HOOFDSTUK 8

Digitaal statistisch onderzoek - SE

Wiskunde A · havo 5



IN DIT HOOFDSTUK LEER JE:

gegevens ordenen en controleren met een spreadsheet;
tabellen, diagrammen en kengetallen maken;
een onderzoek controleerbaar rapporteren.

Wat weet je nog?

Maak deze vragen vóór de nieuwe leerstof. Noteer je berekeningen en redeneringen. Gebruik een schrift voor je antwoorden. De onderzoeksgegevens zijn voor de oefeningen samengesteld.

1 Welke gegevens heb je?

Een fietsenstalling registreert per fiets het type, de aankomsttijd en de verblijfsduur.

a Welke variabele is kwalitatief en welke is kwantitatief? Geef van beide één voorbeeld.

b Bij 18 van 60 fietsen is de stalling langer dan twee uur. Bereken het percentage.

c Een ontbrekende verblijfsduur wordt vervangen door nul. Waarom kan dat het gemiddelde verstoren?

2 Samenvatten en begrenzen

De wachttijden 1, 2, 2, 3 en 12 minuten horen bij vijf bezoekers.

a Bereken gemiddelde, mediaan en spreidingsbreedte.

b Waarom is een cirkeldiagram ongeschikt om de relatie tussen wachttijd en bezoekduur te onderzoeken?

c Een interval is berekend uit een vrijwillige online enquête. Verhelpt een smal interval de selectiviteit?

3 Vraag, resultaat of conclusie?

Een onderzoeker schrijft: We bekeken veertig retourpakketten van maandag. Twaalf waren beschadigd. Ons nieuwe verpakkingsmateriaal voorkomt schade.

a Welk deel beschrijft de opzet en welk deel geeft een resultaat?

b Welke conclusie is wel direct met de aantallen te onderbouwen?

Werkgegevens hoofdstuk 8



Gebruik deze tabellen bij de opdrachten van hoofdstuk 8. Het werkbestand is als spreadsheetbijlage aan de PDF toegevoegd. Open de paperclip rechtsboven deze pagina, of het bijlagenpaneel van je PDF-lezer. Sla het bestand op en werk in een kopie. De tabellen hieronder geven dezelfde brongegevens voor werken op papier.

De gegevens zijn samengesteld voor deze lesopgaven. Een leeg vakje is een ontbrekende waarde; 0 is een werkelijk gemeten waarde.

Brongegevens hoofdstuk 8

Bibliotheek_raw

Tien ruwe registraties van bibliotheekbezoeken. Controleer ID's, categorieën, ontbrekende waarden en onmogelijke tijden.

Bezoek-ID	Dagdeel	Wachten (min)	Bezoek (min)
B01	ochtend	0	42
B02	middag	6	35
B03	Ochtend	3	50
B04	middag	ontbreekt	40
B05	avond	12	30
B06	middag	-4	45
B07	avond	2	60
B08	ochtend	4	55
B09	avond	40	25
B09	avond	40	25

Bezoek (min) is de tijd binnen na het wachten. Deze ruwe tabel bevat bewust gegevens die je moet controleren en opschonen. Bewaar dit bronblad.

Studieplekken

Op één dinsdag zijn in de ochtend, middag en avond steeds de eerste acht aankomende bezoekers gevolgd.

Bezoek-ID	Dagdeel	Wachten_min	Bezoek_min
S01	ochtend	0	30
S02	ochtend	1	35
S03	ochtend	2	40
S04	ochtend	2	45
S05	ochtend	3	50
S06	ochtend	4	55
S07	ochtend	5	60
S08	ochtend	7	65
S09	middag	1	35
S10	middag	3	40
S11	middag	4	45
S12	middag	5	50
S13	middag	6	55
S14	middag	8	60
S15	middag	10	65
S16	middag	15	70
S17	avond	0	40
S18	avond	0	45
S19	avond	1	50
S20	avond	2	55
S21	avond	3	60
S22	avond	4	65
S23	avond	6	70
S24	avond	8	75

Bezoek_min is de totale tijd van aankomst tot vertrek, inclusief wachten. Beschikbare studietijd = Bezoek_min - Wachten_min. Alle 24 registraties zijn volledig.

Wachttijden

Wachttijden in minuten van acht bezoekers per vestiging. Gebruik dezelfde klassen en schaal als je de groepen vergelijkt.

Vestiging	Wachten_min
A	1
A	2
A	2
A	3
A	4
A	4
A	5
A	7
B	1
B	1
B	2
B	2
B	3
B	3
B	4
B	16

8.1 Een dataset voorbereiden in een spreadsheet

Leerdoel

Ik kan een dataset organiseren met één waarneming per rij, één variabele per kolom en herkenbare eenheden en categorieën.

Theorie

Een vaste betekenis per kolom

Een dataset heeft één waarneming per rij. Een ID maakt de rij herkenbaar, maar is geen meetwaarde waarover je een gemiddelde berekent. Kies één variabele per kolom en vermeld de eenheid in de kolomkop.

Een codeboek beschrijft de betekenis van iedere variabele, toegestane categorieën en wat leeg of nul betekent. Bewaar de ontvangen gegevens op een apart tabblad Bron. Werk in een kopie. Zie voorbeeld 8.1A, blz. 598.

Onthoud: Eerst een duidelijke structuur, daarna pas rekenen.

1 Van notitie naar tabel

Een medewerker noteert: bezoek A11, ochtend, 4 minuten wachten, 36 minuten binnen; bezoek A12, middag, 0 minuten wachten, 24 minuten binnen; bezoek A13, ochtend, 8 minuten wachten, 42 minuten binnen.

a Maak vier kolommen met duidelijke namen en zet ieder bezoek op één rij.

b Waarom is een aparte rij per bezoek geschikter dan een aparte kolom per bezoek?

c Beschrijf in een codeboek wat 0 bij $\text{Wachten}_{\text{min}}$ betekent en welke categorieën Dagdeel mag bevatten.

Controleer met de bron

Leerdoel

Ik kan invoerproblemen, ontbrekende waarden en verdachte waarden opsporen en beslissingen daarover documenteren.

Theorie

Controleer met de bron

Zoek ontbrekende waarden, dubbele ID's, onmogelijke waarden en verschillende schrijfwijzen. Een identieke rij kan een dubbele invoer zijn; controleer eerst of het werkelijk dezelfde waarneming is.

Verander een waarde alleen op basis van een vastgelegde bron of meetafspraak. Houd onbekende waarden leeg. Noteer per wijziging de ID, oude en nieuwe waarde en reden. Een uitschieter kan een echte waarneming zijn. Zie voorbeeld 8.1A, blz. 598.

Onthoud: Onbekend is niet hetzelfde als nul.

2 Wat is een invoerfout?

Bekijk de tabel Bibliotheek_raw. De notities vermelden: B06 wachtte 4 minuten; B09 is tweemaal uit hetzelfde formulier overgenomen; de wachttijd van B04 is niet gemeten; B09 wachtte echt 40 minuten.

Bezoek-ID	Dagdeel	Wachten (min)	Bezoek (min)
B01	ochtend	0	42
B02	middag	6	35
B03	Ochtend	3	50
B04	middag		40
B05	avond	12	30
B06	middag	-4	45
B07	avond	2	60
B08	ochtend	4	55
B09	avond	40	25
B09	avond	40	25

a Maak een werkkopie en herstel alleen fouten die je met de notities kunt onderbouwen.

b Waarom verwijder je B09 niet omdat de wachttijd uitzonderlijk groot is?

c Maak een wijzigingslog met oude waarde, nieuwe waarde en reden.

