

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,
Michiel Noback ¹, Fenna Feenstra ¹
¹Hanzehogeschool Groningen

Samenvatting

In dit onderzoek wordt de oogkleurprevalentie onder studenten van Hanzehogeschool Groningen onderzocht en wordt dit vergeleken met andere locaties in Nederland. Het doel is om een dominante oogkleur te identificeren en verschillen tussen de locaties te onderzoeken. Metingen werden uitgevoerd bij 120 proefpersonen op drie locaties met behulp van een referentiekaart. Blootstelling aan zonlicht en het dragen van gekleurde contactlenzen werden meegenomen. De chi-square test analyseerde de gegevens op significante verschillen tussen geslachten en locaties. Er vond tevens een vergelijking plaats met literatuurgegevens en evaluatie van meetfouten. Verwacht werd dat blauwe ogen het meest voorkwamen. De bevindingen dragen bij aan het begrip van genetische variaties in oogkleur en bieden inzicht in de prevalentie van oogkleuren op de Hanzehogeschool te Groningen.

1 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 zich verhoudt tot andere locaties in Nederland. 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 prevalentie van oogkleur in 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 samengestelde referentiekaart, die ook in de repository op `GitHub` te vinden is. De metingen zijn genoteerd in verschillende notitiesoftware (Google Docs, Apple- en Samsung notities en Keep My Notes). Bij 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 ook de prevalenties van oogkleuren op 2 andere locaties (het centrum van Groningen en het centrum van Zwolle) onderzocht en vergeleken met de prevalentie van oogkleur op de Hanzehogeschool Groningen.

Voor dit onderzoek is 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 was in 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 (Katsara en Nothnagel 2019) wordt verwacht dat “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 prevalentie van oogkleur 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.

2 Materialen en Methoden

In de repository op Github <https://github.com/cheyennebrouwer/eyereseach> staat het protocol voor het verzamelen van de oogdata. De analisten hebben verschillende notitiessoftware[5.2] gebruikt voor data notitie: Google Docs door analist 8075, Apple Notities door analist 5609, Keep My Notes door analist 9308 en Samsung Notes door analist 2193. Hierin hebben de analisten op 2 dagen de metingen genoteerd. Op de dag van de metingen zijn de analisten groepjes proefpersonen (2-7) gaan vragen of deze wilden deelnemen aan het onderzoek. Bij instemming is er gevraagd naar geslacht, leeftijd en is per proefpersoon de oogkleur bepaald aan de hand van een referentieformulier, welke ook terug te vinden is in de repository.

De analisten hebben soms twee kleuren ingevuld of een kleurenkenmerk, zoals: “donker” of “licht”. Op dag 1 is er gemeten op de studentencampus, de Hanzehogeschool Groningen locatie Zernikeplein 11 (Van Doorenveste, Groningen 1). Op dezelfde dag is er ook gemeten in het centrum van Groningen (Groningen 2). Op dag 2 is vervolgens gemeten in het centrum van Zwolle (Zwolle). 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 lichtconditie niet constant, met variërende daglichtreflectie, lichttemperaturen en lichtamplitude. Er waren echter geen lichtcondities die de oogkleur bepaling onmogelijk of erg moeilijk maakten, aldus de analisten. Op de 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 één van de analisten en verwerkt, zodat de dubbel genoteerde kleuren kon worden samengevoegd tot enkele kleuren. De verwerkte data is in tidy vorm gezet met de functie `pivot_longer()` uit de tidy package in de programmeertaal R[5.3]. Het resultaat is opgeslagen in een nieuw bestand en zijn analyses op uitgevoerd.

