

My Awesome Catchy Title!

Michiel Noback ¹, Fenna Feenstra ¹, John Doe ²

¹Hanzehogeschool, ²Een ander instituut

Samenvatting

Geef hier je samenvatting in maximaal 150 woorden. Het is een samenvatting van het hele artikel; niet alleen de resultaten! Begin met het belang van dit onderzoek, dan hoe het onderzoek is aangepakt en de belangrijkste resultaten en eindig met de implicaties ervan voor de wetenschap/de maatschappij. Neem nooit figuren of tabellen op in de samenvatting.

1 Introductie op deze template

Dit hoofdstuk niet in je eigen paper toevoegen!

Dit template bevat alle verplichte onderdelen van de paper die je moet schrijven. De titels van de secties mag je NIET wijzigen. Voor elk onderdeel is gegeven wat erin thuis hoort en hoe je dat kan aanpakken. Ook is er aangegeven hoeveel woorden er minimaal en maximaal in de sectie gebruikt mogen worden. In totaal mag je artikel niet minder dan 1000 en niet meer dan 2000 woorden bevatten.

Gebruik de [wordcount plugin](#) om het aantal woorden te tellen van een sectie, en voeg aan het einde een totaal in (zoals al aanwezig in deze template).

1.1 Setup *Chunk*

Aan het begin van je paper kan je een zogenaamde *setup chunk* toevoegen. Hierin kan je het gedrag van `knitr` configureren en de bibliotheken laden die je in je code gebruikt. Hieronder is een voorbeeld van zo'n *setup chunk*.

NB Gebruik altijd een naam voor iedere chunk. Dat maakt het *debuggen* van problemen bij het *knitten* van je RMarkdown document veel gemakkelijker.

1.2 Code “*Chunks*” wel of niet tonen

In principe laat je code nooit zien in een publicatie, behalve als de code expliciet besproken wordt en een essentieel onderdeel vormt van je werk. Je kan code gemakkelijk verbergen door gebruik te maken van de optie `echo=FALSE` in de chunk header. Ook storende output kan je eventueel verbergen door gebruik te maken van `message=FALSE` of `warning=FALSE`. Eventueel kan je all chunks in een keer verbergen door dit in je *setup chunk* te plaatsen: `knitr::opts_chunk$set(echo = FALSE)`.

1.3 Referenties / citeren

Voor alle secties geldt dat je referenties kan gebruiken naar andere publicaties (externe referenties ofwel citatie) of naar secties, tabellen of figuren in je eigen publicatie (interne referenties). Voor externe referenties gebruik je deze notatie: [`@<ref_key>`], bijvoorbeeld [`@xie2013ddrk`]. Hier gebruik ik hem echt: (Xie 2013). De *processing engine* zal in het bibliografie bestand zoals opgegeven in de **yaml header** van dit bestand (`simple_template.bib`) zoeken naar de referentie met deze naam en vervolgens de hele referentie onderaan in het document opnemen. Kijk vooral in het `.bin` document hoe je deze kan gebruiken. Er zijn voor verschillende soorten bronnen verschillende (**annotaties?**); zie in de `.bib` voor enkele voorbeelden.

Interne referenties naar tabellen, figuren, equations en secties kan je maken door deze syntax te gebruiken: `\@ref(fig:<figuur-naam>)` waarbij `<figuur-naam>` de naam is van het chunk waarin de figuur gemaakt wordt. Hetzelfde werkt voor tabellen: `\@ref(tab:<tabel-naam>)`. Een sectie kan je labelen door er `{#label-naam}` achter te zetten. Bijvoorbeeld, dit `\@ref(setup-chunk)` linkt terug naar de sectie over de setup chunk in sectie [1.1].

Zie ook de hoofdstukken in het geweldige boek over [bookdown hier](#) en [hier](#)

Voor de rest: GIYF!

1.4 Taalgebruik

Wetenschappelijk taalgebruik is heel anders dan dagelijks communiceren - laat staan de taal die in online wordt gebruikt in bv Whatsapp of Instagram! Wetenschappelijk taalgebruik is formeel, onpersoonlijk en ondubbelzinnig. Een klein voorbeeld: Je schrijft nooit “ik heb de verandering van zoutconcentratie bij langdurig huilen onderzocht” maar “de verandering van zoutconcentratie in de tijd bij aanhoudend huilen is onderzocht”

Natuurlijk moeten spelling en grammatica (zo goed als) foutloos zijn! RStudio heeft redelijk goede spellingscontrole, ook voor Nederlands (althoewel je deze misschien wel apart moet installeren). Gebruik deze!

