# Topics: Descriptive Statistics and Probability

1. Look at the data given below. Plot the data, find the outliers and find out $\mu, \sigma, \sigma^2$

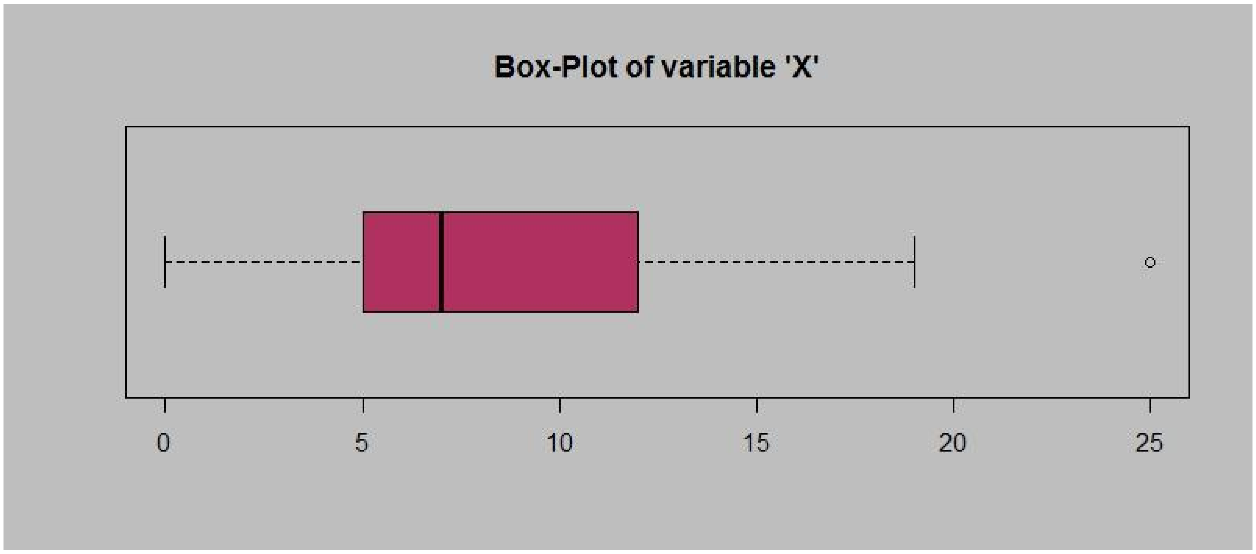
| Name of company | Measure X |
|---|---|
| Allied Signal | 24.23% |
| Bankers Trust | 25.53% |
| General Mills | 25.41% |
| ITT Industries | 24.14% |
| J.P.Morgan & Co. | 29.62% |
| Lehman Brothers | 28.25% |
| Marriott | 25.81% |
| MCI | 24.39% |
| Merrill Lynch | 40.26% |
| Microsoft | 32.95% |
| Morgan Stanley | 91.36% |
| Sun Microsystems | 25.99% |
| Travelers | 39.42% |
| US Airways | 26.71% |
| Warner-Lambert | 35.00% |

µ=33.27133333333333

α =287.1466123809524

=σ=16.945400921222028

2.




Answer the following three questions based on the box-plot above.

(i) What is inter-quartile range of this dataset? (please approximate the numbers) In one line, explain what this value implies.