Meetafspraak: Bezoek (min) is de tijd binnen na het wachten.
De wachttijd is niet in deze bezoekduur inbegrepen.

3 Controle zonder gokken

Een tafel met leveringen bevat: R1, 12 kg; R2, 0 kg; R3, leeg; R4, 120 kg; R4, 120 kg. De weegbon bevestigt 120 kg voor R4. De twee R4-rijen horen bij dezelfde levering. R2 is een lege retourkrat; alleen nettomassa wordt gemeten.

a Welke rijen behoud je en welke waarde blijft onbekend?

b Bereken de gemiddelde bekende nettomassa. Leg uit waarom delen door vier onjuist is.

c Een klasgenoot vervangt 120 door 12 omdat dat beter past. Beoordeel dit.

d Welke controle kan een dubbele ID opsporen zonder twee verschillende leveringen met dezelfde massa te verwarren?

Sorteren en nieuwe kolommen

Leerdoel

Ik kan sorteren, filteren en eenvoudige berekende kolommen gebruiken zonder het verband tussen gegevens van één waarneming te verbreken.

Theorie

Sorteren en nieuwe kolommen

Sorteer hele rijen, zodat de variabelen van een waarneming bij elkaar blijven. Een filter verbergt records; hij verwijdert ze niet. Controleer bij een berekening of gefilterde of verborgen rijen worden meegeteld.

Een formule begint met =. Staat tijd in minuten in B2, dan geeft $=\frac{B2}{60}$ uren. Kopiëren naar de volgende rij verandert B2 in B3. Controleer zelf een eerste, middelste en laatste rij. Zie voorbeeld 8.1B, blz. 599.

Onthoud: Een formule moet ook na kopiëren de juiste rijen gebruiken.

4 De rij moet bij elkaar blijven

Vier vrijwilligers verrichtten een sorteertaak. De tabel geeft duur en aantal dozen.

ID	Minuten	Dozen
V1	30	10
V2	40	12
V3	30	12
V4	40	8

a Voeg een kolom dozen per uur toe. Geef een formule voor de eerste rij als minuten in B2 en dozen in C2 staan.

4 De rij moet bij elkaar blijven (vervolg)

Vier vrijwilligers verrichtten een sorteertaak. De tabel geeft duur en aantal dozen.

ID	Minuten	Dozen
V1	30	10
V2	40	12
V3	30	12
V4	40	8

b Sorteer de volledige tabel van hoge naar lage productie per uur.

c Filter op minstens 10 dozen. Welke vrijwilligers blijven zichtbaar?

5 Een snelle sortering

De oorspronkelijke rijen zijn K1: 10 minuten en 6 dozen; K2: 30 minuten en 9 dozen; K3: 20 minuten en 8 dozen. Noor sorteert alleen de kolom minuten oplopend en laat de andere kolommen staan.

a Schrijf op welke twee onjuiste combinaties hierdoor ontstaan.

b Bereken de echte productie van K2 per uur en de productie volgens de beschadigde tabel.

c Beschrijf hoe je de fout veilig herstelt en voortaan voorkomt.

d Na een filter zijn twee rijen zichtbaar. Is een gewoon GEMIDDELDE over de oorspronkelijke hele kolom altijd het gemiddelde van alleen de zichtbare rijen?

Tellen, delen en controleren

Leerdoel

Ik kan een frequentietabel en procentuele verdeling produceren en de uitkomsten met totalen en een steekproefcontrole valideren.

Theorie

Tellen, delen en controleren

Maak eerst duidelijk wie of wat in de noemer telt. Voor een categoriepercentage deel je het aantal in die categorie door het aantal records met een bekende categorie.

Gebruik voor frequenties AANTAL. ALS (COUNTIF), voor numerieke aantallen AANTAL (COUNT) en voor een gemiddelde GEMIDDELDE (AVERAGE). Namen en scheidingstekens kunnen per programma verschillen. Een telling met de hand blijft een nuttige controle.

Deel iedere frequentie door hetzelfde passende totaal. Een procentnotatie zet 0,25 om in 25%. Bewaar voldoende decimalen tot de einduitkomst. Zie voorbeeld 8.1B, Voorbeelden blz. 2.

Onthoud: Controleer zowel het totaal aantal records als de som van de percentages.

6 Een verdeling controleren

Gebruik de gecorrigeerde Bibliotheek_raw uit opgave 2 , of herstel de bron met deze notities: Ochtend betekent ochtend; B06 is 4; B09 is dubbel; B04 blijft onbekend.

a Maak een frequentietabel voor het dagdeel en controleer het totaal.

b Bereken de verdeling van dagdelen in procenten.

c Hoeveel procent van de bekende wachttijden is hoogstens 4 minuten?

d Maak twee afzonderlijke totalencontroles voor de categorieën en de wachttijdselectie.

7 Zelfstandig met een nieuwe registratie

Een buurtreparatiepunt registreert zes reparaties. Elke ID hoort bij één reparatie. Bij P4 ontbreekt de gemeten tijd; P6 komt tweemaal uit hetzelfde formulier. De monteur bevestigt 90 minuten voor P5. Bij P3 is klein abusievelijk met een hoofdletter geschreven.

ID	Type	Tijd (min)
P1	klein	6
P2	groot	24
P3	Klein	18
P4	groot	
P5	groot	90
P6	klein	12
P6	klein	12

a Zet de gegevens in een spreadsheet met één rij per reparatie en een codeboek.

b Maak een onderbouwd wijzigingslog en bereken het gemiddelde van de bekende tijden.

c Voeg uren toe en sorteer de volledige werkkopie op tijd. Noem de eerste en laatste reparatie met bekende tijd.

d Maak een procentuele verdeling van het reparatietype en controleer deze.

8.2 Data analyseren met ICT

Leerdoel

Ik kan bij een onderzoeksvraag passende frequentietabellen en diagrammen kiezen en met correcte assen, labels en eenheden maken.

Theorie

Een figuur die de vraag beantwoordt

Begin bij het variabelentype en de onderzoeksvraag. Een categorie met een aantal past bij een staafdiagram; een verdeling van meetwaarden bij een histogram of boxplot; twee gekoppelde meetwaarden bij een puntenwolk.

Noem assen, eenheden en groepen. Gebruik bij vergelijking dezelfde grenzen en schaal. Verschillende groepsgroottes vergelijk je meestal met percentages. Zie voorbeeld 8.2A, blz. 603.

Onthoud: De vorm van het diagram moet het bedoelde verschil zichtbaar maken.

1 Wat wil je zichtbaar maken?

Een fietsenmaker wil van twee werkdagen de verdeling van reparatietijden vergelijken. Maandag: 10, 10, 15, 15, 20, 25, 30, 35 minuten. Dinsdag: 5, 10, 10, 10, 15, 20, 25, 65 minuten.

a Maak per dag een frequentietabel met klassen 0 tot 20, 20 tot 40, 40 tot 60 en 60 tot 80 minuten. De linkergrens telt mee.

b Maak met ICT twee histogrammen met dezelfde assen. Leg uit waarom deze weergave bij de vraag past.

c Iemand gebruikt per dag een andere horizontale schaal. Welk vergelijkingsprobleem ontstaat?

