

Empirisch-Wissenschaftliches Arbeiten

Übung zur computergestützten Datenanalyse

© Prof. Dr. Stephan Huber

11. Juli 2024

Inhaltsverzeichnis

Vorwort	1
I. Einleitung	4
1. Die Programmiersprache R	5
2. Wissenschaftliche Texte schreiben	6
2.1. WYSIWYG Anwendungen	6
2.2. Vorteile von codebasierten Anwendungen	7
2.3. Einführung in Quarto	8
2.4. Erste Schritte mit Quarto	9
2.5. APA konformes Manuscript erstellen mit Quarto (apaquarto)	9
2.6. Vorlage zur Hausarbeit mit Quarto	9
II. Anwendungen	10
3. Daten einlesen und aufbereiten	11
4. Zwei-Wege-ANOVA-Modellen	12
4.1. R-Sitzung einrichten	12
4.2. Daten einlesen	12
4.3. Deskriptive Statistik	14
4.3.1. Tabellarisch	14
4.3.2. Grafisch	15
4.4. t-Test	17
4.4.1. One Sample t-test	17
4.4.2. Two sided t-test	18
4.5. ANOVA	20
4.6. Diagnostics	21
4.7. Interaktions Diagramm	21
4.8. Multiple-Vergleichs-Test	23
4.9. Schlussfolgerungen ziehen und Ergebnisse präsentieren	25
5. ANOVA Ergebnisse und Quarto	29
Literatur	30

Abbildungsverzeichnis

4.1. Boxplots	16
4.2. Boxplots mit <code>ggbetweenstats</code>	17
4.3. Grafische Veranschaulichung des Models	24

Tabellenverzeichnis

4.1. Deskriptive Statistiken	15
4.2. ANOVA Ergebnisse	21

Vorwort

💡 Eine PDF-Version dieser Notizen ist [hier verfügbar](#).

Bitte beachten Sie, dass der Inhalt der PDF-Version identisch ist, sie jedoch nicht für das PDF-Format optimiert wurde. Daher könnten einige Teile nicht wie beabsichtigt erscheinen.

Wikipedia sagt:

“Die Psychologie [...] ist eine empirische Wissenschaft”

Diese Unterlagen helfen...

- die Abfolge und die Inhalte der *Übung zur computergestützten Datenanalyse* zu überblicken,
- die Übungsaufgaben zu verstehen und zu bearbeiten und
- die Projektarbeit der Veranstaltung *Empirisch-Wissenschaftliches Arbeiten* erfolgreich zu gestalten.

Die Übung vermittelt...

- Kenntnisse der Programmiersprache R welche eine wissenschaftliche Datenbearbeitung und Datenanalyse ermöglichen.
- Kenntnisse zum programm-basierten Verfassen von wissenschaftlichen Arbeiten (Aufsätze, Bücher, Arbeitspapiere, Hausarbeiten).

Studierende lernen...

- Daten mit der Programmiersprache R und mit Hilfe der integrierten Entwicklungsumgebung RStudio einzulesen, zu bearbeiten und empirisch auszuwerten.
- Empirische Ergebnisse in ein publikationswürdiges Format zu übertragen.
- Einen APA konformes Manuskript mit Quarto, bzw. (R)Markdown, zu erstellen und dies entsprechend zu publizieren.
- Literatur entsprechend wählbaren Zitationsregeln unter Verwendung von Quarto und BibTeX in einen Aufsatz einzuarbeiten.

Studierende sollen...

- Die angeführte Literatur studieren: Ohne eigenständige Vor- und Nachbereitung lassen sich die Programmierkenntnisse nicht erlernen.
- Aktiv um Hilfe bitten: Wenn etwas unklar ist, kann ich individuell während des Kurses versuchen zu helfen. Für eine intensivere Betreuung, bitte ich mich zu kontaktieren, in die [Sprechstunde](#) zu kommen, oder eine außerordentliche Sprechstunde zu vereinbaren. Dies ist möglich und erwünscht.
- Inhaltliche Fragen und Wünsche jederzeit kommunizieren. Es besteht die Möglichkeit diese in das Curriculum aufzunehmen.

Liebe Studierende,

das Erlernen einer Programmiersprache in Verbindung mit empirischen Arbeiten ist eine Herausforderung die Vielen keinen Spaß macht. So ist es nur Verständlich, dass die Sinnhaftigkeit dieses Kurses teilweise von Studierenden angezweifelt wird. Tätigkeiten die keinen Spaß machen, sollten sinnstiftend sein oder zumindest ein monetäres Einkommen sichern. Da das Vorhandensein von empirischen Kenntnissen und einer Programmiersprache in einem Lebenslauf zweifelsfrei in der heutigen Zeit die Vermittlungsfähigkeit und die Verhandlungsposition am Arbeitsmarkt wesentlich verbessern, will ich mich hier kurz bemühen, die Sinnhaftigkeit zu thematisieren.

Ich verstehe die Abneigung gegenüber diesen Kurs: Viele haben sich nicht für ein Studium der Psychologie entschieden, um empirische Methoden und deren computergestützte Umsetzung zu erlernen. In der modernen Welt aber, insbesondere in der psychologischen Forschung, ist ein Verständnis von empirischen Methoden sowie deren computergestützten Umsetzung die praktische Voraussetzung zum Erkenntnisgewinn. Ohne dieses Verständnis verharret man bei rein theoretische und philosophische Überlegungen ohne jede Evidenz. Eine professionell agierende Psychologin und Psychologe, sollte die Fähigkeit besitzen die Literatur in seinem Fach zu begreifen sowie in der Lage ein die Ergebnisse kritisch zu hinterfragen und/oder zu überprüfen.

Ich bemühe mich, die Veranstaltung so attraktiv wie möglich zu gestalten. Ich biete...

- ein ausführliches [Skript zur Programmiersprache R](#) an, welches
 - eine Batterie an [Übungsaufgaben mit Lösungskripten](#) und
 - eine Vielzahl an [interaktive Übungen](#) zum eigenständigen bearbeiten enthält.
- dieses Skript, welches
 - Psychologie-spezifische empirische Inhalte aufgreift und
 - Software vorstellt, welche die Erstellung der Projektarbeit erleichtert.
- mündliche Erklärungen in der Veranstaltung.
- die Möglichkeit spezifische Fragen zu stellen und Unklarheiten anzusprechen.
- individuelle Betreuung während und außerhalb der [Sprechstunde](#).

Wenn sie Vorschläge und Wünsche bezüglich der Inhalte oder der didaktischen Aufbereitung haben, bitte ich diese auszusprechen. Konstruktive Kritik ist sehr willkommen. Ich nehme diese an und ernst. Ob Sie diesen Kurs letztendlich als gelungen betrachten, ist ihrer Wahrnehmung überlassen. Bevor Sie den Kurs aber schlecht evaluieren, bitte ich sie um Folgendes: Fragen Sie sich, ob ihr Wille und ihr Wunsch ausgeprägt genug waren, um sich ernsthaft mit den Inhalten

und den Angeboten auseinanderzusetzen und ob sie mir evtl. die Gelegenheit gegeben haben auf Ihre Wünsche einzugehen.

Abschließend wünsche ich Ihnen viel Freude mit dem Kurs und den angebotenen Unterlagen und Inhalten. Ich freue mich, diesen Kurs halten zu dürfen und zu können. Es ist mir stets eine Freude, den anwesenden Studierenden R, Quarto, BibTeX und Co. erklären zu können. Ich wünsche mir, möglichst Viele mit den dargebotenen Inhalten, das Studium zu bereichern und die Bearbeitung der Projektarbeit sowie der Abschlußarbeit zu erleichtern.

Ihr
Stephan Huber

Lizenz

Dieses Skript wird unter der *Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International* Lizenz veröffentlicht. Das bedeutet, dass die Inhalte wiederverwendet, remixt, behalten, überarbeitet und weiterverbreitet werden können, solange den Autoren angemessenes Credit gegeben wird. Wenn Sie die Originalversion dieses offenen Skripts remixen oder modifizieren, müssen Sie alle Versionen dieses offenen Skripts unter derselben Lizenz weiterverbreiten.

Teil I.

Einleitung

1. Die Programmiersprache R

Ich bitte Sie, studieren sie das Skript *How to use R for data science* [Huber, 2024b].

In den ersten Wochen werden wir uns ausschließlich damit beschäftigen, die Programmiersprache R zu erlernen. Das ist ähnlich mühsam wie das Erlernen einer wirklichen Sprache. Wer keine Lust darauf hat, wird es schwer haben. Ich beispielsweise hatte in der Schule überhaupt keine Lust auf Englisch und Latein. Dementsprechend schlecht waren meine Noten. Ich musste die siebte Klasse wiederholen und bis zum Abitur waren Sprachen für mich ein nötiges Übel. Erst als ich im Studium sah, dass praktisch alle relevanten und für mich interessanten Artikel und Bücher in englischer Sprache verfasst sind, machte das Erlernen der Sprache einen Sinn für mich. Jetzt lehre ich abseits dieses Kurses ausschließlich auf Englisch und publiziere in englischer Sprache. Interesse und Freude sind mächtige Katalysatoren für Erfolg.