De analyses zijn gedaan in RStudio[5.5] en de programmeertaal R, met de libraries[5.4]: plyr, dplyr, ggplot2, ggpubr van de tidyverse collectie, knitr, gmodels en readxl. De data is ingeladen met behulp van `read.table`

uit `readxl`. Met `plyr` en `dyplyr` is dataverwerking gedaan, zodat analyses konden worden uitgevoerd op gefilterde delen van de data. Met `ggplot` zijn vervolgens grafieken gemaakt en deze zijn met `ggpubr` naast elkaar gezet. Voor de analyse van de meetfout tussen de analisten is de functie `chisq.test` van R gebruikt, voor een two-way chi-square test, waarbij de factoren analist en oogkleur zijn gebruikt en waarbij alle gemeten personen met één of meer lege velden zijn verwijderd. Voor de analyse van de verschillen tussen de drie meetlocaties en verschillen tussen geslacht is de data eerst gecategoriseerd als “undefined” wanneer er geen consensus was onder de analisten, om aan te geven dat er geen definitieve oogkleur kon worden bepaald. Vervolgens is de `CrossTable`-functie uit de library `gmodels` voor de chi-square-test en andere statistische waarden gebruikt. Hierbij is de locatie opgedeeld in groepen op basis van identifier en vergeleken met de oogkleur, hetzelfde is gedaan voor het geslacht in relatie tot de oogkleur. Voor de analyse van de vergelijking van de percentages van de gemeten oogkleuren zijn deze uitgerekend en vergeleken met die van de literatuur. Om de verschillen tussen twee metingen in tabellen en grafieken weer te geven, is de laagste meting gedeeld door de hoogste meting en maal 100.

3 Resultaten

Om te onderzoeken welke oogkleuren vaak voorkomen bij studenten van de Hanzehogeschool Groningen, is eerst gekeken of er sprake is van een significante meetfout onder de analisten. Dit is gedaan aan met behulp van een tabel met het aantal waarnemingen per analist en per oogkleur, evenals de afgeleide percentages hieruit. Hieruit is de p-waarde berekend met behulp van een tweezijdige chi-square test. Op deze manier kan worden vastgesteld welke analisten de grootste verschillen vertonen in aantal waarnemingen en of deze verschillen significant zijn. Dit vormt een basis om te bepalen hoe betrouwbaar de gegevens van dit onderzoek zijn. Vervolgens is er een vergelijking gemaakt met de nationale gegevens om de significantie van de verzamelde data te controleren en een vergelijking te kunnen maken tussen de Hanzehogeschool Groningen en het gemiddelde van de Nederlandse bevolking. Daarnaast worden de locatiefactoren belicht en tenslotte de algemene verdeling van de geslachten. Op deze manier wordt onderzocht of een oogkleur vaker voorkomt bij de Hanzehogeschool Groningen en of hierin verschillen zijn tussen de geslachten. Bij het vergelijken van de oogkleurtellingen met de literatuur en bij het onderzoeken van de locatie- en geslachtsfactoren is gebruikgemaakt van data die bepaald is door de meerderheid van de analisten. Als een waarde “undefined” geeft, was er geen sprake van een meerderheid onder de analisten. Bovendien wordt er in de resultaten gesproken over een rechter- en linkeroogkleur, omdat er geen verschillende oogkleuren gevonden zijn bij de proefpersonen.

3.1 Verschil tussen analisten

Tabel 1: Het waargenomen aantal oogkleuren per analist. Linker- en rechteroogkleuren zijn gecombineerd voor overzicht. De analisten zijn tegen elkaar uitgezet om de aantallen te vergelijken.

