**CBA: Practice Problem Set 2**

**Topics: Sampling Distributions and Central Limit Theorem**

1. Examine the following normal Quantile plots carefully. Which of these plots indicates that the data …
2. Are nearly normal?
3. Have a bimodal distribution? (One way to recognize a bimodal shape is a “gap” in the spacing of adjacent data values.)
4. Are skewed (i.e. not symmetric) ?
5. Have outliers on both sides of the center?



Ans.

**I. Nearly Normal**

* **Plot C** shows a straight line with most points closely following it, indicating that the data are nearly normal.

**II. Bimodal Distribution**

* **Plot B** shows a "gap" in the middle with points clustering at two regions, which is a characteristic of a bimodal distribution.

**III. Skewed (Not Symmetric)**

* **Plot A** shows a curve deviating from the straight diagonal line, suggesting skewness in the data (likely right-skewed).

**IV. Outliers on Both Sides of the Center**

* **Plot D** shows points deviating significantly from the main trend line on both ends, indicating outliers on both sides of the center.

1. For each of the following statements, indicate whether it is True/False. If false, explain why.

The manager of a warehouse monitors the volume of shipments made by the delivery team. The automated tracking system tracks every package as it moves through the facility. A sample of 25 packages is selected and weighed every day. Based on current contracts with customers, the weights should have *μ* = 22 lbs. and *σ* = 5 lbs.

1. Before using a normal model for the sampling distribution of the average package weights, the manager must confirm that weights of individual packages are normally distributed.
2. The standard error of the daily average SE() = 1.

Ans.

**1. "Before using a normal model for the sampling distribution of the average package weights, the manager must confirm that weights of individual packages are normally distributed."**

**Answer: False.**  
**Explanation:** The Central Limit Theorem states that the sampling distribution of the sample mean will be approximately normal regardless of the distribution of individual weights if the sample size is large enough (usually n ≥ 30). In this case, the sample size is 25, which is slightly below 30 but still acceptable if the population distribution is not severely non-normal (e.g., no extreme skewness or heavy tails). Hence, it is not necessary to confirm that individual weights are normally distributed.

**2. "The standard error of the daily average SE(x̄) = 1."**

**Answer: True.**  
**Explanation:** The formula for the standard error of the sample mean is:

SE(x̄) = σ / √n

Where:

* σ = 5 lbs (population standard deviation),
* n = 25 (sample size).

Substituting the values:

SE(x̄) = 5 / √25 = 5 / 5 = 1

Therefore, this statement is true.

1. Auditors at a small community bank randomly sample 100 withdrawal transactions made during the week at an ATM machine located near the bank’s main branch. Over the past 2 years, the average withdrawal amount has been $50 with a standard deviation of $40. Since audit investigations are typically expensive, the auditors decide to not initiate further investigations if the mean transaction amount of the sample is between $45 and $55. What is the probability that in any given week, there will be an investigation?
2. 1.25%
3. 2.5%
4. 10.55%
5. 21.1%
6. 50%

Ans.

**Determine the standard error (SE) of the sample mean**  
The formula for the standard error is:  
SE = σ / √n

Where:  
σ = 40 (population standard deviation)  
n = 100 (sample size)

Substitute the values:  
SE = 40 / √100 = 40 / 10 = 4

Thus, the standard error is **4**.

**Convert the sample mean limits to z-scores**  
The z-score formula is:  
z = (x̄ - μ) / SE

Where:  
x̄ = sample mean  
μ = 50 (population mean)  
SE = 4

For x̄ = 45:  
z = (45 - 50) / 4 = -5 / 4 = -1.25

For x̄ = 55:  
z = (55 - 50) / 4 = 5 / 4 = 1.25

Thus, the z-scores for the limits are **-1.25** and **1.25**.

**Find the probability between the z-scores**  
Using the standard normal distribution table:

* The cumulative probability for z = 1.25 is approximately 0.8944.
* The cumulative probability for z = -1.25 is approximately 0.1056.

The probability between z = -1.25 and z = 1.25 is:  
P(-1.25 ≤ z ≤ 1.25) = 0.8944 - 0.1056 = 0.7888

Thus, the probability that the sample mean is between $45 and $55 is **78.88%**.

**Find the probability of an investigation**  
The probability of an investigation occurs when the sample mean is **outside** the range of $45 to $55:  
P(Investigation) = 1 - P(-1.25 ≤ z ≤ 1.25)

P(Investigation) = 1 - 0.7888 = 0.2112

Thus, the probability of an investigation is **21.1%**.

1. The auditors from the above example would like to maintain the probability of investigation to 5%. Which of the following represents the minimum number transactions that they should sample if they do not want to change the thresholds of 45 and 55? Assume that the sample statistics remain unchanged.
2. 144
3. 150
4. 196
5. 250
6. Not enough information

Ans.

**Define the relationship between z-scores and sample size**  
The auditors want the sample mean to fall within the range $45 to $55, which corresponds to z-scores of -z and +z for a given confidence level. For a **5% probability of investigation**, 95% of the data must fall within the range, leaving 2.5% in each tail. From the z-table, the z-score for 95% confidence is approximately **1.96**.

The formula for the z-score is:  
z = (x̄ - μ) / SE

Rewriting the standard error (SE):  
SE = σ / √n

Substituting into the z-score formula:  
z = √n (x̄ - μ) / σ

For x̄ = 45 (lower threshold):  
1.96 = √n (50 - 45) / 40

Simplify:  
1.96 = √n × 5 / 40

√n = (1.96 × 40) / 5

√n = 78.4 / 5

√n = 15.68

**Square both sides and find n**  
n = 15.68²

n = 245.86

**Round up to the nearest whole number**  
Since n must be a whole number and larger sample sizes ensure the required confidence level, round **245.86** up to **250**.

1. An educational startup that helps MBA aspirants write their essays is targeting individuals who have taken GMAT in 2012 and have expressed interest in applying to FT top 20 b-schools. There are 40000 such individuals with an average GMAT score of 720 and a standard deviation of 120. The scores are distributed between 650 and 790 with a very long and thin tail towards the higher end resulting in substantial skewness. Which of the following is likely to be true for randomly chosen samples of aspirants?
2. The standard deviation of the scores within any sample will be 120.
3. The standard deviation of the mean of across several samples will be 120.
4. The mean score in any sample will be 720.
5. The average of the mean across several samples will be 720.
6. The standard deviation of the mean across several samples will be 0.60

Ans.

**Option A:**  
The standard deviation of the scores **within any sample** will not necessarily be **120** because it depends on the sample size and the distribution of the population. However, the **population standard deviation** remains 120. **This statement is incorrect.**

**Option B:**  
The **standard deviation of the mean** across several samples is given by the **standard error (SE):**  
SE = σ / √n

Since the sample size is not provided in the question, the statement that "the standard deviation of the mean across several samples will be 120" is incorrect. **This statement is false.**

**Option C:**  
The **mean score in any sample** will not always be exactly 720. The sample mean can vary around 720 depending on the sample. **This statement is incorrect.**

**Option D:**  
The **average of the means across several samples** is equal to the population mean (μ), which is 720. According to the **Central Limit Theorem**, the sampling distribution of the mean will have a mean equal to the population mean. **This statement is correct.**

**Option E:**  
The **standard deviation of the mean across several samples** (standard error) is given by:  
SE = σ / √n

Without knowing the sample size, we cannot calculate SE, but **0.60** is highly unlikely for a population standard deviation of **120**. **This statement is incorrect.**