

Oogkleurprevalentie onder studenten van de Hanzehogeschool Groningen; Een verkennend onderzoek naar genetische variatie

Willem-Daniël Visser ¹, Cheyenne Brouwer ¹, Demi van 't Oever ¹, Kasthury Inparajah¹
¹Hanzehogeschool Groningen

```
knitr::opts_chunk$set(out.width = '100%', dpi=300) # configuratie voor figuur grootte en resolutie
knitr::opts_chunk$set(echo = FALSE)
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.2      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.0
## v ggplot2    3.4.2      v tibble    3.2.1
## v lubridate  1.9.2      v tidyr     1.3.0
## v purrr      1.0.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(readxl)
library(dplyr)
library(plyr)
```

```
## -----
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
## -----
```

```
##
## Attaching package: 'plyr'
##
## The following objects are masked from 'package:dplyr':
##
##   arrange, count, desc, failwith, id, mutate, rename, summarise,
##   summarize
##
## The following object is masked from 'package:purrr':
##
##   compact
```

```
library(knitr)
library(ggplot2)
library(gmodels)
```

Warning: package 'gmodels' was built under R version 4.4.0

Introductie

De afgelopen decennia is er veel onderzoek gedaan naar genetische variaties tussen populaties, waaronder de haar- en oogkleur. Deze onderzoeken kunnen helpen bij het in kaart brengen van bepaalde fysieke kenmerken in populaties & individuen om bijvoorbeeld de nauwkeurigheid van voorspellingen op basis van DNA te verbeteren.

Het onderzoek op de Hanzehogeschool te Groningen is uitgevoerd om te kijken of er een ‘dominante’ oogkleur op de locatie is en hoe dit tegenover andere locaties in Nederland staat. Onderzoek naar de prevalentie van oogkleuren is belangrijk en relevant om verschillende redenen, waaronder forensisch onderzoek en genetische en medische aspecten.

Uit eerdere onderzoeken is alleen de oogkleurprevalentie van heel Nederland in het algemeen bekend en niet specifiek een regio of gebied. Voor dit onderzoek is er gebruik gemaakt van een protocol dat in de repository op **GitHub** te vinden is. De ogen van 120 personen, exact verdeeld over 3 locaties (Hanzehogeschool Groningen, het centrum van Groningen en het centrum van Zwolle), zijn door 4 analisten gemeten met behulp van een zelf in elkaar gestelde referentiekaart dat tevens ook in de repository op **GitHub** is te vinden. De metingen zijn genoteerd in diverse notitie software (Google Docs, Apple en Samsung notities en Keep My Notes). Voor het meten van de oogkleur is er rekening gehouden met verschillende factoren zoals de invloed van zonlicht op het waarnemen van de oogkleur en het dragen van gekleurde contactlenzen.

Het doel van dit onderzoek is om de prevalentie van oogkleuren op de Hanzehogeschool Groningen te onderzoeken. Hierbij worden de oogkleur prevalentie bij 2 andere locaties (het centrum van Groningen en het centrum van Zwolle) ook onderzocht en vergeleken met de oogkleurprevalentie op de Hanzehogeschool Groningen.

Voor dit onderzoek zijn de data uit het onderzoek geanalyseerd met behulp van de statische methode **chi-square** test. Als eerste is er gekeken of er een aantoonbaar verschil zat in de oogkleur tussen mannen en vrouwen door middel van een statistische test (**chi-square** test) uit te voeren. Daarna is er met behulp van de dezelfde **chi-square** test gekeken of de oogkleuren per locatie van elkaar verschillen. Vervolgens is de data uit het eigen onderzoek vergeleken met data uit een ander literatuuronderzoek en ten slotte is er gekeken of er een meetfout door de analisten is gemaakt.

