STAT 503 – Statistical Methods for Biology

Homework 5

30 Points (32 available). Due before midnight on Tuesday, July 21, 2020

Optional practice: Whitlock and Schluter Chapter 6, practice questions. Answers are in the back of the book.

**Please use complete sentences unless the question is marked with an asterisk (\*).**

**Round answers to 3 significant figures**

1. [12 points] Each of the following scenarios describes a set of variables and a study goal. For each one, (i) identify the response variable, (ii) identify the population parameter of interest, and (iii) state a null and alternative hypothesis for the study. Remember that statistical hypotheses concern the values of population parameters, not sample statistics. Hypotheses may be stated verbally or mathematically, but must be specific. If you use an equation, you must clearly and completely define any symbols that you use except and .
   1. Most of you have probably heard that the typical oral temperature reading for a healthy human adult is 98.6F (37C). Mackowiak et al. (1992, *Journal of the American Medical Association* 268:1578-1580) tested the validity of this statement (and found that it was incorrect).
   2. To prevent rabies in wild raccoon (*Procyon lotor*) populations, oral vaccines are placed in small bait packages and then distributed in locations where raccoons are likely to encounter them. To determine whether raccoons prefer specific baits, a series of 10 randomly selected raccoons were each presented with two treats: a tablespoon of peanut butter, or a tablespoon of marshmallow fluff. The first treat that was eaten was assumed to be preferred (as it turns out, raccoons *really* like marshmallows). The researchers assume there will be no preference.
   3. Corum (2003: <http://style.org/unladenswallow/>) used theoretical calculations to predict the average airspeed velocity of an unladen, European swallow to be 11 m/s. Suppose that you collect data to test this prediction.
2. Suppose that we equip 15 European swallows with accelerometers and measure the average flight speed for each bird. Treating these 15 individual-level averages as our data (so *n* = 15), we obtain the summary statistics m/s and m/s. Follow the steps below to complete the test of null hypothesis that you specified in Question 1c.
   1. [0.5 point] Select a significance level, and briefly explain why you picked this value.
   2. [5 points] Select an appropriate test statistic and calculate its value. Show your work. If appropriate, also calculate the statistic’s degrees of freedom.
   3. [0.5 point] Determine the -value for the hypothesis test. You may use R for this step. Write the code that you use here.
   4. [1 point] Based on your -value, draw a conclusion to reject or not reject the null hypothesis.
   5. [1 point] Estimate the confidence interval that goes with the hypothesis test.
   6. [2 points] Formally report your result in 1-2 sentences (see the sheet on reporting statistical results in either Brightspace > Content > Getting Help or the cheet sheets page on the website).
3. Suppose that the smallest effect size that you consider to be biologically important in Question 2 is 0.5 m/s, and that after performing the hypothesis test, you obtain (this is not the -value that you will get in Question 2).
4. [2 points] Suppose that the power of your test at an effect size of 0.5 m/s is 0.6. Is it reasonable to draw any conclusion about the validity of the null or alternative hypotheses from these results? Why, or why not?
5. [2 points] How would your answer in Question 3a change if the power at an effect size of 0.5 m/s had been 0.9 instead of 0.6? The -value is still 0.23. Please explain your reasoning.
6. [3 points] It is very important to understand the difference between the statistical significance of a finding and its biological importance.
   1. Please explain the difference between statistical significance and biological importance (or "biological significance") in your own words.

* 1. Under what circumstances might a result be statistically significant but biologically unimportant?

1. [3 points] Assuming that is the standard "default" significance level, please recommend the **best** choice of significance level to use in each of the following scenarios, out of the following options: ,, or. You may use each value as many times as you want to, and you are not required to use all of them. To make your selection, consider both the risks and consequences of making a Type I or Type II error.

\_\_\_\_\_ Your alternative hypothesis suggests that a generally accepted and previously well-supported theory is incorrect, and may therefore invite controversy and skepticism.

\_\_\_\_\_ If supported, your alternative hypothesis will provide interesting new insights about your study system, but it represents a plausible extension of existing theory.

\_\_\_\_\_ You are conducting an initial screening of several dozen proteins, in hopes of identifying promising candidates for future research. The current data have small sample sizes, but follow-up studies will use much larger sample sizes.

\_\_\_\_\_ You did not give the significance level much thought prior to analyzing your data. After analyzing your data, you find that .

\_\_\_\_\_ You are checking specimens for contamination. The null hypothesis is that they are not contaminated. Undetected contamination will invalidate future results, but a false positive will trigger time consuming and expensive follow-up tests. You can easily get as many specimens as you need, so power is not limited.

\_\_\_\_\_ You are checking food samples for contamination. The null hypothesis is that they are not contaminated. A false positive will trigger follow-up tests, but undetected contamination could pose a significant public health threat and risk legal liability. Sample size is small and power is limited.