Name (Last, First):
Kunjangada Arun Sushma

Student ID:
002473132

Assignment 8

NEU_COE_INFO6105_Fall2024

Instructions:

1. For answering **programming questions**, please use Adobe Acrobat to edit the pdf file in two steps **[See Appendix: Example Question and Answer]**:
    a. Copy and paste your R or python code as text in the box provided (so that your teaching team can run your code);
    b. Screenshot your R or python console outputs, save them as a .PNG image file, and paste/insert them in the box provided.
    c. Show all work - credit will not be given for code without showing the code in action by including the screenshot of R or python console outputs.
2. To answer **non-programming questions**, please type or handwrite your final answers clearly in the boxes. Show all work - credit will not be given for numerical solutions that appear without explanation in the space above the boxes. **You're encouraged to use R or python to graph/plot the data and produce numerical summaries; please append your code and screenshot of the outputs at the end of your pdf submission.**
3. **[Total 87 pts = 12 + 24 + 12 + 12 + 24 pts + 3 Extra Credit pts]**

Grading Rubric

Each question is worth 3 points and will be graded as follows:

3 points: Correct answer with work shown

2 points: Incorrect answer but attempt shows some understanding (work shown)

1 point: Incorrect answer but an attempt was made (work shown), or **correct answer without explanation (work not shown)**

0 points: Left blank or made little to no effort/work not shown

Reflective Journal [3 pts]

(Copy and paste the link to your live Google doc in the box below)

https://docs.google.com/document/d/1ptEhnYHniNtT1yxDPcvXK7LpJaPGzi80BCSGZZhom7Y/edit?usp=sharing

Part I. The Normal Distribution and Combining Normal Random Variables (4 x 3 = 12 pts)

1) Consider a set of 9000 scores on a national test that is known to be approximately normally distributed with a mean of 500 and a standard deviation of 90.

(a) What is the probability that a randomly selected student has a score greater than 600?
[Write your numerical answer in the box, and your work of deriving and calculating in the empty space]

Answer:

$\boxed{0.1335}$

Mean (μ) = 500 , X → represents the test score.

Standard deviation (σ) = 90

Calculate the z-score for a score of 600:

$$z = \frac{X-\mu}{\sigma} = \frac{600-500}{90} = \frac{100}{90} \simeq 1.11$$

From the z-score table, the area to the left of $z = 1.11$ is approximately 0.8665

$$P(x>600) = 1 - P(z \leq 1.11) = 1 - 0.8665 = 0.1335$$

So, the probability that a randomly selected student has a score greater than 600 is approximately 0.1335

(b) How many scores are there between 450 and 600? **[Write your numerical answer in the box, and your work of deriving and calculating in the empty space]**

Answer: 5209 Scores.

Mean $(\mu) = 500$
Standard deviation $(\sigma) = 90$
$X \rightarrow$ represents the test score

For 450:

$$z = \frac{450-500}{90} = \frac{-50}{90} = -0.56$$

For 600:

$$z = \frac{600-500}{90} = \frac{100}{90} = 1.11$$

For $z = -0.56$:- From the z-table, $P(z \leq -0.56) \simeq 0.2877$

For $z = 1.11$:- From The z-table, $P(z \leq 1.11) \simeq 0.8665$

Probability score between 450 and 600:

$$P(450 \leq X \leq 600) = P(z \leq 1.11) - P(z \leq -0.56)$$
$$= 0.8665 - 0.2877 = 0.5788$$

Number of scores between 450 and 600:

$$9000 \times 0.5788 \simeq 5209$$

(c) Megan needs to be in the top 1% of the scores on this test to qualify for a scholarship. What is the minimum score Megan needs? **[Write your numerical answer in the box, and your work of deriving and calculating in the empty space]**

Answer: 710

Megan needs to be in top 1% to qualify for a scholosship $\Rightarrow$ She needs a score in the 99th percentile

Mean $(\mu) = 500$, Standard deviation $(\sigma) = 90$

$$P(Z \leq z) = 0.99$$
$$z = 2.33$$

$$X = \mu + z \cdot \sigma = 500 + 2.33 \times 90$$
$$= 500 + 209.7$$
$$= 709.7$$
$$\simeq 710$$