Das Schreiben von Code ist für die meisten Studierenden Neuland. Studierende im Jahr 2024 sind zumeist mit dem Smartphone aufgewachsen und demnach sind Sie es gewohnt, ihre Geräte (Smartphone, Tablet, Desktop-PC) ohne zur Hilfenahme einer Programmiersprache zu steuern. Das ist wunderbar: Die grafische Benutzeroberfläche heutzutage erlaubt eine effiziente und intuitive Art der Steuerung mit der Computermaus, durch Wischen, Tippen oder durch Spracheingabe. Leider hat diese Art der Steuerung massive Nachteile beim wissenschaftlich orientierten Arbeiten mit Daten. Insbesondere was die Reproduzierbarkeit der Ergebnisse und die Flexibilität des Arbeitsprozesses anbelangt, stößt man bei Applikationen ohne Code an Grenzen. Die Vor- und Nachteile von Script-basierten Arbeiten werden im Kapitel *The limitations of no-code applications* [Huber, 2024b] ausführlich erläutert.

Zusammenfassend sollten Studierende nach den ersten 5-6 Unterrichtseinheiten folgendes getan haben beziehungsweise erlernt haben:

- Installation von
 - R,
 - RStudio und der
 - gängigsten Pakete.
- Wissen über...
 - den Aufbau von R Skripten.
 - die Verwendung von Funktionen, Objekten und Pakete in R.
 - die grundsätzlichen Eigenheiten der Programmiersprache R.
 - das Ausführen von Code (**Ctrl+Enter**, Klicken von **Run**, oder durch die Funktion `source()`).
 - die Verwendung von Pipes mit dem Pipe Operator (`|>`).
 - die Verwendung von logischen und relativen Operatoren.
 - die Funktionen des Pakets `dplyr` (`filter()`, `select()`, `mutate()`, `summarise()`, etc.)

2. Wissenschaftliche Texte schreiben

Studierende verwenden zum Verfassen wissenschaftlicher Texte gerne Microsoft Word, Apples Pages oder LibreOffice. Diese Textverarbeitungsprogramme zeigen das Dokumentenlayout bereits während des Schreibens an. Dies wird auch als “What you see is what you get” (WYSIWYG) Prinzip bezeichnet. Dieses Prinzip und die entsprechenden Anwendungen sind weit verbreitet und erscheinen vielen als alternativlos. Dies trifft jedoch keineswegs zu: Es gibt eine Vielzahl alternativer Textsatz-Systeme wie LaTeX, Markdown, R Markdown und Quarto, die beachtenswerte Vorteile bieten. Nicht ohne Grund nutzen viele professionell arbeitende Wissenschaftler und Publizisten diese Alternativen. Eine große Anzahl von Doktorarbeiten und wissenschaftlichen Aufsätzen wird mit LaTeX verfasst, und nahezu alle Herausgeber und Verlage arbeiten mit codebasierten Alternativen, die nicht dem WYSIWYG-Prinzip folgen.

Bei codebasierten Alternativen werden die Angaben zum Layout entweder an den Anfang des Textes oder direkt in den Fließtext eingefügt. Das endgültige Dokument wird erst nach der Umwandlung (auch Kompilieren oder Rendern genannt) in ein Dokumentformat wie PDF sichtbar. Dies mag zunächst gewöhnungsbedürftig sein und sicherlich weniger intuitiv als eine WYSIWYG-Benutzeroberfläche. Doch die intuitivste Lösung ist nicht zwangsläufig die beste oder einfachste. Die vielen Studienarbeiten, die ich betreuen durfte, zeigen deutlich, dass die intuitiven Features von MS Word und Pages sich mittel- und langfristig oft als zeitaufwändig erweisen und Nutzer nur unzureichend dabei unterstützen, Fehler beim Verfassen von wissenschaftlichen Arbeiten zu vermeiden. Studierende, die sich gegen eine WYSIWYG-Office-Anwendung entscheiden, sind in der Regel weniger frustriert und erfolgreicher – zumindest trifft dies auf die von mir betreuten Arbeiten zu.

Codebasierte Anwendungen ermöglichen es den Schreibenden, sich auf die eigentliche Textarbeit zu konzentrieren. Die Feinheiten des Formats und die Einhaltung von Zitationsregeln werden größtenteils automatisch von der Software übernommen. Die notwendige Anfangsinvestition, sich etwa mit Quarto vertraut zu machen, macht sich schnell bezahlt, und die Qualität der wissenschaftlichen Texte verbessert sich spürbar.

In den folgenden Unterabschnitten werde ich zunächst die typische Nutzung von WYSIWYG-Anwendungen umreißen. Anschließend werde ich die Vorteile des codebasierten Verfassens von Texten am Beispiel von Quarto beleuchten, um dann zu erläutern, wie das Verfassen von Texten mit Quarto gelingen kann.

2.1. WYSIWYG Anwendungen

Die Nutzung klassischer Textverarbeitungsprogramme wie Microsoft Word oder Apple Pages zum Verfassen wissenschaftlicher Texte ist in der studentischen Welt weit verbreitet. Obwohl diese Programme für alltägliche Schreibprojekte benutzerfreundlich sind, erzeugen sie erheblichen Mehraufwand, um den Ansprüchen wissenschaftlicher Arbeit gerecht zu werden.

Ein erstes Problem ist die Einbindung von Literatur. Die korrekte Formatierung nach verschiedenen Zitierrichtlinien ist oft alles andere als intuitiv und Fehler treten leicht auf. Dies gilt insbesondere, wenn die von der Software bereitgestellten Zitat- und Bibliografiefunktionen

nicht oder nicht richtig genutzt werden. Anstelle externer Zitationsmanager zu nutzen und sich in deren Gebrauch einzuarbeiten, verfassen viele Studierende Zitate und Literaturlisten manuell. Dies führt erfahrungsgemäß zu zahlreichen kleinen und manchmal größeren Fehlern, die vermeidbar wären.

Ein weiterer Schwachpunkt von studentischen Arbeiten ist die Einhaltung spezifischer Formatierungsvorgaben. Akademische Institutionen und Journale fordern oft eine strenge Beachtung von Formatierungsrichtlinien, inklusive der Gestaltung von Titelseiten, Kopf- und Fußzeilen, Seitenrändern und Überschriftenhierarchien. Zwar bieten Word und Pages Vorlagen und Stile an, diese müssen jedoch für jedes Dokument individuell angepasst und oft aufgrund geringfügiger Änderungen im Text modifiziert werden. Wenn eine Formatanpassung erforderlich wird, ist dies meistens nur mit großem Aufwand möglich.

Das Einfügen empirischer Ergebnisse wie statistischer Daten und Grafiken stellt eine zusätzliche Hürde dar. In Word und Pages gestaltet sich der Vorgang häufig manuell: Forschungsdaten müssen aus Statistiksoftware exportiert, als Bilder abgespeichert und anschließend in das Dokument eingebunden werden. Ändert sich etwas an den Daten, so muss dieser mühsame Prozess wiederholt werden, was den Arbeitsaufwand signifikant erhöht und die Fehleranfälligkeit steigert.

2.2. Vorteile von codebasierten Anwendungen

Der traditionelle Ansatz zum Verfassen wissenschaftlicher Texte mittels MS Word oder Pages kann für Studierende zeitaufwendig und fehleranfällig sein. Im folgenden Abschnitt möchte ich Quarto (bzw. R Markdown) vorstellen, eine moderne Alternative, die folgende Vorteile bietet:

- Mit Quarto lassen sich mühelos verschiedene Ausgabeformate generieren. So kann derselbe Text beispielsweise als Website (HTML), Manuskript (PDF, DOCX), Buch (EPUB, PDF) oder in Form von Präsentationsfolien (PDF). Diese Flexibilität erlaubt es, sich eher auf den Inhalt als auf das Format zu konzentrieren.
- Die Formatierung kann in Quarto einfach geändert werden indem bestimmte Vorlagen verwendet werden.
- Literaturreferenzen lassen sich unkompliziert einbinden, während die Einhaltung von Zitierregeln von der Software übernommen wird. Quartos Integration mit Zitationsverwaltungssystemen ermöglicht es, Literaturverweise und Bibliografien effizienter und konsistenter zu handhaben als beispielsweise in Word.
- Querverweise auf Abschnitte, Tabellen und Abbildungen lassen sich leicht erstellen.
- Die Datenanalyse und das Erstellen von Datenoutputs erfolgen direkt in Quarto. Dadurch sind dargestellte Grafiken und Tabellen stets aktuell und manuelles Nachbearbeiten entfällt, wodurch die Reproduzierbarkeit der Ergebnisse gewährleistet ist.
- Forschende können ihre Datenvisualisierungen ohne manuelle Zwischenschritte direkt in den Text einbetten.
- Versionskontrollsysteme wie Git erleichtern die Zusammenarbeit an wissenschaftlichen Dokumenten, da Änderungen nachverfolgbar sind und integriert werden können, ohne auf komplexe und konfliktträchtige Vergleichstools angewiesen zu sein.