	blauw	groen	bruin	hazel
9308	132	20	60	18
2193	125	22	64	19

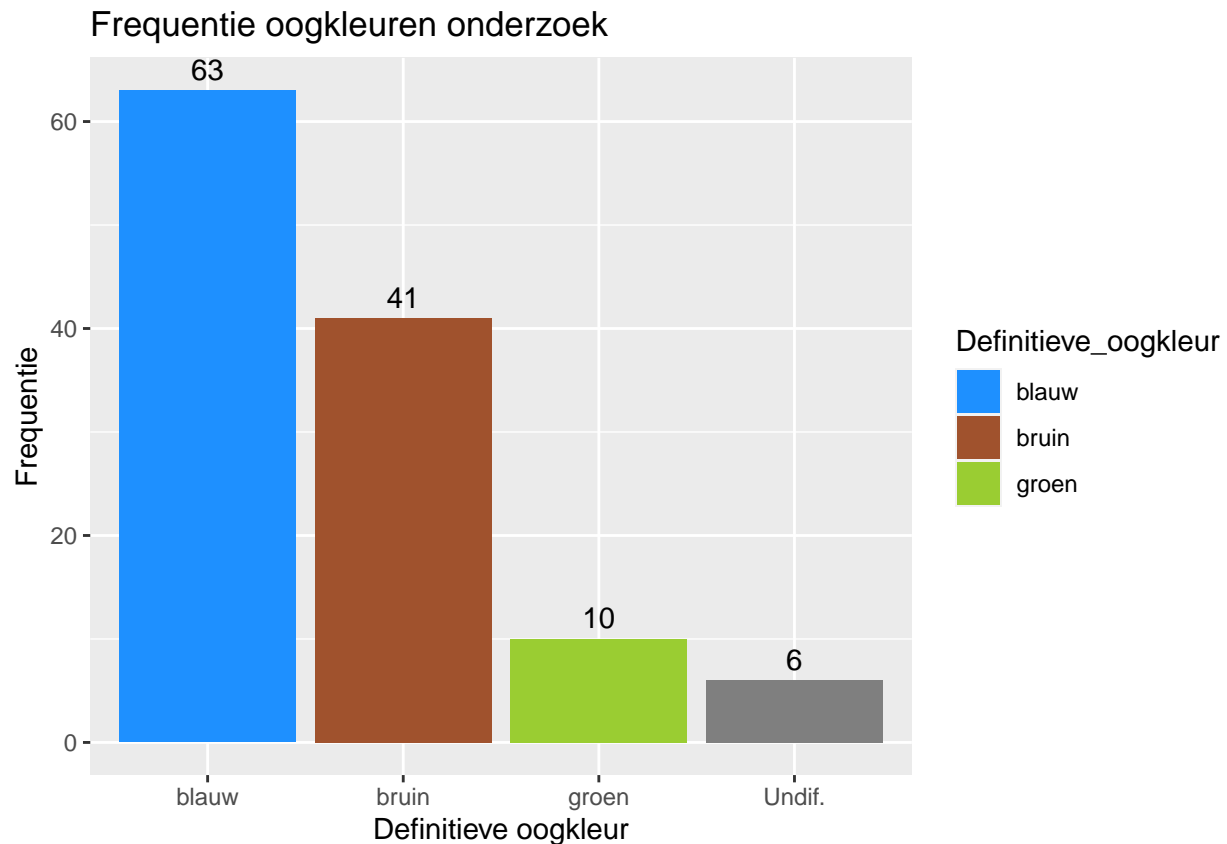
5609	114	38	58	20
8705	110	36	72	12

Tabel 2: Het waargenomen aantal oogkleuren per groep analisten. Linker- en rechteroogkleuren zijn gecombineerd voor overzicht. Analisten 9308 met 2193 zijn een groep, evenals 5609 met 8705. De twee groepen zijn tegen elkaar gezet vanwege de zichtbare overeenkomst tussen de leden van de groep en verschil tussen de twee groepen.

groep	blauw	groen	bruin	hazel
5609 & 8705	112.0	37	65	16.0
9308 & 2193	128.5	21	62	18.5

