

Name: \_\_\_\_\_

**Instructions:**

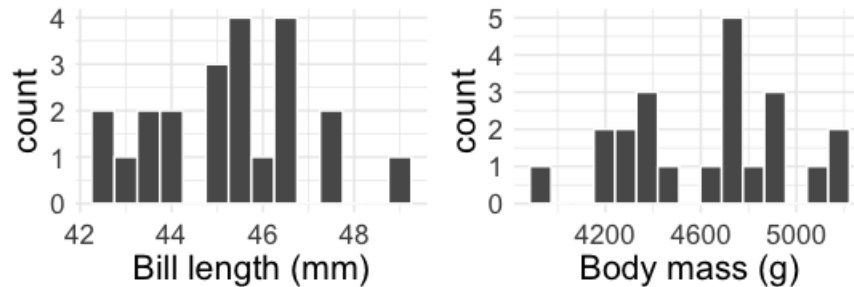
- You will have **75 minutes** to complete this midterm.
- If you use extra paper, you do not need to rewrite the statement of each question, but the question number itself should prominently appear at the start of each problem.
- Solutions must be clearly legible, and you must show all your work and provide adequate justification to receive full credit for a problem.
- The only resources you are allowed to use are a calculator, your single-sided formula sheet, and the provided tables. Your formula sheet will be submitted alongside the midterm.
- Round to the nearest value if using the  $z$  or  $t$  tables.
- A rubric is provided on the back of this page.

By signing below, you agree to adhere Middle College Honor Code pledge: “I have neither given nor received unauthorized aid on this assignment.”:

\_\_\_\_\_  
Signature

- 
1. Gallup randomly polled 635 Democrats and 372 Republicans on their opinion of driving while impaired by marijuana. 20% of Democrats and 35% of Republicans said that driving while impaired by marijuana poses a very serious threat, when compared to alcohol, prescription painkillers, and prescription antidepressants. **In this problem, you may assume conditions for inference are met.**
    - (a) Suppose we want to conduct a hypothesis test for evaluating whether the proportions of Democrats and Republicans who think that driving while impaired by marijuana poses a very serious threat are different. Write down the hypotheses of interest using statistical notation, defining quantities appropriately.
  
  
  
  
  
  
  
  
  
  
    - (b) Calculate the standard error for the hypothesis test you defined in (a).
  
  
  
  
  
  
  
  
  
  
    - (c) Suppose we want to estimate the difference between the proportions of Democrats and Republicans who think that driving while impaired by marijuana poses a very serious threat using a confidence interval. Is the value of the standard error for this confidence interval different than the standard error in the hypothesis test? If yes, explain how and why. *No calculation is required.*

2. We have a random sample of 22 female Gentoo penguins. For each penguin, the body mass (g) and the bill length (mm) were recorded. The histograms of the data are displayed below:



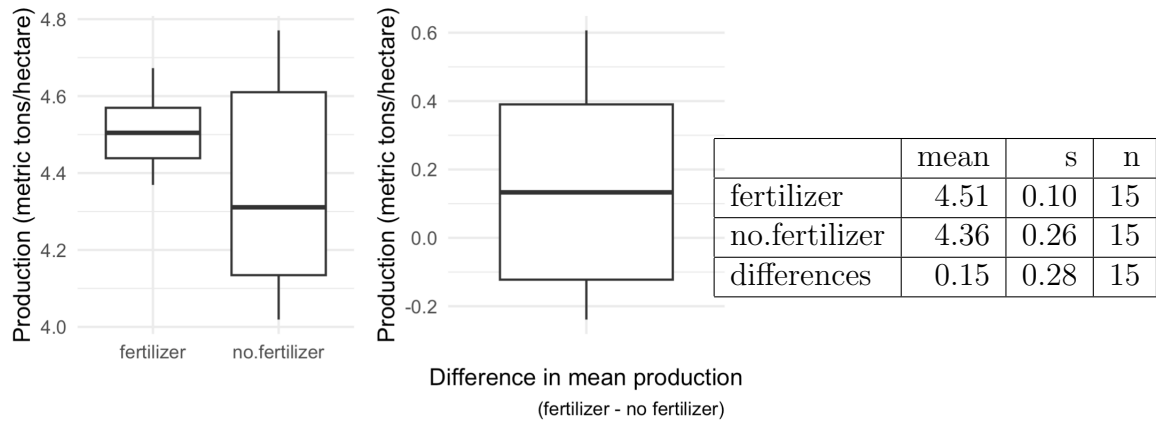
We fit a linear regression model using bill length to understand the behavior of body mass. The model has an  $R^2$  of 0.16. The partial tidy output of the coefficients is:

term	estimate	std.error	statistic	p.value
(Intercept)	983.284	1866.495	0.527	0.604
bill_length	80.449	37.948		

- (a) Is there evidence of a positive linear relationship between bill length and body mass? Use an appropriate method to answer this question. You may assume that conditions are met for this part.

