HCMIU - Probability - Mid-term Test
Semester 2 - Year: 2021 - 2022 - Duration : 90 minutes
Date Modified : Monday, Sep. $22^{\text{nd}}$, 2025

**INSTRUCTION:**

- Each student is allowed one doubled-sized sheet of reference material (size A4 of similar). All other documents and electronic devices are forbidden, except scientific calculators.

---

**Question 1.** (20 points) Consider a Markov Chain $(X_n)_{n\geq 0}$ with the transition matrix

$$P = \begin{array}{c} \text{To} \\ \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.6 & 0 & 0.4 & 0 \\ 0 & 0.6 & 0 & 0.4 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \quad \text{From}$$

(a) Draw the diagram of this Markov chain.

(b) Compute $P\left(X_3 = 2 | X_0 = 1\right)$.

(c) Which state(s) is recurrent and transient? Of all recurrent classes, list them.

**Question 2.** (20 points) Determine the stationary distribution of the Markov chain with transition matrix

$$P = \begin{array}{c} \text{To} \\ \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0.2 & 0.4 & 0.4 \\ 0.3 & 0.2 & 0.5 \\ 0.2 & 0.7 & 0.1 \end{bmatrix} \end{array} \begin{array}{c} \\ \\ 1 \\ 2 \\ 3 \end{array} \quad \text{From}$$

**Question 3.** (10 points) The net worth (in billions of dollars) of a sample of the richest people in the United States is shown.

59 52 28 26 19 19 18 17 17 17

Find the mean, median, mode, variance, and standard deviation for the data.

**Question 4.** (10 points) The numbers of faculty at 32 randomly selected state-controlled colleges and universities with enrollment under 12,000 students are shown below.

$$\begin{bmatrix} 211 & 384 & 396 & 211 & 224 & 337 & 395 & 121 \\ 356 & 621 & 367 & 408 & 515 & 280 & 289 & 180 \\ 431 & 176 & 318 & 836 & 203 & 374 & 224 & 121 \\ 412 & 134 & 539 & 471 & 638 & 425 & 159 & 324 \end{bmatrix}$$

Estimate the mean number of faculty at all state-controlled colleges and universities with enrollment under 12,000 with 95% confidence. Assume $\sigma = 165.1$.

**Question 5.** (10 points) A pizza shop owner wishes to find the 99% confidence interval of the true mean cost of a large plain pizza. How large should the sample be if she wishes to be accurate to within \$0.12? A previous study showed that the standard deviation of the price was 0.26.

**Question 6.** (10 points) A random sample of 250 adults in a medium-size college town were surveyed, and it was found that 110 were regular voters. Estimate the true proportion of regular voters with 95% confidence.

**Question 7.** (10 points) A state executive claims that the average number of acres in western Pennsylvania state parks is less than 2000 acres. A random sample of five parks is selected, and the number of acres is shown.

959 1187 493 6249 541

At a level of significant 0.01, is there enough evidence to support the claim?

**Question 8.** (10 points) Daily weather observations for southwestern Pennsylvania for the first three weeks of January show daily high temperatures as follows: 55, 44, 51, 59, 62, 60, 46, 51, 37, 30, 46, 51, 53, 57, 57, 39, 28, 37, 35, and 28 degrees Fahrenheit. The normal standard deviation in high temperatures for this time period is usually no more than 8 degrees. A meteorologist believes that with the unusual trend in temperatures the standard deviation is greater. At $\alpha = 0.05$, can we conclude that the standard deviation is greater than 8 degrees?

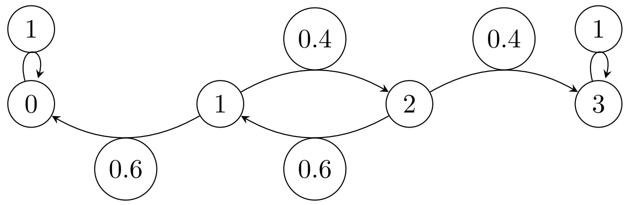
**Question 9.** (10 points) A product developer is interested in reducing the drying time of a primer paint. Two formulations of the paint are tested; formulation 1 is the standard chemistry, and formulation 2 has a new drying ingredient that should reduce the drying time. From experience, it is known that the standard deviation of drying time is 8 minutes, and this inherent variability should be unaffected by the addition of the new ingredient. Ten specimens are painted with formulation 1, and another 10 specimens are painted with formulation 2; the 20 specimens are painted in random order. The two sample average drying times are $\overline{x_1} = 121$ minutes and $\overline{x_2} = 112$ minutes respectively. What conclusions can the product developer draw about the effectiveness of the new ingredient (whether the new ingredient reduces the drying time) , using $\alpha = 0.05$?