Kengetallen met betekenis

Leerdoel

Ik kan met ICT centrum- en spreidingsmaten berekenen en controleren of die voor de gegevens betekenisvol zijn.

Theorie

Kengetallen met betekenis

Bereken een gemiddelde alleen van kwantitatieve waarden met dezelfde betekenis en eenheid. Gemiddelde en standaardafwijking reageren op extreme waarden; mediaan en interkwartielafstand beschrijven het midden robuuster.

Gebruik voor een steekproef de steekproefstandaardafwijking S. Bereken kwartielen in dit hoofdstuk als de medianen van de onderste en bovenste helft; bij oneven n laat je de middelste waarneming bij die helften weg. Leg afwijkende software-instellingen vast. Zie voorbeeld 8.2A, blz. 603.

Onthoud: Controleer de selectie en de functie voordat je de uitkomst vertrouwt.

2 Een gemiddelde vertelt niet alles

Gebruik het tabblad Wachttijden. De waarden zijn A: 1, 2, 2, 3, 4, 4, 5, 7; B: 1, 1, 2, 2, 3, 3, 4, 16 minuten.

a Bereken per vestiging gemiddelde, mediaan en spreidingsbreedte met ICT.

b Bereken de steekproefstandaardafwijking. Controleer één gemiddelde met de hand.

c Kies een centrummaat om een gebruikelijke wachttijd te beschrijven en motiveer je keuze.

d Kun je op basis van het gemiddelde stellen dat iedere bezoeker bij B langer wacht?

Groepen beschrijven

Leerdoel

Ik kan twee groepen beschrijvend vergelijken op ligging, spreiding en verdelingsvorm en een conclusie met concrete gegevens onderbouwen.

Theorie

Groepen beschrijven

Een goede vergelijking noemt ligging, spreiding en verdelingsvorm. Kies concrete waarden: bijvoorbeeld mediaan, interkwartielafstand en een lange staart. Beschrijf welke groep op welk kenmerk hoger, lager of variabeler is.

Een hoger gemiddelde betekent niet dat ieder individu hoger scoort. Bij dit hoofdstuk volstaat een beschrijvende vergelijking; formele verschilregels volgen in hoofdstuk 10. Zie voorbeeld 8.2B, blz. 608.

Onthoud: Een verschil in centrum en een verschil in spreiding zijn verschillende uitspraken.

3 Boxen en een staart

Gebruik dezelfde twee lijsten uit Wachttijden. Neem kwartielen als medianen van de onderste en bovenste helft.

a Bepaal de vijf-getallensamenvatting en teken twee boxplots op één schaal.

3 Boxen en een staart (vervolg)

Gebruik dezelfde twee lijsten uit Wachttijden. Neem kwartielen als medianen van de onderste en bovenste helft.

b Beschrijf het verschil met één centrummaat en twee spreidingskenmerken.

c Een programma tekent 16 als los punt buiten de snor. Is de dataset veranderd?

4 Een loketkeuze

Twee loketten gebruikten dezelfde klokmeting. Deze twaalf tijden per loket komen uit één dinsdagochtend, zonder loting. L: 2, 3, 3, 4, 4, 5, 5, 6, 6, 7, 8, 9. M: 1, 2, 2, 3, 3, 4, 4, 5, 5, 6, 8, 17. Alle tijden zijn minuten.

a Vergelijk gemiddelde en mediaan en controleer de aantallen.

b Maak een passende vergelijkende grafiek en beschrijf ook de spreiding.

c Een klant wil vooral zeer lange wachttijden vermijden. Welke informatie weegt dan mee?

d Waarom kun je hiermee geen algemeen oorzakelijk effect van het loket aantonen?

Van steekproef naar populatie

Leerdoel

Ik kan waar de steekproefopzet en voorwaarden het toelaten een eerder behandeld betrouwbaarheidsinterval verwerken en de conclusie beperken tot wat de gegevens ondersteunen.

Theorie

Van steekproef naar populatie

Een betrouwbaarheidsinterval beschrijft de onzekerheid van een schatting onder passende steekproefaannames. Gebruik bij een voldoende grote, passende steekproef voor een proportie $p \pm 2\sqrt{(\frac{p(1-p)}{n})}$, en voor een gemiddelde $\bar{x} \pm \frac{2S}{\sqrt{n}}$.

Een vrijwillige of selectieve respons wordt niet representatief door een grote omvang. Benoem de populatie, selectie, ontbrekende waarden en beperkingen. Een interval voor een gemiddelde is geen interval met 95% van de meetwaarden. Zie voorbeeld 8.2B, blz. 608.

Onthoud: Controleer de steekproefopzet vóór je een interval interpreteert.

5 **Een schatting voor de maand**

Uit alle bezoeken van een maand zijn 400 bezoeken aselect getrokken. Alle wachttijden zijn beschikbaar; 120 bezoeken duurden langer dan 5 minuten. Behandel de waarnemingen als voldoende onafhankelijk.

a Bereken met het formuleblad een benaderend 95%-interval voor de populatieproportie.

b Beoordeel de claim dat meer dan 40% van alle bezoeken langer dan 5 minuten wachttijd heeft.

c Beschrijf correct wat 95% betrouwbaarheid betekent.

6 Precies maar selectief

Een vrijwillige enquête onder appgebruikers heeft n=2500 en gemiddelde tevredenheid 8,4. Een berekend interval is [8,3;8,5]. Iemand schrijft: alle bezoekers zijn met 95% zekerheid heel tevreden.

a Noem twee verschillende problemen met de conclusie.

b Formuleer een wel houdbare beschrijvende conclusie.

c Stel een verbeterde steekproefopzet voor.

7 Nieuwe gegevens beoordelen

Een aselecte steekproef van 256 afzonderlijke bestellingen heeft volledige leveringstijden. Gemiddelde 42 minuten, steekproefstandaardafwijking 16. Twee bezorggebieden hebben daarnaast deze samenvatting: Noord n=120, mediaan 38, Q1=30, Q3=46; Zuid n=136, mediaan 43, Q1=35, Q3=57. In Noord zijn 24 bestellingen later dan 60 minuten en in Zuid 34.

a Maak een geschikte tabel en een diagram om de aandelen later dan 60 minuten te vergelijken.

b Bereken en vergelijk de interkwartielafstanden.

c Vergelijk de gebieden beschrijvend en benoem wat je over de volledige verdelingsvorm nog niet weet.

d Bereken het 95%-interval voor het gemiddelde van alle bestellingen en interpreteer de reikwijdte.

8.3 Een onderzoek rapporteren

Leerdoel

Ik kan een afgebakende onderzoeksvraag formuleren met passende variabelen, populatie en een verantwoord gegevensplan.

Theorie

Een vraag die je kunt beantwoorden

Een onderzoeksvraag noemt een afgebakende populatie, een variabele en wat je wilt weten: verdeling, aandeel of verschil. Bepaal vóór de dataverzameling wanneer en hoe je meet.

Beschrijf selectie en non-respons afzonderlijk. Een volledig antwoord van een aselect gekozen groep is iets anders dan een grote vrijwillige groep. Zie voorbeeld 8.3A, blz. 609.

Onthoud: Een scherp afgebakende vraag bepaalt welke gegevens nodig zijn.