- (b) Interpret the estimated slope coefficient in context.
- (c) Interpret the intercept in context. Is it a practical interpretation? Explain.
- (d) The standard deviation of the bill **lengths** in this sample was 1.69 mm. The observed average **body mass** was 4627.27 grams. Now, suppose we'd like to focus on the body mass alone. Obtain a 90% confidence interval for the **average body mass** of all female Gentoo penguins. No need to interpret.

3. A research team developed a new fertilizer that is designed to increase wheat output. Unfortunately, the fertilizer has lots of chemicals that are suspected to be toxic. The Federal Drug Administration (FDA) has hired you to determine if the new fertilizer is effective. To collect data, the FDA randomly selected 30 farms in Vermont. Half of these farms were randomly assigned to receive the new fertilizer. The remaining farms did not receive the fertilizer. Each farm recorded their metric tons of wheat per hectare at the end of harvesting season. The data are as follows:



- (a) Conduct a hypothesis test for the FDA's research question at the  $\alpha = 0.05$  level. Make a decision and state a conclusion for the FDA in context.

- (b) It is possible you may have made the wrong decision in (b). What sort of error might you have committed, and why? What is an example of a consequence associated with committing this type of error, in context?

---

**Determine the appropriate test**

4. For each of the following research situations, decide which of the following statistical procedures (a)-(f) would *most likely* be used to answer the research question. Assume all assumptions have been met for the procedure. Note: each statistical procedure will be used **exactly once**.

**Statistical procedures:**

- (a) Test one mean against a hypothesized value.
- (b) Test the difference between two means.
- (c) Test the mean difference for paired sampled.
- (d) Construct confidence interval(s).
- (e) Test one proportion against a hypothesized value.
- (f) Test whether two variables are linearly related.

**Research situations:**

- How does the amount of hours spent studying for this midterm relate to performance on the midterm?
- What percentage of all US adults have fallen in love at least once in their life?
- Do Vermont high school seniors who attend a summer math camp score above the state average on the math section of the SAT?
- Do students in my different STAT 201 sections perform similarly on this midterm?
- How does the percentage of vegan-ism among Vermont students compare to the national rate?
- Do students in my STAT 201 sections perform better on the second midterm than the first?

**Multiple choice**

For each of the following, choose the best answer.

5. A random sample of 100 runners who completed the Cherry Blossom 10 mile run yielded an average completion time of 95 minutes. A 95% confidence interval calculated based on this sample is 92 minutes to 98 minutes. Which of the following is false based on this confidence interval?
- (a) We are 95% confident that the true average finishing time of all runners who completed the Cherry Blossom 10 mile run is between 92 minutes and 98 minutes.
  - (b) 95% of the time the true average finishing time of all runners who completed the Cherry Blossom 10 mile run is between 92 minutes and 98 minutes.
  - (c) The margin of error of this confidence interval is 3 minutes.
  - (d) A 98% confidence interval would include 92 and 98 minutes in its interval.
6. Suppose we have the following set of hypotheses:  $H_0 : p = 0.8$  versus  $H_A : p > 0.8$ , where  $n = 30$  and  $\hat{p}_{obs} = 0.6$ . Suppose we want to use simulation-based methods to test these hypotheses. Which of the following is the correct set up for this hypothesis test? Let red = success; blue = failure;  $\hat{p}_{sim}$  = proportion of reds in simulated samples.
- (a) Place 50 red and 50 blue chips in a bag. Sample, *with replacement*, 30 chips and calculate the proportion of reds. Repeat this many times and calculate the proportion of simulations where  $\hat{p}_{sim} \geq 0.6$ .
  - (b) Place 60 red and 40 blue chips in a bag. Sample, *with replacement*, 30 chips and calculate the proportion of reds. Repeat this many times and calculate the proportion of simulations where  $\hat{p}_{sim} \geq 0.8$ .
  - (c) Place 80 red and 20 blue chips in a bag. Sample, *without replacement*, 30 chips and calculate the proportion of reds. Repeat this many times and calculate the proportion of simulations where  $\hat{p}_{sim} \geq 0.6$ .
  - (d) Place 80 red and 20 blue chips in a bag. Sample, *with replacement*, 30 chips and calculate the proportion of reds. Repeat this many times and calculate the proportion of simulations where  $\hat{p}_{sim} \geq 0.6$ .
7. Suppose in a population, 20% of adults do not have a savings account. What is the expected shape of the sampling distribution of proportions of adults without a savings account in random samples of 60 adults from this population?
- (a) Right-skewed
  - (b) Left-skewed
  - (c) Symmetric
  - (d) Cannot tell with the information provided