---

***END***

**SUGGESTED ANSWER**

**Question 1.** If the probability changes over 3 stages, compute the Markov at 3rd transition, then identify.




From the graph, state 0 and 3 are recurrent - hence they form classes, whereas state 1 and 2 are transient. After 3 transition being made, it gradually reaches the steady state.

$$P^3 = P^2 \cdot P = \begin{array}{c} \text{To} \\ \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.244 & 0 & 0.096 & 0.16 \\ 0.36 & 0.144 & 0 & 0.496 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} \text{From} \implies \boxed{P(X_3 = 2|X_0 = 1) = r_{12}^{(3)} = 0.096}$$

**Question 2.** Linear system for stationary distribution of the Markov Chain, corresponding with transition matrix:

$$\begin{cases} \pi P = \pi \\ \sum_{i=1}^{n} \pi_i = 1 \end{cases} \longrightarrow \begin{cases} 0.2\pi_1 + 0.3\pi_2 + 0.2\pi_3 = \pi_1 \\ 0.4\pi_1 + 0.2\pi_2 + 0.7\pi_3 = \pi_2 \\ 0.4\pi_1 + 0.5\pi_2 + 0.1\pi_3 = \pi_3 \\ \pi_1 + \pi_2 + \pi_3 = 1 \end{cases} \implies \pi = \begin{bmatrix} \pi_1 & \pi_2 & \pi_3 \end{bmatrix} = \begin{bmatrix} \frac{37}{153} & \frac{64}{153} & \frac{52}{153} \end{bmatrix}$$

**Question 3.** Mean: takes the sum of all values, divided by the number of sample. Median is identified by the intermediate value (sorted sample). Mode is the frequency of number appearing in a sample test.

- Mean: 27.2
- Median: 19
- Mode: 17
- Variance: $S^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{N} \approx 233.49$
- Sample standard deviation: $\sigma = S = 15.28$

**Question 4.** 95% confident interval of mean member of faculty

$$\overline{x} \in (\overline{x} - \text{ME} + \overline{x} + ME) \approx \left(364.25 - 1.16 \cdot \frac{165.1}{\sqrt{32}}, 364.25 + 1.16 \cdot \frac{165.1}{\sqrt{32}}\right) \approx (330.39, 398.11)\,, \quad \text{ME} = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

**Question 5.**

$$\text{ME} = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \implies n = \left(z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\text{ME}}\right)^2 = \left(2.57 \cdot \frac{0.26}{0.12}\right)^2 \approx 31$$

In order to be accurate to within \$0.12, the sample should be as large as 31.

**Question 6.** 95% confident interval for true proportion

$$\hat{p} \in \left(p - z_{\frac{\alpha}{2}} \frac{p(1-p)}{n}, p + z_{\frac{\alpha}{2}} \frac{p(1-p)}{n}\right) \approx (0.378, 0.502)$$

**Question 7.** Let $\mu_0$ be the population average of acres in western Pennsylvania

- $\mu_0 \geq 2000$
- $\mu_0 < 2000$

t-test for unknown population standard deviation:

$$t_{\text{obs}} = \frac{\overline{x} - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{1885.8 - 2000}{\frac{2456.3}{\sqrt{5}}} \approx -0.10$$

Since $t_{obs} > -t_{n-1} = -3.747$, the average of acres is larger than 2000 (null hypothesis is not rejected).

**Question 8.** Chi-square test for standard deviation with the following hypotheses:

- $H_0 : \sigma \leq 8$
- $H_1 : \sigma > 8$

$$\chi_{\text{obs}} = \frac{(n-1)S^2}{\sigma^2} = \frac{19 \cdot 11.02^2}{8^2} \approx 36.05 > \chi^2_{\alpha^2, n-1} = 30.144$$

At significant level of 5%, the standard deviation is greater than 8 degrees (Null hypothesis is rejected)

**Question 9.** Let $\mu_1$ and $\mu_2$ be respectively means of drying time of formulation 1 and 2. State the hypotheses and apply two-sample z-score with mean and population standard deviation.

- $H_0 : \mu_1 - \mu_2 \leq 0$: The new ingredient does not reduces the drying time.
- $H_1 : \mu_1 - \mu_2 > 0$: The new ingredient does reduce the drying time

$$z_{\text{obs}} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \approx 1.41$$

At significant level of 5%, the z-score is approximately 1.64, and larger than the observational one. Hence, the null hypothesis is not rejected, which means there is no difference in the reduction of drying time of new ingredient.

---

***END***