**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?

C

1. Have a bimodal distribution? (One way to recognize a bimodal shape is a “gap” in the spacing of adjacent data values.)

B and D.

1. Are skewed (i.e. not symmetric) ?

A, B and D are skewed.

1. Have outliers on both sides of the center?

A and B have outliers



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.

True. A normal model is a statistical model that assumes that data follows a normal distribution. A sampling distribution is the distribution of a test statistic corresponding to a particular dataset (here the mean of package weights) calculated from multiple samples.

The sample size of 25 is relatively small, and since the number of samples is unknown, assessing the normality of individual package weights is important for reliable statistical inferences. (from the central limit theorem)

1. The standard error of the daily average SE() = 1.

The daily average is the mean of the package weights, which according to context is 22lbs.

SE= standard deviation/sqrt(sample size) = 5/sqrt(25) = 1

True. The statement "The standard error of the daily average SE(x̅) = 1" is true based on the provided information.

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%

D. 21.13%

Sample size n =100

Population mean and standard deviation = $50 and $40.

SE= standard deviation/ sqrt(sample size) = 4

Let sample mean be X

P($45<X<$55) = P(Z1<Z<Z2)

Where, Z1= (45-50)/SE=-5/4 and Z2=(55-50)SE = 5/4

Therefore, P($45<X<$55)= P(-5/4<Z<5/4) = 0.7887

Hence, the probability that in any given week, there will be investigation = 1-0.7887= 21.13%

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

D. 250

Having the probability of investigation equal 5% means probability of sample mean falling in between 45 and 55 is 95%.

Therefore, P(z1<z<z2) =0.95.

=>Z2=1.96 and Z1=-1.96

=>(55-50)/(40/sqrt(n))=1.96 or (45-50)/(40/sqrt(n))=1.96

=> n=(1.96\*8)^2 = 245

Therefore, the sample size should be 245 to have the probability of 5%. Since they want to avoid investigation sample size of 250(or even higher) is preferable.

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

Option A isn’t always true as the standard deviation within the sample may vary, especially if the samples are small. The same is the case with Option C.

Options B and E are not true. The standard deviation of the mean across several samples will depend on the sample size and population standard deviation which is equal to std/sqrt(n).

Option D is true, as the law of large numbers suggests that as you take more and more random samples from the same population, the average of the sample means (across all samples) will tend to converge to the population mean of 720. However, individual sample means may vary.