

Oefeningen beschrijvende statistiek

Opgave 1

Flairs grote ontrouwenquête: 33% van de jonge vrouwen heeft ooit haar lief bedrogen

43% van die vrouwen vindt dat hun partner het overspel had kunnen voorkomen.

(<https://www.flair.be/nl/lifestyle/flairs-grote-ontrouwenquete-33-van-de-jonge-vrouwen-heeft-ooit-haar-lief-bedrogen/>)

Een scheve schaats rijden, naast de pot plassen, een slippertje maken: er zijn tientallen synoniemen voor vreemdgaan en opvallend veel mensen die ermee in aanraking komen. Dat blijkt tenminste uit onze grote ontrouwenquête die op dinsdag 28 november in Flair verschijnt.

33% is ontrouw

We vonden 2231 lezeressen tussen 18 en 35 jaar bereid onze ontrouwenquête anoniem in te vullen. Ze beantwoordden vragen over hun eigen ontrouw, wat ze verstaan onder overspel, de oorzaken van ontrouw, verzachtende omstandigheden, enzovoort.

” Slechts 50% van de vrouwen durft zeggen dat ze zelf nooit een scheve schaats zou rijden.

Een van de opvallendste resultaten is dat maar liefst 33% van de jonge vrouwen – ofwel een op drie – toegeeft ooit ontrouw te zijn geweest. 27% van die dames ging all the way, de anderen hebben het over tongzoenen, een online affaire of ongepast gefriemel onder de gordel. Een nog groter percentage, zo'n 55%, zegt te weten of sterk te vermoeden dat ze ooit bedrogen werden. Slechts de helft van de ondervraagde vrouwen durft zeggen dat ze zelf nooit, never, jamais de fout zou maken.

Bespreek populatie en steekproef. Is deze steekproef representatief en betrouwbaar?

Opgave 2

TOM MICHIELSEN | 26 januari 2020 22:28

De Vlaming is over het algemeen tevreden over de VRT, en vindt het erg belangrijk dat de openbare omroep zorgt voor informatie, educatie, cultuur en ontspanning.

Dat blijkt uit een bevraging van de Strategische Adviesraad voor Cultuur (SARC) ter voorbereiding van een advies over de nieuwe beheersovereenkomst, dat er later dit jaar aankomt.

De Strategische Adviesraad voor Cultuur, Jeugd, Sport en Media liet onderzoeksbureau Indiville bijna 1.900 Vlamingen van 16 of ouder ondervragen over het belang van de VRT. Dat gebeurde ook al in 2010 en 2015, en de conclusies van toen blijven ook nu gelden: de Vlaming is over het algemeen tevreden over de openbare omroep. Meer dan de helft (53 procent) geeft de VRT een score van 8 op 10 of hoger, een kwart geeft een 6 of 7 op tien. De overige 17 procent geeft de VRT een score tussen 0 en 5 op tien. Gemiddeld haalt de VRT zo 7,2 op 10.

<https://www.tijd.be/cultuur/televisie/vlaming-blijft-tevreden-over-vrt/10202690.html>
Bepreek populatie, steekproef en onderzochte variabele.

Uit het onderzoeksrapport halen we o.a. volgende informatie:

3.3. Populatie, steekproef en weging

De onderzoekspopulatie bestond uit Nederlandstalige inwoners van Vlaanderen en Brussel van 16 jaar of ouder. Tussen 4 november en 1 december 2019 namen 2081 Vlamingen van 16 jaar of ouder deel aan het onderzoek. De deelnemers aan de online enquête werden gerekruteerd via de onlinepanels van Bpact en Dynata. De deelnemers aan de telefonische enquête werden gerekruteerd via 'random dialing', waarbij ook mobiele nummers werden gecontacteerd. Na data cleaning werden 2070 ingevulde vragenlijsten weerhouden. 1859 respondenten vulden de vragenlijst online in, 211 namen telefonisch deel.

