

## Problem Set 2.5: Categorical variables

ADS2

Semester 2, 2023/24

---

**1.** A candy company produces a new type ice cream with 5 flavours: “Mint”, “Vanilla”, “Chocolate”, “Lemon” and “Orange”. The company sent out samples to 200 people and let them choose their favourite flavour. Here is the result:

Mint	Vanilla	Chocolate	Lemon	Orange
40	32	48	57	23

Based on the result, do you think the company should produce the same amount of ice cream with different flavours?

**1.1** What is your null hypothesis. Use a simulation to generate the curve of chi-square values in this situation and get the p-value. Compare the result with `chisq.test()` result.

**1.2** What is the degree of freedom (df) in the test? Use `rchisq()` to get the distribution of chi-square with the specific df. Does this curve match the curve in 1.1?

**2.** Analyze the 3-way mouse data from the lecture.

	WT	KO
	Male	Female
Alive	40	34
Dead	9	7

**2.1** Input the data into an array in R. Print it out.

**2.2** Apply the chi-square test. Note: You cannot use an 3-dimensional object in the `chisq.test()`. Convert the array to a table object and find the chi-square test result in the summary of the object. Your result should match the one in the lecture.

**2.3** Use `ggplot2` to visualise the data in a dot plot.

**Additional exercise.** Change the question in the mouse experiment. Now we want to know if the survival of mice is dependent on `geneX` and `sex` (without assuming `geneX` and `sex` are independent on each other). What is the null hypothesis now? How do you reshape the data and perform the test?

---

Previous versions by Chaochen Wang and Hugo Samano.

Last update by DJ MacGregor in 2024