Alle volgende onderdelen moet je opnemen in je eigen artikel. Let vooral ook op het minimum

2 Introductie

Je start altijd met de maatschappelijke drijfveer voor je onderzoek; waarom is het van belang dat dit onderzoek is uitgevoerd? Vervolgens bespreek je de achtergronden van je onderzoek. Wat is er al eerder onderzocht op dit vlak en wat mist er juist nog; welke speciale technieken heb je gebruikt? Refereer zorgvuldig naar bronnen die je hebt gebruikt - zie ook paragraaf 7.

Je eindigt de introductie met de doelstelling(en) van jouw onderzoek, en hoe je deze doelstellingen denkt te gaan verwezenlijken; de aanpak. Geef hier ook eventuele hypothesen.

De introductie bevat tussen de 400-800 woorden.

3 Materialen en Methoden

Github repo: https://github.com/RamonReilman/growth_impact_cress Materialen en Methoden beschrijft verhalend wat je hebt gebruikt (data, tools) en wat je hebt gedaan (ontwikkelde methodes). ***Het is essentieel dat dit hoofdstuk je onderzoek reproduceerbaar en valideerbaar maakt.*** Verwijs bij de start van dit hoofdstuk naar je code repository (meestal je github link).

Neem referenties op!

De materialen zijn de meetinstrumenten die je hebt gebruikt, maar ook bijvoorbeeld datasets die je hebt gedownload.

Beschrijf de gebruikte software tools, alsmede de bibliotheken/plugins, met naam, versie, referentie en gebruiksdoel (in dit project). Dit kan eventueel in een tabel aan als die lang is mag het een online bijlage zijn.

Beschrijf bestaande methodologieën met hun relevantie voor je project. Geef het doel, de toepassing en welke software en parameters er zijn gebruik. Voeg eventueel een flowchart toe. Beschrijf de gebruikte statistische methoden.

Beschrijf wat je zelf in het kader van dit onderzoek hebt ontwikkeld aan methodologieën. Geef de naam van scripts/programma's en waar deze (in je repo) te vinden zijn.

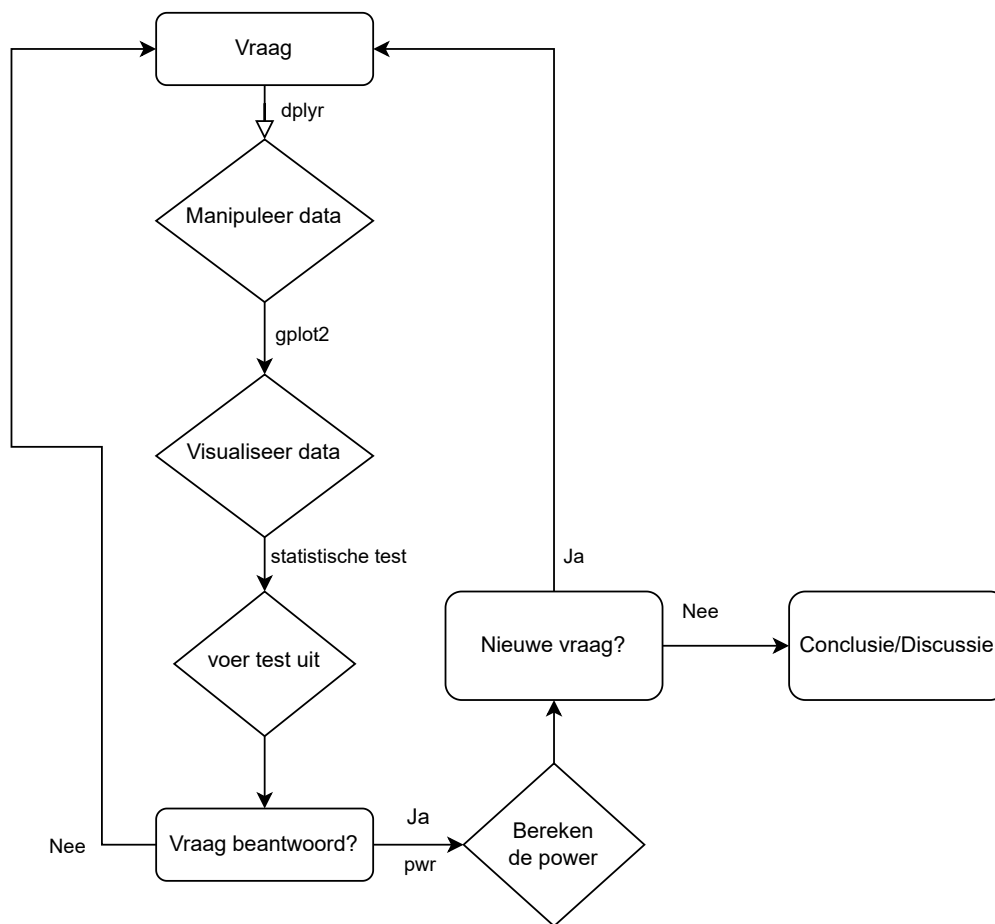
Deze sectie bevat tussen de 400-800 woorden.

