# 4. Statistische uitspraken doen

Boekje 4 havo wiskunde A, domein E: Statistiek

Uitwerkingen

# Inhoud

§ 4.3 Populatieproportie ..... 4
§ 4.4 Populatiegemiddelde ..... 5
§ 4.5 Verschil tussen twee groepen ..... 6
§ 4.6 Samenhang tussen twee kwantitatieve variabelen ..... 9
§ 4.7 Gemengde opgaven ..... 10
§4.9 Lessenserie: Statistiek op een groot gegevensbestand ..... 11
§4.10 Diagnostische computertoets ..... 24

## § 4.3 Populatieproportie

**Opgave 1**
[0,654; 0,746].

**Opgave 2**
[0,090; 0,096].

**Opgave 3**
a. Hoe hoger de betrouwbaarheid, hoe breder het interval.
b. 90%-BI: [0,833; 0,874]; 95%-BI: [0,829; 0,878]; 99%-BI: [0,822; 0,886].

**Opgave 4**
a. Duitse jongeren.
b. [0,747; 0,833].
c. Een leeftijd-geslachtsverdeling in de steekproef die overeenkomt met de bevolkingsstatistiek. Een verdeling over de verschillende schoolniveaus die overeenkomt met landelijke gegevens hierover.
d. Een enquête op een of slechts enkele specifieke plekken in het land.
Een enquête op een of slechts enkele specifieke scholen in het land.

**Opgave 5**
a. Ouderen in Nederland.
b. [0,175; 0,225].
c. n > 25.600.

**Opgave 6**
a. De 3200 fruittelers in Nederland.
b. [0,236; 0,424].
c. Alle fruittelers nummeren van 1 tot en met 3200 en vervolgens 100 willekeurige nummers trekken uit de getallen 1 tot en met 3200.
d. Alleen fruittelers uit een bepaalde regio in de steekproef.
Voornamelijk ‘grote’ of juist ‘kleine’ fruittelers in de steekproef.

## § 4.4 Populatiegemiddelde

**Opgave 7**

a. [4,36; 4,44].
b. [10,82; 11,18].

**Opgave 8**

a. 140.
b. [139,51; 140,49].

**Opgave 9**

a. Onderbouwklassen in het basisonderwijs in Nederland.
b. [23,03; 24,37].
c. [20,17; 21,63], de intervallen overlappen niet.
d. n > 25600.
e. Ministerie van Onderwijs; Centraal Bureau voor de Statistiek.

**Opgave 10**

Soort A: [598,21; 601,79].
Soort B: [1147,11; 1152,89].

**Opgave 11**

Aflezen gemiddelde is ongeveer 163 en standaardafwijking is ongeveer 6.
Daarmee kom je op [162,83; 163,17].

**Opgave 12**

[246,25; 253,75], dus 266 ligt er niet in.

## § 4.5 Verschil tussen twee groepen

**Opgaven 13 en 14**

Bijvoorbeeld:

| | **Jongens** | **Meisjes** | **Totaal** |
|---|---|---|---|
| A | 25 | 25 | 50 |
| B | 25 | 25 | 50 |
| **Totaal** | 50 | 50 | 100 |

phi = 0

| | **Jongens** | **Meisjes** | **Totaal** |
|---|---|---|---|
| A | 0 | 50 | 50 |
| B | 50 | 0 | 50 |
| **Totaal** | 50 | 50 | 100 |

phi = 1

| | **Jongens** | **Meisjes** | **Totaal** |
|---|---|---|---|
| A | 30 | 30 | 60 |
| B | 20 | 20 | 40 |
| **Totaal** | 50 | 50 | 100 |

phi = 0

| | **Jongens** | **Meisjes** | **Totaal** |
|---|---|---|---|
| A | 40 | 10 | 50 |
| B | 10 | 40 | 50 |
| **Totaal** | 50 | 50 | 100 |

phi = 0,6

**Opgave 15**

De voorspelling zou kunnen zijn:

| | |
|---|---|
| Gering | Gering |
| Groot | Groot |
| Groot | Groot |

De waarden voor *phi* zijn:

| | |
|---|---|
| 0 | 0 |
| 1 | 1 |
| 0,723 | 0,877 |

**Opgave 16**

De voorspellingen zouden kunnen zijn:

| | |
|---|---|
| Middelmatig | Gering |
| Gering | Gering |

De waarden voor *phi* zijn:

| | |
|---|---|
| 0,145 | −0,15 |
| 0,023 | 0,004 |

**Opgave 17**

Phi = −0,18, dus het verschil is gering.

**Opgave 18**

Phi = −0,082, dus het verschil is gering.

**Opgave 19**

| | **M** | **V** | **T** |
|---|---|---|---|
| Wel kleurenblind | 65 | 7 | 72 |
| Niet kleurenblind | 535 | 393 | 928 |
| Totaal | 600 | 400 | 1000 |

Phi = 0,17, dus het verschil is gering.

**Opgave 20**

*Phi* = 0,063, dus het verschil is gering.

**Opgave 21**

*Phi* = 0,128, dus het verschil is gering.

**Opgave 22**
MaxV$_{cp}$ ≈ 8,6, dus het verschil is gering.

**Opgave 23**
MaxV$_{cp}$ ≈ 6, dus het verschil is gering.

**Opgave 24**
E = 1,12, dus het verschil is groot.

**Opgave 25**
E = 0,8, dus het verschil is middelmatig.

**Opgave 26**
De boxen overlappen en de medianen liggen niet in de andere box, dus het verschil is middelmatig.

**Opgave 27**
De boxen overlappen elkaar niet, dus het verschil is groot.

**Opgave 28**
E = 0,106, dus het verschil is gering.
De boxen overlappen elkaar en de mediaan van de ene ligt in de box van de andere,
dus het verschil is gering.