Vooraf werden quota vooropgesteld voor leeftijd x geslacht (hard), provincie (hard), opleiding (zacht) en beroep (zacht), waarbij werd gestreefd naar 90% – 105% van de vooropgestelde quota voor de harde quota en naar 70% - 115% van de vooropgestelde aantallen voor de zachte quota, zodat de weegfactoren niet te groot worden. Deze quota werden gebaseerd op de officiële statistieken van de FOD Economie/Statbel. De steekproef werd vervolgens gewogen voor elk van deze variabelen. Tabel 3.1 geeft de percentageverdelingen weer voor de steekproef van de bevraging voor en na weging.

De samenstelling van en kwaliteitscontrole op de surveypanels en de random dialing methode, gecombineerd met het vooropstellen van quota op de belangrijkste sociodemografische variabelen en de weging van de resultaten, samen met de omvang van de steekproef (N=2070), laten toe om de resultaten van de studie te veralgemenen naar de Vlaamse bevolking van 16 jaar en ouder. De maximale foutenmarge bedraagt + of – 2.15% bij een betrouwbaarheid van 95%.

De samenstelling van en kwaliteitscontrole op de surveypanels en de random dialing methode, gecombineerd met het vooropstellen van quota op de belangrijkste sociodemografische variabelen en de weging van de resultaten, samen met de omvang van de steekproef (N=2070), laten toe om de resultaten van de studie te veralgemenen naar de Vlaamse bevolking van 16 jaar en ouder. De maximale foutenmarge bedraagt + of – 2.15% bij een betrouwbaarheid van 95%.

Bespreek populatie en steekproef. Is deze steekproef representatief en betrouwbaar?

Opgave 3:

Ga op zoek op het internet of in kranten en weekbladen naar een artikel waarin de conclusies van een statistisch onderzoek besproken worden. Vraag vervolgens –indien het bestaat – het originele onderzoeksrapport op..(dit moet je Bespreek tenslotte populatie en (de representativiteit) van de steekproef

Opgave 4:Meerkeuzevragen

1. We vroegen aan 100 personen hoe vaak ze ziek geweest waren in het afgelopen jaar. 55 antwoordden 0 keer, 27 zeiden 1 keer, 8 zeiden 2 keer en 12 gaven 3 of meer keer op. Wat is hier de variabele :

- de ondervraagde personen
- de antwoordcategorieën 0, 1, 2 of 3 of meer
- de aantallen 55, 27, 8 en 12
- de ziektefrequentie

2. In een enquête hebben de vragen de antwoordcategorieën: "geheel mee eens", "mee eens", "geen mening", "niet mee eens", "geheel niet mee eens" en "geen mening". Het meetniveau van deze schaal is:

- nominaal
- ordinaal

3. In de frequentietabel met de rondetijden op de 10 km schaatsen is de bovengrens van de klasse met klassenmidden 41

- 42.0
- 41.9
- 41.999...

4. Kan je in een frequentietabel met alleen de klassen, de klassenbreedtes en de frequentiedichtheden de aantallen in iedere klasse terugvinden?

- Dat kan niet.
- Dat kan wel.

5. Als de 10de persoon niet 6 maar 16 fouten had gemaakt (zie voorbeeld bij mediaan), zou de mediaan:

- nog steeds 1.5 zijn geweest;
- iets groter dan 1.5 zijn geworden;
- gelijk aan 2 zijn geworden.

6. In een rapportage over armoede in een stadswijk kun je als representatieve centrale inkomensmaat het beste gebruik maken van:

- de modale klasse
- de mediaan
- het gemiddelde

7. Als de steekproefomvang, n , toeneemt, wordt de steekproefstandaardafwijking, s :

- kleiner.
- niet groter en niet kleiner.

8. Zijn inwoners van Brugge chauvinistischer dan inwoners van Brussel? Dan zouden Bruggelingen minder vaak wedstrijden van Anderlecht moeten bezoeken dan Brusselaars wedstrijden van Brugge. Bij een steekproefonderzoek onder 135 Brugge supporters tijdens een wedstrijd bleek 20% van hen in Brussel te wonen. In een vergelijkbare steekproef van 129 Anderlechtersupporters bleek 5% van hen in Brugge te wonen. De onderzoekers concludeerden, dat Bruggelingen chauvinistischer zijn dan Brusselaars