Om antwoord te krijgen op onze onderzoeksvraag is de data getest, verwerkt en gevisualiseerd met R. R:4.4.0 is een statistische programmeertaal, die gebouwd is om data te verwerken, manipuleren en visualiseren. Er is gebruik gemaakt van een IDE om in te werken: Rstudio:2024.04.1 build 748. (Posit Team, z.d.; R Core Team 2012)

Voor het manipuleren en visualiseren is gebruikt gemaakt van dplyr:1.1.4 en ggplot2:3.5.1. dplyr is een library die gebruikt is om data te manipuleren, verwerking, selecteren etc. Het verwerken van data is een stuk gemakkelijker via deze library, en werkt hand in hand met ggplot2. ggplot2 is gebruikt om figuren te maken. De library zorgt ervoor dat er gemakkelijk en snel, mooie duidelijke figuren gemaakt kunnen worden. (Wickham e.a. 2023; Wickham 2016)

In de logboeken is gebruik gemaakt van pwr:1.3.0 en gridExtra:2.3. pwr kan de power van een test berekenen. Voor dit onderzoek is pwr ook gebruikt om te kijken wat de sample size moet zijn voor het onderzoek. gridExtra is gebruikt om meerdere plotjes op 1 chart te maken, maar dit bleek ook mogelijk te zijn met de facet wrap functie in R.

De gebruikte workflow staat weergegeven in de aankomende flowchart.



flowchart 1, de workflow van het analyseren van de data.

De gebruikte workflow was relatief simpel, we hebben een hoofdvraag. En om die te beantwoorden moesten we kijken naar verschillende variabelen van de plant. Groei, kleur, aantal blaadjes, etc. Dus voor elke variabele was een nieuwe vraag. Deze vragen hebben we verwerkt door de data eerst te manipuleren en daarna visualiseren. Vervolgens is er gekeken naar deze visualisaties om te kijken of er opvallende nummers waren, of mogelijk zelfs duidelijke verschillen tussen factoren. Als er iets opvallend te zien was in de visualisatie zijn er vervolgens statistische testen uitgevoerd op de data. Dit is gedaan om te controleren of er een statistisch significant verschil zit tussen de invloed op de variabele tussen 1 of meer factoren.

Als de gestelde vraag niet beantwoord was, begonnen we overnieuw. Mogelijk met een net wat andere vraag. In het geval dat de vraag beantwoord is, is de power van de test berekend. En is er gekeken naar de volgende stap: Zijn er nog meer vragen? Of kan er een conclusie geformuleerd worden?

Voor het testen van de data maken we gebruik van een aantal verschillende testen.

Dit omdat we verschillende soorten data hebben. Er is gebruik gemaakt van een 2-way t-test. De t-test wordt gebruikt om het verschil te vinden tussen 2 datasets, met 1 factor die verschillend is. Deze is gebruikt voor het testen of er een significant verschil is tussen 2 duplo's. De effect sterkte tussen de 2 duplo's te berekenen is gedaan door cohen's d te berekenen. -2 way chisq test voor ontkieming van de zaadjes.