Hier worden twee tabellen weergegeven met het totaal aantal metingen per oogkleur per analist. Hieruit blijkt dat er geen significant verschil is tussen alle analisten samen (p-waarde = 0,07672, alfa = 0,05). Er is echter wel een significant verschil tussen de analisten 9308 en 8705 (p-waarde = 0,03118, alfa = 0,05). Dit is ook te zien aan de aanzienlijke verschillen tussen de analisten bij de oogkleuren hazel (verschil van $\pm 33,3\%$) en groen (verschil van $\pm 44,6\%$). De oogkleur groen vertoonde ook het grootste verschil tussen het laagste en hoogste aantal metingen: 52,6% verschil. Bovendien had groen ook met een waarde van 9 de oogkleur met de grootste deviatie. Daarnaast is er aan de tabel te zien dat de analisten 5609 en 8705 samen een groep vormen, evenals analisten 9308 en 2193, omdat ze dicht bij elkaar liggen. Hierop is nog een **chi-square** test uitgevoerd, waarbij de p-waarde 0,008905 (alfa = 0,05) is. Dit wijst op een flink significant verschil tussen de twee groepen analisten. Op basis van deze gegevens blijkt er geen ernstige meetfout aanwezig te zijn en kan er een vergelijking worden gedaan met bestaande gegevens over heel Nederland om te kijken of er een sprake is van een systematische meetfout of er een verschil zit tussen Nederland als geheel en de onderzochte locaties.

3.2 Verschil met populatie Nederland



Figuur 1: Barplot met aantallen subjects met een bepaalde oogkleur.

Hier is te zien hoe blauw (52,5%) het grootste deel van de metingen uitmaakt, gevolgd door bruin (34,17%) en groen (8,33%). Zonder de “undefined” waarden meegerekend komt dit neer op de verhoudingen blauw, bruin, groen van respectievelijk 55,26%, 35,96% en 8,77%.

In de kruistabel [3] staat data uit een ander literatuuronderzoek dat is uitgevoerd om informatie over de verdeling van de oog- en haarkleur te verzamelen binnen verschillende populaties. Deze data wordt gebruikt als referentie om de nauwkeurigheid van voorspellingen op basis van DNA te verbeteren (Katsara en Nothnagel 2019). Dit onderzoek bevat 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 vergelijken.

De percentages uit het literatuuronderzoek geven de prevalentie van de 3 oogkleuren (blauw, bruin en groen) aan in Nederland. Het onderzoek suggereert dat de kleur blauw (waar ook grijs onder valt) met 60.9% de meest voorkomende oogkleur is, wat hoger is dan wat het eigen onderzoek aangeeft. Bruin is in Nederland minder voorkomend dan bij het eigen onderzoek, terwijl dit voor de kleur groen andersom is.

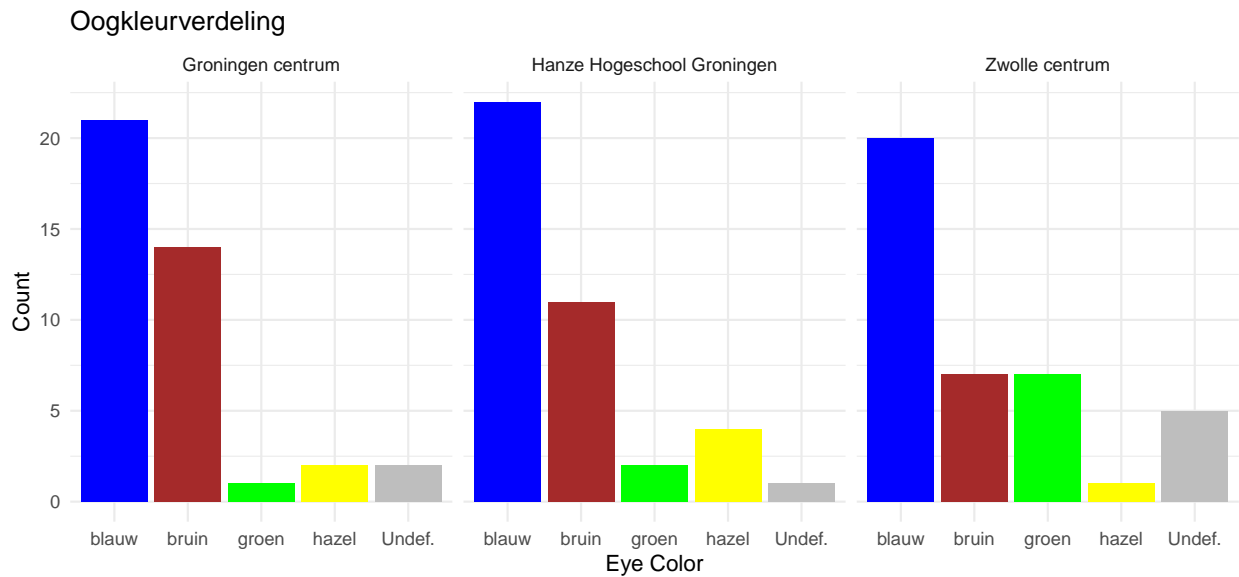
De **chi-square** p-waarde tussen de percentages uit het literatuuronderzoek en het eigen onderzoek is 1.837e-

