**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?
3. Have a bimodal distribution? (One way to recognize a bimodal shape is a “gap” in the spacing of adjacent data values.)
4. Are skewed (i.e. not symmetric) ?
5. Have outliers on both sides of the center?



**Answer:**

1. C
2. B & D
3. A, B, & D
4. A & B
5. 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.
2. The standard error of the daily average SE() = 1.

**Answer:**

1. The statement is False.

Because a sampling distribution is a probability distribution of a statistic obtained from a larger number of samples drawn from a specific population. In this case the samples contain 25 packages and the larger number of samples contain of each such 25 packages taken into different samples (25+25+25+25…and so on). The mean for one these samples is 22lbs and standard deviation of 5lbs which means each individual package is having a weight varying between + or – 5lbs with respect to mean(22lbs). Hence, it is invalid to take a weight of individual packages & confirm that it follows normal distribution before using a normal model for the sampling distribution. The Sample Central Limit Theorem states that the sampling distribution of the samples mean approaches normal distribution as the sample size is large enough.

1. The statement is True.

As Standard Error (SE) = sample standard deviation / square root of (number of sample)

SE = 5 / (25)^1/2

SE = 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%

**Answer:** D

import scipy.stats as stats

import math

import numpy as np

#given data

n = 100 #sample withdrawal

df = n - 1 #degree of freedom

sample\_std\_dev = 40 #sample standard deviation

pop\_mean = 50

#sample mean ampount lies between 45 & 55, therefore we will consider two sample mean

sample\_mean45 = 45

sample\_mean55 = 55

#We will go with t-score as population standard deviation is unknown

t\_score45 = ((sample\_mean45 - pop\_mean) / (sample\_std\_dev/math.sqrt(n)))

t\_score55 = ((sample\_mean55 - pop\_mean) / (sample\_std\_dev/math.sqrt(n)))

#finding the probability

p\_value45 = stats.t.cdf(t\_score45, df)

p\_value55 = stats.t.cdf(t\_score55, df)

prob = p\_value55 - p\_value45

prob #the value of P(45 < x < 55)

prob\_invs = 1 - prob #Subtracting 1 from prob value for finding the probability of investigation

prob\_invs

print('Probability of investigation is', np.round(prob\_invs \* 100,2) , '%')

O/p - Probability of investigation is 21.42 %

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

**Answer:** D

import scipy.stats as stats

import math

#given data

s\_mean = 45

pop\_mean = 50

sample\_std\_dev = 40

#This calculates the critical t-value for a two-tailed test with a 95% confidence level and 249 degrees of freedom.

t\_critical = stats.t.ppf(0.025, df = 249)

t\_critical

#Calculates the required sample size (n) using the formula for the t-test for a given margin of error (5), sample standard deviation, and critical t-value

n = (sample\_std\_dev \* (abs(t\_critical)) / 5) \*\* 2

df = n - 1 #Calculates the degrees of freedom for the t-distribution

print('The Auditors would like to maintain the probability of investigation to 5%, they should sample', round(n,2), 'transactions if they do not want to change the thresholds of 45 to 55')

print(n,df)

O/p -

The Auditors would like to maintain the probability of investigation to 5%, they should sample 248.26 transactions if they do not want to change the thresholds of 45 to 55

248.26083027145847 247.26083027145847

print("Confidence Interval:",(stats.t.interval(0.95, df=df, loc=pop\_mean, scale=sample\_std\_dev/math.sqrt(n))))

O/p -

Confidence Interval: (44.999828238713164, 55.000171761286836)

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:** D

The average of the mean across several samples will be 720