**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?



**I. Normality:**

A normal quantile plot shows a straight line pattern when the data is nearly normal. Points lying close to a straight line indicate the data follows a normal distribution. Deviations from the line indicate departures from normality.

**II. Bimodal Distribution:**

A bimodal distribution typically shows two distinct peaks or modes in the data. On a normal quantile plot, this might appear as a non-linear pattern, possibly with a noticeable gap or shift in the alignment of points along the line.

**III. Skewness:**

Skewness, indicating asymmetry in the data distribution, can be observed on a normal quantile plot when the points deviate from a straight line in a systematic manner. If the points curve upwards or downwards consistently, it suggests skewness in the data.

**IV. Outliers:**

Outliers in a dataset are typically observed as points that significantly deviate from the expected linear pattern. They might appear as isolated points far away from the main cluster of data points on the normal quantile plot.

By observing these characteristics, let's match each type of normal quantile plot to the corresponding description:

I. **Nearly Normal**: A plot where the points closely follow a straight line indicates data that is nearly normal. The deviation from the line is minimal, showing a strong correlation between observed values and theoretical quantiles.

II. **Bimodal Distribution**: A plot with noticeable gaps or shifts in the alignment of points might suggest a bimodal distribution. The presence of two distinct groups or peaks in the data would lead to a deviation from the linear pattern.

III. **Skewed (Not Symmetric)**: A plot where the points systematically curve upwards or downwards, rather than following a straight line, indicates skewness in the data. This deviation from linearity implies an asymmetric distribution.

IV. **Outliers on Both Sides**: Outliers on both sides of the center are depicted as points significantly far away from the main cluster of data points on the normal quantile plot.

Matching these characteristics to the plots will help identify the specific nature of the data distribution in each case.

Top of Form

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.

(i) **True/False: 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.**

False. The Central Limit Theorem (CLT) states that the sampling distribution of the sample mean becomes approximately normal regardless of the underlying distribution of the population when the sample size is sufficiently large (typically >30). Therefore, for large sample sizes, the normality assumption for the individual package weights is less critical. However, if the sample size is small (less than 30) and there are indications that the individual package weights significantly deviate from a normal distribution, relying on the normal model for the sampling distribution might not be ideal.

(ii) **True/False: The standard error of the daily average SE(𝑥̅) = 1.**

False. The formula for the standard error of the sample mean (��(�ˉ)*SE*(*x*ˉ)) is given by��*n*​*σ*​, where �*σ* is the population standard deviation and �*n* is the sample size.

Given: Population standard deviation (�*σ*) = 5 lbs Sample size (�*n*) = 25

Calculating ��(�ˉ)=525=55=1*SE*(*x*ˉ)=25​5​=55​=1 lbs.

So, the statement is True. The standard error of the daily average (��(�ˉ)*SE*(*x*ˉ)) is indeed 1 lb based on the provided population standard deviation and sample size.

1. Top of Form
2. 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?
3. 1.25%
4. 2.5%
5. 10.55%
6. 21.1%
7. 50%

To find the probability of an investigation, let's first consider the range given by the auditors where they won't initiate further investigations: between $45 and $55.

Given: Mean (�*μ*) = $50 Standard deviation (�*σ*) = $40 Sample size (�*n*) = 100

We'll use the Central Limit Theorem (CLT) to approximate the sampling distribution of the sample mean. The mean of the sampling distribution will still be $50, but the standard deviation of the sampling distribution (��(�ˉ)*SE*(*x*ˉ)) will be ��*n*​*σ*​:

��(�ˉ)=��=40100=4*SE*(*x*ˉ)=*n*​*σ*​=100​40​=4

The auditors decide not to initiate investigations if the sample mean (�ˉ*x*ˉ) is between $45 and $55. To find the probability of this, we'll use the z-score formula:

�=Value−MeanStandard Error*Z*=Standard ErrorValue−Mean​

For $45: �45=45−504=−1.25*Z*45​=445−50​=−1.25

For $55: �55=55−504=1.25*Z*55​=455−50​=1.25

Now, we find the probability that the sample mean (�ˉ*x*ˉ) falls between $45 and $55 using the standard normal distribution table or calculator:

�(−1.25<�<1.25)≈�(�<1.25)−�(�<−1.25)*P*(−1.25<*Z*<1.25)≈*P*(*Z*<1.25)−*P*(*Z*<−1.25)

Looking up these values in the standard normal distribution table or using a calculator, you'll find that �(−1.25<�<1.25)≈0.8944*P*(−1.25<*Z*<1.25)≈0.8944.