So, the minimum score Megan needs to be in the top 1% is approximately 710

2) Gus and Cal go bowling every week. Gus's scores are normally distributed with a mean of 175 pins and a standard deviation of 30 pins. Cal's scores are normally distributed with a mean of 150 pins and a standard deviation of 40 pins. Assume that their scores in any given game are independent. Let G be Gus's score in a random game, C be Cal's score in a random game, and D be the difference between Gus's and Cal's scores where D = G – C. What is the probability that Gus will knock down more pins than Cal? **[Write your numerical answer in the box, and your work of deriving and calculating in the empty space]**

Answer: $\boxed{\simeq 69.15\%}$

Gus (G):

$\mu_G = 175$

$\sigma_G = 30$

Cal (C):

$\mu_C = 150$

$\sigma_C = 40$

Since $D = G - C$

Mean of D: $\mu_D = \mu_G - \mu_D = 175 - 150 = 25$

Variance of D: $\sigma_D^2 = \sigma_G^2 + \sigma_C^2 = 30^2 + 40^2$

$= 900 + 1600 = 2500$

Standard deviation D: $\sigma_D = \sqrt{2500} = 50$

Probability $D > 0$,

Z Score for $D = 0$

$$z = \frac{0 - 25}{50} = \frac{-25}{50} = -0.5$$

$z = -0.5$

From table, $P(z \leq -0.5) \simeq 0.3085$

$P(D > 0) = 1 - P(z \leq -0.5) = 1 - 0.3085$

$= 0.6915$

There is approximately 69.15% chance that Gus will knock down more pins than Cal.

Part II. Sampling Distribution of Sample Proportions (8 x 3 = 24 pts)

1) **(12 pts)** Suppose a large candy machine has 15% orange candies. Imagine taking an SRS of 25 candies from the machine and observing the sample proportion $\hat{p}$ of orange candies. [Write your numerical answer in the box, and your work of deriving and calculating in the empty space]

(a) What is the mean of the sampling distribution of $\hat{p}$? Why?

Population proportion $p = 0.15$
Sample size $n = 25$

Answer:

The mean of the sampling distribution of the sample proportion ($\hat{p}$) is equal to the population proportion (p).
In this case, the population proportion of orange candies is 0.15, so the mean of the sampling distribution of $\hat{p}$ is also 0.15

$\mu_{\hat{p}} = p = 0.15$

(b) Find the standard deviation of the sampling distribution of $\hat{p}$. Check to see if the 10% condition is met.

Standard deviation $= \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.15 \times (1-0.15)}{25}} = \sqrt{\frac{0.15 \times 0.85}{25}} \approx 0.07$

Answer:

$\sigma_{\hat{p}} \approx 0.07$

Sample size = 25

To meet 10% Requirement :- $n \leq 0.1 \times$ Population Size

Population Size $\geq 10 \times n = 10 \times 25 = 250$ Candies.

$\therefore$ The 10% condition is met as population size or the number of candies is atleast 250.

(c) Is the sampling distribution of $\hat{p}$ approximately Normal? Check to see if the Large Counts condition is met.

$n \cdot p \geq 10$ $\qquad n \cdot (1-p) \geq 10$ $\qquad$ (Population proportion $p = 0.15$)

Answer:

$n \cdot p = 25 \times 0.15 = 3.75$ $\qquad n \cdot (1-p) = 25 \times 0.85 = 21.25$

Since $n \cdot p = 3.75$ i.e. less than 10, the large count condition is not met.

$\therefore$ Sampling distribution of $\hat{p}$ is not approximately normal.

(d) If the sample size were 225 rather than 25, how would this change the sampling distribution of $\hat{p}$?

Population proportion $p = 225$.

