**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 : Plot c and D are nearly normally distributed

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

Ans : D plot have a bimodal distribution

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

Ans : A and B plot are skewed data is not normally distributed

1. Have outliers on both sides of the center?

Ans : Plot c has outliers on both the sides of the centre



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

Ans : TRUE. 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 as the sample size become bigger (over 30).

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

Ans : TRUE.

Standard error equal to standard deviation divided by square root of sample size = 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 : mean = 50

sample = 100

s.d. = 40

mean transaction amount = 45 – 55

solution : if there is transaction below and above 45 and 55 it means

something is not under control in any given week there will be an

Investigation

To find probability that there is investigation means there is a transaction below 45 and above 55 .

Using T test formula for 45 = x-u/(sd./root n ) = 1.24

Using T test formula for 55 = x-u/(sd./root n ) = -1.24

from pandas and scipy in python

stats.t.cdf(1.24 , df=99) = p for 45 = 0.89

stats.t.cdf(-1.24 , df=99) = p for 55 = 0.1089

area betwwn the shaded portion is p (55) – p ( 45 ) = 0.89-0.11 = 0.78 means 78 %

To find probability that there is investigation means there is a transaction below 45 and above 55

Except the shaded area = 1- 0.78 = 0.22 = 22%~ 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 : Given : 1. probability of investigation to 5%. , 2. thresholds of 45 and 55

Solution :

from above example we know we get 5 % of probability of getting invested means

=( 1-5%)= 95% is the shaded area now it means transaction are in contol

By that we calculated for each sample and get 0.9750 for 45 and 0.0249 for 55 whose sample size is 250

Therefore : 0.9750-0.0249 = 0.9501 = 95% correct answer is 250 option D

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 D is correct

data is distributed with more information at the center and lesser information at the tail. This mean there is higher chance that average of mean of a aspirant that randomly chosen will be 720 that fall in between 650 and 790 at the center.