Q2

30 Points

The total depth of water in a column of the atmosphere if all the water in that column were precipitated as rain is known as the total precipitable water. From previous research, it is well documented that daily measurements of the log of total precipitable water (let's call this LPW) follows a normal distribution (Foster and Bevis, 2003; Fujita and Sato, 2007; NOAA, 2019).

Researchers are interested in estimating the mean LPW in a Peruvian valley to a specific degree of precision for agricultural purposes. Suppose that the variance of LPW in this valley is known to be 6 mm.

Q2.1

7 Points

The researchers wish to estimate the mean LPW with a two-sided 95% confidence interval that has a total length of no more than 1.5 mm. How many samples must be taken in order to do so? You may attach any supporting documents or work if needed.

No files uploaded

Q2.2

7 Points

Suppose we are still only interested in intervals with length no more than 1.5 mm. What two-sided confidence levels can be achieved if we have funding for 50 observations? You may attach any supporting documents or work if needed.

No files uploaded

Q2.3

5 Points

Suppose a different research team (let's call them Research Team B) also set up LPW monitoring stations in this valley, and calculated a two-sided 95% confidence interval of (3.28, 3.80). How many samples did they use to obtain this interval? You may attach any supporting documents or work if needed.

No files uploaded

Q2.4

5 Points

Research Team B interpreted their interval as follows: *"We are 95% confident that the mean LPW in this valley lies between 3.28 and 3.80 mm."*

Explain what it means to be *"95% confident"* in the context of this research problem.

Q2.5

3 Points

Using the same data from Q2.3, Research Team B conducted a hypothesis test to evaluate whether mean LPW was equal to 4 mm. Would their p-value be lower than 0.05, equal to 0.05, higher than 0.05, or can we not say for sure? No explanation needed.

Q2.6

3 Points

Suppose the true LPW in the valley was 3.65 mm. In the hypothesis test from Q2.5, did Research Team B make a type 1 error, a type 2 error, neither type of error, or can we not tell from the information provided? Explain.

Q3

20 Points

The **Ljung-Box test** is a hypothesis test regarding autocorrelation in time series models. The null and alternative hypotheses are as follows:

- H_0: the data do not exhibit serial correlation
- H_1: the data exhibit serial correlation

Under the null hypothesis, the test statistic has a chi-square distribution with certain degrees of freedom. We reject the null hypothesis if the test statistic is greater than some critical value of this chi-square distribution.

Suppose we performed the Ljung-Box test on a dataset, obtained a p-value of 0.014, and rejected the null hypothesis based on a pre-specified α of 0.05. Specify whether each statement below is TRUE or FALSE **in the context of this scenario**.

Q3.1

2 Points

The Ljung-Box test is a one-tailed test.

Q3.2

2 Points

All else being equal, increasing the significance level (α) of the Ljung-Box test would also increase the power of the test.

Q3.3
2 Points

If we were to perform the same experiment repeatedly on new observations, we would conclude that the data exhibit serial correlation approximately 5% of the time.

Q3.4
2 Points

We can determine the exact probability that we have made a type 2 error.

Q3.5
2 Points

We can determine the exact probability that we have made a type 1 error.

Q3.6
2 Points

The probability that we conclude the data exhibit serial correlation when in fact they do not is 0.014.

Q3.7
2 Points

The probability that the data exhibit serial correlation is at least 0.95.

Q3.8
2 Points

It is appropriate to conclude that the data exhibit serial correlation.

Q3.9
2 Points

If there is truly no serial correlation in our data, then we have made a type 1 error.

Q3.10
2 Points

The probability that the data do not exhibit serial correlation is 0.014.

Q4
35 Points

Lead poisoning is a serious, but preventable, cause of morbidity and mortality. In 2014, residents of Flint began to worry about lead levels in their drinking water after the water supply was switched to the Flint River and demanded their water be tested by authorities. In written instructions to residents regarding testing for lead levels, the Michigan Department of Environmental Quality (MDEQ) instructed residents to flush their water (i.e., let taps run) for "at least 5 minutes" before taking a sample to be tested. Under these conditions, the MDEQ did not find sufficient evidence to suggest that lead levels were high enough to be actionable.

