**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:**

Option C is normal. This is because the points roughly form a straight line, meaning the two quantiles are from normal distribution.

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

**Ans:**

Option D , because the values are just 0 or 1 and hence there is a gap that is found in the QQ plot.

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

**Ans:**

Option A. the curve in QQ plot, indicates that the sample data we have is skewed.

1. Have outliers on both sides of the centre?

**Ans:**

Option B. the QQ plat has a straight line at the centre and curves towards the ends, showing that the sample data have more extreme values than expected, preferably the outliers are more.



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. At least an average of 30 packages must be checked every day. Based on central limit theorem, the sampling distribution of the sampling mean approaches a normal distribution if the sample size is larger than 30.

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

**Ans:**

Standard error = standard deviation (*σ)* / square root of sample size (n)

= 5/sqrt (25) => 1

The 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:**

Mean = 50

Standard error = standard deviation (*σ)* / square root of sample size (n)

Standard error = 40/sqrt (100) = 40/10 = 4

Sample (n) = 100 (given), which is greater than 30, thus following a normal distribution, according to central limit theorem.

P (45<x<55) – Probability of not having an investigation:

stats.norm.cdf(x, mean , *σ/(*[square root of (n -1)])

stats.norm.cdf (55, 50, 4)-stats.norm.cdf (45, 50, 4) = 0.788700

therefore probability of having an investigation = 1- P(45<x<55)

= 1- 0.788700

**= 0.211 ~ 21.1%**

**Option D**

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:**

Necessary Sample Size = (Z-score)² \* StdDev\*(1-StdDev) / (margin of error)²

**Option E**

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

**Answer:**

Sample = 40000

Mean = 720

Standard deviation = 120

Standard error = standard deviation / square root(n)

= 120/square root(40000) => 120/200 => 0.6

1. False. Standard deviation of any sample will not be 120 because we don’t know the size of the sample.
2. False. Standard deviation of the mean across several sample will also not be 120. It will be less, around 0.6
3. False, The mean could vary through different samples.
4. True. The average of the mean across several sample will be 720.
5. False.