1 Een onderzoeksvraag aanscherpen

Een bibliotheek vraagt: Zijn onze studieplekken goed? Zij kan aankomsttijd, wachttijd en duur van bezoeken registreren.

a Maak de vraag toetsbaar met één meetbare uitkomst en een afgebakende periode.

b Beschrijf een meetafspraak voor wachttijd.

c Ontwerp een steekproef die ook rustige dagdelen omvat.

d Leg uit waarom tevredenheid en wachttijd niet dezelfde variabele zijn.

De berekening moet terug te volgen zijn

Leerdoel

Ik kan een controleerbare digitale verwerking opleveren met brongegevens, meetafspraken, verwerkingsstappen en passende resultaten.

Theorie

De berekening moet terug te volgen zijn

Een controleerbaar onderzoeksbestand bevat brongegevens, codeboek, verwerkingsstappen en resultaten. Gebruik herkenbare tabbladen en formules. Leg correcties en uitgesloten records vast met hun reden.

Een rapport bevat achtereenvolgens vraag, methode, resultaten, conclusie en beperkingen. Zet tabellen en figuren bij de bewering die ze ondersteunen. Zie voorbeeld 8.3A, blz. 609.

Onthoud: Een ander moet jouw uitkomst uit de bron kunnen reconstrueren.

2 Een controleerbaar rapport

Een leerling levert alleen een screenshot van een grafiek in. In het rapport staat: De mediaan daalde van 7 naar 5 minuten. Vier lege cellen zijn als nul ingevuld; de ruwe gegevens zijn overschreven.

a Noem drie ontbrekende onderdelen die controle onmogelijk maken.

b Waarom mag je de mediaandaling nog niet als bevinding overnemen?

c Maak een korte lijst van wat een vervangende inzending moet bevatten.

d Is een screenshot van een correct diagram nutteloos?

Een onderbouwde conclusie

Leerdoel

Ik kan een conclusie formuleren die de onderzoeksvraag beantwoordt en onderbouwen met kengetallen, diagrammen en waar passend een interval.

Theorie

Een onderbouwde conclusie

Beantwoord de vraag met relevante getallen en een passende figuur. Vermeld eenheden en de onderzochte groep. Kies alleen een interval wanneer de steekproefopzet en voorwaarden dat dragen.

Beschrijf eerst wat de gegevens tonen. Een mogelijke verklaring mag als hypothese worden genoemd, maar is daarmee nog geen aangetoonde oorzaak. Zie voorbeeld 8.3B, blz. 610.

Onthoud: Elke hoofdconclusie krijgt een concrete onderbouwing.

3 Een klein eigen onderzoek

Onderzoek een onpersoonlijk kenmerk dat je zonder namen of gevoelige gegevens kunt meten, bijvoorbeeld het gewicht van dezelfde soort lege verpakking. Kies minstens 30 waarnemingen en een duidelijk afgebakende populatie.

a Formuleer vóór de meting een onderzoeksvraag, meetafspraak en selectieplan.

b Lever de brondata, het codeboek en een wijzigingslog in en voer de analyse met ICT uit.

c Schrijf een conclusie van 100 tot 150 woorden met minstens twee concrete resultaten.

d Beschrijf één selectiebeperking, één meetbeperking en een uitvoerbare verbetering.

Grenzen van het onderzoek

Leerdoel

Ik kan beperkingen door steekproefselectie, meetwijze, ontbrekende waarden en onzekerheid benoemen en een zinvolle verbetering voorstellen.

Theorie

Grenzen van het onderzoek

Noem beperkingen die werkelijk invloed op de conclusie kunnen hebben: selectie, non-respons, meetfouten, ontbrekende waarden en toevalsvariatie. Alleen schrijven dat er meer onderzoek nodig is, is te algemeen.

Verbind iedere verbetering met een probleem. Meer waarnemingen vermindert toevalsvariatie, maar herstelt niet vanzelf een verkeerde selectie of foutieve meetdefinitie. Zie voorbeeld 8.3B, blz. 610.

Onthoud: Maak duidelijk wat je wel kunt concluderen en waarvoor gegevens ontbreken.

4 Studieplekken onderzoeken

Gebruik het bijgevoegde tabblad Studieplekken. Op één dinsdag zijn in de ochtend, middag en avond steeds de eerste acht aankomende bezoekers gevolgd. Wachten_min is de tijd tot een beschikbare plek; Bezoek_min is de tijd van aankomst tot vertrek. Alle 24 registraties zijn volledig. De gegevens zijn samengesteld.

a Formuleer een onderzoeksvraag over verschillen in wachttijd tussen de drie geregistreerde dagdelen. Benoem de onderzochte groep.

b Maak een codeboek, bewaar het bronblad en controleer de data op aantallen, ID's en onmogelijke tijden.

c Bereken per dagdeel gemiddelde, mediaan en spreidingsbreedte van de wachttijd.

d Maak één figuur waarmee je de drie verdelingen eerlijk kunt vergelijken.

e Voeg de beschikbare studietijd per bezoek toe. Bereken per dagdeel het gemiddelde daarvan.

f Schrijf een conclusie van 120 tot 180 woorden met twee kengetallen en een verwijzing naar je figuur.

4 Studieplekken onderzoeken (vervolg)

g Beoordeel of een betrouwbaarheidsinterval voor alle bezoeken van het jaar hier passend is.

h Stel twee concrete verbeteringen voor en lever bron, verwerking en rapport samen in.

5 Een rapport over leenmateriaal

Een centrum wil weten of er in februari genoeg uitleenlaptops waren. Uit een volledige lijst zijn 200 aanvragen aselect gekozen; iedere aanvraag is geregistreerd. Bij 34 aanvragen was direct geen laptop beschikbaar. De uitleenbalie heeft drie medewerkers; geen gegevens over hun afzonderlijke prestaties zijn gemeten.

a Formuleer een passende onderzoeksvraag en benoem populatie en variabele.

b Beschrijf de digitale verwerking waarmee de uitkomst controleerbaar wordt.

c Bereken een 95%-interval en formuleer een onderbouwde conclusie.

d Beoordeel de uitspraak: Het tekort komt door te langzaam werkende medewerkers. Geef een passende vervolgmeting.

Gemengd oefenen

1 Een percentage met twee noemers

Bij een verhuurpunt zijn 80 aanvragen geregistreerd, waarvan 12 zonder bekende wachttijd. Van de bekende tijden zijn 17 langer dan 10 minuten.

a Bereken het percentage langer dan 10 minuten onder de bekende tijden.

b Welke onder- en bovengrens voor alle 80 aanvragen kun je zonder verdere aanname geven?

2 Een oude meting opnieuw gebruiken

Vorig jaar was de gemiddelde wachttijd 6 minuten bij S=3. Dit jaar is in een vergelijkbare aselecte steekproef n=144, gemiddelde 5,5 en S=3,6 gemeten.

a Bereken het 95%-interval voor het gemiddelde van dit jaar.

b Bewijst dit interval dat de populatiegemiddelden verschillen?

3 Twee grootten vergelijken

Groep A heeft 30 van 50 reacties positief; groep B 72 van 120.

a Vergelijk de percentages en de absolute aantallen.

b Ontwerp een diagram waarmee dit direct zichtbaar is.