Tabel 3: Percentages waargenomen oogkleur literatuuronderzoek en eigen onderzoek

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

13 ($\alpha = 0,05$). Om antwoord te krijgen op de vraag of de Hanzehogeschool aanleg heeft voor een bepaalde oogkleur, zijn ook de verschillende locaties met elkaar vergeleken.

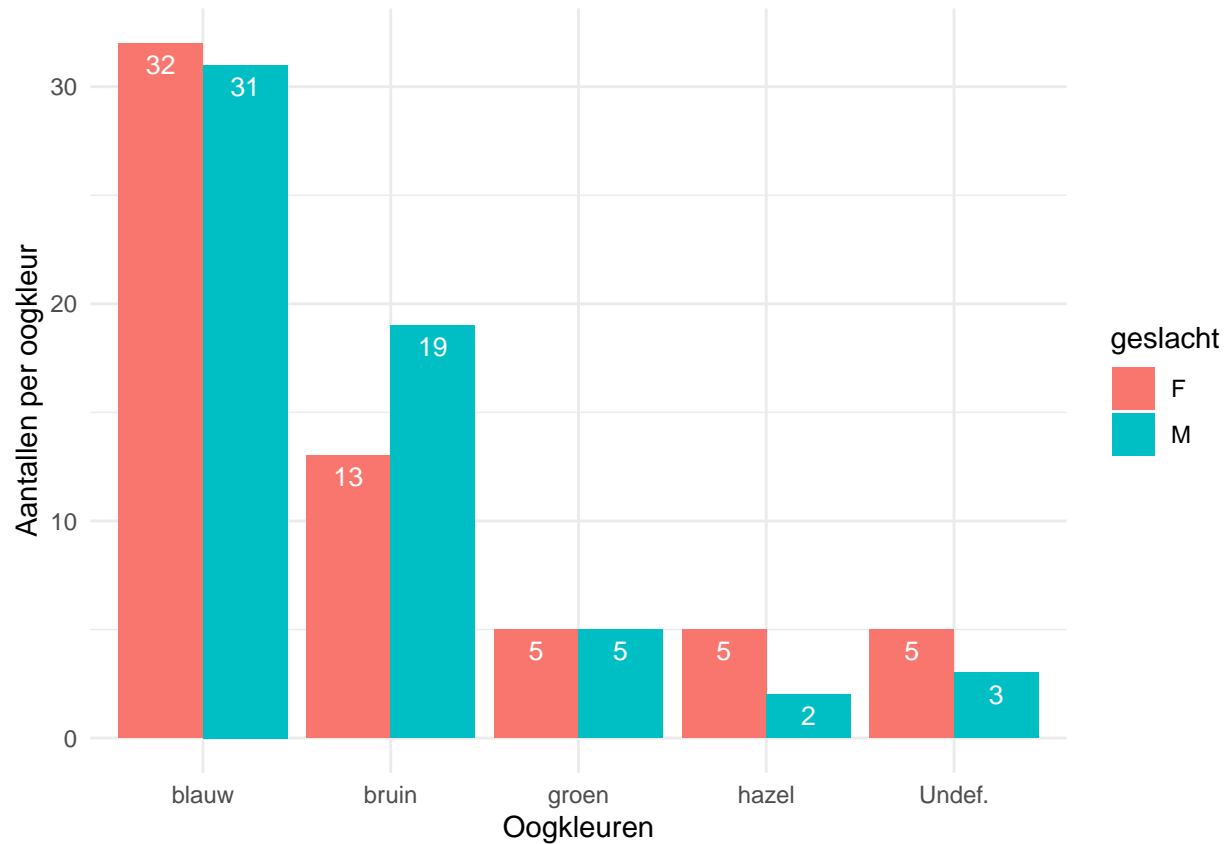
3.3 Verschil tussen locaties



Figuur 2: Drie barplots met de drie meetlocaties en het aantal gemeten oogkleuren.

Het verschil tussen de verschillende locaties is geen significant verschil (p -waarde = 0.08554, $\alpha = 0,05$). Het verschil tussen de bruine en blauwe oogkleur in Groningen is wel significant.

3.4 Verschil tussen mannen en vrouwen



Figuur 3: Barplot met aantal oogkleuren per geslacht van alle meet- locaties samen. Verdeeld onder sexe.

Bij de oogkleuren bruin en hazel is er geen significant verschil tussen mannen en vrouwen (chi-square test p-waarde - 0,57 bij $\alpha = 0.05$). Daarnaast zijn er geen andere grote verschillen tussen sexe en de oogkleuren. Maar dat kan niet worden meegenomen in dit onderzoek.

4 Discussie en Conclusies

Discussie

Om onze onderzoeksvraag te kunnen beantwoorden zijn enkele deelvragen opgesteld. De eerste deelvraag betreft het verschil in oogkleur tussen mannen en vrouwen. Hierbij is geen significant verschil gevonden (p-waarde chi-square test = 0,57). Dit resultaat komt overeen met de verwachting, aangezien oogkleur genetisch bepaald wordt en gekoppeld is aan het X- of Y-chromosoom.