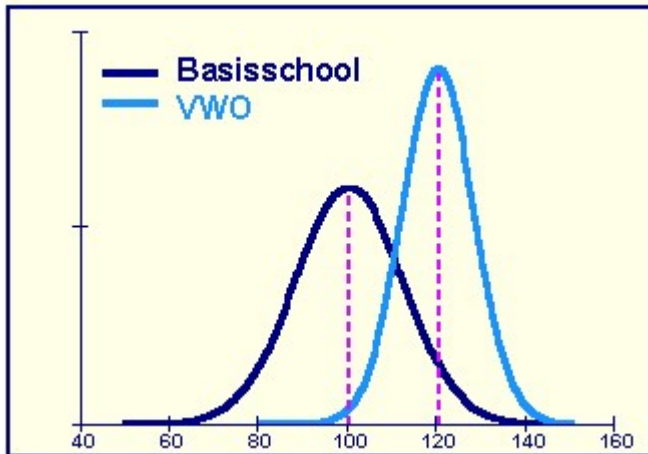
- Dit onderzoek is valide, zodat deze conclusie (met enig voorbehoud vanwege de kleine steekproeven) mag worden getrokken.
- Dit onderzoek is fout opgezet, zodat deze conclusie niet mag worden getrokken.

9. Om de personele inzet te kunnen plannen telt de balie van een ziekenhuis gedurende een week het gemiddelde aantal bezoekers per dag. Het daggemiddelde is

- een populatiegemiddelde;
- een steekproefgemiddelde.

Opgave 6:

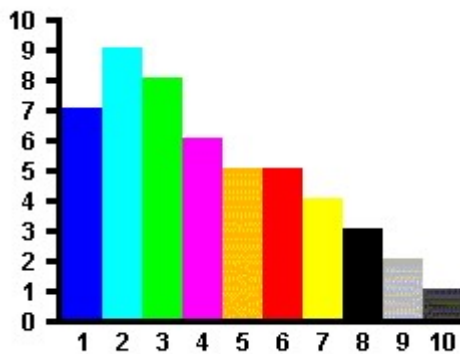
- 1) In de grafiek is de verdeling van het IQ van leerlingen van het VWO (richting wet-wisk in Nederland) vergeleken met die van leerlingen op een lagere school.



De VWO-leerlingen hebben een gemiddeld IQ van 120. De spreiding is vrij klein, de uitkomsten liggen zo dicht bij elkaar dat een IQ kleiner dan 100 of groter dan 140 bijna niet vóórkomt. Op de lagere school is het gemiddelde IQ 100 en de spreiding veel groter: de variatiebreedte is daar ongeveer 80.

