MATH 32 – MIDTERM EXAM 2 – NOVEMBER 14, 2024

Instructions

- For the following questions, **fully explain your reasoning**. If you aren't sure how to complete a problem, explain what challenges you're running into. In all cases, your **reasoning is more important** than the final numerical result!
- No books, notes, calculators, phones, electronic devices, or outside assistance are allowed on the exam.
- There are four questions on the exam and several pages for extra work space. If you use an extra page, please **clearly mark** where your work is.

As a student at UC Merced, I will not give or receive any assistance on this exam. I will follow the academic integrity policies of this exam, the course, and the university.

Name: Aditi Srinivas

Signature:

Student ID: 228413345

Useful information

- Chebyshev's inequality says that for any distribution, the probability that an outcome is more than k standard deviations away from the mean is at most $1/k^2$:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}. \qquad P(|X-\mu| \geq a) \leq \frac{Var(x)}{a^2}$$

- The following table summarizes the critical z-scores for the normal distribution, corresponding to $P(Z \geq z_p) = p$:

Probability	0.1	0.05	0.025	0.01
z-Score	1.282	1.645	1.960	2.326
	80	90	95	99

- The following table summarizes the right critical values $t_{m,p}$ of the t-distribution with m degrees of freedom:

m	0.1	0.05	0.025	0.01
1	3.078	6.314	12.706	31.821
2	1.886	2.920	4.303	6.965
3	1.638	2.353	3.182	4.541
4	1.533	2.132	2.776	3.747
5	1.476	2.015	2.571	3.365

Extra Work Space

1. ($2+2+2+2=8$ points) Lauren is trying to determine whether UC Merced students prefer eating at the Pav or at a food truck. To do so, she asks 100 students which dining location they prefer. You may assume that each student prefers exactly one of the two options and that different students' preferences are independent.

(a) Explain how you can estimate the standard deviation for one student's preference.

Since there is only two options, you can say this is a Bernoulli distribution. For a Bernoulli distribution the variance is $p(1-p)$, and standard deviation is $\sqrt{\text{variance}}$. We can assume that ~~the~~ $p=0.5$, as someone has a 50% chance of choosing either the Pav or food truck. You would get a var of 0.25 so SD = 0.5.

(b) Lauren would like to set up a hypothesis test to determine which option is more popular. She thinks that students prefer both options equally. Formulate null and alternative hypotheses for this situation.

$H_0: \mu = 0.5$ μ = percentage of students who like the Pav

$H_a: \mu \neq 0.5$

(c) Complete the following statement, and briefly explain your answer.

If students prefer both options equally, then there is approximately a 95% chance that the number of surveyed students who prefer the Pav is between 40 and 60.

In a normal distribution, 95% is two standard deviations away from the mean. In this case the mean is 50. We can you z-score to find the upper and lower bound.

$$\frac{x-0.5}{\frac{0.5}{\sqrt{100}}} = 2 \qquad x-0.5 = 2(0.05) \qquad x = 0.1+0.5 \qquad x = 0.6 \qquad \frac{0.5}{\sqrt{100}} = \frac{0.5}{10} = 0.05$$

(d) In Lauren's survey, she finds that 70 students prefer going to the food trucks. Test her hypothesis from (b) using a 5% significance level.

$$z = \frac{0.7-0.5}{0.05} = \frac{0.2}{0.05} = 4$$

A z-score of 4 results in a p-value way less than 0.05, so we can reject the null. This means students do not prefer food trucks and the Pav equally.

2. $(3 + 5 = 8$ points)

(a) Brad is writing a calculus exam and is trying to come up with some integration problems. He writes down the probability density function f of a distribution and finds the following values for various integrals:

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

$$\int_{1}^{9} f(x)\,dx = 0.70$$

$$\int_{-\infty}^{\infty} x \cdot f(x)\,dx = 5 \;= E[x]$$

He also finds that the probability distribution he is working with has variance 4. Explain how you know at least one of his computed integrals is incorrect.