Lesetipp

Für Interessierte empfehle ich den Onlinekurs *Introduction to Reproducible Publications with RStudio*, der explizit erläutert, wie man empirisch nachvollziehbar arbeitet. Eine etwas kompaktere Einführung bietet [Bauer and Landesvatter \[2023\]](#) und das Standardwerk zum

Thema stammt von Gandrud [2020].

2.3. Einführung in Quarto

Quarto kann in RStudio genutzt werden, um APA-konforme Texte zu erstellen. Gehen Sie dazu bitte wie folgt vor:

- Installieren Sie R und R Studio.
- Installieren Sie Quarto folgendermaßen:

```
install.packages("quarto")
```

- Installieren Sie das Paket `tinytex`, um PDF-Dateien zu generieren:

```
install.packages("tinytex")  
tinytex::install_tinytex()
```

- Es ist zudem ratsam, weitere Pakete zu installieren, die später benötigt werden könnten:

```
if (!require(pacman)) install.packages("pacman")  
pacman::p_load(knitr, rmarkdown, papaja)
```

- Eignen Sie sich Kenntnisse in Markdown an. Markdown ist eine leichtgewichtige Markup-Language zur Formatierung von Klartext. Sie ist eine essenzielle Fähigkeit für die effektive Nutzung von Quarto. Beginnen Sie damit, ausreichend Markdown für die Strukturierung Ihrer Arbeit zu erlernen, einschließlich Überschriften, Listen, Links und Codeblöcken. Markdown ist schnell zu erlernen; ich empfehle dazu den Besuch von www.markdowntutorial.com und das Durcharbeiten der interaktiven Lektionen sowie des Abschnitts *Markdown Basics* auf quarto.org.
- Machen Sie sich vertraut mit Quarto. Als Lektüre dient Telford [2023]: *Enough Markdown to Write a Thesis*, welcher fast alles abdeckt, was für das akademische Schreiben hilfreich ist. Alternativ finden Sie umfassende Informationen zur Arbeit mit Quarto direkt auf der Webseite quarto.org/docs/guide.

i Quarto und R markdown

Quarto ist ein relativ neues Werkzeug und kann als Nachfolger von R Markdown betrachtet werden. Die meisten R Markdown-Dokumente sind mit Quarto kompatibel. Allerdings bietet Quarto einige verbesserte Funktionen gegenüber R Markdown, die die Benutzerfreundlichkeit steigern. Einen detaillierten Überblick über die Unterschiede und Gemeinsamkeiten zwischen den beiden Plattformen finden Sie in [diesem Artikel] (<https://quarto.org/docs/faq/rmarkdown.html>).

2.4. Erste Schritte mit Quarto

- Öffnen Sie RStudio.
- Wählen Sie “File” -> “New File” -> “Quarto Document” und dann “Create”.
- Speichern Sie die neue Datei in einem leeren Ordner und definieren Sie diesen Ordner als Ihr Arbeitsverzeichnis.
- Klicken Sie auf “Render”.
- Besuchen Sie die Webseite [Markdown Basics](#), fügen Sie etwas Markdown in Ihr Dokument ein und klicken Sie erneut auf “Render”.
- Klicken Sie auf den Pfeil neben dem “Render”-Knopf. Hier können Sie andere Dateiformate auswählen und diese generieren. Probieren Sie es aus.
- Konsultieren Sie die Webseite PDF Basics und ergänzen Sie Ihren Header mit den dort gefundenen Informationen.
- Versuchen Sie das Paper von [Huber and Rust \[2016\]](#), das Sie [hier](#) finden, in Ihrem Dokument zu zitieren.
 - Klicken Sie dazu auf “Visual”,
 - gehen Sie an die Stelle im Text, an der Sie das Paper zitieren möchten, und wählen Sie “Insert” -> “Citation”.
 - Suchen Sie im Kontextmenü mithilfe der entsprechenden DOI (<https://doi.org/10.1177/1536867X160>) nach dem Paper und fügen Sie es ein.
- Um mit dem APA Version 7-Stil zu zitieren, schreiben Sie folgendes in den YAML-Header:

```
csl: "https://www.zotero.org/styles/apa"
```

- Wählen Sie einen anderen Zitierstil von www.zotero.org/styles. Rendern Sie dann das Dokument erneut und beobachten Sie die Unterschiede.

2.5. APA konformes Manuscript erstellen mit Quarto (apaquarto)

Um ein APA konformes Manuscript zu erstellen, empfiehlt es sich, die *Quarto Extension* `apaquarto` zu benutzen. Wie das geht wird [hier](#) genau beschrieben. Durch die Verwendung der Vorlage werden alle APA Regeln automatisch berücksichtigt. Da auch APA viel Spielraum lässt und jeder Gutachter Sonderwünsche hat, erlaubt es `apaquarto`, eine Vielzahl von Einstellungen. Beispielsweise kann in der Preamble (YAML header) [die Sprache geändert](#) werden oder der [allgemeine Stil des Dokumentes](#) verändert werden.

2.6. Vorlage zur Hausarbeit mit Quarto

Ich habe eine Vorlage erstellt, die Sie zur Erstellung ihrer Hausarbeit verwenden können. Sollte bei Ihnen etwas nicht funktionieren oder sie Hinweise zur Verbesserung haben, freue ich mich über eine Nachricht. Um die Vorlage zu verwenden, folgen Sie bitte den anweisungen auf meinem GitHub Account im Repository `temp_apa_de`:

https://github.com/hubchev/temp_apa_de

Teil II.

Anwendungen

3. Daten einlesen und aufbereiten

Dieses Dokument beschreibt exemplarisch die Datenaufbereitung der Datei ‘Dataset 71.txt’. Alle Schritte werden mit R durchgeführt. Der Code ist in die Quarto Dateien eingebettet. Die Ergebnisse sind vollständig replizierbar. Der verwendete Code kann wieder und anderweitig verwendet werden.

Tipp

Die PDF Datei kann hier heruntergeladen werden: https://github.com/hubchev/ewa/raw/main/ss_24/read_in_71/doc_read_in_71.pdf.

Um die komplette Arbeit zu replizieren und gegebenenfalls auf einen anderen Datensatz anzuwenden, kann das Repository “ewa” von meinem GitHub Account heruntergeladen werden. Alle entsprechenden Dateien befinden sich im Verzeichnis “ewa/ss_24/read_in_71”. Hier ist der Link zu dem entsprechenden Repository: <https://github.com/hubchev/ewa/>

Wie das alles im Detail von statten geht, wurde in der Übung behandelt.

4. Zwei-Wege-ANOVA-Modellen

Dies ist eine Übung, bei der Datenmanagement mit den `dplyr`-Funktionen `pivot_longer`, `rename` und `bind_rows` geübt wird. Außerdem zeige ich, wie eine ANOVA-Analyse mit R durchgeführt und in Quarto veranschaulicht werden kann. Dabei beziehe ich mich auf den Inhalt von [Childs et al. \[2021, Kapitel 27\]](#).

Unser Ziel ist es, zu lernen, wie man mit Zwei-Wege-ANOVA-Modellen in R arbeitet, anhand eines Beispiels aus einem Pflanzenwettbewerbsexperiment. Der Arbeitsablauf ist sehr ähnlich wie bei der Einweg-ANOVA in R. Wir beginnen mit dem Problem und den Daten und arbeiten dann durch die Modellanpassung, die Bewertung der Annahmen, den Signifikanztest und schließlich die Darstellung der Ergebnisse.