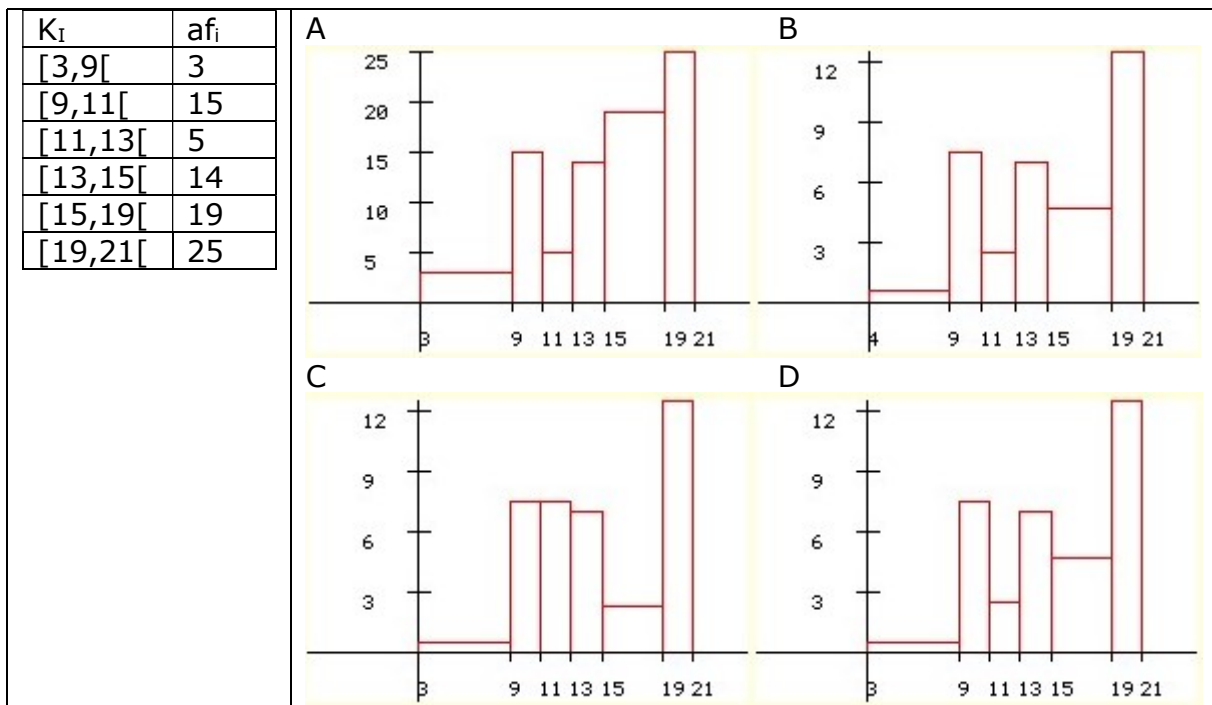
Verklaar Deze verschillen

- 2) Het volgend histogram geeft de punten weer van 50 leerlingen op een test (op 10) van Frans



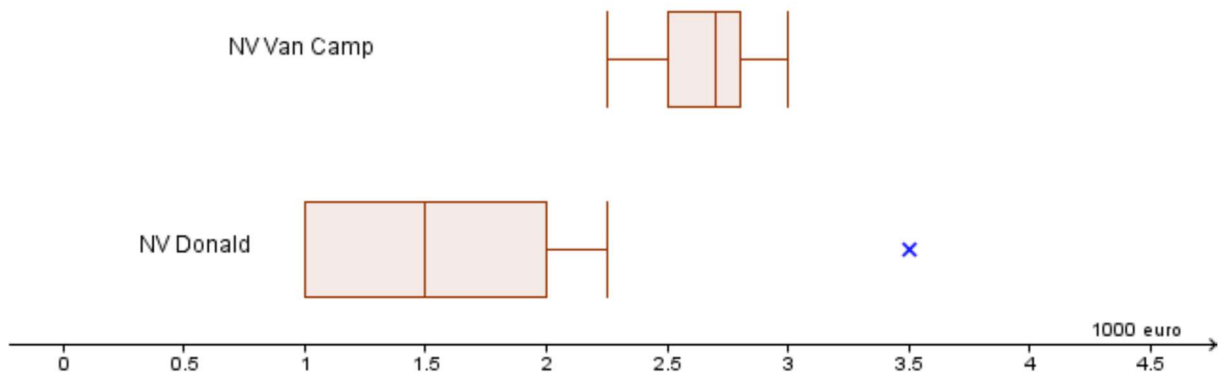
Bereken de drie centrummaten, de standaardafwijking en in de interkwartielafstand

3) Welk histogram komt overeen met de volgende frequentietabel? Verklaar je antwoord:



4) In de onderstaande boxplots worden de inkomens (in €) van de werknemers van twee bedrijfjes (de n.v. Donald en de n.v. Van Camp) met elkaar vergeleken:
Welke bewering is fout?

- a) In de n.v. Donald is Q_1 ook het laagst voorkomende inkomen
 - b) Het mediane inkomen is het grootst in de n.v. Van Camp
 - c) De interkwartielafstand is groter in de n.v. Donald
 - d) Voor het opstellen van deze grafiek werden de gegevens van 20 werknemers gebruikt
 - e) Er komen uitschieters voor bij de n.v. Donald
 - f) 75% van de werknemers bij de n.v. Donald verdient 2000€ of minder
 - g) In de n.v. Van Camp is 3000€ het hoogst waargenomen inkomen
 - h) De interkwartielafstand in de n.v. Donald is gelijk aan 1000€
 - i) Het laagste inkomen in de n.v. Van Camp is gelijk aan het hoogste inkomen in de n.v. Donald
- Verklaar je antwoord



