

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.)

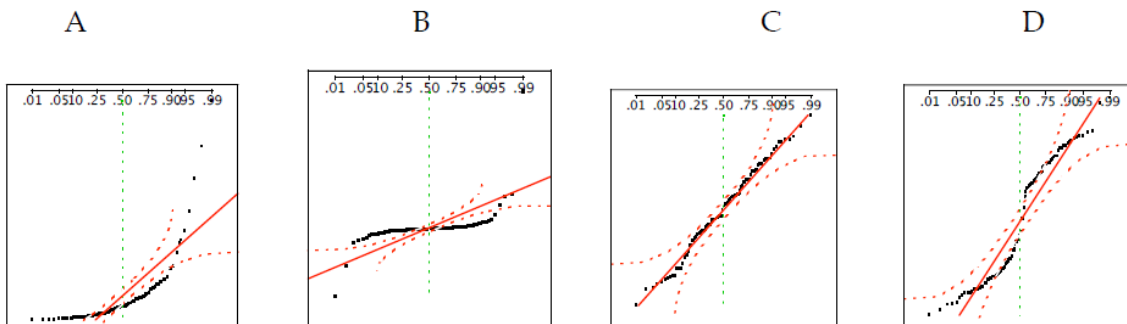
B and D

- III. Are skewed (i.e. not symmetric) ?

A ,B and D

- IV. Have outliers on both sides of the center?

A and B



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.

False

it is not necessary for the weights of individual packages to be normally distributed in order to use a normal model for the sampling distribution of the average package weights, as long as the sample size is large enough. In this case, with a sample size of 25 packages, the manager can reasonably assume normality of the sampling distribution of the sample mean, even if the weights of individual packages do not follow a normal distribution

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

True

Standard error is based on standard deviation and sample size

Standard error=standard deviation/sample_size0.5 = 5/25**0.5 = 1**

if either of these values change, the standard error will also change

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%

Option D is correct

P_mean=50

std=40

n=100

(45-P_mean)/(std/n0.5)**

-1.25

(55-P_mean)/(std/n0.5)**

1.25

from scipy import stats

stats.norm.cdf(1.25)-stats.norm.cdf(-1.25)

0.7887004526662893

***100**

(1-0.7887)*100

21.130000000000003

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

Option D is correct

P_mean=50

std=40

confidence=0.05

z=stats.norm.ppf(confidence/2)

z

-1.9599639845400545

n=((std*z)/(P_mean-45))2**

n

245.85336452442414

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.

True

The average of the sample means across several samples is expected to be equal to the population mean due to the Central Limit Theorem.

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

True

The reason is that the standard deviation of the mean across several samples, given a sample size of 40000, is 0.60

The standard deviation of the mean across several samples, known as the standard error of the mean = std/sample_size0.5**

=120/400000.5**

= 120/200

= 0.6