Op basis van eerdere (literatuur)onderzoeken (ref) wordt er verwacht dat de oogkleur “blauw” de meest voorkomende oogkleur zal zijn op de Hanzehogeschool te Groningen. Bovendien wordt vermoed dat er aantoonbare verschillen zijn in oogkleur tussen mannen en vrouwen, evenals variaties in oogkleurprevalentie op verschillende locaties. Er wordt niet verwacht dat er significante verschillen zullen zijn tussen het eigen onderzoek en een ander literatuuronderzoek, noch tussen de analisten onderling met betrekking tot meetfouten van de oogkleuren.

Materialen en Methoden

In de repository op Github <https://github.com/cheyennebrouwer/eyeresearch> staat het protocol voor het verzamelen van de oogdata. De analisten hebben verschillende notitie software gebruikt voor data notitie: Google Docs[@ref(g-docs)] door analist 8075, Apple notities[@ref(a-notes)] door analist 5609, Keep My Notes[@ref(keep-my-notes)] door analist 9308 en [meer]. Hierin hebben de analisten op 2 dagen de metingen genoteerd. Op de dag van de metingen zijn de analisten subject groepjes(2-7) gaan vragen of deze wouden meedoen met het onderzoek. Bij instemming is er gevraagd naar sexe en leeftijd en zijn de analisten de ogen van de subject met zijn allen tegelijk gaan bekijken waarbij 1 of meer van de analisten er een referentie blaadje bij hielden. Deze staat ook in de Github. De subjecten zijn 1 voor 1 gevraagd of ze mee wouden doen en naar bijhorende gegevens, tot alle subjecten zijn gemeten en de volgende groep werd aangesproken.

De analisten vulden soms twee kleuren in of een kleur karakteristiek, zoals: donker of licht. Op dag 1 is er op de studentencampus, de Hanzehogeschool Groningen; Zernikeplein 11, binnen in het gebouw “Van Doornveste” gemeten (Groningen 1). Op dezelfde dag is er in het centrum van Groningen (Groningen 2) gemeten. Op een dag 2 is vervolgens in het centrum van Zwolle (Zwolle) gemeten. Op alle drie locaties was het zonnig weer op het moment van metingen. Ook was het onder schooltijd voor de meeste MBO en HBO studenten. Op locatie Groningen 1 was de licht conditie niet constant met variërende daglicht reflectie, licht temperaturen en licht amplitude. Er waren geen licht condities die de oogkleur bepaling onmogelijk/erg moeilijk heeft gemaakt, aldus de analisten. Op locaties Groningen 2 en Zwolle is niet rechtstreeks richting de zon gekeken bij het bepalen van de oogkleur. Wel is er zowel in zon als schaduw gemeten.

Toen de metingen klaar waren werden alle metingen verzameld door 1 van de analisten en verwerkt zodat de dubbele kleuren door de analisten genoteerd resulteren tot enkele kleuren. De verwerkte data is in tidy vorm gezet met de functie `pivot_longer()` uit de tidy package[@ref(tidyr)] in de programmeertaal R[@ref(r-lang)]. Het resultaat is opgeslagen in een nieuw bestand en zijn analyses mee gedaan.

De analyses zijn gedaan in R-studio[@ref(r-studio)] en de programmeertaal R, met de libraries: `plyr`[@ref(plyr)], `dyplyr`[@ref(dplyr)], `ggplot2`[@ref(ggplot2)] en `knitr`[@ref(knitr)] van de tidyverse collectie. Met `plyr` en `dyplyr` is dataverwerking gedaan, zo kon er analyse worden gedaan op gefilterde delen van de data. Met `ggplot` zijn vervolgens grafieken gemaakt. Voor de analyse van de meetfout tussen de analisten is de functie `chisq.test` van R gebruikt, voor een two-way chi-square test, tussen met de factoren analist en oogkleur, met daarvoor alle gemeten personen met een of meer lege velden weggehaald. De analyse voor de verschillen tussen de drie meet locaties en verschillen tussen sexe is eerst de data zonder meerderheid onder de analisten als “undefined” neergezet, om zo aan te geven dat er geen definitieve oogkleur valt te bepalen. Toen werd de `CrossTable` voor de chi-square test en nog andere statistische waarden gebruikt. Hierbij is de locatie in groepjes gezet op basis van subject id en opgezet tegen de oogkleur. En de sexe ook opgezet tegen de oogkleur. Als analyse voor de vergelijking van de percentages van de gemeten kleuren zijn de deze uitgerekend en vergeleken met die van de literatuur. Voor de verschillen tussen twee metingen te zien in tabellen en grafieken is het laagste meting gedeeld door het hoogste meting en maal 100.

