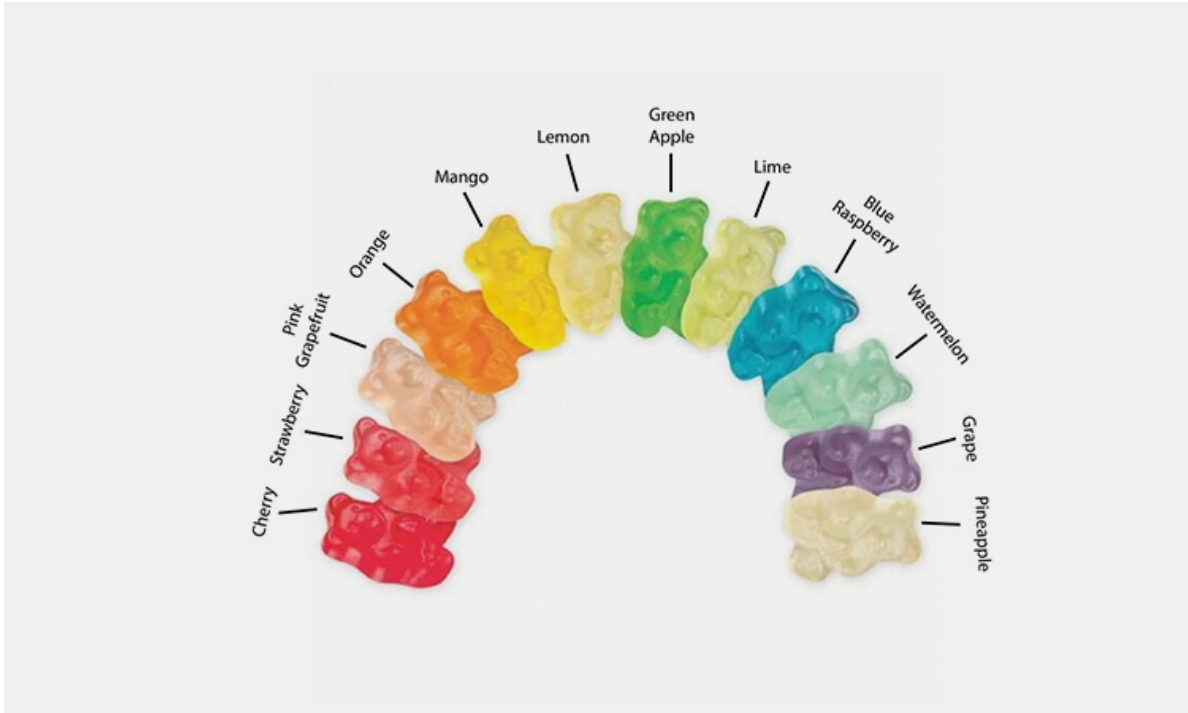# Lab 4

Part I. Fisher's Exact (Gummy) Test

For this question, we will be putting your taste buds to a test. This question will require working with a partner and eating Albanese 12 flavor gummy bears:

Each select a bag that contains 11 gummy bears. There will be 6 of one flavor, which is labeled on the bag, and 5 other gummy bears. The goal is to see how well you can identify (artificial) fruit flavors.

1. You have 6 of same flavor, use the first one as a taster.

2. Mix the remainder of the bag up. Cover your eyes and grab a gummy bear. Your goal is to determine whether this is the flavor you initially tasted. Your partner's task will

be to record the flavor you've tried along with the guess. For this experiment we will consider the response to be binary (specific flavor or not).

3. Continue this process until you have tasted and guessed for all 10 gummy bears. Your partner should not reveal your results until all 10 gummy bears have been tasted.

4. It is not strictly necessary that you make 5 guesses for each flavor, but the math will be much simpler if you do so.

1. What is the hypergeometric distribution and how is it useful in this setting? (4 points)

2. Construct at 2-by-2 contingency table to display your results from the study. (4 points)

3. If you are an excellent gummy bear flavor taster, what would you expect your results for Q2 to look like? (2 points)

4. If you are a terrible gummy bear flavor taster, what would you expect your results for Q2 to look like? (2 points)

5. Use Fisher's exact test either `fisher.test()` or by hand and report on your tasting skills (4 points)

Part II.

Recall the coinfection data from last week's lab (Synchronized seasonal excretion of multiple coronaviruses coincides with high rates of coinfection in immature bats).

We also collected data from adult, and sub adult flying foxes. Complete data can be seen at https://www.nature.com/articles/s41467-025-61521-7/tables/4.

6

6.1 (6 points)

Use a χ^2 test to separately determine whether there seems to be unusual co-infection patterns for the two clades for each of the three age classes. Summarize your findings.

6.2 (4 points)

Write a paragraph describing the results from 1.1 in the context of the scientific research question.

7. (6 points)

The previous question focused on the conditional, on age, relationships. Now compute the marginal relationship for coinfection that combines all age classes. Summarize your findings.

8. (4 points)

Simpson's paradox is concerned with settings where an overall effect disappear's or changes directions when considering specific groups (conditional relationships). Is this data an example of Simpson's paradox, why or why not?