Help and Hints for Assignment 4

DISCLAIMER: These helps and hints are intended as a guide only—they may contain typos or errors. Ultimately, you are responsible for the accuracy and correctness of your own answers. For further help, reach out to one of the TAs and/or the professor.

- MS 7.118 - pg 364
  a) The `sample()` function is probably the easiest way to do this. Consider the following code:
     - ```
       # Select random 35 numbers from 1 to the number of the rows in example_dataframe
       numbers = sample(1:116, size = 35, replace = FALSE)
       # select 35 species
       sample = example_dataframe[numbers,]
       ```
     - You'll have to change `example_dataframe` to the relevant data frame
     - Check and make sure there are only 116 rows
  b) Answers will vary. Expect a large variance.
  c) A confidence interval is a range in which we are confident—to a certain degree—that the true population parameter lies.
  d) Carefully think about the meaning of a 95% confidence interval; it is not just a measure of our feelings about the data, but rather of the likelihood that an observed sample is statistically representative of the true population. Consider: if we took 100 samples of the same size from the bird data, then constructed confidence intervals for each of those 100 samples, how many of those 100 confidence intervals would contain the true mean? (Hint: $1 - \alpha$)
  e) This should only require some small edits of the code. Remember, we can access columns of a data frame by using a dollar sign: `example_dataframe$column_name`
  f) This is very similar to Lab 11, Task 5.
  g) Consider the point estimate from Lab 11, Task 5. It is the *difference* between two proportions. The confidence interval, then, describes the range of the *magnitude of that difference*.
- MS 7.120 – pg 365
  a) This problem is very similar to Lab 11, Task 3. We have two samples of different sizes and we're trying to estimate the *mean difference* between them.
  b) This problem is very similar to Lab 11, Task 6. Again, we have two different sample sizes and we want to estimate the *ratio* of one variance to another. If ratio is 1, then the variances don't differ, so to answer the last part: does the confidence interval include 1? If not, then we have *statistically significant evidence* that the variances are different.
- MS 7.128 – pg 367
  a) See Theorem 6.11, and remember: $Z = \frac{(Y-\mu)}{\sigma}, \mu = 0$
       ```

b) We want to *pivot* about Y to estimate population variance. We can do this with χ^2. Rearrange $P\left(\chi^2_{\left(1-\frac{\alpha}{2}\right)} \leq \frac{Y^2}{\sigma^2} \leq \chi^2_{\left(\frac{\alpha}{2}\right)}\right) = 1-\alpha$ so that σ^2 is the subject of the inequality.

- MS 8.24
  a) Remember, almost always: equality is in the null hypothesis. The null hypothesis typically assumes "business as usual" or no effect due to outside influence.
  b) These values are provided on the table in the book.
  c) The rejection region should be reported as "$t < \{some\ value\}$ and $t > \{some\ value\}$". You can use R to calculate these values or the t tables in Appendix B. These are equivalent to $\pm\, t_{\left(\frac{\alpha}{2}\right)}$
  d) Compare the calculated t value (-1.02) to your result in part (c).
  e) Consider the use of the statistic and the role α plays in these calculations.
- MS 8.28
  a) A t test is most appropriate here. Remember $t = \frac{\bar{y}-\mu_{population}}{\frac{s}{\sqrt{n}}}$ (you can calculate this by hand). You'll need to look up $\pm\, t_{\left(\frac{\alpha}{2}\right)}$ in Appendix B or calculate in R and compare the t statistic you just calculated to see if lies outside the acceptance range $\left(-t_{\left(\frac{\alpha}{2}\right)}, t_{\left(\frac{\alpha}{2}\right)}\right)$.
  b) Rearrange $t = \frac{\bar{y}-\mu_{population}}{\frac{s}{\sqrt{n}}}$ for $\bar{y}$, then substitute in the values of $\pm\, t_{\left(\frac{\alpha}{2}\right)}$ to find a rejection region for $\bar{y}$. Then, find the probability of $\bar{y}$ falling in that rejection region. Your final probability calculation should look like $P(t < \{some\ value\}) + P(t > \{some\ value\}) = 0.1212$
- MS 8.44
  a) This problem is similar to Lab 11, Task 3. Use the samples for small sample sizes. You can rearrange the equations for t to get the test statistic like this: $t = \frac{(\bar{y}_1-\bar{y}_2)-(\mu_1-\mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_i}+\frac{1}{n_2}\right)}}$, where, for the null hypothesis, $\mu_1 - \mu_2 = 0$.
- MS 8.84
  a) This is an F test, as in Lab 11, Task 6. Instead of making a statement about Exercise 8.33.a, just state your conclusion about the variances of the two turbines' heat rates.
  b) Follow process from part (a).
- MS 8.99
  a) Again, consider an F test: a ratio of variances where a value of 1 indicates the same variance. (Equality is in the null.)
  b) Again, see Lab 11, Task 6.
  c) The rejection region will be the upper tail of the F distribution. In R, you can find this with `qf(0.025,4,5, lower.tail = FALSE`. This should be reported as "the rejection region is $F > \{some\ value\}$. (Minus the curly braces, of course.)
  d) You can use pf() with the result from part (c)
  e) Is the null hypothesis rejected? Is there enough evidence to indicate a variation between the two locations for the given alpha?
  f) Consider the assumptions of the F test.
- MS 8.104

a) This is another t test, where the hypotheses are concerned with a difference of the means.

- Question 10

  a) Follow the logic of the problem. We've made the necessary adjustment before in previous lab assignments. The best way to return multiple values is to add to the list in the `return` line. You'll have to add the theoretical sample calculation for the blue numbers.