**Opgave 29**
a. Phi = 0,18 dus het verschil is gering.
b. Uit de tabel blijkt dat in lesgroep 1 29% een onvoldoende (dus lager dan 5,5) heeft.
   Dit komt overeen met de grafiek.
   Uit de tabel blijkt dat in lesgroep 2 14% een onvoldoende (dus lager dan 5,5) heeft.
   Dit komt overeen met de grafiek.
c. Lesgroep 2, want het percentage onvoldoendes is lager (zie tabel) en de grafiek ligt rechts ten opzichte van lesgroep 1.
d. MaxV$_{cp}$ ≈ 25, dus het verschil is middelmatig.

**Opgave 30**
a. In figuur 1 omdat de grafiek van de jongens rechts ligt ten opzichte van de grafiek van de meisjes.
   In figuur 2 zitten er in de klasse 0-5 uur meer meisjes dan jongens, in de andere klassen zitten juist meer jongens dan meisjes.
   In figuur 3 staat dat het gemiddelde voor de jongens 8,7 is en voor de meisjes 3,9.
b. MaxV$_{cp}$ ≈ 38, dus het verschil is middelmatig.
c. E = 0,69 dus het verschil is middelmatig.

**Opgave 31**

Bloeddruk is een kwantitatieve variabele, dus de effectgrootte gebruiken of de boxplots vergelijken. Het meest geschikt is de effectgrootte. Daarvoor eerst het gemiddelde en de standaardafwijking van beide groepen uitrekenen.

Neem aan: onder de 90 -> de waarde is 87,5

Dan kom je op:

| | **Non-user** | **User** |
|---|---|---|
| Gemiddelde | 117,1 | 120,6 |
| Standaardafwijking | 12,77 | 12,01 |

E = 0,28, dus het verschil is gering.

## § 4.6 Samenhang tussen twee kwantitatieve variabelen

**Opgave 32**
De samenhang lijkt zwak positief, dus de correlatie zou ongeveer 0,1 kunnen zijn.

**Opgave 33**
Matig positieve samenhang.

**Opgave 34**
a. TV = 14,31.
b. TV = 14,91.
c. TV = 16,11.

**Opgave 35**
a. S = 40000.
b. A = 667.

**Opgave 36**
De puntenwolk suggereert een matig negatieve samenhang, dus hoe hoger het inkomen per persoon in een staat, hoe minder sociale en gezondheidsproblemen er zijn in de betreffende staat.

**Opgave 37**
Als je een samenhang vindt tussen twee (kwantitatieve) variabelen, hoeft er nog geen causale relatie (=oorzaak-gevolgrelatie) te bestaan tussen deze twee variabelen.

## § 4.7 Gemengde opgaven

**Opgave 38**

a. Jongeren in Nederland.
b. Ordinaal.
c. [0,900; 0,948].
d. Breder.

**Opgave 39**

a. Mensen die een facelift laten uitvoeren.
b. 50.
c. [2,874; 3,326].
d. n > 1024.

**Opgave 40**

a. $MaxV_{cp} = 6$.
b. De boxplots zijn hetzelfde.
c. E = 0,09.
d. Alle maten geven aan dat het verschil gering is.

**Opgave 41**

a. Phi = −0,013, dus het verschil is gering.
b. E = 0,253, dus het verschil is gering.

**Opgave 42**

a. Leeftijd: kwantitatief; herkomst: nominaal; type school: ordinaal; wiskundecijfer: kwantitatief.
b. E = 0,545, dus het verschil is middelmatig.
c. Andere lesmethoden, andere docenten.

**Opgave 43**

a. Phi = 0,09, dus het verschil is gering.
b. $MaxV_{cp} = 36$, dus het verschil is middelmatig.
c. Ten aanzien van de werkwijze geldt dat de variabelen niet kwantitatief zijn terwijl dit wel zou moeten voor een puntenwolk en een trendlijn.
   Ten aanzien van de conclusie geldt dat de gevonden samenhang nog niet hoeft te betekenen dat er sprake is van een oorzaak-gevolgrelatie.

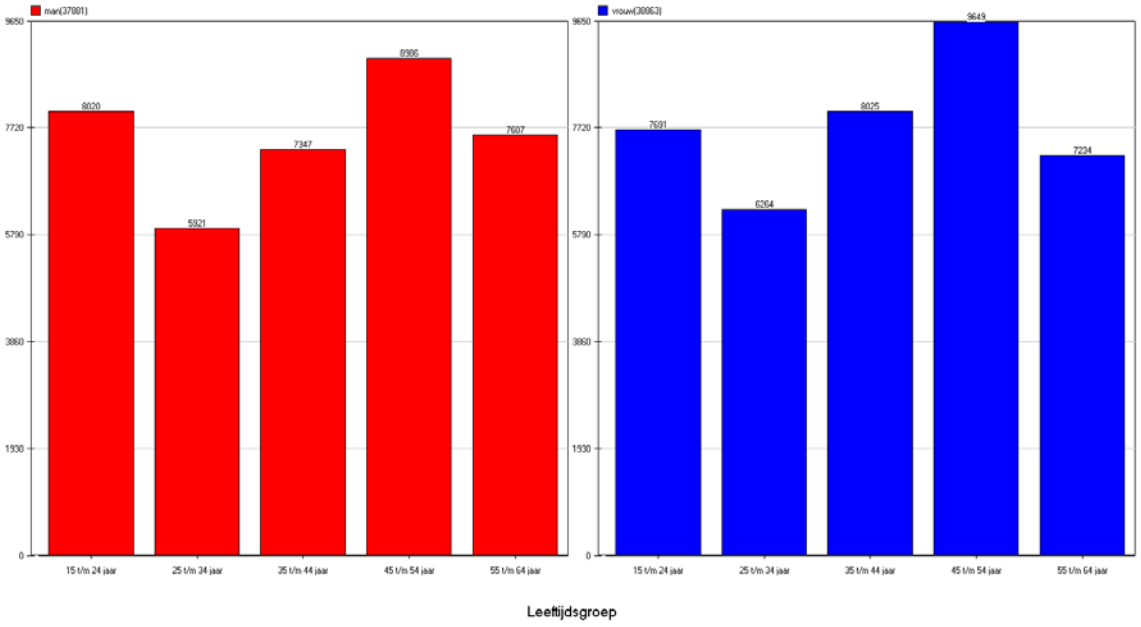
## §4.9 Lessenserie: Statistiek op een groot gegevensbestand