Voor het onderzoek zijn ook een paar functies gemaakt in R, deze zijn te vinden in: [Functies](#)

De volgende functies zijn gemaakt en gebruikt:

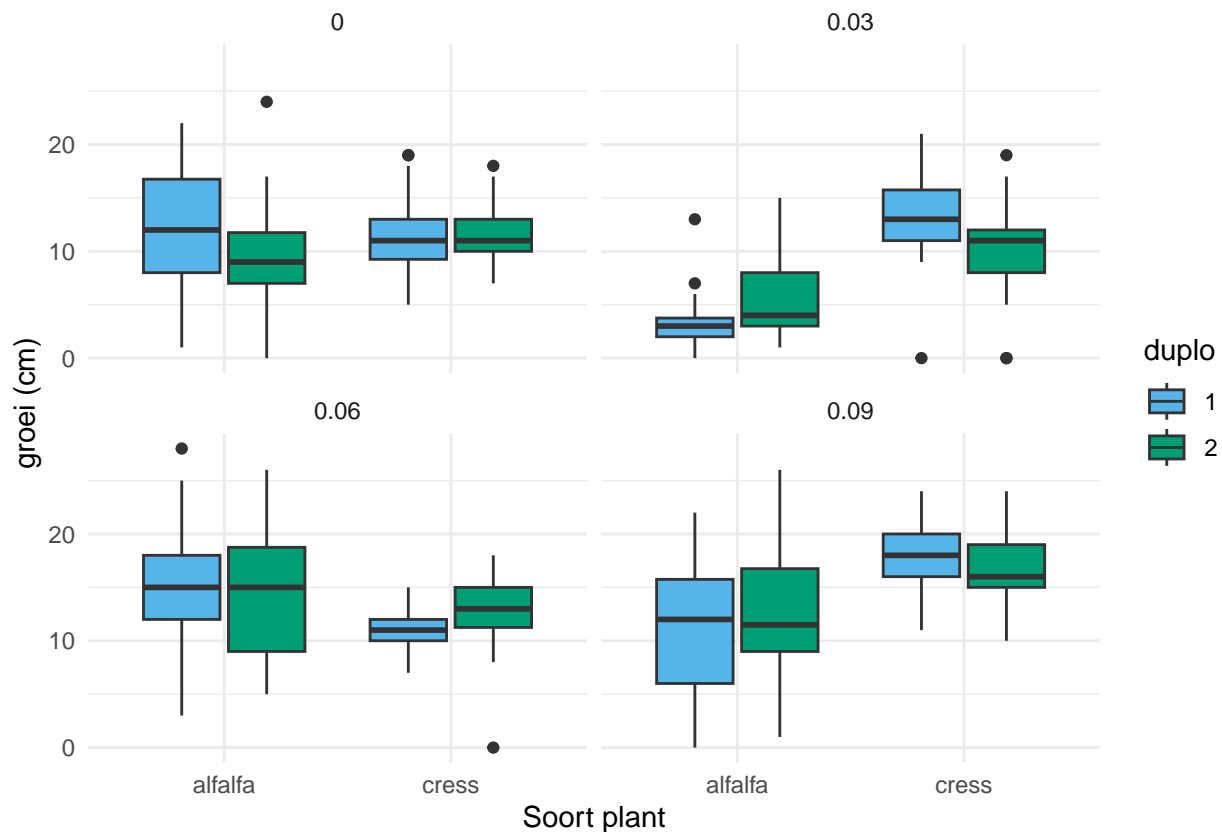
-duplo_test, een functie die gemakkelijk het verschil tussen de 2 duplo's test.

-cohens_d_duplo, een functie die cohen's d berekend voor 2 gegeven duplo's.

4 Resultaten

4.1 Duplo's

Aan het einde van de datacollectie is de lengte en de kleur van de plaatjes bepaald van 800 plantjes verspreid over 2 soorten en 4 verschillende zout oplossingen. Omdat er met duplos gewerkt is is er eerst gekeken naar verschil tussen duplo's (figuur 1).



Figuur 1: spreiding van groei (cm) van de 2 soorten over elke oplossing. Deze grafiek geeft de spreiding van de groei van de plantjes weer voor elke oplossing en soort plant. waarbij de x-as de platen soort is en de y-as de groei in cm. Ook zijn de duplo's van elkaar gesplitst, zo kan het verschil tussen deze duplo's gevisualiseerd worden. Uitschieters zijn weergegeven als de zwarte punten.

De spreiding van groei tussen de duplo's van tuinkers bij een oplossing van 0.09% en 0% is niet heel groot, de boxen liggen zo goed als naast elkaar. Bij 0.06 en 0.03 van de tuinkers blijkt het verschil iets groter. Bij alfalfa zijn de boxen bij 0.09 0.06% niet helemaal even groot, maar de gemiddelden zijn wel gelijk. bij 0 en

0.03% wijken de boxen en gemiddelden meer af van elkaar.