4 Een verandering in de meting

Een wachttijd werd eerst vanaf binnenkomst gemeten en later pas vanaf aanmelden bij de balie.

a Waarom is een lager gemiddelde niet automatisch een verbetering van de service?

b Hoe maak je een vervolgvergelijking bruikbaar?

Gericht verder oefenen

5 Herstel de structuur

Een tabel heeft kolommen persoon 1, persoon 2 en persoon 3. In iedere kolom staan achtereenvolgens leeftijd en reistijd: 16 jaar en 20 minuten; 17 jaar en 35 minuten; 16 jaar en 25 minuten.

a Herschik dit tot één rij per persoon.

b Bereken de gemiddelde reistijd en controleer de eenheid.

6 Van telling naar verdeling

Bij 25 registraties zijn de categorieën fiets 10, lopen 5, bus 8 en onbekend 2.

a Bereken de percentages onder de 23 bekende vervoermiddelen.

b Hoe rapporteer je de twee onbekende waarden?

7 Schrijf minder stellig

Na één middagmeting met mediaan 4 minuten staat in een rapport: Onze wachttijd is altijd 4 minuten.

a Welke twee beperkingen ontbreken?

b Herschrijf de conclusie.

Nieuwe controle

8 Een nieuwe gegevenscontrole

Van zeven unieke bestellingen hebben zes een bekende massa: 100, 150, 150, 200, 250 en 350 gram. De zevende massa is onbekend.

a Bereken de gemiddelde bekende massa en het aandeel bekende massa's van minstens 200 gram.

b Een rapport noemt een gemiddelde van 171,4 gram. Verklaar de vermoedelijke fout.

9 Een nieuwe onderzoeksconclusie

Uit alle meldingen van een maand zijn 225 meldingen aselect gekozen. Alle reactietijden zijn gemeten: gemiddelde 18 minuten, S=6.

a Bereken het 95%-interval voor het gemiddelde en benoem de populatie.

b Een manager zegt dat daarmee 95% van de meldingen binnen 18,8 minuten is beantwoord. Beoordeel dit.

Leerdoelenoverzicht

8.1 · Leerdoel 1: Ik kan een dataset organiseren met één waarneming per rij, één variabele per kolom en herkenbare eenheden en categorieën.

8.1 · Leerdoel 2: Ik kan invoerproblemen, ontbrekende waarden en verdachte waarden opsporen en beslissingen daarover documenteren.

8.1 · Leerdoel 3: Ik kan sorteren, filteren en eenvoudige berekende kolommen gebruiken zonder het verband tussen gegevens van één waarneming te verbreken.

8.1 · Leerdoel 4: Ik kan een frequentietabel en procentuele verdeling produceren en de uitkomsten met totalen en een steekproefcontrole valideren.

8.2 · Leerdoel 1: Ik kan bij een onderzoeksvraag passende frequentietabellen en diagrammen kiezen en met correcte assen, labels en eenheden maken.

8.2 · Leerdoel 2: Ik kan met ICT centrum- en spreidingsmaten berekenen en controleren of die voor de gegevens betekenisvol zijn.

8.2 · Leerdoel 3: Ik kan twee groepen beschrijvend vergelijken op ligging, spreiding en verdelingsvorm en een conclusie met concrete gegevens onderbouwen.

8.2 · Leerdoel 4: Ik kan waar de steekproefopzet en voorwaarden het toelaten een eerder behandeld betrouwbaarheidsinterval verwerken en de conclusie beperken tot wat de gegevens ondersteunen.

8.3 · Leerdoel 1: Ik kan een afgebakende onderzoeksvraag formuleren met passende variabelen, populatie en een verantwoord gegevensplan.

8.3 · Leerdoel 2: Ik kan een controleerbare digitale verwerking opleveren met brongegevens, meetafspraken, verwerkingsstappen en passende resultaten.

8.3 · Leerdoel 3: Ik kan een conclusie formuleren die de onderzoeksvraag beantwoordt en onderbouwen met kengetallen, diagrammen en waar passend een interval.

8.3 · Leerdoel 4: Ik kan beperkingen door steekproefselectie, meetwijze, ontbrekende waarden en onzekerheid benoemen en een zinvolle verbetering voorstellen.

Eindevaluatie

Kijk je werk na met de uitwerkingen. Noteer per leerdoel een opgave die goed ging en een opgave die nog oefening vraagt.

Werk het passende onderdeel opnieuw door. Maak daarna een nieuwe controleopgave zonder je eerdere uitwerking erbij. Controleer of je aanpak nu klopt.

Eindevaluatie

Kijk na met de uitwerkingen. Kies bij een fout de bijbehorende opgaven onder Oefen opnieuw. Maak daarna de opgaven onder Controleer zonder hulp.

Toetsvraag	Leerdoel	Oefen opnieuw	Controleer
Test jezelf opgave 10a-b	Datastructuur en broncontrole	Gericht verder oefenen opgave 5; 8.1 opgave 3a-c	Nieuwe controle opgave 8a-b
Test jezelf opgave 10c	Formulekolom en hele rijen sorteren	8.1 opgave 4a-c; 8.1 opgave 5a-c	8.1 opgave 7c
Test jezelf opgave 10d en 11a	Frequenties, passende noemer en diagram	Gericht verder oefenen opgave 6; 8.2 opgave 1a-c	8.1 opgave 7d; 8.2 opgave 7a
Test jezelf opgave 11b-c	Ligging en spreiding beschrijvend vergelijken	8.2 opgave 2a-d; 8.2 opgave 3a-c	8.2 opgave 7b-c
Test jezelf opgave 11d en 12c	Betrouwbaarheidsinterval berekenen en interpreteren	8.2 opgave 5a-c; 8.2 opgave 6a-b	Nieuwe controle opgave 9a-b
Test jezelf opgave 12a-b-d	Controleerbaar rapport en begrensde conclusie	Gericht verder oefenen opgave 7; 8.3 opgave 1a-d en 2a-d	8.3 opgave 5a-b-d

Voorbeeld 8.1A

Een ontbrekende temperatuur

Bij drie metingen zijn de temperaturen 8 °C, leeg en 14 °C.

a Hoe leg je de ontbrekende waarde vast en bereken je het gemiddelde van de bekende waarden?

1. Wat zoek ik?

Het gemiddelde van de werkelijk gemeten temperaturen.

2. Welke basis en aanpak kies ik, en waarom?

Een lege meting is onbekend; ik bewaar haar apart van nul.

3. Berekening

$\frac{(8 + 14)}{2} = 11$.

4. Antwoord in context

Het gemiddelde van twee bekende metingen is 11 °C.

5. Controle

11 ligt tussen 8 en 14. Delen door drie zou een niet-gemeten nul toevoegen.

Voorbeeld 8.1B

Van bronregistratie naar een controleerbaar percentage

Je leert gegevens veilig corrigeren, een formule kopiëren, hele rijen sorteren en filteren, en het juiste totaal gebruiken. Voer de stappen zelf uit in Excel of LibreOffice Calc. De gegevens zijn voor dit voorbeeld samengesteld.

1. Wat zoek ik?

Welk percentage van de bezoeken met een bekende wachttijd wacht langer dan 5 minuten? We meten vanaf aanmelden tot de bezoeker een studieplek krijgt. Direct een plek krijgen is 0 minuten; niet gemeten blijft onbekend.