**LES 1**

a. Het bestand bevat 76746 records en de onderstaande 10 variabelen. In de laatste kolom staat het meetniveau van de variabele.

| | | |
|---|---|---|
| 1. | Geslacht | nominaal. |
| 2. | Leeftijdsgroep | ordinaal. |
| 3. | Herkomst | nominaal. |
| 4. | Beroepsbevolking | nominaal. |
| 5. | Arbeidsduur per week | ordinaal. |
| 6. | Meer of minder willen werken | ordinaal. |
| 7. | Bereidheid tot werken | nominaal. |
| 8. | Beroepsrichting | nominaal. |
| 9. | Opleidingsniveau | ordinaal. |
| 10. | Onderwijsrichting | nominaal. |

b. Staafdiagram voor leeftijdsgroep gesplitst op geslacht:




c. Kruistabel van opleidingsniveau met daarin percentages:

| **Opleidingsniveau** | **Freq.** | **Perc.** |
|---|---|---|
| Laag | 24566 | 32,14 |
| Midden | 31089 | 40,67 |
| Hoog | 20779 | 27,19 |
| **Totaal** | 76434 | 100% |

Je kunt aflezen dat 27,19% hoog is opgeleid.

Om na te gaan of er een verschil is tussen mannen en vrouwen kunnen we deze tabel splitsen op geslacht. Dat geeft onderstaande tabel:

| **Geslacht** | **Man** | | **Vrouw** | | |
|---|---|---|---|---|---|
| **Opleidingsniveau** | **Freq.** | **Perc.** | **Freq.** | **Perc.** | **Totaal** |
| Laag | 12055 | 31,97 | 12510 | 32,31 | 24565 |
| Midden | 15130 | 40,12 | 15959 | 41,21 | 31089 |
| Hoog | 10525 | 27,91 | 10253 | 26,48 | 20778 |
| **Totaal** | 37710 | 100% | 38722 | 100% | 76432 |

Je ziet dat de percentages voor mannen en vrouwen nauwelijks van elkaar verschillen, dus er is nauwelijks verschil tussen mannen en vrouwen voor wat betreft hun opleidingsniveau.

d.

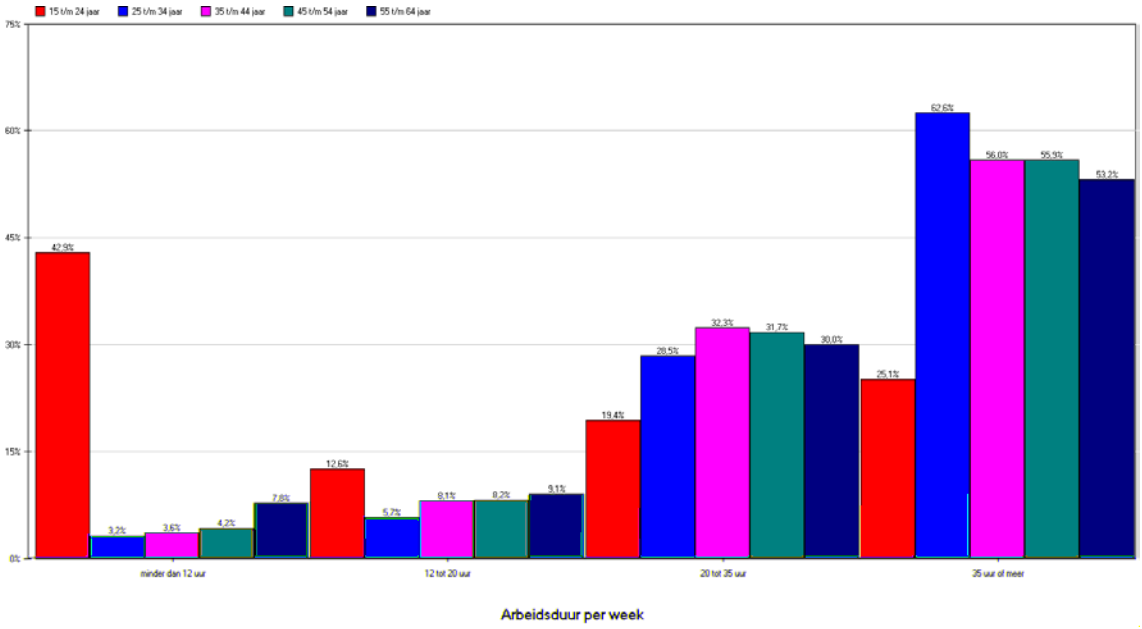
| **Herkomst** | **Freq.** | **Perc.** |
|---|---|---|
| Autochtoon | 63079 | 82,19 |
| Westers allochtoon | 6222 | 8,11 |
| Turken en Marokkanen | 2511 | 3,27 |
| Antillianen, Arubanen en Surinamers | 2187 | 2,85 |
| Overig niet-westerse allochtonen | 2677 | 3,49 |
| Onbekend | 70 | 0,09 |
| **Totaal** | 76746 | 100% |

e. Kruistabel van *meer of minder willen werken* en *geslacht* met daarin percentages:

| **Geslacht** | **Man** | | **Vrouw** | | |
|---|---|---|---|---|---|
| Meer of minder willen werken | **Freq.** | **Perc.** | **Freq.** | **Perc.** | **Totaal** |
| Meer willen werken | 1696 | 4,48 | 2723 | 7,01 | 4419 |
| Minder willen werken | 2051 | 5,41 | 2010 | 5,17 | 4061 |
| Niet meer of minder willen werken | 23819 | 62,88 | 18263 | 46,99 | 42082 |
| Vraag niet gesteld | 10315 | 27,23 | 15867 | 40,83 | 26182 |
| **Totaal** | 37881 | 100% | 38863 | 100% | 76744 |

