STAT 503 – Statistical Methods for Biology

Homework 5

24 Points (26 available). Due at 11:59 PM on Monday, Nov. 2, 2020

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

**Round answers to 3 significant figures**

1. [6 points] Each of the following scenarios describes a set of variables and a study goal. Please identify a (i) null and (ii) alternative hypothesis for each study. Each hypothesis should be stated verbally and as an equivalent mathematical equation. Your answer should unambiguously identify the variable and parameter that you are making a hypothesis about. Remember to define any symbols that you use except and .
   1. Average body temperature in adult humans is often stated to be 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. Can a snake be right- or left-handed? To investigate, Roth (2003; *Animal Behavior* 66:337-341) observed the coiling behavior of cottonmouth snakes (*Agkistrodon piscivorus*). On average, each snake was observed 60 different times, and on each occasion, the direction of the snake’s coil was classified as either clockwise or counterclockwise. Please state your hypotheses for a single snake. It would show “handedness” if it coils in one direction more often than the other.
   3. A team of researchers is studying blood glucose concentrations in subjects at two different locations, and for logistical reasons, they have chosen to analyze specimens locally instead of sending all of the samples to a single lab. To make sure that their analyses are consistent, they collect blood samples from 30 subjects, and then divide each sample into two parts, labeled A and B. For each specimen, part A is sent to one lab, and part B is sent to the other. Then they calculate the difference (A – B) for each subject.

All biological organisms are physiologically adapted to specific climatic conditions, and their fitness declines if they live in a climate that is too cold, too warm, or too dry. As a result, plants, animals, and other organisms are only found at certain latitudes and elevations. For example, different species of trees occur at different elevations on mountain slopes, and as elevation increases, temperatures eventually become cold enough that tree seedlings cannot survive and establish, resulting in the alpine tree line.

In theory, increases in global mean temperatures due to climate change should increase the elevation at which plants and animals encounter limiting climatic conditions. As a result, the highest elevation at which a species is found should increase over time. This phenomenon is called an *elevation range shift*. Chen et al. (2011, *Science* 333:1024-1026) collected published data on elevation range shifts in 1367 individual species of organisms, and then grouped the species by taxon (birds, mammals, butterflies, plants, etc.) and geographic area. This resulted in taxon-area groups. For each group, they calculated the average elevation range shift in meters per decade. Positive values indicate an uphill shift, and negative values indicate downhill shift. You can find these data on the website in the file **Chen\_et\_al\_2011\_elevation.csv**. Please use it to answer the following questions.

1. [1 point] What is the sample unit for this analysis?
2. [1 point] State anull and alternative hypothesis that will allow you to use to test the prediction that on average, organisms' elevational ranges have shifted uphill over time. You may use either a one-tailed or two-tailed alternative, as you feel is appropriate (either is acceptable; your answers to later questions will be graded on the basis of your decision here).
3. [1 point] In Question 3, you selected either a one-tailed or a two-tailed alternative hypothesis. Please (a) identify which alternative you chose to use (I'm looking for consistency with the hypothesis you actually stated), and (b) briefly justify your decision. Note that they are *both* justifiable, but for different reasons.
4. Perform the following steps to complete a single sample -test for your hypotheses:
   1. [1 point] State the value of that you will use for this test, and briefly explain why you selected this value?
   2. [1 point] List the assumptions of the -test.
   3. [2 points] Examine a quantile plot to verify the model's distributional assumptions. Include your plot here (with caption). Based on the graph, do you feel that the -test is appropriate for these data? Why, or why not (please cite specific evidence from the graph)?
   4. \*[2 points] The sample mean, standard deviation, and sample size for these data are m/decade, m/decade, and . Please calculate the -score by hand and report it here, along with the appropriate degrees of freedom (show your work).
   5. \*[1 point] Use the t-score that you calculated in 5d and the pt() function in R to find the -values for each of the following alternative hypotheses:
      1. The average range shift differs from zero.
      2. The average range shift is positive.
      3. The average range shift is negative.
   6. [1 point] Calculate the confidence interval that corresponds to this test (show your work).
   7. [2 points] Write a short statement to formally report your results and conclusion. Your statement should include all of the relevant statistical information, indicate whether or not you reject the null hypothesis, and give a biological interpretation of the result in the context of the research question. You should one use ONE of the -values that you calculated in 5e. See *Guidelines for reporting statistical results*, on the website’s Cheat Sheets page, for guidance.

1. [2 points] Corum (2003: <http://style.org/unladenswallow/>) used theoretical calculations to predict that the average airspeed velocity of an unladen, European swallow should be 11 m/s. Suppose that we equipped 15 European swallows with accelerometers and measured the average flight speed for each bird. Treating these 15 individual-level averages as our data (so *n* = 15), we obtained the summary statistics m/s and m/s. As a result, and . We consider any difference in velocity > 0.5 m/s to be biologically important.
2. 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?
3. How would your answer in Question 6a 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 0.3018. Please explain your reasoning.
4. [1 point] In your own words, please explain the difference between statistical significance and biological importance (or "biological significance").

1. [3 points] Please select the **best** choice of significance level to use in each of the following scenarios, out of the following options: ,, or. To make your selection, consider both the risks and consequences of making a Type I and a Type II error, and assume that represents the standard significance level that would be used under normal, “business-as-usual” conditions. You may use each value as many times as you want to, and you are not required to use all of them.

Your alternative hypothesis suggests that a generally accepted and previously well-supported theory is incorrect. Rejecting the null 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 good candidates for future research. The current data have small sample sizes, but follow-up studies will use much larger sample sizes.

Before analyzing your data, you did not give the significance level much thought. After analyzing your data, you find that .

You are checking reagents 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 for contamination. The null hypothesis is that the food is not contaminated. A false positive will trigger follow-up testing, but undetected contamination would pose a significant public health threat and risk legal liability. Your sample size is small and power is limited.

1. \*[1 point] Please paste your code here. Remember to organize it and mark the code for each question.