Statistics and R short course

Session 4 - Practical

Vester Gunsaru

28 January 2026

Session 4 - Practical

Exercise 1

Take the iris dataset that we worked with during Sessions 1 and 2.

Explain how you would, in a formal statistical way, compare the following:

- `Petal.Width` between the flower species `virginica` and `setosa`.
- `Sepal.Length` between all 3 flower species.

For each comparison, state which test you will use (there may be more than one valid option!), state the null and alternative hypotheses, do the test and interpret the results.

Exercise 2

In a drug trial, researchers are assessing overall in-hospital mortality as the primary outcome. The new drug is compared against the standard-of-care treatment (SOC). Patients are randomised 1:1 to the new drug and SOC. At trial conclusion, the researchers observe that out of 250 SOC patients, 61 have died and out of 250 patients on the new drug arm, 48 have died.

Perform a statistical test to conclude whether or not there is a difference between the new drug and the SOC. State the test you use, the null and alternative hypotheses, perform the test and interpret the results.

Exercise 3

Test whether the 2 variables from Table 1 below are independent or not. State the test you use, the null and alternative hypotheses, do the test and interpret the results.

Table 1: Summary of patient outcomes for different health centers.

	alive	dead
Hospital1	92	29
Hospital2	54	15
Hospital3	31	3

What about when you repeat your analysis for Table 2 below?

Table 2: Summary of patient outcomes for different health centers.

	alive	dead

Hospital1	920	290
Hospital2	540	150
Hospital3	310	30

Comment on the results from your analyses for both tables.

Exercise 4

Researchers want to estimate the effect of malnutrition in early childhood on body height at adult age. For this, the researchers recruit former participants of a childhood malnutrition cohort study. Recruited participants are known to have been malnourished or not at any point in the previous cohort study.

Assume a standard deviation of 6cm for height, power of 90%, significance level of 5% and equal group sizes (malnourished and non-malnourished groups). How many former cohort study participants will need to be recruited, if the researchers want to be powered to detect an average difference in height of 2cm or more?