2. Welke basis en aanpak kies ik, en waarom?

Ik gebruik alleen de bezoeken waarvan de wachttijd bekend is als basis. Ik bewaar de bron, verantwoord de correcties en controleer de telling met formules én een filter. Zo blijven onbekende tijden onbekend en blijven de gegevens van ieder bezoek bij elkaar.

3. Berekening

Bewaar de bron

Maak een leeg bestand en noem het eerste werkblad Bron. Zet de vier kolomkoppen in A1:D1 en de onderstaande 26 registraties in A2:D27. Een leeg vak in de tabel blijft ook in de spreadsheet echt leeg: typ daar geen streepje, nul of woord.

ID	Dagdeel	Wachten (min)	Bezoek (min)
V01	ochtend	0	42
V02	middag	6	35
V03	ochtend	3	50
V04	middag		40
V05	avond	12	30
V06	middag	-4	45
V07	avond	2	60
V08	ochtend	4	55
V09	avond	18	35
V10	middag		38
V11	ochtend	1	24
V12	avond	7	44
V13	middag	3	41
V14	ochtend	5	45

ID	Dagdeel	Wachten (min)	Bezoek (min)
V15	avond		50
V16	middag	6	36
V17	Ochtend	2	33
V18	middag	8	48
V19	avond	3	40
V20	ochtend		45
V21	middag	9	35
V22	avond	4	50
V23	ochtend	5	44
V24	middag		42
V25	avond	7	48
V17	Ochtend	2	33

Bronnotities: V06 wachtte volgens het meetformulier 4 minuten. De twee regels met V17 zijn kopieën van hetzelfde formulier. Ochtend en ochtend betekenen hetzelfde dagdeel. De wachttijd van V04, V10, V15, V20 en V24 is niet gemeten. Voor V09 bevestigt het meetformulier 18 minuten.

Maak een werkkopie en verantwoord de wijzigingen

Dupliceer het blad Bron en noem de kopie Verwerking. Laat Bron ongewijzigd. Herstel op Verwerking C7 van -4 naar 4 en B18 van Ochtend naar ochtend. Verwijder alleen de laatste dubbele V17-regel, rij 27. Er blijven 25 bezoeken over in A2:D26. De 18 minuten van V09 blijven staan: een lange tijd is niet automatisch een invoerfout.

ID / cel	Oude invoer	Nieuwe invoer	Reden
V06 / C7	-4	4	Meetformulier bevestigt 4 min
V17 / B18	Ochtend	ochtend	Eén schrijfwijze per categorie
V17 / rij 27	Tweede identieke regel	Verwijderd	Zelfde formulier dubbel ingevoerd

Codeboek: ID is een uniek bezoeknummer. Dagdeel is ochtend, middag of avond. Wachten is de tijd tot een plek: 0 is gemeten, leeg is onbekend. Bezoek is de tijd van aankomst tot vertrek. Beide tijden staan in minuten en zijn niet negatief.

Maak een berekende kolom die onbekend niet in nul verandert

Typ Wachten (uur) in E1. Typ in E2 de formule =ALS(C2="";"";C2/60). Kopieer deze met de vulgreep naar E3:E26. ALS controleert eerst of C2 leeg is: dan blijft ook de uitkomst leeg. Anders rekent de formule minuten om naar uren. Geef de kolom een getalnotatie met twee decimalen; bewaar de precieze formulewaarde.

Controle	Invoer in C	Uitkomst in E
V01	0 min	0 uur
V02	6 min	0,10 uur
V04	Leeg	Leeg
V09	18 min	0,30 uur

Bereken aantallen en het percentage

Zet de volgende labels in G2:G7 en de bijbehorende formules in H2:H7. De functies zijn voor een Nederlandstalige Excel- of Calc-installatie. In een Engelse installatie heten ALS, AANTALARG, AANTAL en AANTAL.ALS respectievelijk IF, COUNTA, COUNT en COUNTIF. De gebruikte instelling kan ; of , als scheidingsteken vragen.

Cel	Wat tel of bereken je?	Formule	Uitkomst
H2	Geregistreerde bezoeken	=AANTALARG(A2:A26)	25
H3	Bekende numerieke wachttijden	=AANTAL(C2:C26)	20
H4	Onbekende wachttijden	=H2-H3	5
H5	Langer dan 5 minuten	=AANTAL.ALS(C2:C26;">5")	8
H6	Bekend en hoogstens 5 minuten	=H3-H5	12
H7	Aandeel langer dan 5 onder bekende tijden	=H5/H3	0,40

Geef H7 een percentage-opmaak: 0,40 wordt 40%. Vermenigvuldig de formule dan niet nogmaals met 100. AANTAL telt alleen de 20 numerieke wachttijden; een echte 0 wordt meegeteld. De vijf ontbrekende metingen vormen geen vijf korte wachttijden.

Sorteer volledige bezoeken

Selecteer A1:E26, dus alle kolommen van de 25 bezoeken. Kies Gegevens > Sorteren. Geef aan dat de eerste rij kolomkoppen bevat en sorteer op Wachten (min), van groot naar klein. Vergroot de selectie als het programma daarover vraagt. Sorteer nooit alleen C: dan komen wachttijden bij de verkeerde ID te staan.

Controleer de eerste twee bekende tijden: V09 hoort bij 18 minuten en V05 bij 12 minuten. In hun andere kolommen moeten nog steeds respectievelijk avond/35 minuten bezoek en avond/30 minuten bezoek staan. De formule in E moet voor V09 nog steeds 0,30 uur geven.

Filter en controleer je telling

Selecteer A1:E26. Zet in Excel Gegevens > Filter aan; kies in Calc Gegevens > AutoFilter. Open het pijltje bij Wachten (min). Gebruik een numeriek filter groter dan 5; in Calc kun je daarvoor Standaardfilter gebruiken. Er moeten acht bezoeken zichtbaar blijven.

Bezoeken die zichtbaar blijven	Tijden in minuten
V02, V05, V09, V12	6, 12, 18, 7
V16, V18, V21, V25	6, 8, 9, 7

Een filter verbergt de andere rijen en verwijdert ze niet. De formules in H2:H7 blijven over hun volledige bereik rekenen: H3 blijft 20 en H5 blijft 8. Wil je een gemiddelde van uitsluitend de zichtbare rijen, dan is een gewoon GEMIDDELDE over de hele kolom daarvoor niet voldoende. Wis het filter zodra je de acht ID's hebt gecontroleerd.

4. Antwoord in context

Van de 20 geregistreerde bezoeken met een bekende wachttijd wacht 40% langer dan 5 minuten. Van vijf andere bezoeken kennen we de wachttijd niet. Daarom mag je 40% niet zonder meer als het percentage voor alle 25 bezoeken presenteren.

5. Controle

Controle: 8 + 12 = 20 en 20 + 5 = 25. De filtercontrole geeft dezelfde acht lange wachttijden als de formule. Voor alle 25 bezoeken ligt het mogelijke aandeel tussen $\frac{8}{25} = 32\%$ en $\frac{13}{25} = 52\%$, omdat de vijf onbekende tijden kort of lang kunnen zijn.

Probeer nu zelf

Bereken in een nieuwe cel het aandeel bekende wachttijden dat langer is dan 7 minuten. Controleer de uitkomst met een filter en schrijf één zin met de juiste onderzochte groep. Kijk pas daarna naar de uitwerking.

