**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?

Ans: C

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

Ans:B

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

Ans:A,C,D

1. Have outliers on both sides of the center?

Ans: A



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.

Ans: True as in this case, at least 30 sample packages must be selected and weighed everyday. Based on the central limit theorem, the sampling distribution of the sample mean approach normal distribution when the sample size become bigger (over 30).

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

Ans: True as SE = SD/sqrt(n) = 5/sqrt(25) = 1

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: given n = 100

Mu = $50 SD =$40

Sandard Error SE = SD/sqrt(n) = 40/sqrt(100) = 4

P(45<x<55) : in python probability can be calculated using stats.norm.cdf(x,mu,SE)

probability of the mean transaction amount of sample mean in b/w 45, 55 is: 0.7887

Probability of the mean transaction amount of sample mean not in b/w 45,55 is: 1- 0.7887

= 0.2113 = 21.13%

I.e probability of investigation is 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

Ans:

As we know mu = 50 SD = 40 , standard error changes when the no of transactions changes

Using trial and error method let us assume

1. n = 144, SE = 40/sqrt(144) = 3.3

P(45<x<55) = 87.03%

probability of investigation = 100- p(45<x<55) = 12.97 as it is greater than 5 n cannot be 144.

2 .n = 150, SE = 40/sqrt(144) = 3.3 (nearly)

P(45<x<55) = 87.03%

probability of investigation = 100- p(45<x<55) = 12.97 as it is greater than 5 n cannot be 150.

3. n = 196, SE = 40/sqrt(144) = 2.8

P(45<x<55) =92.59%

probability of investigation = 100- p(45<x<55) = 7.41 as it is greater than 5 n cannot be 150.

4.n = 250, SE = 40/sqrt(144) = 2.5

P(45<x<55) =95.45%

probability of investigation = 100- p(45<x<55) = 4.5% as it is less than and nearly 5 n we can perform 250 experiments to maintain probability of investigation to 5%

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?

Ans: given mu = 720, SD = 120, SEM = SD/sqrt(n) =0.6

1. The standard deviation of the scores within any sample will be 120.

Ans: False, Standard deviation of the scores within any sample will not be 120 as we don’t know the exact sample size

1. The standard deviation of the mean of across several samples will be 120.

Ans:False, SD of mean across several samples will not be 120, it will be probably around 0.6

1. The mean score in any sample will be 720.

Ans: False , It totally depends on the sample size and the data available in particular sample

1. The average of the mean across several samples will be 720.

Ans:It can be true but we should know the mean of all the samples within that sample size

1. The standard deviation of the mean across several samples will be 0.60

Ans:True - As the standard error is 0.6 and the standard error is known as the standard deviation of sample mean