De tweede deelvraag gaat over verschil in prevalentie van oogkleur tussen de drie verschillende locaties. Op alle locaties bleek de dominante oogkleur blauw te zijn, behalve in Zwolle. In Zwolle werden relatief meer proefpersonen met een groene oogkleur gevonden, evenals een lagere prevalentie van bruine en hazelkleurige ogen. Er werd echter geen significant verschil gevonden (chi-square test p-waarde 0,086 bij $\alpha 0,05$). Dit

ligt binnen de verwachtingen, omdat oogkleur niet vastzit aan de plek waar iemand woont. De p-waarde ligt echter maar net boven de grens van 0,05. Mogelijk zal de p-waarde met een grotere steekproefgrootte nog wel lager uitvallen.

Bij de derde deelvraag is gekeken naar verschillen tussen de gemeten oogkleurprevalentie in dit onderzoek en uit bestaand wetenschappelijk onderzoek (Katsara en Nothnagel 2019). Het literatuuronderzoek vermeldde alleen percentages zonder de steekproefgrootte (n) te vermelden. Ook voor deze vraag is een ‘chi-square’ test gebruikt, maar werden percentages in plaats van absolute getallen gebruikt. Uit beide onderzoeken bleek dat prevalentie van groene ogen vergelijkbaar was. Er is wel verschil in de percentages waargenomen van blauwe en bruine oogkleuren. Een mogelijke verklaring hiervoor kan zijn dat het literatuuronderzoek betrekking had op heel Nederland, terwijl in dit onderzoek enkel steekproeven in Groningen en Zwolle zijn uitgevoerd. Het literatuuronderzoek wees uit dat mensen uit het noorden vaker blauwe ogen hebben (60,9%) dan de rest van het land. Dit werd niet teruggevonden in het eigen onderzoek, waar 52,1% van de proefpersonen blauwe ogen had.