4.1. R-Sitzung einrichten

```
rm(list = ls())

if (!require(pacman)) install.packages("pacman")
pacman::p_load(tidyverse, rstatix, ggpubr, agricolae, tinytable, rempsyc,
               knitr, kableExtra, ggstatsplot, papaja, janitor, apa)
```

4.2. Daten einlesen

Pflanzen haben einen optimalen Boden-pH-Wert für ihr Wachstum, und dieser variiert zwischen den Arten. Folglich würden wir erwarten, dass wenn wir zwei Pflanzen im Wettbewerb zueinander unter verschiedenen pH-Werten anbauen, der Effekt des Wettbewerbs je nach Boden-pH-Wert unterschiedlich ausfallen könnte. In einer aktuellen Studie wurde das Wachstum des Grases *Festuca ovina* (Schaf-Schwingel) im Wettbewerb mit der Besenheide *Calluna vulgaris* (Heidekraut) in Böden mit unterschiedlichen pH-Werten untersucht. *Calluna* ist gut angepasst, auf sehr sauren Böden wie dem Millstone Grit und den Hochmoorflächen um Sheffield zu wachsen. *Festuca* wächst auf Böden mit einem viel breiteren pH-Bereich. Wir könnten die Hypothese aufstellen, dass *Calluna* in sehr sauren Böden ein besserer Konkurrent von *Festuca* sein wird als in mäßig sauren Böden. Hier sind die Daten: Die Spalten `pH 3.5` und `pH 5.5` enthalten das Gewicht, die Spalte `Condition` enthält die Anwesenheit oder Abwesenheit von *Calluna*.

Dies ist ein vollständig faktorielles Zwei-Wege-Design. Die Gesamtanzahl der unterschiedlichen Behandlungsgruppen beträgt $2 \times 2 = 4$. Für jede der Behandlungen gab es 5 Messwerte bzw. Pflanzen, was insgesamt $2 \times 2 \times 5 = 20$ Beobachtungen ergibt. Hier sind die vorliegenden Daten:

4. Zwei-Wege-ANOVA-Modellen

```
data_present <- data.frame(  
  Condition = rep(c("Calluna Present"), each = 5),  
  `ph_3_5` = c(2.76, 2.39, 3.54, 3.71, 2.49),  
  `ph_5_5` = c(3.21, 4.10, 3.04, 4.13, 5.21),  
  check.names = FALSE  
)  
data_present
```

	Condition	ph_3_5	ph_5_5
1	Calluna Present	2.76	3.21
2	Calluna Present	2.39	4.10
3	Calluna Present	3.54	3.04
4	Calluna Present	3.71	4.13
5	Calluna Present	2.49	5.21

```
data_absent <- data.frame(  
  Condition = rep(c("Calluna Absent"), each = 5),  
  `ph_3_5` = c(4.10, 2.72, 2.28, 4.43, 3.31),  
  `ph_5_5` = c(5.92, 7.31, 6.10, 5.25, 7.45),  
  check.names = FALSE  
)  
data_absent
```

	Condition	ph_3_5	ph_5_5
1	Calluna Absent	4.10	5.92
2	Calluna Absent	2.72	7.31
3	Calluna Absent	2.28	6.10
4	Calluna Absent	4.43	5.25
5	Calluna Absent	3.31	7.45

Um diese zwei Datensätze zu kombinieren, verwende ich die Funktion `bind_rows` (siehe [R Dokumentation](#)):

```
data <- bind_rows(data_present, data_absent)  
data
```

	Condition	ph_3_5	ph_5_5
1	Calluna Present	2.76	3.21
2	Calluna Present	2.39	4.10
3	Calluna Present	3.54	3.04
4	Calluna Present	3.71	4.13
5	Calluna Present	2.49	5.21
6	Calluna Absent	4.10	5.92
7	Calluna Absent	2.72	7.31
8	Calluna Absent	2.28	6.10
9	Calluna Absent	4.43	5.25
10	Calluna Absent	3.31	7.45

4. Zwei-Wege-ANOVA-Modellen

Um diesen Datensatz nun im sogenannten *Long-Format* darzustellen, verwende ich die Funktion `pivot_longer`. Dieses Format hat bei der Verwendung einiger Befehle vorteile. Wie zwischen dem *Long-Format* und den *Wide-Format* gewechselt werden kann, bitte ich [Wickham and Grolemund \[2023\]: 5.3 Lengthening data](#) zu entnehmen.

```
festuca <- data |>
  pivot_longer(cols = starts_with("pH"), names_to = "ph", values_to = "weight") |>
  rename(calluna = Condition) |>
  mutate(across(c(calluna, ph), as.factor))

festuca
```

```
# A tibble: 20 x 3
  calluna      ph    weight
  <fct>      <fct>    <dbl>
1 Calluna Present ph_3_5    2.76
2 Calluna Present ph_5_5    3.21
3 Calluna Present ph_3_5    2.39
4 Calluna Present ph_5_5    4.1
5 Calluna Present ph_3_5    3.54
6 Calluna Present ph_5_5    3.04
7 Calluna Present ph_3_5    3.71
8 Calluna Present ph_5_5    4.13
9 Calluna Present ph_3_5    2.49
10 Calluna Present ph_5_5    5.21
11 Calluna Absent ph_3_5    4.1
12 Calluna Absent ph_5_5    5.92
13 Calluna Absent ph_3_5    2.72
14 Calluna Absent ph_5_5    7.31
15 Calluna Absent ph_3_5    2.28
16 Calluna Absent ph_5_5    6.1
17 Calluna Absent ph_3_5    4.43
18 Calluna Absent ph_5_5    5.25
19 Calluna Absent ph_3_5    3.31
20 Calluna Absent ph_5_5    7.45
```

4.3. Deskriptive Statistik

Um Aussagen über die Beziehung des pH-Werts mit der Pflanzenart tätigen zu können, sollte zunächst ein deskriptiver Blick auf die Daten getätigt werden. Lassen Sie uns also auf den Mittelwert und die Standardabweichung der vier Gruppen blicken. Dies kann tabellarisch oder grafisch geschehen.

4.3.1. Tabellarisch

Dies geht flexibel mit den Funktionen `group_by` in Kombination mit `summarize`:

```
summary_stats <- festuca |>
  group_by(calluna, ph) |>
  summarize(
    mean = mean(weight),
    sd = sd(weight)
  ) |>
  ungroup()

summary_stats
```

```
# A tibble: 4 x 4
  calluna      ph    mean    sd
  <fct>      <fct> <dbl> <dbl>
1 Calluna Absent ph_3_5  3.37 0.904
2 Calluna Absent ph_5_5  6.41 0.945
3 Calluna Present ph_3_5  2.98 0.609
4 Calluna Present ph_5_5  3.94 0.869
```

Wenn wir schließlich eine publikationswürdige Tabelle haben wollen, geht das wie folgt:

```
```{r , echo=FALSE, warning=FALSE, message=FALSE}
#| label: tbl-desc_calluna
#| tbl-cap: Deskriptive Statistiken
#| tbl.align: left

summary_stats |>
 kable()
```
```

Das Ergebnis, ist in Tabelle 4.1 zu sehen.

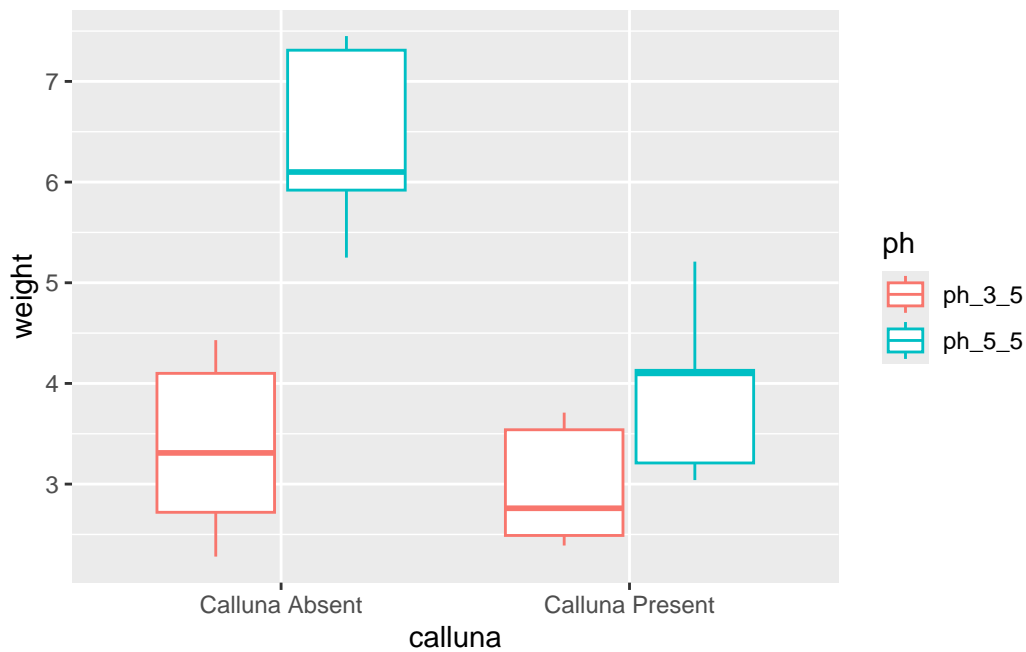
| calluna | ph | mean | sd |
|-----------------|--------|-------|-----------|
| Calluna Absent | ph_3_5 | 3.368 | 0.9042511 |
| Calluna Absent | ph_5_5 | 6.406 | 0.9451614 |
| Calluna Present | ph_3_5 | 2.978 | 0.6089089 |
| Calluna Present | ph_5_5 | 3.938 | 0.8685448 |