Je kunt aflezen dat 4,48% van de mannen meer wil werken.

f. Staafdiagram van de arbeidsduur per week in procenten gesplitst op leeftijdsgroep:




Je ziet dat de arbeidsduur per week voor jongeren anders is dan voor ouderen.
Bijvoorbeeld: 42,9% van jongeren werkt minder dan 12 uur per week; voor de ouderen is dit 7,8%.
Ander voorbeeld: 25,1% van de jongeren werkt 35 uur of meer per week, voor de ouderen is dit 53,2%.

g. Je maakt een kruistabel van onderwijsrichting en beroepsrichting met percentages en vindt daarin dat ongeveer 50% van de personen die opgeleid zijn tot leraar werken als docent of staf onderwijs. Omdat de kruistabel nogal groot is om in deze uitwerking volledig op te nemen, is hieronder eerst een selectiefilter toegepast om zo de personen te selecteren met onderwijsrichting leraren. Dit geeft een selectie van 3631 personen. Voor deze groep is een frequentietabel gemaakt van de beroepsrichting.

| **Beroepsrichting** | **Freq.** | **Perc.** |
|---|---|---|
| Geen werkkring | 654 | 18,01 |
| Docenten en staf onderwijs | 1828 | 50,34 |
| Agrarisch, exact | 23 | 0,63 |
| Technisch | 126 | 3,47 |
| Transport, communicatie en verkeer | 51 | 1,40 |
| Medisch en paramedisch | 100 | 2,75 |
| Economisch, administratief, commercieel | 375 | 10,33 |
| Juridisch, bestuurlijk, openbare orde, taal en cultuur | 67 | 1,85 |
| Gedrag en maatschappij | 119 | 3,28 |
| Persoonlijke en sociale verzorging | 161 | 4,43 |
| Management en algemeen | 107 | 2,95 |
| Onbekend | 20 | 0,55 |
| **Totaal** | 3631 | 100% |

Je kunt aflezen dat ongeveer de helft (50%) van de personen die opgeleid zijn tot leraar werkzaam is als docent of staf onderwijs.
Een behoorlijk deel van de personen die opgeleid zijn tot leraar heeft geen werkkring (ongeveer 18%) en ongeveer 10% is werkzaam in de richting economisch, administratief, commercieel.

**LES 2**

a.

| Meer of minder willen werken | Freq. | Perc. |
|---|---|---|
| Meer willen werken | 322 | 8,98 |
| Minder willen werken | 353 | 9,84 |
| Niet meer of minder willen werken | 2637 | 73,54 |
| Vraag niet gesteld | 274 | 7,64 |
| **Totaal** | 3586 | 100% |

Je kunt aflezen dat 9,84% minder wil werken.

b. Maak eerst de variabele *herkomst2* en selecteer daarna via het selectiefilter de werkzame en werkloze beroepsbevolking.
Maak vervolgens een frequentietabel van de beroepsbevolking gesplitst op *herkomst2*:

| HERKOMST2 | Allochtoon | | Autochtoon | | |
|---|---|---|---|---|---|
| **Beroepsbevolking** | **Freq.** | **Perc.** | **Freq.** | **Perc.** | **Totaal** |
| Werkzame beroepsbevolking | 7806 | 88,45 | 43181 | 94,47 | 50987 |
| Werkloze beroepsbevolking | 1019 | 11,55 | 2529 | 5,53 | 3548 |
| Niet beroepsbevolking | 0 | 0,00 | 0 | 0,00 | 0 |
| N.v.t. | 0 | 0,00 | 0 | 0,00 | 0 |
| Totaal | 8825 | 100% | 45710 | 100% | 54535 |

Je kunt aflezen dat 11,55% van de allochtone beroepsbevolking werkloos is.
Voor de autochtone beroepsbevolking is dit 5,53%.

c. Kentallen van *leeftijd*:

| Variabele | LEEFTIJD |
|---|---|
| Aantal waarnemingen | 76746 |
| Gemiddelde | 40,6 |
| Mediaan | 40,0 |
| Modus | 50 |
| Minimum | 20 |
| Maximum | 60 |
| SDn-1 | 14,11 |
| SDn | 14,11 |
| VARn-1 | 199,02 |
| VARn | 100,02 |
| Maximum | 60 |

d. Kentallen van *arbeidsduur2*:

| **Variabele** | **ARBEIDSDUUR2** |
|---|---|
| Aantal waarnemingen | 57343 |
| Gemiddelde | 29,7 |
| Mediaan | 38 |
| Modus | 38 |
| Minimum | 6 |
| Maximum | 38 |
| SDn-1 | 10,68 |
| SDn | 10,68 |
| VARn-1 | 114,14 |
| VARn | 114,14 |

Merk op dat bij de berekening van de kentallen van *arbeidsduur2* alleen gebruikt wordt gemaakt van de records waarvan een *arbeidsduur2* bekend is.

**LES 3**

a. Selecteer eerst via het selectiefilter de beroepsbevolking.
Maak vervolgens een frequentietabel van de beroepsbevolking met percentages:

| **Beroepsbevolking** | **Freq.** | **Perc.** |
|---|---|---|
| Werkzame beroepsbevolking | 51048 | 93,50 |
| Werkloze beroepsbevolking | 3551 | 6,50 |
| Niet beroepsbevolking | 0 | 0,00 |
| N.v.t. | 0 | 0,00 |
| **Totaal** | 54599 | 100% |

Van de beroepsbevolking (n = 54599) is 3551 werkloos.
De proportie werklozen in de steekproef (p) is dus 3551 / 54599 = 0,065.
Het 95%-betrouwbaarheidsinterval kan berekend worden met:

Dit geeft het interval [0,063; 0,067].

b. Selecteer eerst de personen werkzaam in het onderwijs via het selectiefilter.
Reken daarna de kentallen van *leeftijd* uit.

| **Variabele** | **LEEFTIJD** |
|---|---|
| Aantal waarnemingen | 3586 |
| Gemiddelde | 43,8 |
| Mediaan | 50,0 |
| Modus | 50 |
| Minimum | 20 |
| Maximum | 60 |
| SDn-1 | 12,72 |
| SDn | 12,72 |
| VARn-1 | 161,92 |
| VARn | 161,88 |

De steekproef bestaat uit 3586 personen (n).
Het steekproefgemiddelde is 43,8 jaar ( ) en de standaardafwijking is 12,72 (S).
Het 95%-betrouwbaarheidsinterval voor de gemiddelde leeftijd in de populatie werkzaam in het onderwijs wordt gegeven door:

Dit geeft het interval [43,4;44,2].

