Lab 6 - Factorial Designs

ENVX2001 Applied Statistical Methods

Semester 1, 2026

Learning outcomes

At the end of this practical students should be able to:

- use R to analyse experiments with a factorial treatment structure where the experimental design is a CRD or RCBD;

All of the data for this practical is in the **Data6.xlsx** file.

Exercise 1 - More Treatment Structures (Walk-through)

A microbiologist conducted an experiment to assess the survival of *Salmonella typhinerium* when subjected to various treatments. A factorial treatment design was used, the treatments being various combinations of sorbic acid, pH, and water activity. The density of Salmonella seven days after the treatment began was recorded. The Salmonella data is in the form:

$$log_e(density/ml)$$

, to 2 decimal places.

The data is found in the **Salmonella** sheet in the **Data6.xlsx** file.

There is a slight **trick** to analysing this experiment which will improve your ability to analyse factorial experiments in general.

The table below shows you how to specify the treatment structure and individual interaction term in a factorial design involving 2 factors, `A` and `B`.

term	command
2-way factorial	A*B
2-way interaction	A:B

Question 1.1

(i) Import the data into R and describe the data using numerical and graphical summaries. In particular does the exploratory data analysis (EDA) show any difference in *S.typhinerium* density between the levels of each of the treatment factors (a) sorbic acid (b) pH (c) water activity? Is there evidence of an interaction using interaction plots. No need to look at marginal means as the interaction plots summarise these.

Tip

Some hints for the EDA are:

- in many experiments we will have treatment factors which have a numerical value (as is the case here) but for ANOVA-type problems we are comparing the mean response between each value so we need to tell R the data type is a `factor`. If we don't, it will fit a regression model which is Topics 7-9. This will result in the wrong type of analysis.
- use the `tapply` function in conjunction with the `summary` function;
- boxplots for individual levels of a factor are likely to be the most informative as compared to numerical summaries they visually show the spread of the observations as well. Example R code is below.

```
CODE
boxplot(Density~pH,data=salm)
```

- to create interaction plots use the `interaction.plot` or `emmip` function. You will need one of these for each 2-way interaction in your model.

```
CODE
#
```

Question 1.2

(ii) Write out the model you are fitting in terms of main effects and interactions, words are fine. Note, that each of these will have an associated statistical hypotheses. Obtain an ANOVA for these data, including appropriate interaction terms in the model.

What do you notice in the ANOVA table when your run the full 3-way factorial ANOVA? Use the `summary` function to extract the ANOVA table.

Run the model again but only include the interactions involving two factors: **do not include the three factor interaction term**. Why are we not including the three factor interaction?

Tip

Look at the data in Excel and **manually** calculate the mean `Density` for when `pH` = 5.0, `Activity` = 0.78 and `Sorbic.Acid` = 0. How many observations were used to calculate the mean?

The model in words is:

Question 1.3

(iii) Test that the model assumptions have been met.

```
CODE
#
```

Question 1.4

(iv) What are the significant effects in the model? For the significant effects we conduct post-hoc tests the Tukey's Test to determine which pairs are significantly different. Write overall conclusions, in terms of the statistical hypothesis testing and in terms of the biological description of the experiment.

```
CODE
#
```

Exercise 2 - More practice

An experiment was performed to study the control of potato blight on potatoes. A factorial treatment structure was employed with 3 Varieties in combination with 5 Chemical treatments. The experiment was conducted using a randomised complete block design (RCBD) with three blocks. The response is yield of potatoes (lbs).

Provide graphical summaries of the data, analyse the experiment, conduct any post-hoc tests if the results are significant. The data is found in the **Potato** sheet in the **Data6.xlsx** file.

```
CODE
#
```