Answer:

$\sigma_p = \sqrt{\frac{(1-p) \cdot p}{225}} = \sqrt{\frac{0.85 \times 0.15}{225}} \approx 0.0238$

$n \cdot p = 225 \times 0.15 = 33.75$

$n \cdot (1-p) = 225 \times 0.85 = 191.25$

Since both condition are now satisfied, the sampling distribution of $\hat{p}$ be approximately normal with smaller standard devation $\Rightarrow$ more precise.

2) **(12 pts)** The Harvard College Alcohol Study finds that 67% of college students support efforts to "crack down on underage drinking". Does this result hold at a large local college? To find out, college administrators surveyed an SRS of 100 students and found that 62 support a crackdown on underage drinking. What is the probability that the proportion in an SRS of 100 students is 0.62 or less? Does this refute the claim made by Harvard? **Explain using the 4 step process.**

Answer:

Population proportion $p = 0.67$

Sample size $n = 100$

Observed sample proportion $\hat{p} = 0.62$

Step 1: $\mu_{\hat{p}} = p = 0.67$

step 2: $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.67 \times 0.33}{100}} \approx 0.047$

Step 3: $z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}} = \frac{0.62 - 0.67}{0.047} \approx -1.06$

Step 4: Using z-table $P(z \leq -1.06)$ -1.06 or less is approximately 0.1446 or 14.46%

Since there is probability of 14.46% on getting sample proportion 0.62 or less we can conclude result does not strongly refute the Harward Claim.

Shape: Since both samples are sufficiently large ($n_N = 250$ & $n_S = 200$), the sampling distribution of the difference $\hat{p}_N - \hat{p}_S$ will be approximately normal, as condition of Large counts is met!

North High School: $n_N \cdot p_N = 250 \times 0.18 = 45 \geq 10$, $n p(1-p_N) = 250 \times 0.82 = 205 \geq 10$

South High School: $n_S p_S = 200 \times 0.22 = 44 \geq 10$, $n_S(1-p_S) = 200 \times 0.78 = 156 \geq 10$.

Part III. Sampling Distribution of a Difference in Sample Proportions (12 pts)

In a single town, there are two high schools: North and South (each high school has more than 2000 students in it). At North High School, the principal says that they have 18% of students arriving tardy each day. At South High School, the principal claims that they have 22% of students arriving tardy each day. The North principal plans to take an SRS of 250 students to see how many are tardy, and the South principal plans to take an SRS of 200 students to see how many are tardy. Let p_N represent the proportion of students arriving late at North and p_S represent the proportion of students arriving late a South.

$p_N = 0.18$, $n_N = 250$ $\quad p_S = 0.22$, $n_S = 200$

(a) Describe the shape, center, and spread of the sampling distribution of $\hat{p}_N - \hat{p}_S$.

Answer: Spread $= \sigma_{\hat{p}_N - \hat{p}_S} = \sqrt{\dfrac{p_N(1-p_N)}{n_N} + \dfrac{p_S(1-p_S)}{n_S}} = \sqrt{\dfrac{0.18 \times 0.82}{250} + \dfrac{0.22 \times 0.78}{200}}$

$= \sqrt{0.0005904 + 0.000858} = \sqrt{0.0014484} \simeq 0.038$

Spread $\simeq 0.038$

Center $= \mu_{\hat{p}_N - \hat{p}_S} = p_N - p_S = 0.18 - 0.22 = -0.04$

(b) The principals say that if the proportion of students tardy in South's sample is 10% or higher than the proportion of students tardy in North's sample, they will implement a new tardy policy. What is the probability of this happening? Use the sample distribution you created in part (a).

Answer:

$P(\hat{p}_S - \hat{p}_N) \geq 0.10)$

$$z = \frac{(\hat{p}_S - \hat{p}_N) - (\mu_{\hat{p}_S - \hat{p}_N})}{\sigma_{\hat{p}_S - \hat{p}_N}}$$

$$z = \frac{0.10 - (\cancel{-}\,0.04)}{0.038}$$

$$= \frac{0.06}{0.038} \simeq 1.58$$

$\mu_{\hat{p}_S - \hat{p}_N} = p_S - p_N = 0.22 - 0.18 = 0.04$

$\sigma_{\hat{p}_S - \hat{p}_N} = \sqrt{\dfrac{p_S(1-p_S)}{n_S} + \dfrac{p_N(1-p_N)}{n_N}}$