**LES 4**

a. Selecteer de beroepsbevolking via het selectiefilter. Maak een frequentietabel van de beroepsbevolking met percentages en splits deze op opleidingsniveau:

b.

| **Opleidingsniveau** | **Laag** | | **Midden** | | **Hoog** | | |
|---|---|---|---|---|---|---|---|
| **Beroepsbevolking** | **Freq.** | **Perc.** | **Freq.** | **Perc.** | **Freq.** | **Perc.** | **Totaal** |
| Werkzame beroepsbevolking | 11403 | 90,23 | 22091 | 93,66 | 17318 | 95,52 | 50812 |
| Werkloze beroepsbevolking | 1235 | 9,77 | 1496 | 6,34 | 813 | 4,48 | 3544 |
| Niet beroepsbevolking | 0 | 0,00 | 0 | 0,00 | 0 | 0,00 | 0 |
| N.v.t. | 0 | 0,00 | 0 | 0,00 | 0 | 0,00 | 0 |
| **Totaal** | 12638 | 100% | 23587 | 100% | 18131 | 100% | 54356 |

Maak een 2x2-tabel:

| | **Laag** | **Hoog** | **Totaal** |
|---|---|---|---|
| Werkzaam | 11403 | 17318 | 28721 |
| Werkloos | 1235 | 813 | 2048 |
| **Totaal** | 12638 | 18131 | 30769 |

Volgens de vuistregels is het verschil dus gering.

c. Selecteer personen met de beroepsrichting onderwijs en maak een cumulatieve frequentietabel van de leeftijdsgroepen:

d.

| Leeftijdsgroep | Freq. | Perc. | Cumul. | Cumul.% |
|---|---|---|---|---|
| 15 t/m 24 jaar | 294 | 8,20 | 294 | 8,20 |
| 25 t/m 34 jaar | 752 | 20,97 | 1046 | 29,17 |
| 35 t/m 44 jaar | 664 | 18,52 | 1710 | 47,69 |
| 45 t/m 54 jaar | 1032 | 28,78 | 2742 | 76,46 |
| 55 t/m 64 jaar | 844 | 23,54 | 3586 | 100,00 |
| **Totaal** | 3586 | 100% | 3586 | 100% |

Selecteer personen met een andere beroepsrichting en maak een cumulatieve frequentietabel van de leeftijdsgroepen:

| Leeftijdsgroep | Freq. | Perc. | Cumul. | Cumul.% |
|---|---|---|---|---|
| 15 t/m 24 jaar | 15417 | 21,07 | 15417 | 21,07 |
| 25 t/m 34 jaar | 11433 | 15,63 | 26850 | 36,70 |
| 35 t/m 44 jaar | 14709 | 20,11 | 41559 | 56,81 |
| 45 t/m 54 jaar | 17604 | 24,06 | 59163 | 80,87 |
| 55 t/m 64 jaar | 13997 | 19,13 | 73160 | 100,00 |
| **Totaal** | 73160 | 100% | 73160 | 100% |

Nu kun je kijken naar het cumulatieve percentageverschil:
15 t/m 24 jaar: 12,87%
25 t/m 34 jaar: 7,53%
35 t/m 44 jaar: 9,12%
45 t/m 54 jaar: 4,41%
55 t/m 64 jaar: 0%

Het maximale cumulatieve percentageverschil is gelijk aan 12,87%.
Volgens de vuistregels is er dan sprake van een gering verschil.

b. Splits op leeftijdsgroep:

| Leeftijds-groep | 15 t/m 24 jaar | | 25 t/m 34 jaar | | 35 t/m 44 jaar | | 45 t/m 55 jaar | | 55 t/m 65 jaar | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Beroeps-bevolking** | **Freq.** | **Perc.** | **Freq.** | **Perc.** | **Freq.** | **Perc.** | **Freq.** | **Perc.** | **Freq.** | **Perc.** | **Totaal** |
| Werkzame beroeps-bevolking | 5521 | 86,56 | 10049 | 93,95 | 12586 | 94,72 | 14859 | 95,01 | 8033 | 93,43 | 51048 |
| Werkloze beroeps-bevolking | 857 | 13,44 | 647 | 6,05 | 702 | 5,28 | 780 | 4,99 | 565 | 6,57 | 3551 |
| Niet beroeps-bevolking | 0 | 0,00 | 0 | 0,00 | 0 | 0,00 | 0 | 0,00 | 0 | 0,00 | 0 |
| N.v.t. | 0 | 0,00 | 0 | 0,00 | 0 | 0,00 | 0 | 0,00 | 0 | 0,00 | 0 |
| **Totaal** | 6378 | 100% | 10696 | 100% | 13288 | 100% | 15639 | 100% | 8598 | 100% | 54599 |

Van de jongeren is 13,44% werkloos en van de ouderen is dit 6,57%.