Resultaten

Om er achter te komen welk oogkleur de Hanzehogeschool Groningen aanleg voor heeft is er eerst gekeken of er een significante meetfout is gevormd. Dit is met behulp van een tabel met het aantal opnamen per analist per oogkleur en een aantal percentages die hieruit gehaald kunnen worden en de p-waarde resulterend van een tweezijdige chi-square test. Zo kunnen kan er worden gekeken welke analisten de grootste verschillen in het aantal opnamen en of het verschil significant is en dus basis kan vormen voor advies hoe de betrouwbaar data van dit onderzoek is. Dan word ook het verschil met Nederland breed behandeld voor extra controle op de significantie van deze onderzoeksdata en als een extra vergelijking hoe de Hanzehogeschool opweegt tegen het gemiddelde van de Nederlandse bevolking. Vervolgens worden de locatie factoren aan het licht gebracht, de sexe verdeling per locatie en tenslotte de sexe verdeling over het algemeen. Om zo de vragen of de Hanze een aanleg heeft voor een bepaalde oogkleur en of er een sexe verschil bestaat. De vergelijking met de oogkleur tellingen met de literatuur, de locatie en sexe factoren gebruikten data door meerderheid van de analisten bepaald. Bij een “Nd.”-waarde was er geen sprake van een meerderheid onder de analisten. Ook word bij deze statistiek gesproken over de rechter oogkleur over de linker, omdat er geen verschil is gevonden tussen deze twee.

Verskil tussen analisten

Hier zijn 2 tabellen weergegeven met het totaal aantal metingen per oogkleur per analist. Hieruit blijkt er geen significant verschil te zijn tussen alle de analisten samen(p-waarde = 0,07672, alfa = ,05). Tussen de analisten 9308 en 8705 was wel een significant verschil (p-waarde = 0,03118, alfa = ,05). Dit zie je ook aan de flinke verschillen tussen de analisten bij de oogkeuren hazel(~33,3% verschil) en groen(~44,6%). Oogkleur groen had ook het grootste verschil tussen de laagste aantal metingen en hoogste metingen: 52,6% verschil.

Groen was ook met 9 de oogkleur met de grootste deviatie. Daarnaast is te zien aan de tabel dat de twee analisten 5609 en 8705 en de twee analisten 9308 en 2193 beide groepen vormen, aangezien ze dicht bij elkaar liggen. Hierop is nog een **chi-square** test op los gelaten. Hier is de p-waarde 0,008905 (alfa = ,05). Dus is er een redelijk flink significant verschil tussen de bovenste twee en onderste twee analisten. Uit dit alles blijkt er geen ernstige meetfout te zitten en kan er een vergelijking worden gedaan met de bestaande gegevens over heel Nederland om te kijken of er een systematische meetfout is of er een verschil zit tussen Nederland breed en de locaties die hier zijn onderzocht.

Vershil met populatie Nederland

Hier is te zien hoe blauw(52,5 %) een meerder deel van de metingen uitmaakt, gevolgd door bruin(34,16667 %) en groen(8,333333 %). Zonder de NA-waarden meegerekend komt dit neer op de verhoudingen blauw, bruin, groen (55,26%, 35,96% en 8,77%).