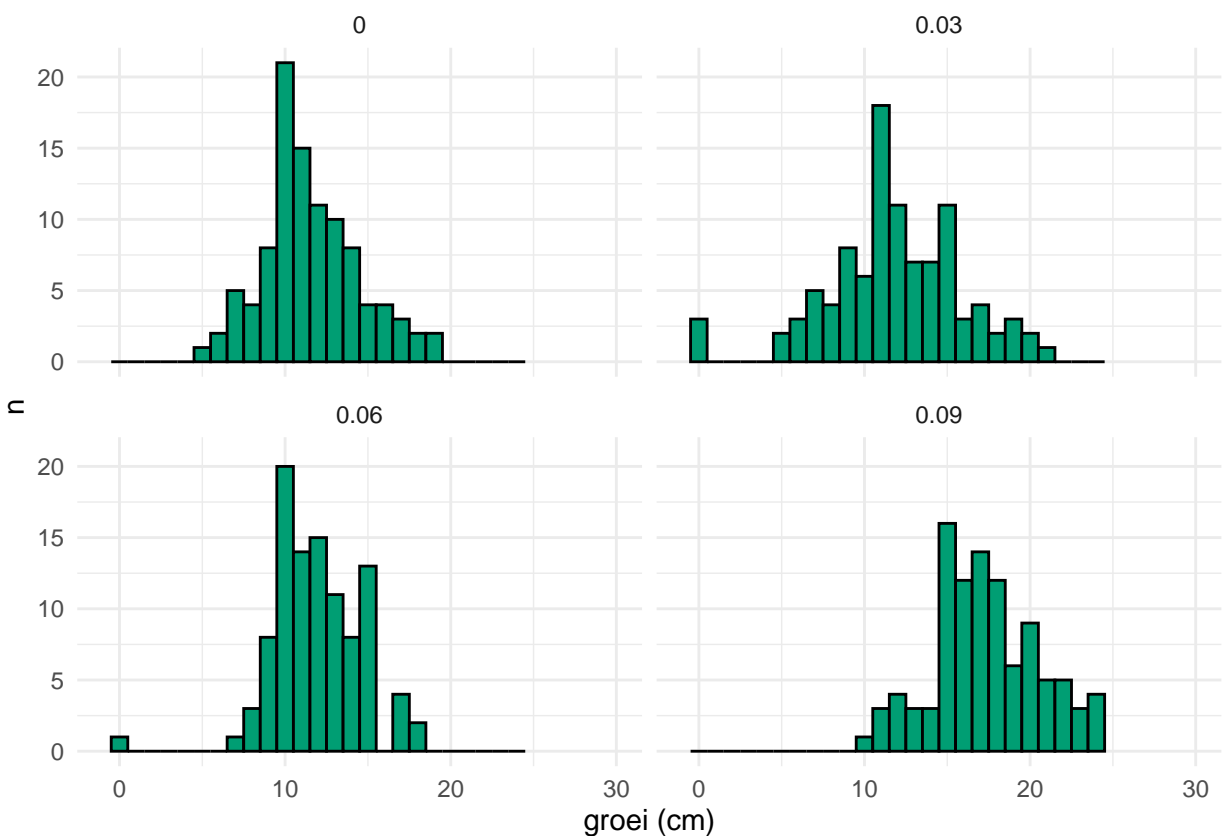
Bij een aantal duplo's is het niet duidelijk of er een significant verschil tussen zit, om deze reden zijn er t-testen uitgevoerd voor de volgende duplo's: alfalfa en cress bij oplossing 0.03, cress bij oplossing 0.06, en alfalfa bij 0%.

Voor alle 4 de testen geldt $p < 0.05$, dus er zit bij deze een significant verschil tussen de geteste duplos. Om te bepalen of dit verschil relevant is, is Cohen's d berekend.

Cohen's $|d|$ voor cress 0.03% en alfalfa 0.03% ligt rond de 0.85 er kan dus gesproken worden over een groot verschil tussen deze duplo's. Cohen's $|d|$ voor cress bij 0.06% ligt rond de 0.68, er kan dus gesproken worden over een matig/groot verschil tussen deze duplo's. Cohen's $|d|$ voor alfalfa 0% ligt rond de 0.5, er kan dus over een matig verschil gesproken worden tussen deze duplo's.