However, these instructions were in contravention to the EPA guidance that samples taken must be at the "first draw" (i.e., immediately after turning on the tap). Citizen scientists were concerned that there might be differential lead levels in water samples depending on how long the tap was run for before taking the sample. In 2015, Flint residents sent three water samples to Virginia Tech to be analyzed: the first draw was immediately after turning on the water, the second draw was 45 seconds later, and the third draw was 2 minutes later.

The `flint.csv` file contains these data. Two relevant variables are as follows:

- `lead`: lead levels in parts per billion (ppb)
- `draw`: which draw the sample was taken at (first, second, or third)

You may access this dataset by cloning the following repository:

Q4.1

5 Points

Construct a 95% confidence interval for the proportion of first draw samples which had lead levels over 15 ppb. You may attach any supporting documents, work, or code if needed.

No files uploaded

Q4.2

15 Points

Formally and comprehensively evaluate whether there is a difference in lead levels in water samples depending on how long the tap was flushed for (first, second, or third draw). If there is a significant difference, determine where the differences are. You may attach any supporting documents, work, or code if needed.

No files uploaded

Q4.3

10 Points

According to EPA guidelines, if over 10% of consumer first draw water samples surveyed have lead levels over 15 ppb, then the water supply is dangerously high in lead and is considered actionable. Formally and comprehensively evaluate whether first draw water samples fall under EPA actionable guidelines among these samples. You may attach any supporting documents, work, or code if needed.

No files uploaded

Q4.4

5 Points

Summarize your scientific findings from Q4.1 - Q4.3 in a paragraph suitable for a public health policy lay audience. Avoid use of statistical terminology such as "p-value" or the names of specific tests you used. Be sure to communicate your findings in the context of policy insights.

Q5

15 Points

Gilchrist et al. (*JAMA Oncology*, 2020) evaluated data from a study that evaluated risk of dying from cancer in a cohort of US adults aged 45 and older. The following table describes some baseline characteristics of the patients in their study, stratified by whether they died of cancer-related causes or not.

Characteristic	Death		*P* value
	No (n = 7734)	Yes (n = 268)	
Baseline data[a]			
Age, mean (SD), y	63.4 (8.5)	68.2 (8.1)	<.001
Male, No. (%)	3511 (45.4)	157 (58.6)	<.001
Black race, No. (%)	2444 (31.6)	68 (25.4)	.03
Region of residence, No. (%)			
No belt or buckle	3519 (45.5)	126 (47)	.71
Stroke buckle	1663 (21.5)	44 (16.4)	
Stroke belt	2552 (33)	98 (36.6)	
Educational level, No. (%)			
Less than high school	480 (6.2)	17 (6.3)	.50
High school graduate	1732 (22.4)	57 (21.3)	
Some college	2057 (26.6)	87 (32.5)	
College graduate	3465 (44.8)	107 (39.9)	
Current smoker, No. (%)	804 (10.4)	45 (16.8)	.001
Alcohol consumption, No. (%)			
None	4339 (56.1)	147 (54.7)	.20
Moderate	3032 (39.2)	100 (37.4)	
Heavy	356 (4.6)	21 (7.9)	

Q5.1

8 Points

Suppose the authors were interested in assessing whether there was a relationship between dying of cancer and educational level in this dataset. Describe a test that would be appropriate for answering this question. Specify the null and alternative hypotheses as well as the distribution of the test statistic under the null hypothesis. No need to perform the test yourself; given the p-value in the table, answer this question in the context of the original data.

Q5.2

7 Points

Suppose the authors wanted to evaluate whether there was any difference in cancer-related death associated with differences in baseline demographic characteristics (in particular, based on age, sex, race, region, educational level, smoking, and alcohol consumption patterns as provided in the table). If we assume that the individual tests they performed for each of these components were independent, what would their family-wise type 1 error rate be if each individual test was evaluated at the nominal α = 0.05 level?