161]

Bas: [3.327; 7.836]

Christine: [-0.925; 9.587]

Daan: [2.618; 6.045]

13. Een fondsmanager heeft twee fondsen in zijn beheer. Pieter wil in een van beide fondsen investeren. Het gemiddelde rendement van beide fondsen is ongeveer gelijk aan elkaar en normaal verdeeld. Pieter is risicomijdend en besluit in het fonds met de laagste variantie te investeren. De fondsmanager geeft Pieter een willekeurige steekproef van 12 observaties betreffende het rendement van Fonds A, met een standaarddeviatie van 10,3%, en een willekeurige steekproef van 17 observaties betreffende het rendement van Fonds B, met een standaarddeviatie van 25,2%.

Is er voldoende statistisch bewijs om te veronderstellen dat de populatie varianties van het rendement van de twee fondsen van elkaar verschilt? Gebruik een significantieniveau van 10% om dit te toetsen.

Antwoorden MC vragen

1. A
2. B
3. C
4. B
5. B
6. B
7. D
8. C
9. B
10. A
11. A

Antwoorden open vragen

12.a. X: temperatuur (onafhankelijke variabele). y: vraag naar hamburgers (afhankelijke variabele). (2 punten). Temperatuur is de onafhankelijke variabele omdat dit de vraag naar hamburgers kan beïnvloeden (andersom is dit niet mogelijk). (2 punten)

12.b

De gemiddelde vraag is:

$$\hat{y} = \frac{1}{9} \sum_{i=1}^9 y_i = \frac{110+92+81+45+73+103+60+53+139}{9} = 84 \quad (1 \text{ punt})$$

De gemiddelde temperatuur is:

$$\hat{x} = \frac{1}{9} \sum_{i=1}^9 x_i = \frac{25+22+23+11+17+28+13+14+27}{9} = 20 \quad (1 \text{ punt})$$

De steekproefvariantie van de temperatuur, s_x^2 , kan berekend worden door:

$$s_x^2 = \frac{1}{9-1} \left(\sum_{i=1}^9 x_i^2 - 9 \hat{x}^2 \right) = \frac{1}{8} (25^2 + \dots + 27^2 - 9 \cdot 20^2) = \frac{1}{8} (3926 - 3600) = 40,75 \quad (2 \text{ punten})$$

De helling van de kleinste kwadratenlijn is $b_1 = S_{xy} / s_x^2 = 176,5 / 40,75 \approx 4,331$. (2 punten)

Het snijpunt met de y-as is $b_0 = \bar{y} - b_1 \bar{x} = 84 - 4,331 \cdot 20 \approx -2,626$. (2 punten)

Het geschatte lineaire verband is dus $\hat{y} = -2,626 + 4,331x$. (2 punten)

12.c Als de temperatuur met een graad Celsius toeneemt, gaat de vraag naar hamburgers gemiddeld/verwacht 4,331 hamburgers omhoog. (3 punten, 2 punten aftrek als gemiddeld of verwacht of iets dergelijks mist).

$$12.d \quad s_e^2 = \frac{SSE}{n-2} = \frac{1198,22}{7} \approx 171,174 \quad (2 \text{ punten})$$

Aan de hand daarvan kun je s_{b1} berekenen door

$$s_{b1} = \frac{s_e}{\sqrt{(n-1)s_x^2}} = \frac{\sqrt{171,174}}{\sqrt{8 \times 40,75}} \approx 0,7246 \quad s_{b1}^2 = 0,7246^2 = 0,0525 \quad (2 \text{ punten})$$

12.e (11 punten)

We willen testen of $\beta_1 \neq 0$

Hypothese

$$H_0 \beta_1 = 0$$

$$H_1 \beta_1 \neq 0$$

Significatieniveau is $\alpha = 0,05$

Toetsingsgrootheid. Test uitvoeren: $t = \frac{b_1 - \beta_1}{s_{b1}}$ Dan vind je $t = 4,331/0,7246 = 5,977$

Distributie. Wanneer de nulhypothese waar is volgt deze test statistiek een t verdeling met $n-2 = 7$ vrijheidsgraden.

Verwerpsgebied bepalen. We verwerpen H_0 wanneer $t < -t_{\alpha/2, v} = -t_{0,025, 7} = -2,365$

Of wanneer: $t > t_{\alpha/2, v} = t_{0,025, 7} = 2,365$

Conclusie. Omdat $t = 5,977$ is en dit in het verwerpsgebied ligt is er voldoende statistisch bewijs om de nulhypothese te verwerpen. Er is dus voldoende statistisch bewijs dat er een lineair verband is tussen de vraag naar hamburgers en de temperatuur.

12.f Uit het antwoord bij c) blijkt dat de hellingscoëfficiënt anders is dan 0 (significant is) op een 5% significantieniveau, daarom kan 0 niet in het 95% betrouwbaarheidsinterval zitten. Daarom is het interval van Christine onjuist (2 punten). Verder moet het interval symmetrisch zijn rond de geschatte hellingscoëfficiënt, dus rond 4,331. Bas' interval is dit duidelijk niet, daarom is dit interval onjuist (2 punten). Arends interval combineert de fouten van C en B (non-symmetrisch en 0 zit in het interval) en kan daarom ook niet juist zijn (2 punten, een van deze twee fouten opnoemen voor uitsluiting van deze mogelijkheid is voldoende). Daan's interval is symmetrisch rond 4,331 en het bevat geen 0, daarom moet het het juiste zijn. (2 punten, ook zonder redenering)

13. (10 punten)

s_A^2 : steekproefvariantie rendement fonds A;

s_B^2 : steekproefvariantie rendement fonds B;

nA : aantal observaties fonds A;

nB : aantal observaties fonds B.

Hypotheses

De hypothesen zijn als volgt: $H_0: \frac{\sigma_A^2}{\sigma_B^2} = 1$ $H_1: \frac{\sigma_A^2}{\sigma_B^2} \neq 1$

Significatieniveau

Het significantieniveau is $\sigma = 0,10$

Test statistiek

Omdat $s_B^2 > s_A^2$ is de teststatistiek als volgt: $F = \frac{s_B^2}{s_A^2} = \frac{25,2^2}{10,3^2} \approx 5,986$

De verdeling van de teststatistiek onder de nulhypothese

Onder de nulhypothese is de test-statistiek F -verdeeld met $v_B = n_B - 1 = 17 - 1 = 16$ vrijheidsgraden in de teller en $v_A = n_A - 1 = 12 - 1 = 11$ vrijheidsgraden in de noemer.

Verwerpingsgebied

De nulhypothese wordt verworpen als $F > F_{v_B, v_A, \alpha/2} = F_{16, 11, 0,05} = 2,701$

Statistische conclusies

Omdat de waarde van de test-statistiek in het verwerpingsgebied ligt, is er voldoende statistisch bewijs om de nulhypothese te verwerpen.

Conclusie in woorden

Er is voldoende bewijs om aan te nemen dat de varianties van de rendementen gemaakt door fonds A en fonds B verschillen.