Mathematical Modeling Virus (MMV)

Mathematical Modeling Virus (MMV) is a viral infection that can cause a serious disease called MMD. MMD patients are unable to control their natural urge to make mathematical models that attempt to describe biological phenomena. The rapid MMV test gives a positive result: (1) 100% of the time for people with the virus, and (2) 5% of the time for people without the virus. A certain population has a prevalence of 1%.

Q1: You pick a person from this population at random, and test them, and the test is positive. What is the probability that they have MMV?

This can be solved using Bayesian Theorem:

$$\text{P(MMD | Positive Test)} = \frac{\text{P(Positive Test | MMV)} \times \text{P(MMV)}}{\text{P(Positive Test)}}$$

Note that P(A|B) represents the probability of A occurring, given that B is true. From the prompt, we know that P(Positive Test | MMD) is 1, P(MMD) = 0.01, and that P(Positive Test | No MMD) = 0.05.

$$\text{P(Positive Test)} = \text{P(Positive Test | MMV)} \times \text{P(MMV)} + \text{P(Positive Test | No MMV)} \times (1 - \text{P(MMV)}$$
$$\text{P(Positive Test)} = 1 \times 0.01 + 0.05 \times (1 - 0.01) = 0.0595$$

Now that we have calculated the probability of a positive test occurring, we can calculate the probability of having MMV given they received a positive test:

$$\text{P(MMV | Positive Test)} = \frac{(1) \times (0.01)}{(0.0595)} = 0.168$$

Therefore, there is a 0.168 or 16.8% chance that the individual, picked at random, has MMV given that the test is positive.

Note, this problem could easily be solved using a frequentist approach as well. Given the information provided, I have chosen to use a Bayesian approach because it provides a more concrete estimate (i.e. more powerful) by taking more assumptions into account.

Q2: You learn that your friend has a positive rapid test for MMV. What do you tell them?

For this question, I will assume that my friend has only taken one rapid test. I would tell my friend what I know about MMV: (1) they may have a false positive, which occurs 5% of the time, (2) the virus is 100% effective for detecting people with the virus, (3) there is a 1% prevalence that they have MMV. I would instruct my friend to take a second test. Here, we will assume the two tests are indpendnent of one another (i.e. taking the first test does not affect the results of the second test):

$$\text{P(Positive Test}_1 \cap \text{Positive Test}_2 \text{ | No MMD)}$$
$$= \text{P(Positive Test}_1 \text{ | No MMD)} \bullet \text{P(Positive Test}_2 \text{ | No MMD)}$$
$$= (0.05)(0.05)$$
$$= 0.0025$$

Therefore, if my friend did not have MMD and took a second test, there would be a 0.0025 or 2.5% chance that both tests come back positive. Hopefully, this is enough to reassure my friend that they are healthy.

Formulate a hypothesis about your data and discuss how you would test it statistically.

The following paragraph can be found in the README.md file within the directory ShaneSeheult/QMEE on Github.

In eastern temperate North America, big brown bats (Eptesicus fuscus) generally give birth to twin pups. The bats mate in the later summer and the fall. Females store sperm during the winter and when they emerge from hibernation in the early spring they ovulate and give birth about two months later. The purpose of this study was to assess and look for differences in the birth dates of pups born to females from two captive populations. One population were caught at a maternity colony in southern Ontario, Canada about three weeks before parturition. The second population of pups were born to females that been in captivity for an extended period during hibernation and prior to parturition. I will assess if pups from newly captive wild-caught females have similar birth dates (Julian Date) compared to pups born to females that had been living in captivity for a longer period. I will also compare these populations to see if they differ in the proportion of same-sex siblings (i.e., male-male or female-female siblings) and different-sex siblings (i.e. male-female).

For each pup in this experiment (n = 58), I have recorded: (1) the pup's sex, (2) the pup's birthday (both as year-month-day and Julian Date format), and (3) the sex of their sibling (if applicable). Pups were noted as being born to moms caught in a maternity colony in Southern Ontario, Canada roughly three weeks before parturition (*Wild-Caught*; n = 41) or born to females that had been in captivity for an extended period during hibernation and prior to parturition (*Captive*; n = 17).

Hypotheses:
My null hypothesis (h_0) is that there is no difference between birthdays of pups born to wild-caught or captive bats. In other words, the pups are drawn from the same population. My alternative hypothesis (h_1) is that the difference in birthdays between the two groups is not equal to 0 and that there is an effect of being born to a mother that has spent extended time in captivity.

Method:
Of course, I do not believe that the effect will be *exactly* zero. Instead, I want to see if the effect size is (1) large or small, (2) positive or negative, and (3) whether the confidence intervals cross 0. One possibility is to conduct a two-sample t-test, examine the confidence interval for the differences between the means of the two groups, and determine if the two groups differ based on whether the confidence intervals cross 0. Although suffice, this test alone would not tell me whether the two groups are statistically equivalent to one another, given an equivalence region.

To examine this, I would run an Equivalence Test on my data. Specifically, I would conduct Two One-Sided Tests (TOST) Equivalence Testing, which can be done using the TOSTER::t_TOST function in R. I will evaluate the effect size of group using confidence intervals and see whether these confidence intervals cross 0.

To test the equivalence, I would set *equivalence bounds*, which are borne out of what we know prior to conducting this experiment. Equivalence bounds (i.e. cut-offs) are hard to determine. One way to do this is to define the smallest effect size of interest —relevant to the research question— and base your equivalence bounds off that. The hypotheses are formulated around these equivalence bounds: the null hypothesis (h_0) would be that there is an effect that is at least as large as $\pm$ the equivalence bound and the alternative hypothesis (h_1) there is not an effect that is at least as large as $\pm$ the equivalence bound.

Using this method will produce four possible outcomes:

(1) *The effect is statistically equivalent and different*.
- The confidence intervals do not cross 0 and do not cross the equivalence bounds.

(2) *The effect is statistically equivalent and not different.*
- The confidence intervals do cross 0, but do not cross the equivalence bounds.

(3) *The effect is not statistically equivalent and different.*
- The confidence intervals do not cross 0, but do cross the equivalence bounds.

(4) *The effect is not statistically equivalent and not different.*
- The confidence intervals do cross 0 and cross the equivalence bounds.

The p-values produced from this test will provide clarity to the results how sure we are that we have observed an effect whereas the confidence intervals will show what we believe is occurring. I believe this will take somewhat of a Bayesian perspective because setting equivalence bounds makes an assumption about the population. In other words, what is a reasonable difference in my measure —in this case, birthday— that can be considered meaningful.