In het kruistabel hieronder staat data uit een ander literatuuronderzoek dat is uitgevoerd om informatie over de verdeling van de oog- en haarkleur te verzamelen over populaties en deze data als een blue print te gebruiken om de nauwkeurigheid van voorspellingen op basis van DNA te verbeteren. [Katsara2019tc] Dit onderzoek heeft meerdere datasets (haarkleur en oogkleur per land), waaruit 3 percentages zijn gehaald voor vergelijking met dit onderzoek.

Omdat de data uit het literatuuronderzoek in percentages zijn, zijn ook de percentages van de waarnemingen van de oogkleur uit het eigen onderzoek berekend om daarmee data met elkaar te vergeleken.

Tabel 1: Percentages waargenomen oogkleur literatuuronderzoek en eigen onderzoek

	Percentage.Blaauw	Percentage.Bruin	Percentage.Groen
Literatuuronderzoek	60.1%	21.7%	11.4%
Eigenonderzoek	52.5%	34.17%	8.33%

De percentages uit het literatuuronderzoek geven de prevalentie van de 3 oogkleuren (blauw (waar ook grijs onder is geschaard), bruin (waar ook hazel onder is geschaard) en groen) aan in Nederland. De kleur blauw (60.9%) was de meest voorkomende oogkleur volgens de literatuur dan het eigen onderzoek suggereert. Bruin is in heel Nederland minder prevalent dan bij het eigen onderzoek en bij de kleur groen is dit andersom.

De **chi-square** p-waarde tussen de percentages uit het literatuuronderzoek en eigen onderzoek is 1.837e-13 (alfa = .05). Om antwoord te krijgen of de Hanze aanleg heeft voor een bepaalde oogkleur zijn ook de locaties vergeleken.

Vershil tussen locaties

Qua locaties heeft de hanze 5-10% meer mensen met blauwe ogen dan de andere locaties, dit vergeleken met een afwijking van 9,3% van de literatuur tegenover alle blauw opnames en een gemiddelde afwijking van 24% tussen alle waarden, kunnen deze waarnemingen toeval zijn. In Zwolle zijn veel meer groene oogkleuren gevonden dan op de Hanze in Groningen.

Tenslotte zijn in Groningen naast blauw ook de meeste hazel kleurige ogen gevonden. Viermaal zoveel dan in Zwolle.

Het verschil tussen verschillende locaties is geen significant verschil (p-waarde = 0.08554, alfa = 0,05). Het verschil tussen de bruine en blauwe oogkleur in Groningen ziet er wel significant uit.

Vershil tussen mannen en vrouwen

Bij de kleuren bruin en hazel is er geen significant verschil tussen mannen en vrouwen (bij $\alpha = 0.05$). Daarnaast zijn er geen andere grote verschillen tussen sexe en de oogkleuren. Maar dit kan niet worden meegenomen voor dit onderzoek.

Discussie en Conclusies

Conclusie

Uit de resultaten is het volgende uit te concluderen:

- Bij deelvraag 1: “Is er een aantoonbaar verschil in oogkleur tussen mannen en vrouwen?” kan H0: “Er is geen aantoonbaar verschil in oogkleur tussen mannen en vrouwen” worden aangenomen en H1: “Er is een aantoonbaar verschil in oogkleur tussen mannen en vrouwen” worden verworpen. De p-waarde uit de **chi-square** test is $p=0,08$ en is boven de α van $0,05$. De p-waarde uit deze chi square is hier niet significant. Er is dus geen aantoonbaar verschil in oogkleur tussen mannen en vrouwen.
- Bij deelvraag 2: “Is er een aantoonbaar verschil in oogkleur tussen de verschillende locaties waarop gemeten is?” kan H0: “Er is geen aantoonbaar verschil in oogkleur tussen de verschillende locaties waarop gemeten is” worden aangenomen en H1: “Er is een aantoonbaar verschil in oogkleur tussen de verschillende locaties waarop gemeten is” worden verworpen. Aan de p-waarde van $0,08$ uit de **chi-square** test is zichtbaar dat er geen significant verschil is tussen de oogkleuren per locatie.
- Bij deelvraag 3: “Komen de gegevens overeen met de bekende literatuur voor populatie oogkleur bij Nederlanders?” kan H0: “De gegevens komen overeen met de bekende literatuur gegevens voor populatie oogkleur bij Nederlanders” worden verworpen en H1: “De gegevens komen niet overeen met de bekende literatuur voor populatie oogkleur bij Nederlanders” worden aangenomen. De p-waarde uit de **chi-square** test ($p=1,837e-13$) laat zien dat er een significant verschil zit tussen de data uit het literatuuronderzoek en eigen onderzoek. De gegevens komen dus niet overeen met de bekende literatuur voor populatie oogkleur bij Nederlanders.
- Bij deelvraag 4: “Is er een meetfout ontstaan door metingen met verschillende analisten?” kunnen H0: “Er is geen meetfout ontstaan door metingen met verschillende analisten” & H1: “Er is een meetfout ontstaan door metingen met verschillende analisten” deels worden aangenomen. De p-waarde van de Pearson’s **Chi-square** test is $p=0,07672$ (met $\alpha = 0,05$) er is dus geen significant verschil tussen de analisten. Echter, er zijn nog meetfouten aanwezig die vermindert kunnen worden.

Discussie

Enkele aandachtspunten voor verdere discussie naar aanleiding van de conclusie zijn:

- Bij het beantwoorden van deelvraag 1: “Is er een aantoonbaar verschil in oogkleur tussen mannen en vrouwen?” is uit het kruistabel een p-waarde van $0,57$ uitgekomen. Deze waarde zit boven de α van $0,05$ dus is er geen significant verschil in de oogkleuren tussen mannen en vrouwen. Echter, bij de kleur “bruin” lijkt er wel een verschil te zitten, maar dit verschil is niet significant genoeg.
- Bij het onderzoek op 3 verschillende locaties (met elk 40 deelnemers, gelijk verdeeld tussen mannen en vrouwen) was er een dominantie aanwezigheid van de oogkleur “blauw” op alle locaties, behalve in Zwolle. In Zwolle waren er relatief meer personen met een groene oogkleur en afwijkende hoeveelheden in bruine- en hazelkleurige ogen. Op de 2 andere locaties waren deze hoeveelheden bruin- en hazelkleurige ogen hoger. De p-waarde geeft echter aan dat er geen significant verschil is in oogkleuren tussen de locaties.

- Bij deelvraag 3 is de **chi-square** test uitgevoerd met percentages in plaats van aantallen, vanwege de ontbrekende gegevens in het literatuuronderzoek. In het literatuuronderzoek waren alleen percentages vermeld, zonder vermelding van de steekproefomvang (n). Daarom zijn de percentages gebruikt voor de analyse, hoewel er normaal gesproken aantallen worden gebruikt bij een **chi-square** test.
- Bij zowel het literatuur als het eigen onderzoek is de oogkleur “groen” in gelijke percentages waargenomen. Echter, er is een verschil tussen de percentages waargenomen “blauwe” en “bruine” oogkleuren. Met name bij de oogkleur “bruin” lijkt dit verschil groter te zijn. Dit verschil kan mogelijk worden veroorzaakt doordat het literatuuronderzoek de oogkleurverdeling in heel Nederland vertegenwoordigt, terwijl het eigen onderzoek voornamelijk plaatsvond in het noorden en oosten van Nederland waardoor de data alleen die gebieden vertegenwoordigt.
- Het eigenonderzoek is voornamelijk uitgevoerd in het noorden (Hanzehogeschool Groningen & Groningen stad) en oosten (Zwolle) van Nederland. Uit onderzoek blijkt dat noorderlingen vaker blauwe ogen hebben dan de rest van het land, maar dit verschil werd niet weerspiegeld in het eigenonderzoek (52.1% blauw) in vergelijking met het literatuuronderzoek (60.9% blauw). Het eigenonderzoek dat voornamelijk in het noorden plaatsvond had opvallend genoeg een lager percentage blauwkleurige ogen dan het literatuuronderzoek (dat heel Nederland vertegenwoordigt).
- Bij deelvraag 4 was er geen significant verschil tussen de analisten, maar er lijkt toch een aanzienlijk verschil te zijn tussen analist 9308 en 8705 in de tabel met het aantal waarnemingen van oogkleuren. Een **chi-square** test tussen deze twee analisten toont een significant verschil aan ($p = 0.03118$, $\alpha = 0.05$).