$Var(x) = E(x^2) - E(x)^2$

$4 = x - 25$

$x = 29$

$\int_{-\infty}^{\infty} x^2 \cdot f(x)\,dx = \int_{0}^{\infty} x^2 \cdot f(x)\,dx + \int_{-\infty}^{0} x^2 \cdot f(x)\,dx = 29$

(b) The power output of a solar panel is approximately 300 watts with a standard deviation of 100 watts.

(i) Use the Central Limit Theorem to estimate the probability that a solar farm consisting of 100 panels produces at least 32,000 watts of electricity. Is it more or less than 10%?

$E[x] = 300 \quad E[X_{100}] = 30000 \quad Var(x) = 10{,}000 \quad Var(X_{100}) = 1{,}000{,}000$

$z = \frac{32{,}000 - 30{,}000}{\sqrt{1{,}000{,}000}} = \frac{2000}{1000} = 2$

$P(\bar{X}_{100} \geq 32{,}000) = P(x \geq 320) = P(z > 2) \approx P(z > 1.96) = 0.025$

$\boxed{\text{It is less than } 10\%}$

(ii) Electra wants to know whether your estimate could apply to 100 solar panels on her roof. Are there any issues in using the CLT there?

No. CLT needs to be IID, and different parts of the roof could get different amounts of sunlight so it wouldn't be IID.

3. $(4+4=8$ points)

(a) Janet is studying a squirrel population in Yosemite Valley and wants to know how many acorns they store for the winter. She observes 10 squirrels and finds that the largest acorn stash has 200 acorns. Initially, she predicted that each squirrel stores between 100 and 300 acorns for the winter with each value being equally likely. Is her observation consistent with her initial assumption?

$300-200 = 100 \leftarrow 2 \; SD$

$50 \leftarrow 1 \; SD$

$$\frac{200-100}{50} = \frac{100}{50} = 2$$

2500

$$P(|A_n - 200| \geq 0.5) \leq \frac{1}{k^2} = \frac{1}{(0.01)^2} = 1000$$

$k(50) = 0.5$

$k = \frac{0.5}{50} = \frac{5}{500} = \frac{1}{100}$

$k = 0.01$

0.01
0.01

0.0001

$\boxed{\text{not consistent}}$

(b) Rocky is studying Janet's hypothesis. He sets up a rejection rule that makes sure a Type I error has a probability of only 5%. He's not sure if he can calculate the probability of a Type II error, though. Explain why this isn't possible.

4. $(2+3+3=8$ points) A scientist is trying to measure the concentration of a pollutant at different points along a river. She gathers 36 samples from various points along the river.

(a) How is the variance of the sampling average $\overline{X}_{36}$ related to the variance of the underlying distribution? Is it larger, smaller, or the same?

The variance is smaller. If you find the variance of an average sample, you divide the variance of the underlying distribution by the number of samples. In this case, $Var(\overline{X}_{36}) = \frac{Var(X)}{36}$.

(b) She knows that her measurement process has a standard deviation of 6 mg/mL and she believes that the average pollutant concentration in the river is 30 mg/mL. She wants to know how likely it is that her sample average will be above 36 mg/mL. Circle all of the probabilities which are possible answers to her question.

50% 10% 3% 1% 0%

(c) Her graduate student researcher measured 4 samples from the river and found that the average concentration was 28 mg/mL and the sample standard deviation was 3 mg/mL. He has a null hypothesis that the average pollutant concentration in the river is 30 mg/mL and an alternative hypothesis that it is not 30. If he uses his data to test the hypothesis at a 5% significance level, what conclusion can he draw?

$z = \frac{28-30}{3/\sqrt{4}} = \frac{2}{\frac{3}{2}} = 2\left(\frac{2}{3}\right) = \frac{4}{3} = 1.33$

$2P(z > 1.83) \approx 2P(z > 1.282) = 2(0.1) = 0.2$

$0.2 > 0.05$ so fail to reject null.

1.33 is between 1.645 and 1.282. 1.645 would result in a p-value of 0.1 and 1.282, 0.2, both of which are greater than our significance level.

Extra work space