$= \sqrt{\dfrac{0.22 \times 0.78}{200}} + \sqrt{\dfrac{0.18 \times 0.82}{250}}$

$= \sqrt{0.000858} + \sqrt{0.0005904}$

$= \sqrt{0.0014484} \simeq 0.038$

$P(z \geq 1.58)$ from z table the area left of $z = 1.58$ is approximately 0.9429

$1 - 0.9429 = 0.0571$

The probability that South's sample proportion is 10% or more above North's sample proportion is approximately 0.0571 or about 5.71%.

Part IV. Sampling Distribution of Sample Means (2 x 6 = 12 pts)

1) (6 pts) The Wechsler Adult Intelligence Scale (WAIS) is a common "IQ test" for adults. The distribution of WAIS scores for people over 16 years of age is approximately normal with mean 100 and standard deviation 15. What is the probability that the average WAIS score of an SRS of 60 people is 105 or higher? [Write your numerical answer in the box, and your work of deriving and calculating in the empty space]

Answer:

$\mu = 100 \quad \sigma = 15 \quad n = 60$

SEM (Standard Error of mean) $= \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{60}} \simeq 1.9365$

$z = \frac{\text{Sample Mean} - \text{Population Mean}}{\text{SEM}} = \frac{105 - 100}{1.9365} \simeq 2.58$

From z table probability corresponding to $z = 2.58$ is about 0.995

$P(z \geq 2.58) = 1 - 0.995 = 0.005$

The probability that the average WAIS score of an SRS of 60 people is 105 or higher is approximately 0.005 or **0.5%**.

2) (6 pts) A company that owns and services a fleet of cars for its sales force has found that the service lifetime of disc brake pads varies from car to car according to a Normal distribution with mean $\mu = 55{,}000$ miles and standard deviation $\sigma = 4500$ miles. The company installs a new brand of brake pads on 8 cars. The average life on the pads on these 8 cars turns out to be 51,800 miles. What is the probability that the sample mean lifetime is 51,800 miles or less if the lifetime distribution is unchanged? (The company takes this probability as evidence that the average lifetime of the new brand of pads is less than 55,000 miles.) [Write your numerical answer in the box, and your work of deriving and calculating in the empty space]

Answer:

$\mu = 55{,}000$ miles $\quad \sigma = 4{,}500$ miles $\quad n = 8$ observed sample mean $= 51{,}800$ miles

$\text{SEM} = \frac{\sigma}{\sqrt{n}} = \frac{4500}{\sqrt{8}} \simeq 1590.99$

$z = \frac{\text{Sample Mean} - \text{Population mean}}{\text{SEM}} = \frac{51800 - 55000}{1590.99} = -2.01$

The probability corresponding to $z = -2.01$ is about 0.0222

The probability that sample mean lifetime is 51,800 miles or less is approximately 0.0222 or 2.22%.

Part V. Sampling Distribution of a Difference in Sample Means (24 pts)

1) **(18 pts)** For the following situations, match the distribution type with the corresponding situation. Each distribution will be used only once.

A. Normal Probability Distribution
B. Combining Normal Random Variables Distribution
C. Sampling Distribution for Sample Proportions
D. Sampling Distribution for a Difference in Sample Proportions
E. Sampling Distribution for Sample Means
F. Sampling Distribution for a Difference in Sample Means

Since it involves comparing the means of 2 groups.

F 1) A researcher is comparing the effectiveness of two different medications in reducing blood pressure. They randomly assign 50 patients to receive Medication A and 50 patients to receive Medication B. What is the probability that the difference in the mean reduction in blood pressure between the two groups is less than 5 mmHg?

Because total weight is the sum of weights of individual products.

B 2) In a factory, the weights of products produced follow a normal distribution with a mean of 500 grams and a standard deviation of 20 grams. If a package contains 10 of these products, what is the probability that the total weight of the package is less than 4950 grams?

