Q01

Online communication. A study suggests that the average college student spends 10 hours per week communicating with others online. You believe that this is an underestimate and decide to collect your own sample for a hypothesis test. You randomly sample 60 students from your dorm and find that on average they spent 13.5 hours a week communicating with others online. A friend of yours, who offers to help you with the hypothesis test, comes up with the following set of hypotheses.

Define a random variable X as the number of hours students spend communicating online.

Indicate any errors you see.

$$H_{\text{Null}} : \bar{X} < 10 \text{ hours}$$
$$H_{\text{Alternative}} : \bar{X} > 13 \text{ hours}$$

Q02

Mental health. The 2010 General Social Survey asked the question: "For how many days during the past 30 days was your mental health, which includes stress, depression, and problems with emotions, not good?". Based on responses from 1,151 US residents, the survey reported a 95% confidence interval of 3.40 to 4.24 days in 2010. (a) Interpret this interval in context of the data. (b) What does "95% confident" mean? (c) If a new survey were to be done with 500 Americans, would the standard error of the estimate be larger, smaller, or about the same (assuming the standard deviation has remained constant since 2010)?

Q03

In the early 1990's, researchers in the UK collected data on traffic flow, number of shoppers, and traffic accident related emergency room admissions on Friday the 13th and the previous Friday, Friday the 6th.

A summary of the number of cars that passed 10 intersections is below

Statistic	6th	13th	13th minus 6th
Mean	128,385	126,550	-1,835
Std Dev.	7,259	7,664	1,176
Number of intersections	10	10	10

We will use a two-sample t-test to compare the mean number of cars passing intersections on the 6th and the 13th. The two-sample t-test is defined as

$$t_{\text{two-sample}} = \frac{\bar{X} - \bar{Y}}{S} \tag{1}$$

where X is a random variable counting the number of cars that pass an intersection on the 6th, Y counting the number of cars that pass an intersection on the 13th. We can compute the standard deviation (S) for the difference between X and Y with the following formula.

$$S = \sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}} \tag{2}$$

where s_X and s_Y are the standrad deviations of the random variables X and Y, and n_X and n_Y are the number of observations collected for random variables X and Y. The degrees of freedom for the above test is the smaller of $n_X - 1$ and $n_Y - 1$.

(a) Please define a statistical hypothesis that evaluates whether the number of cars passing intersections on Friday the 6th is different than the number on Friday the 13th. (b) Calculate the test statistic and the p-value. (c) What is the conclusion of the hypothesis test? (d) Interpret the p-value in this context.

Q04

In recent years, widespread outbreaks of avian influenza have posed a global threat to both poultry production and human health. One strategy being explored by researchers involves developing chickens that are genetically resistant to infection. In 2011, a team of investigators reported in *Science* that they had successfully generated transgenic chickens that are resistant to the virus.

As a part of assessing whether the genetic modification might be hazardous to the health of the chicks, hatch weights between transgenic chicks and non-transgenic chicks were collected. Does the following data suggest that there is a difference in hatch weights between transgenic and non-transgenic chickens?

Use a two-sample t test like in Q03.

Statistic	transgenic chicks (grams)	non-transgenic chicks (grams)
Mean	45.14	44.99
Std Dev.	3.32	4.57
Number of chicks	54	54

Q05 (programming!)

Please compute a 80%, 90%, and 95% confidence interval for the daily number of new confirmed cases of COVID-19 for the state of Pennsylvania. Use case counts from the COVIDtracking project from 2020-08-01 to 2020-11-01.

Here is a code snippit to help get you started.

```
import pandas as pd
paCounts = pd.read_csv("https://api.covidtracking.com/v1/states/pa/daily.csv")
paCountsFromAug = paCounts.loc[(paCounts.date>=20200801) & ((paCounts.date<=20201101)),:]
```