4.3.2. Grafisch

Boxplots bieten einen guten Einblick in die Häufigkeitsverteilung, ohne die Grafik zu überfrachten. Bei wenigen Beobachtungen, wie in unserem Fall, können sie aber problematisch sein da die Datengrundlage (5 Beobachtungen pro Boxplot) nicht ersichtlich ist, siehe Abbildung 4.1.

```
ggplot(data = festuca, aes(x = calluna, y = weight, colour = ph)) +
  geom_boxplot()
```

Abbildung 4.1.: Boxplots



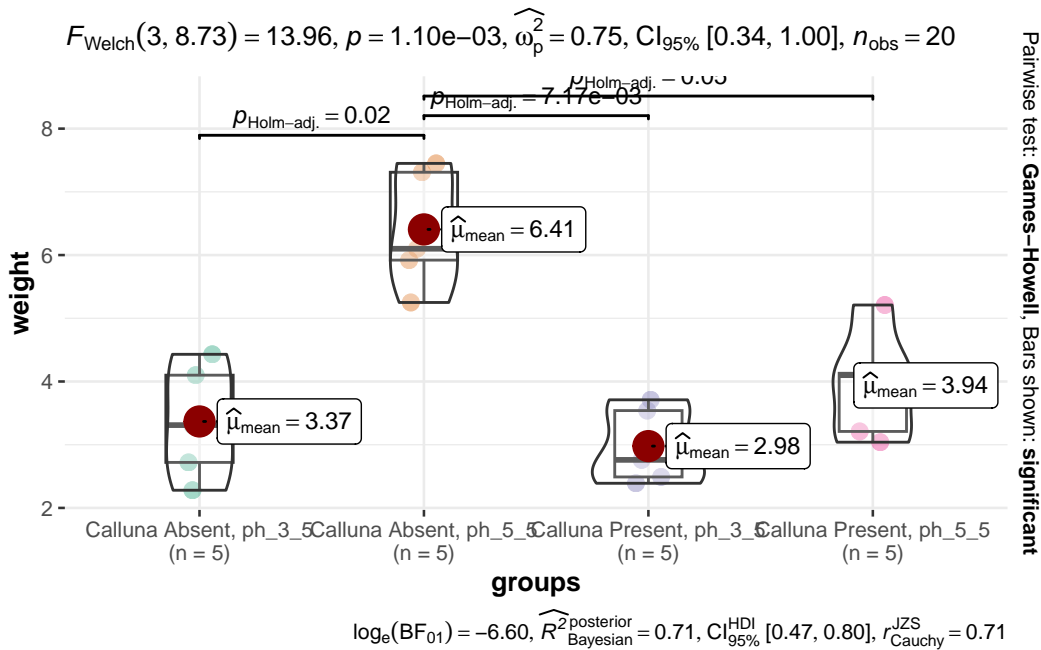
Mit der Funktion `ggbetweenstats` aus dem Paket `ggstatsplot` können die einzelnen Beobachtungen und die statistischen Test zu den Mittelwertvergleichen angezeigt werden, siehe [Abbildung 4.2](#).

```
festuca_group <- festuca |>
  mutate(groups = paste(calluna, ph, sep = ", "))

plt <- ggbetweenstats(
  data = festuca_group,
  x = groups,
  y = weight
)

plt
```

Abbildung 4.2.: Boxplots mit ggbetweenstats



4.4. t-Test

4.4.1. One Sample t-test

```
t.test(festuca$weight, mu = 4)
```

One Sample t-test

```
data: festuca$weight
t = 0.49085, df = 19, p-value = 0.6292
alternative hypothesis: true mean is not equal to 4
95 percent confidence interval:
 3.436939 4.908061
sample estimates:
mean of x
 4.1725
```

```
t.test(festuca$weight, mu = 1)
```

One Sample t-test

```
data: festuca$weight
t = 9.0273, df = 19, p-value = 2.664e-08
alternative hypothesis: true mean is not equal to 1
```

95 percent confidence interval:

3.436939 4.908061

sample estimates:

mean of x

4.1725

4.4.2. Two sided t-test

```
t.test(data$ph_3_5, data$ph_5_5, paired = FALSE)
```

Welch Two Sample t-test

data: data\$ph_3_5 and data\$ph_5_5

t = -3.6529, df = 13.013, p-value = 0.002917

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-3.1811156 -0.8168844

sample estimates:

mean of x mean of y

3.173 5.172

```
t.test(data_absent$ph_3_5, data_absent$ph_5_5, paired = FALSE)
```

Welch Two Sample t-test

data: data_absent\$ph_3_5 and data_absent\$ph_5_5

t = -5.1934, df = 7.9844, p-value = 0.0008343

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-4.387422 -1.688578

sample estimates:

mean of x mean of y

3.368 6.406

```
t.test(data_present$ph_3_5, data_present$ph_5_5, paired = FALSE)
```

Welch Two Sample t-test

data: data_present\$ph_3_5 and data_present\$ph_5_5

t = -2.0237, df = 7.1669, p-value = 0.08173

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-2.0764347 0.1564347

sample estimates:

mean of x mean of y

2.978 3.938

4. Zwei-Wege-ANOVA-Modellen

```
t.test(data_absent$ph_3_5, data_present$ph_5_5, paired = FALSE)
```

Welch Two Sample t-test

```
data: data_absent$ph_3_5 and data_present$ph_5_5
t = -1.0165, df = 7.9871, p-value = 0.3392
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.8633899  0.7233899
sample estimates:
mean of x mean of y
  3.368    3.938
```

```
t.test(data_present$ph_3_5, data_absent$ph_5_5, paired = FALSE)
```

Welch Two Sample t-test

```
data: data_present$ph_3_5 and data_absent$ph_5_5
t = -6.8177, df = 6.8324, p-value = 0.0002776
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -4.622899 -2.233101
sample estimates:
mean of x mean of y
  2.978    6.406
```

```
t.test(weight ~ ph, data = festuca)
```

Welch Two Sample t-test

```
data: weight by ph
t = -3.6529, df = 13.013, p-value = 0.002917
alternative hypothesis: true difference in means between group ph_3_5 and group ph_5_5 is not equal to 0
95 percent confidence interval:
 -3.1811156 -0.8168844
sample estimates:
mean in group ph_3_5 mean in group ph_5_5
      3.173           5.172
```

```
t.test(weight ~ calluna, data = festuca)
```

Welch Two Sample t-test

```
data: weight by calluna
```

4. Zwei-Wege-ANOVA-Modellen

```
t = 2.2371, df = 12.893, p-value = 0.04359
alternative hypothesis: true difference in means between group Calluna Absent and group Calluna Present
95 percent confidence interval:
 0.04785705 2.81014295
sample estimates:
mean in group Calluna Absent mean in group Calluna Present
          4.887              3.458
```

```
results <- festuca |>
  group_by(calluna, ph) |>
  summarise(t_test = list(t.test(weight)$p.value)) |>
  ungroup()
results
```

```
# A tibble: 4 x 3
  calluna      ph      t_test
  <fct>      <fct>   <list>
1 Calluna Absent ph_3_5 <dbl [1]>
2 Calluna Absent ph_5_5 <dbl [1]>
3 Calluna Present ph_3_5 <dbl [1]>
4 Calluna Present ph_5_5 <dbl [1]>
```

4.5. ANOVA

Verwenden Sie dieses Modell, um die ANOVA zu berechnen: `weight ~ ph + calluna + ph:calluna`

```
festuca_model <- aov(weight ~ ph + calluna + ph:calluna, data = festuca)
anova(festuca_model)
```

Analysis of Variance Table

Response: weight

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|------------|----|---------|---------|---------|---------------|
| ph | 1 | 19.9800 | 19.9800 | 28.1792 | 7.065e-05 *** |
| calluna | 1 | 10.2102 | 10.2102 | 14.4001 | 0.00159 ** |
| ph:calluna | 1 | 5.3976 | 5.3976 | 7.6126 | 0.01397 * |
| Residuals | 16 | 11.3446 | 0.7090 | | |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Diese Ergebnis können publikationswürdig mit den Funktionen `apa_print` aus dem Paket `papaja` und `kable` dargestellt werden, siehe Tabelle 4.2:

```
`{r, echo=FALSE, eval=TRUE, message=FALSE, warning=FALSE}
#| label: tbl-festuca_model
#| tbl-cap: "ANOVA Ergebnisse"
```