So, the probability of not initiating an investigation, which is the complement of this probability, is 1−0.8944=0.10561−0.8944=0.1056 or 10.56%.

Therefore, the closest option is C. 10.55%.

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

To maintain a 5% probability of investigation while keeping the thresholds at $45 and $55, the auditors need to adjust the sample size.

We previously found that for a sample size of 100 transactions, the probability of investigation was approximately 10.55%. They want to reduce this probability to 5% while maintaining the same thresholds.

When working with the normal distribution and sample means, we can utilize the z-score formula to find the required sample size to achieve a specific probability.

Given: Previous sample size (�old*n*old​) = 100 Desired probability (�new*P*new​) = 5%

To find the new sample size, we'll use the formula for the standard error of the sample mean (��(�ˉ)*SE*(*x*ˉ)):

��(�ˉ)=��*SE*(*x*ˉ)=*n*​*σ*​

As the thresholds remain unchanged and the auditors want to achieve a 5% probability, the z-scores for these thresholds should remain the same.

From the standard normal distribution table or calculator: For $45 and $55 thresholds, the z-score is approximately 1.645 (which corresponds to a 5% probability on each tail of the distribution).

Now, using the formula for the standard error and setting the z-score equation with the new sample size:

�=55−50��(�ˉ)=1.645*z*=*SE*(*x*ˉ)55−50​=1.645

Solving for ��(�ˉ)*SE*(*x*ˉ): ��(�ˉ)=51.645≈3.04*SE*(*x*ˉ)=1.6455​≈3.04

3.04=40�new3.04=*n*new​​40​

Solving for �new*n*new​: �new=(403.04)2≈175.56*n*new​=(3.0440​)2≈175.56

Rounding up to the nearest whole number, the minimum number of transactions required for the auditors to maintain a 5% probability of investigation without changing the thresholds is 176.

None of the options provided match this calculated value. It seems there might be an oversight in the options given or in the calculations performed.

1. Top of Form
2. 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?
3. The standard deviation of the scores within any sample will be 120.
4. The standard deviation of the mean of across several samples will be 120.
5. The mean score in any sample will be 720.
6. The average of the mean across several samples will be 720.
7. The standard deviation of the mean across several samples will be 0.60

Let's break down each statement to determine its likelihood in randomly chosen samples from the population of GMAT aspirants:

A. **The standard deviation of the scores within any sample will be 120.** False. The standard deviation within a sample tends to decrease as the sample size increases. It will likely be less than the population standard deviation of 120 in any single sample.

B. **The standard deviation of the mean across several samples will be 120.** False. The standard deviation of the sample means, also known as the standard error of the mean (SEM), is calculated as ��*n*​*σ*​, where �*σ* is the population standard deviation and �*n* is the sample size. The SEM decreases as the sample size increases, so it will be less than 120 for larger sample sizes.

C. **The mean score in any sample will be 720.** Likely True. The sample mean tends to approximate the population mean. As the samples are randomly chosen from the population, the mean of each sample is expected to be close to the population mean of 720.

D. **The average of the mean across several samples will be 720.** Likely True. The concept of the Central Limit Theorem states that as more samples are taken, the average of their sample means will tend toward the population mean. Therefore, the average of the means across several samples is likely to be close to 720.

E. **The standard deviation of the mean across several samples will be 0.60.** False. The standard deviation of the sample means (SEM) is given by ��*n*​*σ*​, not 0.60. It will be influenced by the population standard deviation (�*σ*) and inversely proportional to the square root of the sample size (�*n*), but it won't be exactly 0.60 unless �*n* is known and specifically calculated.

In summary, statements C and D are likely to be true for randomly chosen samples of aspirants from the population described.

1. Top of Form