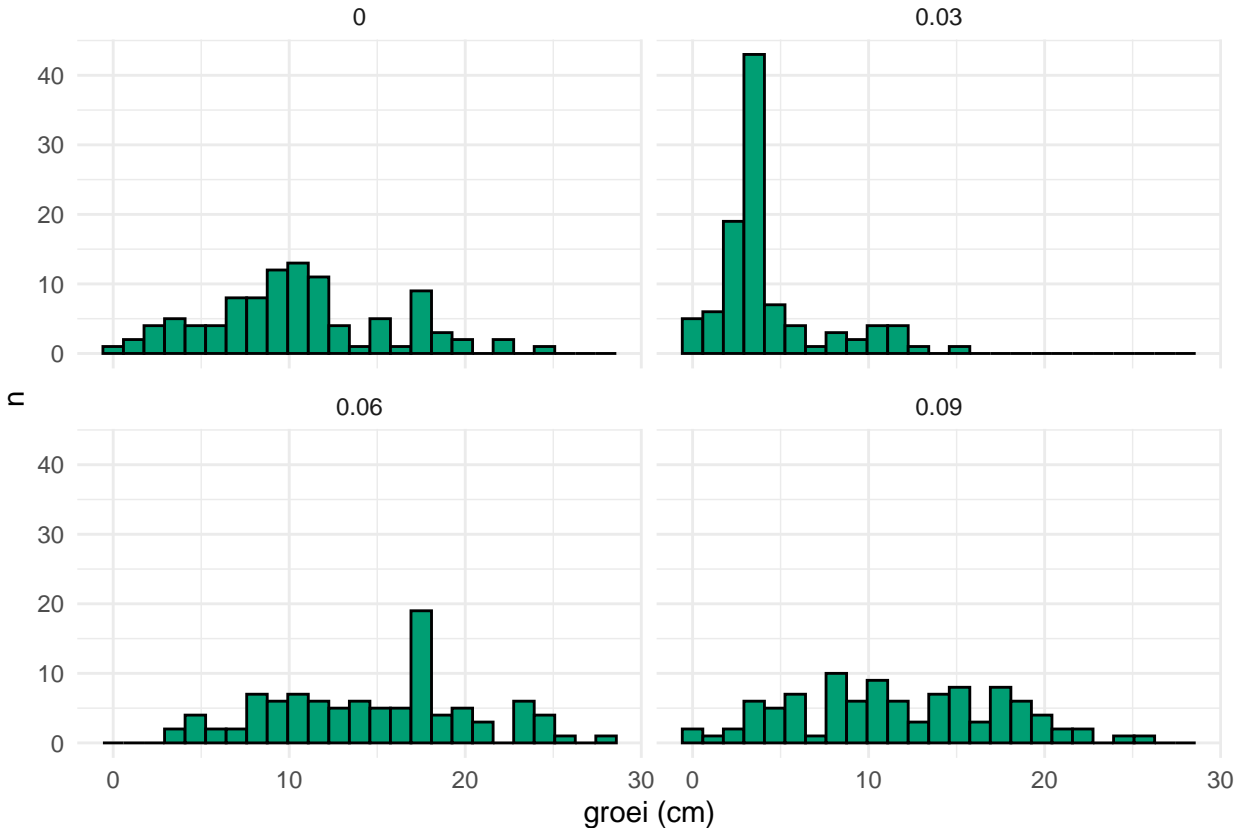
4.2 Groei van de planten

Om te bepalen of iets invloed heeft gehad op de groei van een plant is het een goede eerste stap om naar de groei te kijken van een plant. De gemeten groei van elke plant kan gevonden worden in figuur 2 en 3.



Figuur 2: Verdeling van groei in cm van cress voor elke oplossing. Deze grafiek geeft de verdeling weer van de groei van cress voor elke oplossing. Waarbij de x-as de groei in cm is, en de y-as hoeveel deze groei voorkomt. Boven elke grafiek staat een nummer, deze representeert de oplossing.

Wat zichtbaar is uit deze grafiek is dat bij 0.09% de verdeling een stuk meer naar rechts licht dan bij de andere oplossingen. de verdeling van groei bij 0, 0.03 en 0.06 blijkt allemaal wel in de buurt te liggen van elkaar. Alleen is de spreiding bij 0.06 een stuk minder breed.