4. Zwei-Wege-ANOVA-Modellen

Tabelle 4.2.: ANOVA Ergebnisse

| | Effect | |
|---|-------------|---|
| 1 | (Intercept) | $F(1, 16) = 491.08, p < .001, \eta^2 = .97$ *** |
| 2 | ph | $F(1, 16) = 28.18, p < .001, \eta^2 = .64$ *** |
| 3 | calluna | $F(1, 16) = 14.40, p = .002, \eta^2 = .47$ ** |
| 4 | ph:calluna | $F(1, 16) = 7.61, p = .014, \eta^2 = .32$ * |

```
apa_anova <- apa_print(festuca_model)
knitr::kable(apa_anova$table, booktabs=T)
````
```

### 4.6. Diagnostics

Lesen Sie [Childs et al. \[2021\]: 27.5 Diagnostics](#). Außerdem ist [diese Seite](#) einen Blick wert. Dort finden Sie einige für die ANOVA-Diagnostik hilfreiche R-Funktionen.

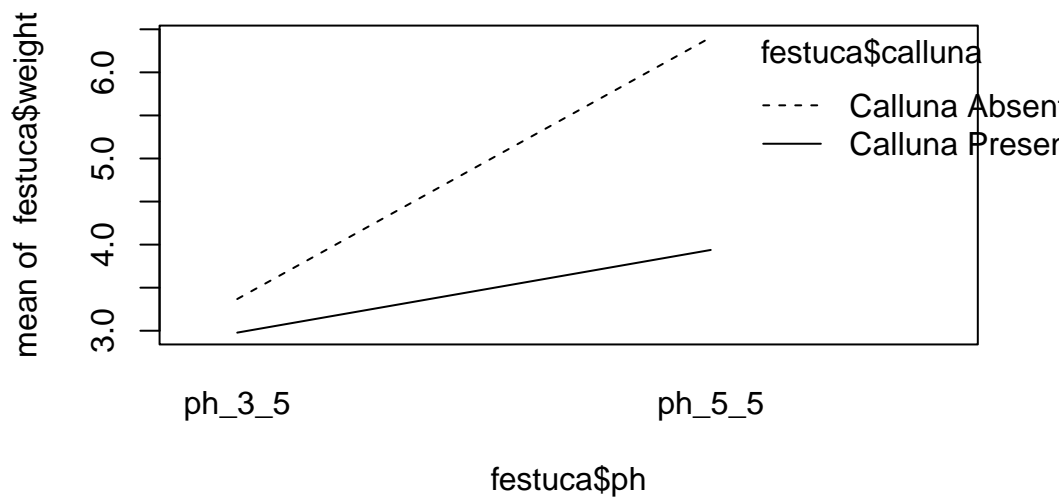
### 4.7. Interaktions Diagramm

Ein Interaktionsdiagramm illustriert, wie zwei oder mehr unabhängige Variablen gemeinsam die abhängige Variable beeinflussen. Es hilft dabei, Wechselwirkungen zwischen Faktoren visuell darzustellen und besser zu verstehen, ob der Effekt einer unabhängigen Variablen von der Ausprägung einer anderen abhängt. Dies ist besonders wichtig, um mögliche Interaktionen identifizieren und interpretieren zu können, die in einer ANOVA-Analyse auftreten.

So ein Diagramm kann mit der Funktion `interaction.plot` erstellt werden:

```
interaction.plot(festuca$ph, festuca$calluna, response = festuca$weight)
```

#### 4. Zwei-Wege-ANOVA-Modellen

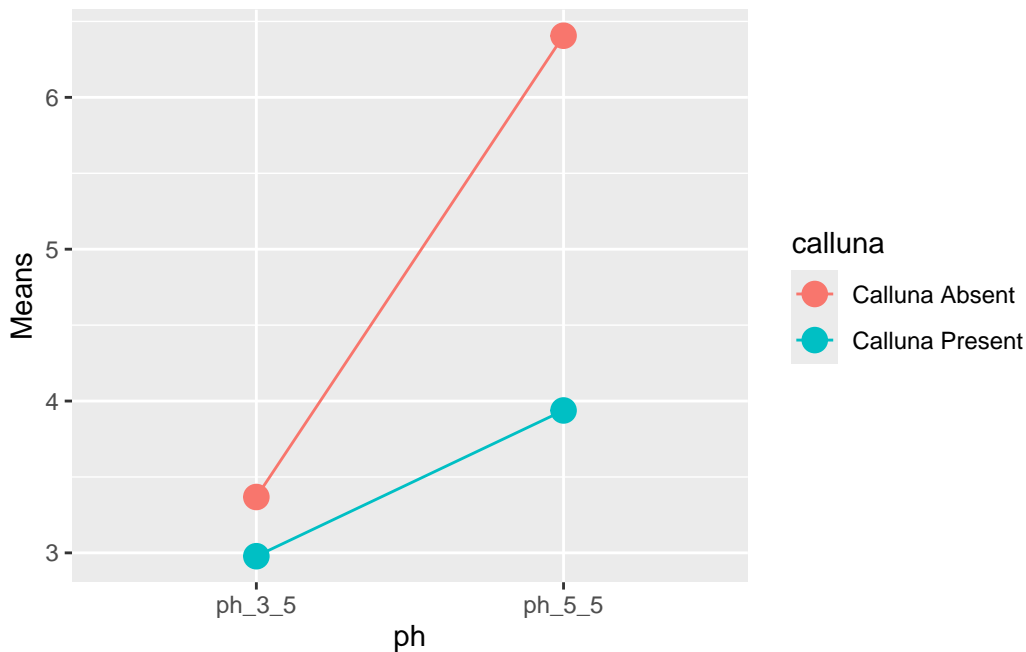


Hier ist eine viel ansprechendere und flexiblere Methode, um Interaktionsdiagramme mithilfe der tidyverse-Funktionen zu erstellen:

```
step 1. calculate means for each treatment combination
festuca_means <-
 festuca |>
 group_by(calluna, ph) |> # <- remember to group by *both* factors
 summarise(Means = mean(weight))
```

```
step 2. plot these as an interaction plot
ggplot(festuca_means,
 aes(x = ph, y = Means, colour = calluna, group = calluna)) +
 geom_point(size = 4) + geom_line()
```

#### 4. Zwei-Wege-ANOVA-Modellen



Bitte lesen Sie [Childs et al. \[2021\]: 27.6.1](#) und berücksichtigen Sie Abbildung 4.3.

### 4.8. Multiple-Vergleichs-Test

Ein Multiple-Vergleichs-Test, wie der TukeyHSD-Test, wird verwendet, um nach einer ANOVA-Analyse die Unterschiede zwischen den Gruppenpaaren genauer zu untersuchen. Er hilft dabei, festzustellen, welche spezifischen Gruppen sich signifikant voneinander unterscheiden, indem er alle möglichen Paarvergleiche berücksichtigt. Dies ist besonders nützlich, um nach signifikanten Ergebnissen aus der ANOVA detailliertere Erkenntnisse zu gewinnen.

