**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?**

Solution -  **C**

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

Solution - **D**

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

Solution - **B**

1. **Have outliers on both sides of the center?**

Solution - **A**



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 manager must confirm that weights of individual packages are normally distributed.**

**Answer:** The statement is false

Because the individual values do not need to be normally distributed as long as the sample size condition is satisfied

(n>10 |K4|)

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

**Answer:** The statement is true.

The standard error of the mean is given by the formula ​SE(X​)=σ/√n. For this​ example, this value is equal to

​SE(X​)=5/25

=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%**

**Solution –**

import numpy as np

import pandas as pd

import matplotlib.pyplot

import seaborn as sns

from scipy import stats

from scipy.stats import norm

# For No investigation P(45<X<55)

# For Investigation 1-P(45<X<55)

# find z-scores at x=45; z=(s\_mean-P\_mean)/(p\_SD/sqrt(n))

z=(45-50)/(40/100\*\*0.5)

z

=-1.25

# find z-scores at x=55; z=(s\_mean-P\_mean)/(p\_SD/sqrt(n))

z=(55-50)/(40/100\*\*0.5)

z

=1.25

# For No investigation P(45<X<55) using z\_scores = P(X<50)-P(X<45)

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

= 0.7887004526662893

stats.norm.interval(0.7887,loc=50,scale=40/(100\*\*0.5))

=(45.00000495667348, 54.99999504332652)

# For Investigation 1-P(45<X<55)

1-0.7887

=0.21130000000000004

**Answer: D. 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**

**Solution:**

z=(x-mean)/sigma/sqrt(n); z-test, because sd is given for the long term

=+/- 5/40/sqrt(100) =+/- 1.25

The probability of z between those values is 0.7887, so probability of an investigation is 1-0.7887, or

0.2113. -For 5%, z has to be +/-1.96

so 1.96=(5)\*sqrt(n)/40

sqrt(n)=15.68

n=245.86 or 246

qnorm(0.975) # 1.95996398454005

50 = 55 + 1.959964 \* (40/ √n) =245~approx. **250**

**Answer: 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?**

**Ans** :-

**The correct option is D, that is – The average of the mean across several samples will be 720.** Data is distributed with leptokurtic kurtosis mean more data at the center and lesser data 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.