Problem set 3

## Section 5.1 problems

### Problem 1: Identify the parameter

For each of the following situations, state whether the parameter of interest is a mean or a proportion.

1. A survey reports that hand sanitizer products show a 255% increase in revenue within a 1-year period.
2. A NBC News survey reports that 60 percent of Americans support keeping stay-at-home restrictions to fight coronavirus.
3. A Duke student survey asks whether or not students have left Durham for the summer.
4. In a survey, people who ride the bus are asked whether or not they use the online schedule.
5. In a survey, people who ride the bus are asked how many times they used the online schedule last year.
6. The bus company reports that the number of daily riders shows a 12% increase since last month.

### Problem 2: Wearing masks

In a random sample 823 adults in the United States, 461 say they plan to always wear a mask in public in the next two months.

1. What population is under consideration in the data set?
2. What parameter is being estimated?
3. What is the point estimate for the parameter?
4. What is the name of the statistic can we use to measure the uncertainty of the point estimate?
5. Compute the value from part (d) for this context.
6. An NBC news host thinks the value is actually 50%. Should she be surprised by the data?
7. Suppose the true population value was found to be 58%. If we use this proportion to recompute the value in part (e) using instead of , does the resulting value change much?

### Problem 3: Repeated sea hare samples.

Aplysia vaccaria, coloquially known as the California black sea hare, can grow to be up to 75 cm (29 in) in length, making it the largest gastropod in the world. About 13% of sea hares grow to weigh over 20 pounds. Marine scientists randomly sample 30 hares and check if those hares exceed 20 pounds. They repeat this 1,000 times and build a distribution of sample proportions.

1. What is this distribution called?
2. Would you expect the shape of this distribution to be symmetric, right skewed, or left skewed? Explain your reasoning.
3. Calculate the variability of this distribution.
4. What is the formal name of the value you computed in (c)?
5. Suppose the scientists decide to sample again, this time collecting 110 sea hares per sample, and they again collect 1,000 samples. They build a new distribution of sample proportions. How will the variability of this new distribution compare to the variability of the distribution when each sample contained 30 observations?

## Section 5.2 problems

### Problem 4: Nuclear opposition

A Gallup poll conducted in 2016 fount that 54% of the American public oppose nuclear energy. This marked the first time that public opposition to nuclear power in the United States had achieved a majority in the 23 years of Gallup polling on the subject. The standard error for this estimate was 4.0%, and a normal distribution may be used to model the sample proportion. Construct a 99% confidence interval for the fraction of the American public who oppose nuclear energy, and interpret the confidence interval in context.

### Problem 5: Mental health

An survey of Duke grad students asked the question: For how many days over the past 30 days was your mental health negatively impacted by your studies? Based on responses from 953 Duke grad students, the survey reported a 95% confidence interval of 6.31 to 8.54 days in 2019.

1. Interpret this interval in context of the data.
2. What does “95% confident” mean? Explain in the context of the application.
3. Suppose the researchers think a 99% confidence level would be more appropriate for this interval. Will this new interval be smaller or wider than the 95% confidence interval?
4. If a new survey were to be done with 500 Duke grad students, do you think the standard error of the estimate would be larger, smaller, or about the same.

## Section 5.3 problems

### Problem 6: Nuclear opposition part II

Using the Gallup poll from Problem 4, identify each of the following statements as true or false. Provide an explanation to justify each of your answers.

1. The data provide statistically significant evidence that more than half of the American public opposes nuclear energy. Use a significance level of .
2. Since the standard error is 4%, we can conclude that 96% of the American public were included in the study.
3. If we want to reduce the standard error of the estimate, we should collect less data.
4. If we construct a 90% confidence interval for the percentage of the American public that opposes nuclear energy, this confidence interval will be wider than a corresponding 99% confidence interval.

### Problem 7: Climate change

An article on climate change suggests that 30% of college-aged Americans don’t believe in global warming. You think that this may be wrong and decide to collect your own sample for a hypothesis. From a random sample of 780 college-aged Americans, you find that 28% of them don’t believe in climate change. Someone in your study group offers to help you with setting up the hypothesis test and comes up with the following hypothesis. Indicate any errors you see.

### Problem 8: Waiting at an Urgent Care

The Duke Orthopoedic Urgent Care center reports a 95% confidence interval for the mean waiting time at their facility of (44 minutes, 97 minutes). Answer the following questions based on this interval.

1. A local newspaper claims that the average waiting time at this Urgent Care exceeds 2 hours. Is this claim supported by the confidence interval? Explain your reasoning.
2. The Dean of Medicine at this center claims the average wait time is 1.1 hours. Is this claim supported by the confidence interval? Explain your reasoning.
3. Without actually calculating the interval, determine if the claim of the Dean from part (b) would be supported based on a 99% confidence interval?

### Problem 9: Working backwards

You are given the following hypotheses:

We know that the sample size is 1,238. For what sample proportion would the p-value be equal to 0.01? Assume that all conditions necessary for inference are satisfied.

### Problem 10: Which is higher?

In each part below, there is a value of interest and two scenarios (I and II). For each part, report if the value of interest is larger under scenario I, scenario II, or whether the value is equal under the scenarios.

1. The standard error of when (I) or (II) .
2. The margin of error of a confidence interval when the confidence level is (I) 90% or (II) 80%.
3. The p-value for a Z-statistic of 2.5 calculated based on a (I) sample with or based on a (II) sample with .
4. The probability of making a Type 2 Error when the alternative hypothesis is true and the significance level is (I) 0.01 or (II) 0.05.

## Chapter problems

### Problem 11: True or false

Determine if the following statements are true or false, and explain your reasoning. If false, state how it could be corrected.

1. If a given value (for example, the null hypothesized value of a parameter) is within a 95% confidence interval, it will also be within a 90% confidence interval.
2. Increasing the significance level () will increase the probability of making a Type 1 Error.
3. Suppose the null hypothesis is and we fail to reject . Under this scenario, the true population proportion is 0.2.
4. With large sample sizes, the magnitude of the difference between the null value and the observed point estimate, a difference often called the effect size, needs to be quite large in order to be identified as statistically significant.

### Problem 12: Atlantic salmon

It is believed that about 30 percent of Atlantic salmon survive to spawn again (these are called repeat spawners). In a random sample of 205 Atlantic salmon, 49 are repeat spawners Conduct a hypothesis test for the following question: do these data provide evidence that the 30% value is inaccurate?

### Problem 13: CLT for proportions

CLT for proportions. Define the term “sampling distribution” of the sample proportion, and describe how the shape, center, and spread of the sampling distribution change as the sample size increases when p = 0.1. (Note this is OIS Problem 5.34.)

### Problem 14: Same observation, different sample size

Suppose you conduct a hypothesis test based on a sample where the sample size is , and arrive at a p-value of . You then refer back to your notes and discover that you made a careless mistake, the sample size should have been . Will your p-value increase, decrease, or stay the same? Explain.

<https://www.cnbc.com/2020/03/03/coronavirus-hand-sanitizer-sales-surge-leading-to-price-hikes.html>

<https://www.nbcnews.com/politics/meet-the-press/poll-six-10-support-keeping-stay-home-restrictions-fight-coronavirus-n1187011>

<https://news.gallup.com/poll/190064/first-time-majority-oppose-nuclear-energy.aspx>