Because it involves the mean height of a sample

E 3) In a study on the heights of sunflowers, it's found that the heights follow a normal distribution with a mean of 150 centimeters and a standard deviation of 20 centimeters. A researcher randomly selects a sample of 25 sunflowers from a field. What is the probability that the average height of the sample is greater than 155 centimeters

Since it involves a single value from a normally distributed variable

A 4) A company manufactures light bulbs with a mean lifespan of 800 hours and a standard deviation of 50 hours. What is the probability that a randomly selected light bulb will last more than 850 hours.

Since it involves comparing proportion between 2 Samples.

D 5) In a study comparing the effectiveness of two different advertising strategies, Strategy A and Strategy B, a marketing team wants to determine if there's a significant difference in the proportion of customers who make a purchase after seeing the ad. In a sample of 200 customers exposed to Strategy A, 40 customers made a purchase. In a sample of 250 customers exposed to Strategy B, 65 customers made a purchase. What is the probability that the difference in the proportion of customers who made a purchase between the two strategies is greater than 0.05?

As it focuses on the propotion of a single sample.

C 6) A survey was conducted to see the proportion of smartphone users who prefer iPhone. A research team randomly selects 800 smartphone users. Among them, 520 users prefer iPhone. What is the probability that in another random sample of 800 smartphone users, the proportion preferring iPhone is less than 0.6?

2) **(6 pts)** In a clinical trial investigating the effectiveness of a new medication for allergies in dogs, 60 dogs were randomly assigned to two groups: 30 dogs went to Group M, where they received the new medication, while 30 dogs went to Group P, where they received a placebo. After a month of treatment, the severity of allergy symptoms in each dog was assessed using a standardized scale. The mean score for Group M was 3.5 units with a standard deviation of 1.2 units, while the mean score for Group P was 4.2 units with a standard deviation of 1.5 units. A higher score means that more allergy symptoms were observed. What is the probability that the difference in the mean scores between the two groups (Group M – Group P) is greater than 0?

Answer:

Group M: $\mu_M = 3.5$, $\sigma_M = 1.2$, $n_M = 30$

Group P: $\mu_P = 4.2$, $\sigma_P = 1.5$, $n_P = 30$

Random: Both represent random sample of dogs

$$\mu_{\bar{x}_M} - \mu_{\bar{x}_P} = \mu_{x_M} - \mu_{x_P} = 3.5 - 4.2 = -0.7$$

Independent: $30 \leq 0.10$ (All doges recieving new medication) & $30 \leq 0.10$ (All dogs recieving a placebo)

$$\sigma_{\bar{x}_M} - \sigma_{\bar{x}_P} = \sqrt{\frac{\sigma_M^2}{n_M} + \frac{\sigma_{x_P}^2}{n_P}} = \sqrt{\frac{1.2^2}{30} + \frac{1.5^2}{30}} = \sqrt{\frac{1.44}{30} + \frac{2.25}{30}} = 0.3507135583 \approx 0.35$$

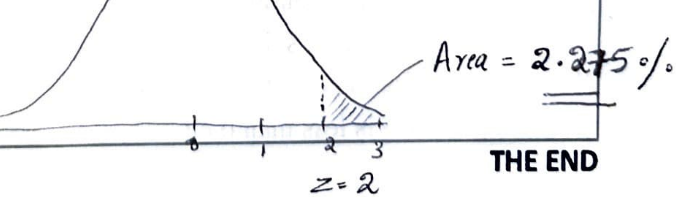
Normal: Both $30 \geq 30$ & $30 \geq 30$ CLT tells approximately normal.

$$P(\mu_{\bar{x}_M} - \mu_{\bar{x}_P} > 0) \text{ when } \mu_{\bar{x}_M} - \mu_{\bar{x}_P} \sim N(-0.7, 0.35)$$

$$= P\left(z > \frac{0 - (-0.7)}{0.35}\right) = 2$$

$$P(z > 2) \approx 0.02275 \qquad 1 - 0.97725$$

The probability that the difference mean score between Group M and Group P is greater than 0 is approximately 0.02275 or 2.275%.



THE END