Figuur 3: Verdeling van groei in cm van alfalfa voor elke oplossing. Deze grafiek geeft de verdeling weer van de groei van alfalfa voor elke oplossing. Waarbij de x-as de groei in cm is, en de y-as hoeveel deze groei voorkomt. Boven elke grafiek staat een nummer, deze representeert de oplossing.

Wat zichtbaar is uit deze grafiek is dat de data van de alfalfa veel meer is verspreid dan bij tuinkers. Dit kon ook al gelezen worden uit figuur 1. Als we deze oplossingen vergelijken met elkaar blijkt er niet veel verschil tussen te zijn. Oplossing 0.06 is de enige die hier wel buiten valt, deze oplossing is de enige met een grote piek. Als we de 2 soorten met elkaar vergelijken zien we snel dat tuinkers minder spreiding heeft dan alfalfa.

Om te bepalen of soort, zout en de interactie van deze 2 variabelen invloed heeft op de groei van de planten is een 2-way anova test uitgevoerd.

Hieruit blijkt dat er een significante interactie is tussen de zout oplossingen en planten soort. ($p = 2e-16 < 0.05$).

Fisher's eta2 van licht, vocht en de interactie van deze 2 variabelen is berekend om de effect sterkte te bepalen.

De interactie tussen oplossing en soort plant heeft een sterk effect op de groei van de planten (eta2.interactie ~ 0.13). De verschillende percentages zout hebben een sterk effect gehad op de groei van de planten (eta2.oplossing

~ 0.2). De 2 soorten hebben een matig effect gehad op de groei van de planten ($\eta^2_{\text{soort}} \sim 0.06$).

De pwr van deze test is berekend.

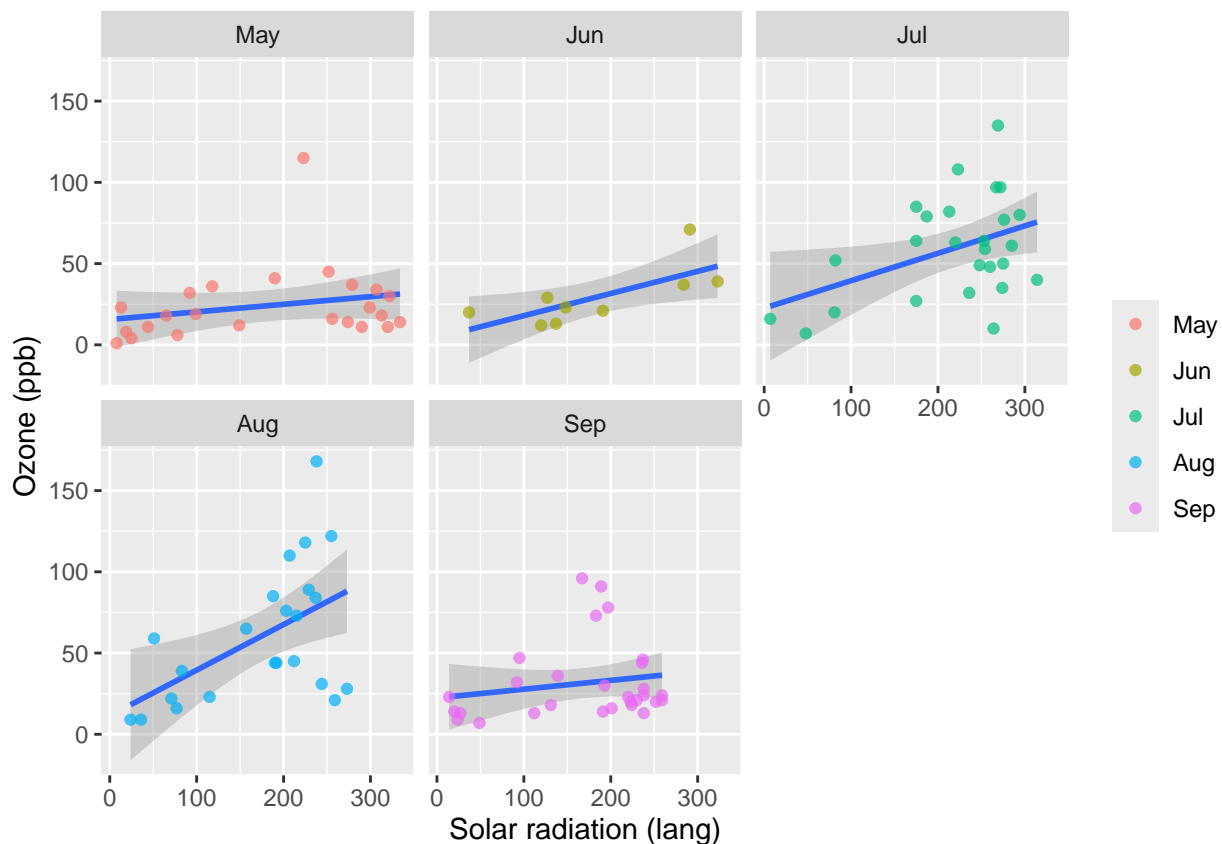
Uit deze berekening is een power van $\sim 96\%$ gekomen.