De vierde deelvraag heeft betrekking op de mogelijke meetfout tussen de analisten. Er werd geen significant verschil tussen de analisten gevonden. Wel bleek er een significant verschil te zijn tussen analisten 9308 en 8705 (p-waarde 0,031 bij alfa 0,05). Dit verschil kan mogelijk verklaard worden door een foutief gebruik van de referentiekaart die is gebruikt, of door verschillen in interpretatie van de kleur per analist. [Valenta e.a. (2021)] (Bompas en ORegan 2006)

Aan het eind van het onderzoek is het protocol en de steekproefgrootte nogmaals beoordeeld. Het protocol werd als voldoende beoordeeld door de analisten. Hoewel de steekproefgrootte aanvankelijk een groot aantal lijkt, blijkt het achteraf toch best een klein aantal te zijn. Een aantal p-waardes zijn niet significant, maar zijn bijna wel significant. Een grotere steekproefgrootte zou hier verschil in kunnen maken. De locaties waarop is gemeten, waren Groningen en Zwolle. Beide locaties beschikken over een hogeschool die regionaal studenten aantrekt. Daarom namen niet alleen proefpersonen deel die daadwerkelijk op die locaties wonen, maar ook proefpersonen uit de omgeving. Het is echter mogelijk dat als er ook locaties in het zuiden en westen waren opgenomen in het onderzoek, er andere resultaten zouden zijn gevonden.

Conclusie

Er is geen significant verschil gevonden in prevalentie van oogkleur tussen geslachten en de onderzochte locaties. Ook tussen alle 4 de analisten bleek geen significant verschil aantoonbaar te zijn. Tussen de analisten 9308 en 8705 bleek wel een significant verschil aanwezig te zijn, hier is dus een meetfout ontstaan. De gegevens die uit dit onderzoek gekomen zijn, zijn vergeleken met al bekende gegevens over de prevalentie van oogkleur in Nederland. Ook hier is geen significant verschil gevonden.

5 Online bijlagen

5.1 Github bijlagen

README

- <https://github.com/cheyennebrouwer/eyereseach/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/eyeresearch/blob/main/analyse/Achtergrondinformatie_oogkleur.Rmd

Literatuuronderzoek

- https://github.com/cheyennebrouwer/eyeresearch/blob/main/analyse/True_colors__A_literature_review_on_the_spatial_dis.pdf
- [https://github.com/cheyennebrouwer/eyeresearch/blob/main/analyse/Ecology and Evolution - 2021 - Valenta - In the eye of the beholder Is color classification consistent among human.pdf](https://github.com/cheyennebrouwer/eyeresearch/blob/main/analyse/Ecology_and_Evolution_-_2021_-_Valenta_-_In_the_eye_of_the_beholder_Is_color_classification_consistent_among_human.pdf)

Logboeken

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

5.2 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.keeptomynotes.app/>

Samsung Notes

Gebruikte versie: 4.4.07.24

Referentie: <https://play.google.com/store/apps/details?id=com.samsung.android.app.notes&hl=en&gl=US>

5.3 R

Gebruikte versie: 4.3.0

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

5.4 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>

ggpubr

Gebruikte versie: 0.6.0

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

gmodels

Gebruikte versie: 2.18.1.1

Referentie: <https://www.rdocumentation.org/packages/gmodels/versions/2.18.1.1>

readxl

Gebruikte versie: 1.4.2

Referentie: <https://www.rdocumentation.org/packages/readxl/versions/1.4.2>

5.5 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

- Bompas, Aline, en J Kevin ORegan. 2006. 'Evidence for a Role of Action in Colour Perception'. *Perception* 35 (1): 65–78. <https://doi.org/10.1068/p5356>.
- Katsara, Maria-Alexandra, en Michael Nothnagel. 2019. 'True colors: A literature review on the spatial distribution of eye and hair pigmentation'. *Forensic Science International: Genetics* 39 (maart): 109–18. <https://doi.org/10.1016/j.fsigen.2019.01.001>.
- Valenta, Kim, Sally L. Bornbusch, Yan-Daniel Jacques, en Omer Nevo. 2021. 'In the eye of the beholder: Is color classification consistent among human observers?' *Ecology and Evolution* 11 (20): 13875–83. <https://doi.org/10.1002/ece3.8093>.