Voorbeeld 8.2A

Een wachttijdverdeling met ICT beschrijven

Gebruik de gecorrigeerde 25 bezoeken uit voorbeeld 8.1B. Dit voorbeeld verbindt centrum, spreiding en een histogram. De onderzoeksvraag is: hoe zijn de bekende wachttijden van deze geregistreerde bezoeken verdeeld? Het gaat om deze oefenregistratie, niet om een aselecte steekproef van alle bezoekers van het jaar.

1. Wat zoek ik?

Ik zoek hoe de bekende wachttijden van deze geregistreerde bezoeken verdeeld zijn: waar het centrum ligt, hoe groot de spreiding is en welke vorm het histogram heeft.

2. Welke basis en aanpak kies ik, en waarom?

Ik gebruik de twintig bekende tijden; de vijf lege cellen vormen geen metingen van nul minuten. Ik bereken centrum en spreiding met het juiste bereik en maak een histogram met vaste, even brede klassen. De kwartielen bepaal ik volgens de afspraak van dit hoofdstuk, zodat mijn berekeningen en figuur dezelfde betekenis hebben.

3. Berekening

Leg de analyse op een eigen werkkopie vast

Wis filters. Dupliceer Verwerking en noem de kopie Analyse. De 25 bezoeken staan nog steeds in A2:E26. Sorteren, nieuwe formules en het diagram doe je op deze kopie. Bewaar Bron en Verwerking, zodat je iedere waarneming kunt terugzoeken.

Bereken centrum en spreiding met het juiste bereik

Zet de labels in G2:G8 en de formules hieronder in H2:H8. Het bereik C2:C26 bevat 20 getallen en vijf lege cellen. De gebruikte functies negeren die lege cellen, maar nemen de gemeten 0 wel mee. Rond alleen de getoonde uitkomsten af.

Cel	Kengetal	Formule	Uitkomst
H2	Aantal bekende tijden	=AANTAL(C2:C26)	20
H3	Som van de tijden	=SOM(C2:C26)	109 min
H4	Gemiddelde	=GEMIDDELDE(C2:C26)	5,45 min
H5	Mediaan	=MEDIAAN(C2:C26)	4,50 min
H6	Minimum	=MIN(C2:C26)	0 min
H7	Maximum	=MAX(C2:C26)	18 min
H8	Spreidingsbreedte	=H7-H6	18 min

Controle met de hand

De 20 bekende tijden, oplopend: 0, 1, 2, 2, 3, 3, 3, 4, 4, 4, 5, 5, 6, 6, 7, 7, 8, 9, 12, 18. De som is 109; dus het gemiddelde is $\frac{109}{20}$ = 5,45 minuten. De middelste tijden, de 10e en 11e, zijn 4 en 5; dus de mediaan is $\frac{(4 + 5)}{2}$ = 4,5 minuten.

Een formule die 109 door 25 deelt geeft 4,36 minuten. Daarmee behandel je de vijf onbekende tijden ten onrechte als nul. Dat is geen gemiddelde van de bekende metingen.

Maak de kwartielafspraak zichtbaar

Sorteer A1:E26 volledig op Wachten (min), van klein naar groot. Controleer dat de 20 numerieke tijden in C2:C21 staan en de vijf lege tijden onderaan. Controleer V01 bij 0 en V09 bij 18. In dit hoofdstuk bepalen we kwartielen als de medianen van de onderste en bovenste helft; daarom gebruiken we hier niet automatisch de knop Kwartiel.

Cel	Maat	Formule na correct sorteren	Uitkomst
H10	Eerste kwartiel Q1	=MEDIAAN(C2:C11)	3 min
H11	Derde kwartiel Q3	=MEDIAAN(C12:C21)	7 min
H12	Interkwartielafstand	=H11-H10	4 min
H13	Steekproefstandaardafwijking S	=STDEV.S(C2:C26)	4,12 min

De middelste 50% loopt van 3 tot 7 minuten. De standaardafwijking S wordt hier als spreidingsmaat van de bekende waarnemingen berekend; dit maakt de registratie op zichzelf nog geen geschikte steekproef voor een betrouwbaarheidsinterval. STDEV.S is beschikbaar onder dezelfde naam in recente Excel- en Calc-versies. In een Engelse installatie heten SOM, GEMIDDELDE en MEDIAAN SUM, AVERAGE en MEDIAN.

Bouw een frequentietabel met vaste klassengrenzen

Maak in J1:N1 de kolomkoppen Klasse, Ondergrens, Bovengrens, Frequentie en Aandeel. Vul in J2:L5 de vier klassen hieronder in. De linkergrens telt mee, de rechtergrens niet: een tijd van 5 minuten komt dus in 5 tot 10. Alle klassen zijn 5 minuten breed.

Rij	Klasse	Ondergrens K	Bovengrens L	Frequentie M	Aandeel N
2	0 tot 5	0	5	10	50%
3	5 tot 10	5	10	8	40%
4	10 tot 15	10	15	1	5%
5	15 tot 20	15	20	1	5%

Typ in M2: =AANTALLEN.ALS($C$2:$C$26;">="&K2;$C$2:$C$26;"<"&L2;$C$2:$C$26;"<>"). Kopieer naar M3:M5. De eerste twee voorwaarden selecteren de klasse; de derde sluit lege cellen expliciet uit. De $-tekens houden het gegevensbereik vast, terwijl K2 en L2 bij kopiëren K3 en L3 worden. In een Engelse installatie heet de functie COUNTIFS.

Typ in N2 =M2/$H$2 en kopieer naar N3:N5. Geef N2:N5 een percentage-opmaak. Controleer met =SOM(M2:M5) dat de frequenties optellen tot 20 en met =SOM(N2:N5) dat de aandelen optellen tot 100%.

Maak een histogram waarvan de grenzen controleerbaar zijn

Gebruik de berekende frequentietabel; laat het programma niet opnieuw zelf klassengrenzen kiezen. Maak hiervoor in J8:K12 een compacte diagramtabel: J8 Klasse, K8 Frequentie; vul J9:J12 met de vier klasseteksten en K9:K12 met =M2, =M3, =M4 en =M5.