Online bijlagen

README

- <https://github.com/cheyennebrouwer/eyeresearch/blob/main/README.md>

Protocol

- https://github.com/cheyennebrouwer/eyeresearch/blob/main/protocol/protocol_oogkleur.Rmd

Materialen

- Oogkleurreferentiekaart: https://github.com/cheyennebrouwer/eyeresearch/blob/main/protocol/oogkleur_refferenties.png

Ruwe data eigen onderzoek

- Oogkleur_simpel2: https://github.com/cheyennebrouwer/eyeresearch/blob/main/analyse/data/Oogkleur_simpel2.csv
- Oogkleur definitief (met __OK...): https://github.com/cheyennebrouwer/eyeresearch/blob/main/analyse/data/Oogkleur_simpel_definitief.csv
- Oogkleur definitief (zonder __OK...): https://github.com/cheyennebrouwer/eyeresearch/blob/main/analyse/data/Oogkleur_simpel.csv
- Oogkleur data uitgebreid: https://github.com/cheyennebrouwer/eyeresearch/blob/main/ruwe_data/Oogkleur_uitgebreid.xlsx

Achtergrondinformatie oogkleuren

- https://github.com/cheyennebrouwer/eyersearch/blob/main/analyse/Achtergrondinformatie_oogkleur.Rmd

Literatuuronderzoek

- https://github.com/cheyennebrouwer/eyersearch/blob/main/analyse/True_colors_A_literature_review_on_the_spatial_dis.pdf

Logboeken

- <https://github.com/cheyennebrouwer/eyersearch/tree/main/logs>

Wordcount

#Voeg aan het eind een woord-telling in:

```
#{r wordcount, message=False} #wordcountaddin:::text_stats() #
```

Notitie software

Google Docs

Gebruikte versie: 1.23.162.05.90

Referentie: <https://play.google.com/store/apps/details?id=com.google.android.apps.docs.editors.sheets&pli=1>

Apple notities

Gebruikte versie: geen data

Referentie: <https://apps.apple.com/us/app/notes/id1110145109>

Keep My Notes

Gebruikte versie: 1.80.180

Referentie: <https://www.keepmynotes.app/>

R

Gebruikte versie: 4.3.0

Referentie: <https://www.r-project.org/>

R libraries

plyr

Gebruikte versie: 1.8.8

Referentie: <https://www.rdocumentation.org/packages/plyr/versions/1.8.8>

dplyr

Gebruikte versie: 1.1.2

Referentie: <https://www.rdocumentation.org/packages/dplyr/versions/1.0.10>

ggplot2

Gebruikte versie: 3.4.2

Referentie: <https://www.rdocumentation.org/packages/ggplot2/versions/3.4.2>

knitr

Gebruikte versie: 1.43

Referentie: <https://www.rdocumentation.org/packages/knitr/versions/1.43>

tidyr

Gebruikte versie: 1.3.0

Referentie: <https://cran.r-project.org/web/packages/tidyr/index.html>

Software Tools

R-studio

Gebruikte versie: 2023.03.0 (Windows en MacOS versies)

Gebruiksreden: Het ontwikkelen van R code voor de analyse van de data.

Referentie: <https://posit.co/products/open-source/rstudio/>

Referenties

Een lijst van referenties wordt hier automagisch toegevoegd.