Er is een log lineaire test uitgevoerd op de verdeling van kleur over soort en oplossing.

Presenteer je resultaten in een logische volgorde. Beschrijf wat er te zien is en werk toe naar het beantwoorden van je doelstelling. Beschrijf indien mogelijke de logica van de keuze voor opeenvolgende experimenten. In elke paragraaf worden de feitelijke conclusies gegeven (bv 'de vergelijking laat zien dat het gemiddelde van groep A significant afwijkt van het gemiddelde van groep B, met een p-waarde van ... Dit komt niet overeen met de in de literatuur (REF) beschreven waardes.')

Gebruik zo veel mogelijk figuren om je informatie uit je resultaten over te brengen. Gebruik tabellen wanneer figuren minder geschikt zijn. In de tekst worden figuren en tabellen geïntroduceerd, besproken en de belangrijkste aspecten toegelicht.

Voorzie je figuren van een nummer en een beschrijvende titel. Zorg voor correcte as-labels (eenheid en grootte), legenda en bijschrift. Hier is een voorbeeldje.



Figuur 4: Ozon concentraties geplot tegen zonlicht intensiteit, vergeleken over verschillende maanden. Lineaire modellen (blauwe lijnen) en confidence intervals (grijze gebieden) zijn hieraan toegevoegd.

Geef tabellen bovenaan een titel en bijschrift die de inhoud beschrijft en onderaan voetnoten die kolomnamen of specifieke waardes verklaren.

Deze sectie bevat tussen de 600-1200 woorden.

5 Discussie en Conclusies

Formuleer je conclusie door eerst in te zoomen op je eigen data en daarna uit te zoomen. Zoom in door je resultaten samen te vatten. Zoom uit om de waarde van je werk te beoordelen, door je bijvoorbeeld de volgende vragen te stellen:

- Kunnen mijn resultaten gebruikt worden in het werkveld?
- Wat betekenen ze voor het werkveld?
- Zijn mijn data betrouwbaar?

Bespreek de resultaten zodanig dat je ze ter discussie stelt, wees kritisch. Vergelijk je resultaten met de literatuur of eerder ontwikkelde data. Geef aanbevelingen voor een vervolg en staaf je aanbevelingen door de impact op wetenschappelijk of maatschappelijk vlak te beschrijven.

Kom ten slotte altijd terug op de doelstelling (en hypothesen).

Deze sectie bevat tussen de 400-800 woorden.

6 Online bijlagen

Vaak zijn online bijlagen vele malen groter dan het eigenlijke artikel. Wees nooit bang om te veel aan bijlagen aan te bieden. Je kan hierbij denken aan

- de ruwe data
- de code voor dataverwerking
- de code voor analyse
- aanvullende figuren en tabellen

Natuurlijk is een git(hub) repo daar de beste plek voor! Zorg ervoor de je repo logisch is ingericht met goede Readme document(en). Ook de code zelf is waar nodig natuurlijk goed gedocumenteerd.

6.1 Wordcount

Voeg aan het eind een woord-telling in:

Method	koRpus	stringi
Word count	2383	2291
Character count	14261	14316
Sentence count	218	Not available
Reading time	11.9 minutes	11.5 minutes

7 Referenties

Een lijst van referenties wordt hier automagisch toegevoegd.

- Posit Team. z.d. *Rstudio: An IDE for R*. Posit. <https://posit.co/products/open-source/rstudio/>.
- R Core Team. 2012. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org/>.
- Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, en Davis Vaughan. 2023. *dplyr: A Grammar of Data Manipulation*. <https://dplyr.tidyverse.org>.
- Xie, Yihui. 2013. *Dynamic Documents with R and knitr*. Boca Raton, FL: CRC Press.