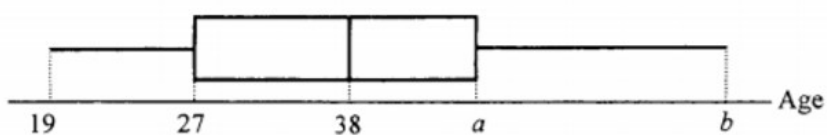
5)

3) Mean of 100 observations is 45. It was later found that two observations 19 and 31 were incorrectly recorded as 91 and 13. The correct mean is

- A) 44.0
- B) 44.46
- C) 45.00
- D) 45.54

6)

The box-and-whisker diagram below shows the distribution of the ages of the clerks in team X of a company. It is given that the range and the inter-quartile range of this distribution are 43 and 21 respectively.



- (a) Find a and b . (3 marks)
- (b) There are 5 clerks in team Y of the company and three of them are of age 38. It is given that the range of the ages of the clerks in team Y is 20. Team X and team Y are now combined to form a section. The manager of the company claims that the range of the ages of the clerks in the section and the range of the ages of the clerks in team X must be the same. Do you agree? Explain your answer. (2 marks)

7)

25) The standard deviation of 25 numbers is 40. If each of the numbers is increased by 5, then the new standard deviation will be [DCE 1995]

- A) 40
- B) 45
- C) $40 + \frac{21}{25}$
- D) None of these

8)

15) The number of observations in a group is 40. If the average of first 10 is 4.5 and that of the remaining 30 is 3.5, then the average of the whole group is [AMU 1992; DCE 1996]

- A) $\frac{1}{5}$
- B) $\frac{15}{4}$
- C) 4
- D) 8

9)

17) The mean age of a combined group of men and women is 30 years. If the means of the age of men and women are respectively 32 and 27, then the percentage of women in the group is

- A) 30
- B) 40
- C) 50
- D) 60

10)

Alle leerlingen van eenzelfde klas leggen 3 toetsen af voor het vak aardrijkskunde. Elke toets staat op 40 punten. Het klasgemiddelde van de tweede toets ligt 20 % hoger dan bij de eerste toets. In vergelijking met de eerste toets ligt het klasgemiddelde van de derde toets dan weer 10 % lager. Over de drie toetsen samen is het klasgemiddelde 31 op 40. Wat is het klasgemiddelde op 40 bij de tweede toets?

- <A> 30
- 32
- <C> 34
- <D> 36

11)

Suppose that n observations x_1, x_2, \dots, x_n have sample mean \bar{x} and sample variance $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$. Prove that s^2 can be written as $s^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)$.

Antwoorden meerkeuzevragen

Antw. 1: de variabele is de vraag naar de ziektefrequentie.

Antw. 2: De eerste 4 categorieën zijn weliswaar ordinaal, maar de laatste past niet in het rijtje en dat maakt de hele schaal nominaal.

Antw. 3: De bovengrens nadert in de limiet tot 42.0, dus 41.999... is de beste aanduiding van de bovengrens.

Antw. 4: dat kan nl. door de frequentiedichtheden te vermenigvuldigen met de klassenbreedtes.

Antw 5: hetzelfde

Antw. 6: Ja, inkomensverdelingen zijn vrijwel altijd rechtsscheef en dan is de mediaan de beste centrummaat.

Antw. 7: De noemer, $n - 1$, in de formule voor s wordt wel groter, maar dat wordt de teller, $\sum d^2$, ook. De steekproefstandaardafwijking is een schatting van de populatiestandaardafwijking en dat is vanzelfsprekend een constante .

Antw. 8: Fout Op deze manier zijn de populaties van *supporters* van Brugge respectievelijk van Anderlecht met elkaar vergeleken. De onderzoeksvraag daarentegen gaat over de populaties van *bewoners* van Brugge respectievelijk Anderlecht.

Antw. 9

Het gemeten daggemiddelde is een *schatting* van het gemiddelde van de populatie van alle (toekomstige) dagen en is dus een steekproefgemiddelde.