c. Hef de splitsing op leeftijdsgroep op en splits op opleidingsniveau:

| Opleidingsniveau | Laag | | Midden | | Hoog | | |
|---|---|---|---|---|---|---|---|
| **Beroepsbevolking** | **Freq.** | **Perc.** | **Freq.** | **Perc.** | **Freq.** | **Perc.** | **Totaal** |
| Werkzame beroepsbevolking | 11403 | 90,23 | 22091 | 93,66 | 17318 | 95,52 | 50812 |
| Werkloze beroepsbevolking | 1235 | 9,77 | 1496 | 6,34 | 813 | 4,48 | 3544 |
| Niet beroepsbevolking | 0 | 0,00 | 0 | 0,00 | 0 | 0,00 | 0 |
| N.v.t. | 0 | 0,00 | 0 | 0,00 | 0 | 0,00 | 0 |
| **Totaal** | 12638 | 100% | 23587 | 100% | 18131 | 100% | 54356 |

Van de laagopgeleiden is 9,77% werkloos en van de hoogopgeleiden is dit 4,48%.

d. Selecteer op beroepsbevolking en jongeren via het selectiefilter.
Frequentietabel van beroepsbevolking en splitsen op opleidingsniveau:

| **Opleidingsniveau** | **Laag** | **Mid2.3(ew 9** |
|---|---|---|

Selecteer van beroepsbevolking en ouderen via het selectiefilter.
Frequentietabel van beroepsbevolking gesplitst op opleidingsniveau:

| | Laag | | Midden | | Hoog | | |
|---|---|---|---|---|---|---|---|
| **Beroepsbevolking** | **Freq.** | **Perc.** | **Freq.** | **Perc.** | **Freq.** | **Perc.** | **Totaal** |
| Werkzame beroepsbevolking | 2239 | 93,10 | 2914 | 92,74 | **2831** | 94,30 | 7984 |
| Werkloze beroepsbevolking | 166 | 6,90 | 228 | 7,26 | **171** | 5,70 | 565 |
| Niet beroepsbevolking | 0 | 0,00 | 0 | 0,00 | 0 | 0,00 | 0 |
| N.v.t. | 0 | 0,00 | 0 | 0,00 | 0 | 0,00 | 0 |
| **Totaal** | 2405 | 100% | 3142 | 100% | 3002 | 100% | 8549 |

Maak een 2x2 tabel:

| | **Jong-laag** | **Oud-hoog** | **Totaal** |
|---|---|---|---|
| **Werkzaam** | 1865 | 2831 | 4696 |
| **Werkloos** | 423 | 171 | 594 |
| **Totaal** | 2288 | 3002 | 5290 |

Het verschil ligt dus precies op de grens tussen gering en middelmatig.

**LES 7**

Frequentietabel van opleidingsniveau gesplitst op leeftijdsgroep:

| **Leeftijds-groep** | **15 t/m 24 jaar** | | **25 t/m 34 jaar** | | **35 t/m 44 jaar** | | **45 t/n 54 jaar** | | **55 t/m 64 jaar** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Opleidings-niveau** | **Freq.** | **Perc.** | **Freq.** | **Perc.** | **Freq.** | **Perc.** | **Freq.** | **Perc.** | **Freq.** | **Perc.** | **Totaal** |

Om na te gaan of dit verschil gering, middelmatig of groot is, berekenen we phi.

2x2-tabel:

| | **Man** | **Vrouw** | **Totaal** |
|---|---|---|---|
| **Niet hoog** | 27185 | 28469 | 55654 |
| **Hoog** | 10525 | 10253 | 20778 |
| **totaal** | 37710 | 38722 | 76432 |

Het verschil is dus gering.

Een alternatieve uitwerking is om het maximale cumulatieve percentageverschil te berekenen.

Tabel met cumulatieve percentages:

| | **Man** | **Vrouw** | **Verschil** |
|---|---|---|---|
| **Laag** | 32,0 | 32,3 | 0,3 |
| **Midden** | 72,1 | 73,5 | 1,4 |
| **Hoog** | 100 | 100 | 0 |

Het maximale cumulatieve percentageverschil is 1,4%. Het verschil is dus gering.

Er is wel een opvallend verschil voor de verschillende leeftijdsgroepen.
Dit wordt zichtbaar als een kruistabel maakt voor opleidingsniveau en geslacht en deze splitst op leeftijdsgroep.

Kruistabel voor leeftijdsgroep 15 t/m 24 jaar:

| | **Geslacht** | | |
|---|---|---|---|
| **Opleidingsniveau** | **Man** | **Vrouw** | **Totaal** |
| Laag | 4437 | 3843 | 8280 |
| Midden | 3087 | 3080 | 6167 |
| Hoog | 485 | 761 | 1246 |
| **Totaal** | 8009 | 7684 | 15693 |

% hoogopgeleide mannen: 6
% hoogopgeleide vrouwen: 10

Kruistabel voor leeftijdsgroep 25 t/m 34 jaar:

| | **Geslacht** | | |
|---|---|---|---|
| **Opleidingsniveau** | **Man** | **Vrouw** | **Totaal** |
| Laag | 1260 | 1047 | 2307 |
| Midden | 2617 | 2576 | 5193 |
| Hoog | 2029 | 2622 | 4651 |
| **Totaal** | 5906 | 6245 | 12151 |

% hoogopgeleide mannen: 34
% hoogopgeleide vrouwen: 42

Kruistabel voor leeftijdsgroep 35 t/m 44 jaar:

| | Geslacht | | |
|---|---|---|---|
| **Opleidingsniveau** | **Man** | **Vrouw** | **Totaal** |
| Laag | 1746 | 1718 | 3464 |
| Midden | 3022 | 3628 | 6650 |
| Hoog | 2533 | 2639 | 5172 |
| **Totaal** | 7301 | 7985 | 15286 |

% hoogopgeleide mannen: 35

% hoogopgeleide vrouwen: 33

Kruistabel voor leeftijdsgroep 45 t/m 54 jaar:

| | Geslacht | | |
|---|---|---|---|
| **Opleidingsniveau** | **Man** | **Vrouw** | **Totaal** |
| Laag | 2353 | 2759 | 5112 |
| Midden | 3599 | 4277 | 7876 |
| Hoog | 2982 | 2572 | 5554 |
| **Totaal** | 8934 | 9608 | 18542 |

% hoogopgeleide mannen: 33

% hoogopgeleide vrouwen: 27

Kruistabel voor leeftijdsgroep 55 t/m 64 jaar:

| | Geslacht | | |
|---|---|---|---|
| **Opleidingsniveau** | **Man** | **Vrouw** | **Totaal** |
| Laag | 2259 | 3143 | 5402 |
| Midden | 2805 | 2398 | 5203 |
| Hoog | 2496 | 1659 | 4155 |
| **Totaal** | 7560 | 7200 | 14760 |

% hoogopgeleide mannen: 33

% hoogopgeleide vrouwen: 23

Onder ouderen is het percentage hoogopgeleide mannen dus groter dan het percentage hoogopgeleide vrouwen, maar onder jongeren is dat andersom.

