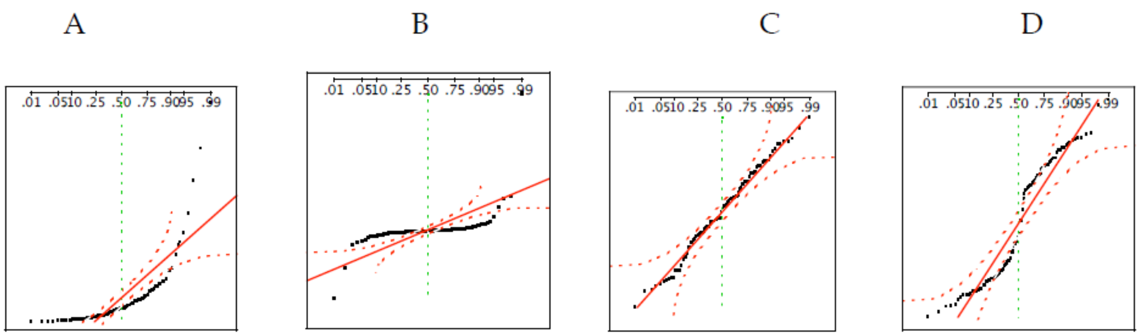
CBA: Practice Problem Set 2
Topics: Sampling Distributions and Central Limit Theorem

1. Examine the following normal Quantile plots carefully. Which of these plots indicates that the data …
   I. Are nearly normal?
   II. Have a bimodal distribution? (One way to recognize a bimodal shape is a "gap" in the spacing of adjacent data values.)
   III. Are skewed (i.e. not symmetric) ?
   IV. Have outliers on both sides of the center?



Ans:
   I. Option C follows the normal distribution.
   II. Option D follows the bimodal distribution.
   III. Option A follows the skewness.
   IV. Option B follows the outliers on both sides of the center.

2. 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 $\mu = 22$ lbs. and $\sigma = 5$ lbs.

   (i) 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.

   (ii) The standard error of the daily average $SE(\bar{x}) = 1$.

Ans:

   i. To use the normal model for the sampling distribution of the average package weights, it generally assumed that the individual weight are normally distributed in the population. This assumption allows us to apply statistical techniques based on the normal distribution.

ii. Standard Deviation =5

Sample (n)=25

Standard error=Standard deviation/sqrt(n)
=5/sqrt(25)
= 5/5
=1
Therefore, the S.E will be 1 on the daily average.

Option: 1 and 2 are true

3. 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?

   A. 1.25%
   B. 2.5%
   C. 10.55%
   D. 21.1%
   E. 50%

Ans:
Mean=50
Standard deviation=40
N=100
X1=45
X2=55

First, we will calculate the standard error of the sample mean:
S. E=Standard deviation/sqrt(N)
S. E=40/sqrt (100)
S. E=40/10
S. E=4

Now we will find the Z_score for X1 and X2:
Z_X1=X1-mean/S.E
=45-50/4

=-5/4
=-1.25

Z_X2=X2-mean/S.E
=55-50/4
=5/4
=1.25

Now we will find the probability for X1 and X2

From scipy import stats
(stats.norm.cdf(45,50,4))
0.105

From scipy import stats
Stats.norm.cdf(55,50,4)
0.8943

To find the probability between the 45 and 55:
From scipy import stats
Stats.norm.cdf(45,50,4)-stats.norm.cdf(55,50,4)
0.7887

To get the probability for the given random week:

```
1-(stats.norm.cdf(55,50,4)-stats.norm.cdf(45,50,4))
```

0.2112 or 21.12%

Option:4

4. 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.

   A. 144
   B. 150
   C. 196
   D. 250
   E. Not enough information

Ans:
Mean=50
Standard deviation=40

N=?
Z=1.645 for 90% confidence interval
X!=45 and 55

We need to find the sample mean for the intervals of
X<45 or X>55

For X<45

Z=X-mean/standard deviation/sqrt(N)
Z=45-50/40/sqrt(n)
Z=-5/40/sqrt(n)
1.645=-5/40/sqrt(n)
n=$(5/1.645)^2$ * 1600/25
n=196

Option: C

5. 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?

    A. The standard deviation of the scores within any sample will be 120.
    B. The standard deviation of the mean of across several samples will be 120.
    C. The mean score in any sample will be 720.
    D. The average of the mean across several samples will be 720.
    E. The standard deviation of the mean across several samples will be 0.60

Ans:

A. The statement is false, while the population of the standard deviation vary on the size taken, so hence the standard deviation of the scores within any sample will not be 120.

B. The statement is false, as the standard deviation of the mean is depending upon the standard deviation and the sample size of the data set, so the standard deviation may be less than 120 or not equal to 120.

C. The statement is true, the mean score of the any sample will be equal to 720.

D. The statement is true, as the average mean of the sample will be 720.

E. This statement is false, The standard deviation of the sample means across several samples depends on the population standard deviation and the sample size but is not equal to 0.60

Option: C & D are correct.