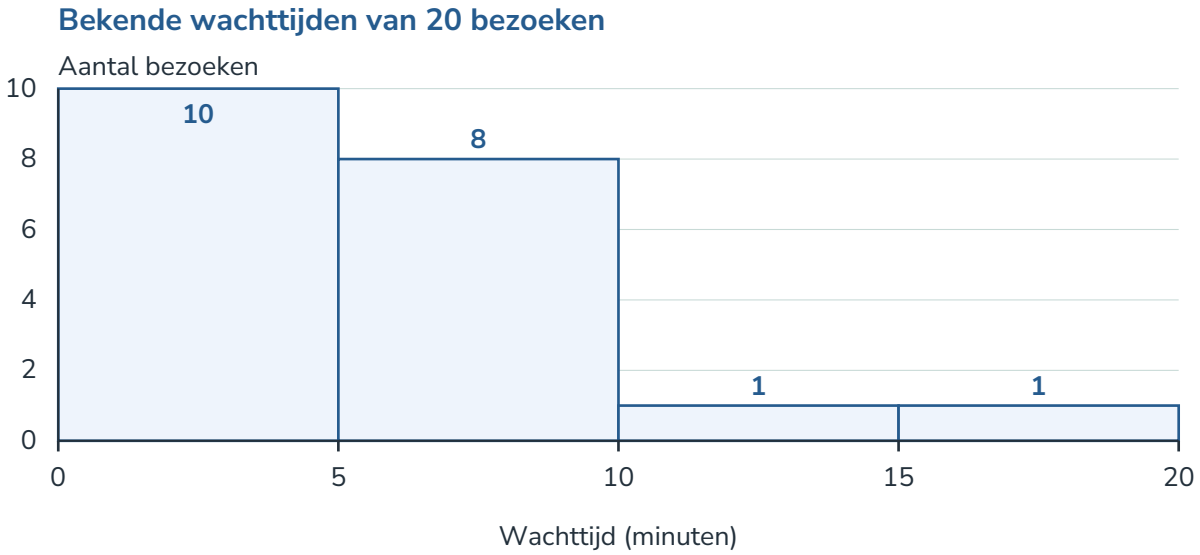
Stap	Excel	LibreOffice Calc
Selecteer	J8:K12	J8:K12
Voeg diagram toe	Invoegen > Kolomdiagram > 2D-kolom	Invoegen > Diagram > Kolom; 2D
Gegevens instellen	Gebruik eerste kolom als categorieën	Gegevensreeksen in kolommen; eerste rij en eerste kolom als label
Staven laten aansluiten	Gegevensreeks opmaken: tussenruimte/breedte van tussenruimte 0%	Gegevensreeks opmaken > Opties: afstand/tussenruimte 0%
Verticale schaal	Minimum 0, maximum 10, hoofdstap 2	Minimum 0, maximum 10, hoofdinterval 2

Kies als titel Bekende wachttijden van 20 bezoeken. De horizontale as heet Wachttijd (minuten), de verticale as Aantal bezoeken. Gebruik vier even brede, aansluitende staven met hoogten 10, 8, 1 en 1. Verwijder een overbodige legenda; kies geen 3D-effect. Omdat de klassen hier allemaal even breed zijn, mogen de hoogten de frequenties weergeven.

Een histogram toont de verdeling van één meetvariabele, geen reeks afzonderlijke bezoek-ID’s. Houd de klassen in numerieke volgorde. Een lijndiagram zou ten onrechte een opeenvolging tussen de bezoeken suggereren. De vijf onbekende tijden zijn geen klasse bij 0 minuten: noem ze in het onderschrift.

Onderschrift bij je figuur

Wachttijd vanaf aanmelden tot een beschikbare studieplek, in minuten. Van 25 bezoeken zijn 20 wachttijden bekend en vijf onbekend. Klassen zijn links inclusief en rechts exclusief. De weergegeven gegevens zijn voor de oefening samengesteld.



4. Antwoord in context

Van de 20 bekende wachttijden is 50% korter dan 5 minuten en 90% korter dan 10 minuten. De mediaan is 4,5 minuten; het gemiddelde is 5,45 minuten. De twee langere tijden van 12 en 18 minuten trekken het gemiddelde omhoog. Het histogram heeft een staart naar rechts. De middelste 50% van de bekende tijden ligt tussen 3 en 7 minuten.

Dit is een beschrijving van de bekende tijden in deze registratie. Over de vijf ontbrekende tijden weten we niet of ze kort of lang zijn. Zonder passende selectie en informatie over die ontbrekende waarden doen we geen uitspraak over alle bezoeken van het jaar en berekenen we hier geen betrouwbaarheidsinterval.

5. Controle

Controle	Verwachte uitkomst
Aantal getallen en handmatig aantal tijden	20
Som / aantal	$\frac{109}{20} = 5{,}45$
Middelste twee na sorteren	4 en 5; mediaan 4,5
Frequenties optellen	10 + 8 + 1 + 1 = 20
Percentages optellen	50 + 40 + 5 + 5 = 100%
Klassecontrole op een grens	5 hoort bij 5 tot 10, niet bij 0 tot 5
Extreme waarde	V09 blijft 18 min; bronformulier bevestigt de meting

Probeer nu zelf: een andere maat en indeling

Maak op een nieuwe kopie twee even brede klassen: 0 tot 10 en 10 tot 20 minuten, met dezelfde afspraak over de grenzen. Bereken beide frequenties en aandelen met formules en maak een histogram. Welke informatie is in deze grovere figuur minder goed zichtbaar?

Probeer nu zelf: gevoeligheid voor één lange tijd

Verander alleen op deze proefkopie de bevestigde tijd van V09 van 18 naar 28 minuten. Dit is een gevoeligheidsproef, geen correctie van de bron. Bereken opnieuw gemiddelde, mediaan en spreidingsbreedte. Pas de klassengrenzen aan zodat iedere bekende tijd in precies één klasse valt.

Bewaar het oorspronkelijke analyseblad en de proefkopie met duidelijke namen. Lever bron, wijzigingslog, verwerkingsformules, frequentietabel en diagram samen in; alleen een screenshot bewijst niet hoe je de uitkomsten hebt gemaakt.

Voorbeeld 8.2B

Een gemiddelde schatten

Een passende aselecte steekproef van 100 ritten geeft een gemiddelde van 28 minuten en S=5 minuten.

a Bereken en interpreteer het 95%-interval voor de gemiddelde ritduur.

1. Wat zoek ik?

Een interval voor het populatiegemiddelde.

2. Welke basis en aanpak kies ik, en waarom?

Met het formuleblad gebruik ik $\bar{x} \pm \frac{2S}{\sqrt{n}}$.

3. Berekening

$28 \pm 2 \times \frac{5}{\sqrt{100}} = 28 \pm 1$.

4. Antwoord in context

Het interval is [27;29] minuten voor het gemiddelde van de ritten in de onderzochte populatie.

5. Controle

De marge heeft de eenheid minuten en wordt kleiner bij een grotere steekproef; dit is geen bereik voor individuele ritten.

Voorbeeld 8.3A

Van mening naar meetbare vraag

Een team vraagt: Is onze afhaalservice snel? De tijd tussen aanmelden en overhandigen kan worden gemeten.

a Formuleer een toetsbare onderzoeksvraag met een meetafspraak.

1. Wat zoek ik?

Een vraag met een meetbare uitkomst en duidelijke populatie.

2. Welke basis en aanpak kies ik, en waarom?

Ik kies de wachttijd, een grens en een afgebakende week.

3. Berekening

Vraag: Welk aandeel afhalers in week 12 wacht langer dan 4 minuten? Meet vanaf aanmelden tot overhandigen.

4. Antwoord in context

Selecteer bezoeken over alle openingsuren volgens een vast aselect plan.

5. Controle

De grens is vastgelegd vóór de meting; direct geholpen betekent 0 minuten.

Voorbeeld 8.3B

Een conclusie aanscherpen

Van 150 vrijwillige bezoekers van een tentoonstellingsapp geven 120 een positief oordeel.

a Schrijf een conclusie die niet verder gaat dan de gegevens.

1. Wat zoek ik?

Een beschrijving van deze respondenten.

2. Welke basis en aanpak kies ik, en waarom?

Ik bereken het aandeel en bewaak de selectie.

3. Berekening

$\frac{120}{150} \times 100 = 80\%$.

4. Antwoord in context

80% van de 150 deelnemende appgebruikers oordeelt positief. Dit is geen bewezen percentage onder alle tentoonstellingsbezoekers.

5. Controle

Wie de app niet gebruikt of niet antwoordt kan anders oordelen; een grotere vrijwillige groep verhelpt dat niet vanzelf.