Name - Niranjan Nevase
Batch - 5th June Data Science

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 …

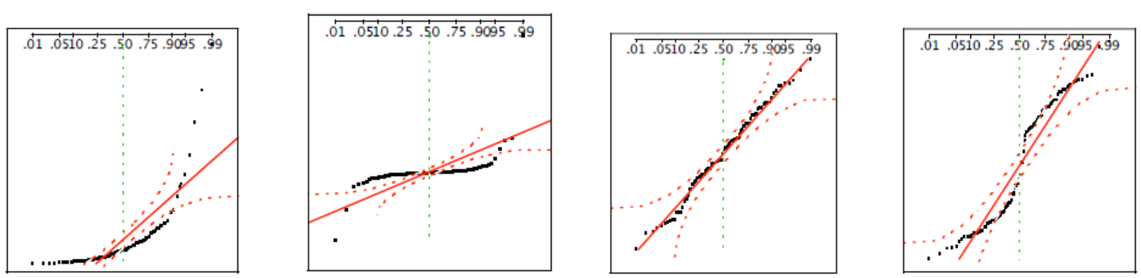
   I. Are nearly normal?
      - **C**

   II. Have a bimodal distribution? (One way to recognize a bimodal shape is a "gap" in the spacing of adjacent data values.)
      - **B & D**

   III. Are skewed (i.e. not symmetric) ?
      - **A, B & D**

   IV. Have outliers on both sides of the center?
      - **A & B**

A B C D



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 μ = 22 lbs. and σ = 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.
      - **False.** The manager does not need to confirm that the weights of individual packages are normally distributed before using a normal model for the sampling distribution of the average package weights. The Central Limit Theorem states that the sampling

distribution of the sample mean will be approximately normal for large sample sizes (n >= 30), regardless of the distribution of the population

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

- **True**. As SE(Standard Error) = sample standard deviation / Square root of (number of sample) SE = 5 / (25)^1/2 SE = 1

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 – D. 21.1%**

```
In [43]:
# qns no 03
# Z-score for the values range 45 to 55
Z1=(45-50)/40
Z2=(55-50)/40
print ("the Z-score range is :",Z1,"to",Z2 )

the Z-score range is : -0.125 to 0.125

In [44]:
#probability that
2*stats.norm.cdf(0.125,50,40)

Out[44]: 0.2124433347629998

In [41]:
#the probability that in any given week, there will be an investigation is
## 21.1%
```

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.

- 144
- 150
- 196
- 250
- Not enough information

- **Ans – D. 250**

```
In [52]:  #qns no 04
          print("the Z-score for 5% pprobability is",round(stats.norm.ppf(.025),3))

          the Z-score for 5% pprobability is -1.96
```

$Z = x - \mu / (\sigma / \sqrt{n})$
$-1.96 = 5 / (40/\sqrt{n})$
$\sqrt{n}=1.9*40/5=15.7$
N= 247 ≈ **250**

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 – E.**
     Standard error = sigma / (n)^0.5 = standard deviation / (sample size)^0.5
     = 120 / (40000)^0.5 = 0.6