## §4.10 Diagnostische computertoets

**Opgave 1 (3 punten)**

| Herkomst | Freq. | Perc. |
|---|---|---|
| Autochtoon | 63079 | 82,19 |
| Westers allochtoon | 6222 | 8,11 |
| Turken en Marokkanen | 2511 | 3,27 |
| Antillianen, Arubanen en Surinamers | 2187 | 2,85 |
| Overig niet-westerse allochtonen | 2677 | 3,49 |
| Onbekend | 70 | 0,09 |
| **Totaal** | 76746 | 100% |

82,19% van de mensen in het gegevensbestand is dus autochtoon.

**Opgave 2 (3 punten)**

Kentallen van de variabele *arbeidsduur2*:

| Variabele | ARBEIDSDUUR2 |
|---|---|
| Aantal waarnemingen | 57343 |
| Gemiddelde | 29,7 |
| Mediaan | 38 |
| Modus | 38 |
| Minimum | 6 |
| Maximum | 38 |
| SDn-1 | 10,68 |
| SDn | 10,68 |
| VARn-1 | 114,14 |
| VARn | 114,14 |

De gemiddelde arbeidsduur is 29,7 uur en de standaardafwijking is 10,68 uur.

**Opgave 3 (4 punten)**

Kruistabel voor leeftijdsgroep en geslacht met optie totaal-percentages:

| | Geslacht | | |
|---|---|---|---|
| **Leeftijdsgroep** | **Man** | **Vrouw** | **Totaal** |
| 15 t/m 24 jaar | 10,45% | 10,02% | 20,47% |
| 25 t/m 34 jaar | 7,72% | 8,16% | 15,88% |
| 35 t/m 44 jaar | 9,57% | 10,46% | 20,03% |
| 45 t/m 54 jaar | 11,71% | 12,57% | 24,28% |
| 55 t/m 64 jaar | 9,91% | 9,43% | 19,34% |
| **Totaal** | 49,36% | 50,64% | 100% |

Je kunt aflezen dat 10,02% + 8,16% = 18,18% vrouw is en jonger dan 35 jaar.

**Opgave 4 (6 punten)**

Kruistabel van beroepsrichting en geslacht met aantallen:

| | Geslacht | | |
|---|---|---|---|
| **Beroepsrichting** | **Man** | **Vrouw** | **Totaal** |
| Geen werkkring | 7939 | 11876 | 19815 |
| Docenten en staf onderwijs | 1302 | 2284 | 3586 |
| Agrarisch, exact | 1350 | 487 | 1837 |
| Technisch | 8670 | 835 | 9505 |
| Transport, communicatie en verkeer | 2735 | 574 | 3309 |
| Medisch en paramedisch | 738 | 3799 | 4537 |
| Economisch, administratief, commercieel | 9560 | 9855 | 19415 |
| Juridisch, bestuurlijk, openbare orde, taal en cultuur | 1548 | 983 | 2531 |
| Gedrag en maatschappij | 785 | 1795 | 2580 |
| Persoonlijke en sociale verzorging | 1601 | 5328 | 6929 |
| Management en algemeen | 1153 | 676 | 1829 |
| Onbekend | 500 | 371 | 871 |
| **Totaal** | 37881 | 38863 | 76744 |

Laten we hier aannemen dat met gezondheidszorg de beroepsrichting medisch en paramedisch wordt bedoeld.
In deze beroepsrichting werken 4537 mensen, waaronder 3799 vrouwen.
Dus n=4537 en p=3799/4537=0,837.
Het 95%-betrouwbaarheidsinterval voor de proportie vrouwen werkzaam in de gezondheidszorg wordt dan gegeven door:

, dat is .

Dus het gevraagde 95%-betrouwbaarheidsinterval is [0,826; 0,848].

**Opgave 5 (6 punten)**

Selectiefilter op beroepsrichting medisch of paramedisch.

Kentallen van leeftijd:

| **Variabele** | **LEEFTIJD** |
|---|---|
| Aantal waarnemingen | 4537 |
| Gemiddelde | 42,4 |
| Mediaan | 40 |
| Modus | 50 |
| Minimum | 20 |
| Maximum | 60 |
| SDn-1 | 12,09 |
| SDn | 12,09 |
| VARn-1 | 146,15 |
| VARn | 146,11 |

Het 95%-betrouwbaarheidsinterval voor de gemiddelde leeftijd van mensen die werkzaam zijn in de gezondheidszorg is dan:

, dus [42,04; 42,76].

**Opgave 6 (24 punten: 8 punten per uitspraak)**

Selectiefilter aanhouden.

a. **Arbeidsduur**

We berekenen de effectgrootte. Immers, we kijken naar het verschil tussen twee groepen (man/vrouw) op een kwantitatieve variabele (*arbeidsduur2*).

| **Geslacht** | **Man** | **Vrouw** |
|---|---|---|
| Aantal waarnemingen | 738 | 3799 |
| Gemiddelde | 34,7 | 26,4 |
| Mediaan | 38,0 | 28 |
| Modus | 38 | 28 |
| Minimum | 6 | 6 |
| Maximum | 38 | 38 |
| SDn-1 | 6,74 | 8,36 |
| SDn | 6,73 | 8,36 |
| VARn-1 | 45,40 | 69,88 |
| VARn | 45,34 | 69,86 |

De effectgrootte is dan:

Volgens de vuistregels is het verschil tussen de twee groepen dan groot.
De uitspraak wordt dus ondersteund door de gegevens.

b. **Onderwijsniveau**

We berekenen het maximale cumulatieve percentageverschil.
Immers, we kijken naar het verschil tussen twee groepen op een ordinale variabele.