```
TukeyHSD(festuca_model, which = 'ph:calluna')
```

```
Tukey multiple comparisons of means
95% family-wise confidence level
```

```
Fit: aov(formula = weight ~ ph + calluna + ph:calluna, data = festuca)
```

```
$`ph:calluna`
```

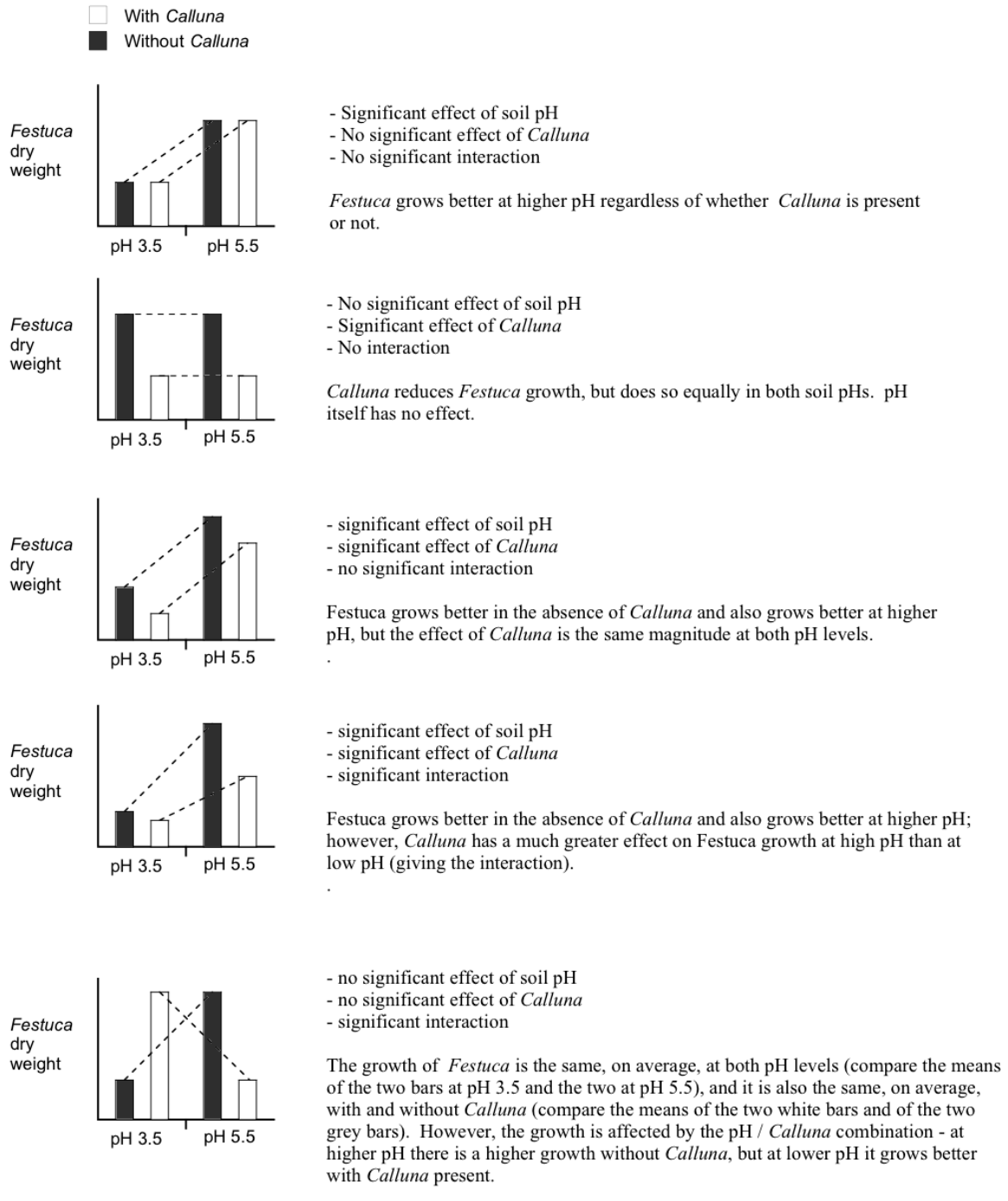
|                                               | diff   | lwr        | upr        |
|-----------------------------------------------|--------|------------|------------|
| ph_5_5:Calluna Absent-ph_3_5:Calluna Absent   | 3.038  | 1.5143518  | 4.5616482  |
| ph_3_5:Calluna Present-ph_3_5:Calluna Absent  | -0.390 | -1.9136482 | 1.1336482  |
| ph_5_5:Calluna Present-ph_3_5:Calluna Absent  | 0.570  | -0.9536482 | 2.0936482  |
| ph_3_5:Calluna Present-ph_5_5:Calluna Absent  | -3.428 | -4.9516482 | -1.9043518 |
| ph_5_5:Calluna Present-ph_5_5:Calluna Absent  | -2.468 | -3.9916482 | -0.9443518 |
| ph_5_5:Calluna Present-ph_3_5:Calluna Present | 0.960  | -0.5636482 | 2.4836482  |

|                                              | p adj     |
|----------------------------------------------|-----------|
| ph_5_5:Calluna Absent-ph_3_5:Calluna Absent  | 0.0001731 |
| ph_3_5:Calluna Present-ph_3_5:Calluna Absent | 0.8826936 |
| ph_5_5:Calluna Present-ph_3_5:Calluna Absent | 0.7117913 |
| ph_3_5:Calluna Present-ph_5_5:Calluna Absent | 0.0000443 |

#### 4. Zwei-Wege-ANOVA-Modellen

Abbildung 4.3.: Grafische Veranschaulichung des Modells



#### 4. Zwei-Wege-ANOVA-Modellen

```
ph_5_5:Calluna Present-ph_5_5:Calluna Absent 0.0014155
ph_5_5:Calluna Present-ph_3_5:Calluna Present 0.3079685
```

```
HSD.test(festuca_model, trt = c("ph", "calluna"), console = TRUE)
```

```
Study: festuca_model ~ c("ph", "calluna")
```

```
HSD Test for weight
```

```
Mean Square Error: 0.709035
```

```
ph:calluna, means
```

|                        | weight | std r     | se          | Min  | Max  | Q25  | Q50  | Q75  |
|------------------------|--------|-----------|-------------|------|------|------|------|------|
| ph_3_5:Calluna Absent  | 3.368  | 0.9042511 | 5 0.3765727 | 2.28 | 4.43 | 2.72 | 3.31 | 4.10 |
| ph_3_5:Calluna Present | 2.978  | 0.6089089 | 5 0.3765727 | 2.39 | 3.71 | 2.49 | 2.76 | 3.54 |
| ph_5_5:Calluna Absent  | 6.406  | 0.9451614 | 5 0.3765727 | 5.25 | 7.45 | 5.92 | 6.10 | 7.31 |
| ph_5_5:Calluna Present | 3.938  | 0.8685448 | 5 0.3765727 | 3.04 | 5.21 | 3.21 | 4.10 | 4.13 |

```
Alpha: 0.05 ; DF Error: 16
```

```
Critical Value of Studentized Range: 4.046093
```

```
Minimum Significant Difference: 1.523648
```

```
Treatments with the same letter are not significantly different.
```

|                        | weight | groups |
|------------------------|--------|--------|
| ph_5_5:Calluna Absent  | 6.406  | a      |
| ph_5_5:Calluna Present | 3.938  | b      |
| ph_3_5:Calluna Absent  | 3.368  | b      |
| ph_3_5:Calluna Present | 2.978  | b      |

### 4.9. Schlussfolgerungen ziehen und Ergebnisse präsentieren

Hier sind einige Code-Beispiele, wie die oben gezeigten Diagramme viel schöner gestaltet werden könnten.

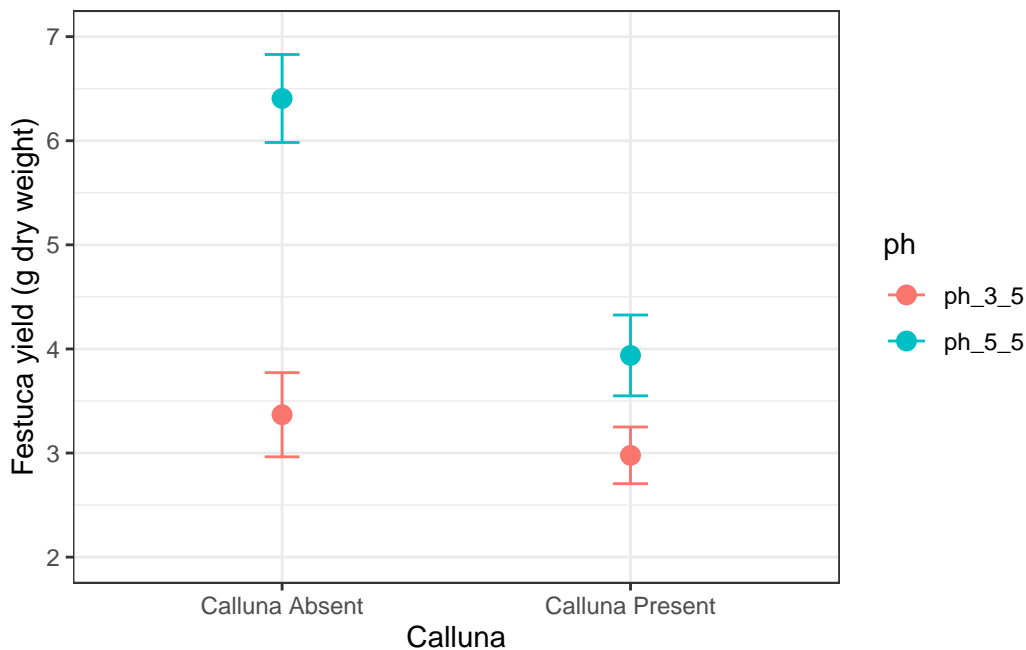
```
step 1. calculate means for each treatment combination
festuca_stats <-
 festuca |>
 group_by(calluna, ph) %>% # <- remember to group by the two factors
 summarise(means = mean(weight), SEs = sd(weight)/sqrt(n()))
```

```
`summarise()` has grouped output by 'calluna'. You can override using the
`.groups` argument.
```

#### 4. Zwei-Wege-ANOVA-Modellen

