

# Guidelines for coursework 2

Methods of Scientific Working (for Crop Sciences) (3502-440)

21 Jan 2025

## Table of contents

1	General information	1
2	The dataset	1
3	Objectives of the data science project	1
4	Structure of the report	2
5	Grading criteria	2
5.1	Submission of the coursework 2 paper . . . . .	3

## 1 General information

The goal of coursework 2 is to conduct a small data analysis projects that includes a complete workflow from obtaining raw data, cleaning and importing them, making a simple statistical analysis, creating plots and writing a short report. By doing this work you will practice the key principles and basic aspects of data science taught in class.

## 2 The dataset

The dataset to be analyses is described in the following paper: Evan B. Craine, et al. (2023) A comprehensive characterization of agronomic and end-use quality phenotypes across a quinoa world core collection. *Frontiers in Plant Science*. <https://doi.org/10.3389/fpls.2023.1101547>. The data to be analyses are measurements of seed traits of 360 quinoa accessions cultivated in a field trial in the USA. You find the data in the Supplementary Material in the ZIP file Presentation 1.zip with the file name SM Data.csv.

Each student will be assigned the data for one trait. You find the assignment on the course website next to the link for this file.

## 3 Objectives of the data science project

You are expected to carry out the following steps and implement them in a fully reproducible document in the RMarkdown Quarto format. Your analysis involves the following steps:

1. Import data into an R data frame using the readr package of tidyverse.
2. Check the proportion of missing data and report them.
3. Plot a histogram with the distribution of your data set.

4. Carry out a statistical test to check whether the data are normally distributed.
5. Plot the correlation of your data set with yield (trait `Yield_g`) and thousand seed weight (trait `TSW`) and conduct a regression analysis between your trait and the two yield traits. Report the regression equations.

## 4 Structure of the report

The analysis report should consist of the following sections:

- A header with your immatriculation number (do not include your name to avoid any bias during grading) and the dataset number.
- A short description of the data (number of individuals, type of data, units)
- A short description of the analysis, i.e., what the analysis is about and what you did (including choice of parameters etc)
- The code that you used in proper code chunks. The code chunks should also be displayed in proper context in the PDF (or HTML).
- A brief description of the results which summarises which information can be retrieved from the results and how it can be interpreted.
- Figures and/or tables that show the results, properly referenced in the text sections and with proper caption and correct annotation (e.g., have axis labels). Avoid redundancy.
- Each analysis should be fully repeatable with the given information.

A discussion section, in which you summarise all your results and discuss them in context with each other. It should put the results in context (i.e., what could be a reason that the data are not normally distributed?). Also check whether the correlations you observe correspond to the trait correlations reported in the publication. Justify any of your conclusions with your analysis results.

The console output should not be printed to the PDF. The report should be concise! Three pages in the PDF format is sufficient.

## 5 Grading criteria

The following criteria are used for the evaluation of the coursework:

1. You submit an `.qmd` and `.pdf` formatted file. To produce the pdf file you must install LaTeX on your machine with the command `quarto install tinytex` in the terminal of the RStudio program.
2. The `.qmd` must be fully executable without any error message
3. The analyses must be done with the tidyverse R package system (e.g., using the `ggplot` library for plotting).
4. Only the output that is relevant for the report is printed (use the corresponding options of the code chunks). Avoid any cluttering output.
5. Overall layout and structure of report according to template
6. Figures and tables are appropriate and meaningful
7. Length and structure introduction, data cleaning, analysis and discussion sections
8. Typing errors and grammar
9. Quality of writing and appropriate use of jargon terms
10. Correct and complete references
11. Connections between data cleaning and data analysis is correct
12. Results of analysis are discussed in a sensible and meaningful manner

The paper is evaluated according to the formal criteria described above and then ranked among all term papers with respect to the criteria. **All term papers will be checked for plagiarism!** We usually write down some comments on the term paper and make them available to you after grading to give you some feedback.

### 5.1 Submission of the coursework 2 paper

Please upload the coursework 2 paper to the Coursework-2 folder on ILIAS as a single pdf.

Name the .qmd file as 123456\_coursework2.qmd where 123456 is your immatriculation number.

Name the accompanying .pdf file accordingly. Do not use your name in the report to avoid any conscious or unconscious bias in the grading process.

Please remember to submit as a separate PDF the signed declaration of plagiarism, which you can download from here: [Link](#) It is fine if you print it out, sign it and take a scan with your smartphone or tablet.

**The due date for submission is on the [main page](#) of the course.**