Frequentietabel van opleidingsniveau gesplitst op geslacht en met cumulatieve percentages:

| **Geslacht** | **Man** | | | | **Vrouw** | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Opleidings-niveau** | **Freq.** | **Perc.** | **Cumul.** | **Cum.%** | **Freq.** | **Perc.** | **Cumul.** | **Cum.%** | **Totaal** |
| Laag | 39 | 5,28 | 39 | 5,28 | 246 | 6,49 | 246 | 6,49 | 285 |
| Midden | 205 | 27,78 | 244 | 33,06 | 2230 | 58,84 | 2476 | 65,33 | 2435 |
| Hoog | 494 | 66,94 | 738 | 100,00 | 1314 | 34,67 | 3790 | 100,00 | 1808 |
| **Totaal** | 738 | 100% | 738 | 100% | 3790 | 100% | 3790 | 100% | 4528 |

Het maximale cumulatieve percentageverschil zit bij het middenniveau en is 32,27%.
Volgens de vuistregels is het verschil tussen de twee groepen dan middelmatig. De uitspraak wordt niet ondersteund door de gegevens.

c. **Proportie die meer zou willen werken**

We berekenen phi. Immers, we kijken naar het verschil tussen twee groepen op een nominale variabele met twee mogelijke uitkomsten.

Frequentietabel van ‘meer of minder willen werken’ gesplitst op geslacht met percentages:

| **Geslacht** | **Man** | | **Vrouw** | | |
|---|---|---|---|---|---|
| **Meer of minder willen werken** | **Freq.** | **Perc.** | **Freq.** | **Perc.** | **Totaal** |
| Meer willen werken | 50 | 6,78 | 391 | 10,29 | 441 |
| Minder willen werken | 61 | 8,27 | 265 | 6,98 | 326 |
| Niet meer of minder willen werken | 607 | 82,25 | 2892 | 76,13 | 3499 |
| Vraag niet gesteld | 20 | 2,71 | 251 | 6,61 | 271 |
| **Totaal** | 738 | 100% | 3799 | 100% | 4537 |

Volgens de vuistregels is het verschil dan gering.
De uitspraak wordt dus niet ondersteund door de gegevens.

**Opgave 7 (6 punten)**

Gevraagde kruistabel met rijpercentages:

| **Arbeidsduur per week** | **Meer willen werken** | **Minder willen werken** | **Niet meer of minder willen werken** | **Vraag niet gesteld** | **Totaal** |
|---|---|---|---|---|---|
| Minder dan 12 uur | | | | 100,00 % | 100 % |
| 12 tot 20 uur | 18,46% | 2,92% | 78,46% | 0,15% | 100% |
| 20 tot 35 uur | 12,44% | 7,41% | 79,90% | 0,25% | 100% |
| 35 uur of meer | 1,80% | 10,54% | 87,34% | 0,33% | 100% |
| **Totaal** | | | | | |

Je ziet bijvoorbeeld dat van de mensen die 12 tot 20 uur per week werken 18,46% meer zou willen werken. Van de mensen die 20 tot 35 uur per week werken is dit 12,44% en van degenen die 35 uur of meer werken is dit 1,80%.
Van de mensen die 35 uur of meer per week werken, zou 10,54% minder willen werken, terwijl dit van degenen die 12 tot 20 uur per week werken slechts 2,92% is.
Kortom, de samenhang is juist beschreven.

**Opgave 8 (6 punten)**

Puntenwolk maken met leeftijd op de x-as en arbeidsduur2 op de y-as.

De correlatiecoëfficiënt is -0,058.

Dit duidt op een zwakke negatieve samenhang tussen *leeftijd* en *arbeidsduur*.

**Opgave 9 (12 punten)**

Selectiefilter op geslacht = vrouw.

Kruistabel van onderwijsrichting en beroepsbevolking:

| | Beroepsbevolking | | | | |
|---|---|---|---|---|---|
| **Onderwijsrichting** | **Werkzame beroepsbevolking** | **Werkloze beroepsbevolking** | **Niet beroepsbevolking** | **N.v.t.** | **Totaal** |
| Algemeen | 4640 | 560 | 6910 | | 12110 |
| Leraren | 1774 | 75 | 537 | | 2386 |
| Sociale wetenschappen, communicatie en kunst | 1857 | 135 | 586 | | 2578 |
| Economie, commercieel, management en administratie | 4112 | 337 | 1326 | | 5775 |
| Juridisch, bestuurlijk, openbare orde | 608 | 27 | 97 | | 732 |
| Wiskunde, natuurwetenschap, informatica | 381 | 24 | 126 | | 531 |
| Techniek | 649 | 80 | 325 | | 1054 |
| Agrarisch en milieu | 294 | 29 | 115 | | 438 |
| Gezondheidszorg, soc. dienstverlening en verzorging | 7725 | 435 | 3469 | | 11629 |
| Horeca, toerisme, vrije tijd, transport en logistiek | 777 | 66 | 217 | | 1060 |
| Onbekend | 338 | 32 | 200 | | 570 |
| **Totaal** | 23155 | 1800 | 13908 | 0 | 38863 |

We zien dat 3469 vrouwen een opleiding hebben genoten in de richting gezondheidszorg, sociale dienstverlening en verzorging. Dit komt overeen met ongeveer 4,5% van het totale gegevensbestand. We weten hier niet precies hoeveel mensen met een leeftijd tussen 15 en 65 in Nederland wonen, maar 4,5% hiervan lijkt inderdaad een groot arbeidspotentieel.