```
step 1. calculate means for each treatment combination
festuca_stats <-
 festuca |>
 group_by(calluna, ph) %>% # <- remember to group by the two factors
 summarise(means = mean(weight), ses = sd(weight)/sqrt(n()))
```

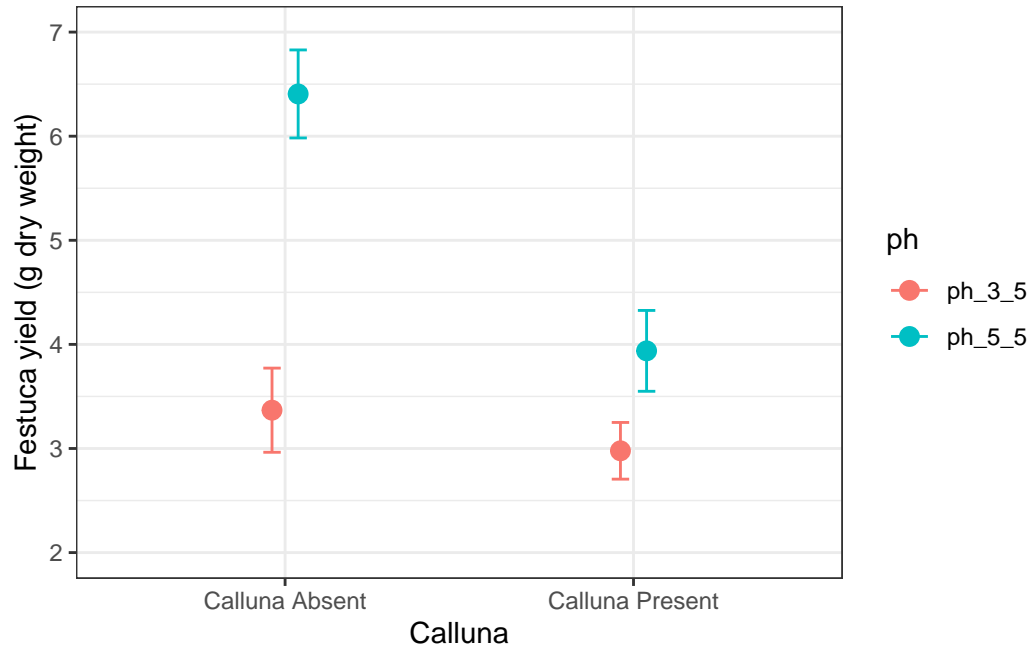
```
step 2. plot these as an interaction plot
ggplot(festuca_stats,
 aes(x = calluna, y = means, colour = ph,
 ymin = means - ses, ymax = means + ses)) +
 # this adds the mean
 geom_point(size = 3) +
 # this adds the error bars
 geom_errorbar(width = 0.1) +
 # controlling the appearance
 scale_y_continuous(limits = c(2, 7)) +
 xlab("Calluna") + ylab("Festuca yield (g dry weight)") +
 # use a more professional theme
 theme_bw()
```



```
define a position adjustment
pos <- position_dodge(0.15)
make the plot
ggplot(festuca_stats,
 aes(x = calluna, y = means, colour = ph,
 ymin = means - ses, ymax = means + ses)) +
 # this adds the mean (shift positions with 'position =')
 geom_point(size = 3, position = pos) +
 # this adds the error bars (shift positions with 'position =')
 geom_errorbar(width = 0.1, position = pos) +
 # controlling the appearance
```

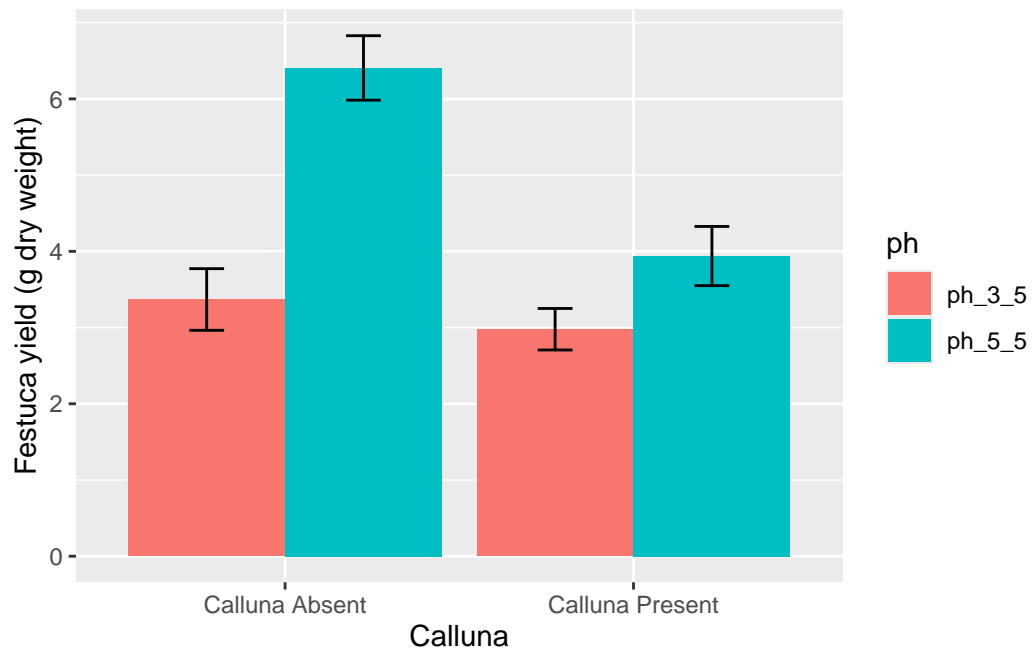
#### 4. Zwei-Wege-ANOVA-Modellen

```
scale_y_continuous(limits = c(2, 7)) +
xlab("Calluna") + ylab("Festuca yield (g dry weight)") +
use a more professional theme
theme_bw()
```



```
ggplot(festuca_stats,
 aes(x = calluna, y = means, fill = ph,
 ymin = means - ses, ymax = means + ses)) +
this adds the mean
geom_col(position = position_dodge()) +
this adds the error bars
geom_errorbar(position = position_dodge(0.9), width=.2) +
controlling the appearance
xlab("Calluna") + ylab("Festuca yield (g dry weight)")
```

#### 4. Zwei-Wege-ANOVA-Modellen





## 5. ANOVA Ergebnisse und Quarto

In den folgenden zwei Abschnitten präsentiere ich zwei Dokumente. Beide Dokumente zeigen exemplarisch auf, wie ANOVA Analysen mit R durchgeführt und mit Hilfe von Quarto veranschaulicht werden können.

Um die dargestellten Ergebniss zu replizieren und den Code gegebenenfalls auf einen anderen Datensatz anzuwenden, kann das Repository “ewa” von meinem GitHub Account heruntergeladen werden. Alle entsprechenden Dateien befinden sich im entsprechenden Unterverzeichnis “ewa/ss\_24”. Hier ist der Link zu dem entsprechenden Repository: <https://github.com/hubchev/ewa/>

### Tipp

Die PDF Datei kann hier heruntergeladen werden: [https://github.com/hubchev/ewa/raw/main/ss\\_24/desc\\_aov/desc\\_aov.pdf](https://github.com/hubchev/ewa/raw/main/ss_24/desc_aov/desc_aov.pdf).

Die dazu gehörende Quarto Datei sowie alle sonstigen Dateien, sind auf meinem GitHub Account zu finden [Huber, 2024a]: <https://github.com/hubchev/ewa/>

Wie das alles im Detail von statten geht, wurde in der Übung behandelt.

# Literatur

- Paul C. Bauer and Camille Landesvatter. Writing a reproducible paper with rstudio and quarto. Technical report, 2023. URL <https://doi.org/10.31219/osf.io/ur4xn>.
- Dylan Z. Childs, Bethan J. Hindle, and Philip H. Warren. Aps 240: Data analysis and statistics with r. online, 2021. URL <https://dzchilds.github.io/stats-for-bio>.
- Christopher Gandrud. *Reproducible research with R and R studio*. Chapman and Hall/CRC, 3 edition, 2020.
- Stephan Huber. Empirisch-wissenschaftlich arbeiten (ewa). GitHub repository, 2024a. URL <https://github.com/hubchev/ewa>.
- Stephan Huber. How to use R for data science, 2024b. URL <https://hubchev.github.io/ds/>.
- Stephan Huber and Christoph Rust. Calculate travel time and distance with openstreetmap data using the open source routing machine (osrm). *The Stata Journal*, 16(2):416–423, 2016.
- Richard J Telford. Enough markdown to write a thesis, 9 2023. URL <https://biostats-r.github.io/biostats/quarto/>.
- Hadley Wickham and Garrett Grolemund. R for data science (2e), 2023. URL <https://r4ds.hadley.nz/>.