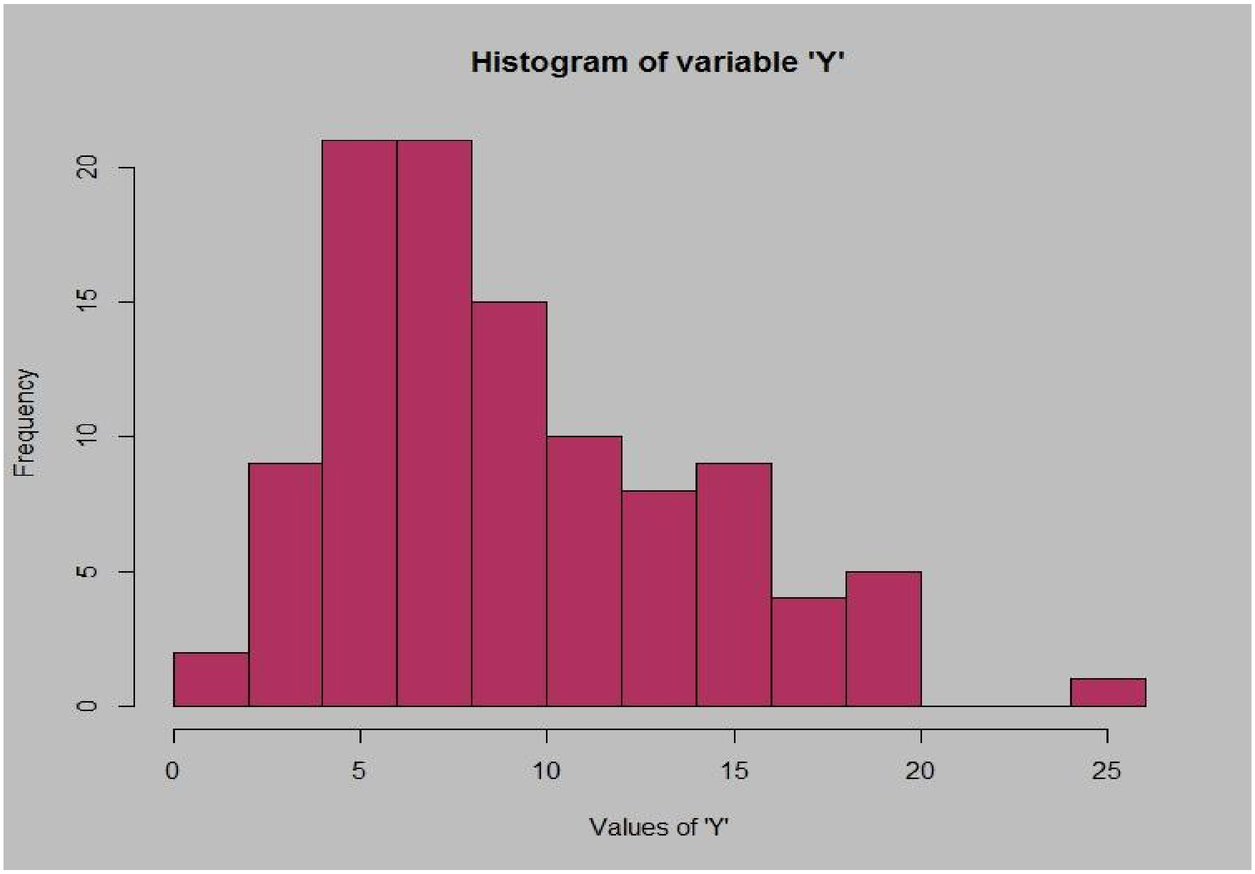
(ii) What can we say about the skewness of this dataset?

(iii) If it was found that the data point with the value 25 is actually 2.5, how would the new box-plot be affected?

**i)Answer= IQR = 12-5 =7, this represents the range which contains 50% of the data points**

**II)Answer= Right skewed**

**Ans= 2.5 will be not considered an outlier .The boxplot will start from 0 and send at 20 in representation.**




Answer the following three questions based on the histogram above.

(i) Where would the mode of this dataset lie?

(ii) Comment on the skewness of the dataset.

(iii) Suppose that the above histogram and the box-plot in question 2 are plotted for the same dataset. Explain how these graphs complement each other in providing information about any dataset.

**Answer: Median in boxplot and mode in Histogram**

**Histogram provides the frequency distribution so we can see how many time each data point is occurring however, boxplot provides the quintile distribution i.e.50%data lies between** %5

3. AT&T was running commercials in 1990 aimed at luring back customers who had switched to one of the other long-distance phone service providers. One such commercial shows a businessman trying to reach Phoenix and mistakenly getting Fiji, where a half-naked native on a beach responds incomprehensibly in Polynesian. When asked about this advertisement, AT&T admitted that the portrayed incident did not actually take place but added that this was an enactment of something that "could happen." Suppose that one in 200 long-distance telephone calls is misdirected. What is the probability that at least one in five attempted telephone calls reaches the wrong number? (Assume independence of attempts.)

**Answer: Median in boxplot and mode in Histogram**
**Histogram provides the frequency distribution so we can see how many time each data point is occurring however, boxplot provides the quintile distribution i.e.50%data lies between** %5

4. Returns on a certain business venture, to the nearest $1,000, are known to follow the following probability distribution

| x | P(x) |
|---|---|
| -2,000 | 0.1 |
| -1,000 | 0.1 |
| 0 | 0.2 |
| 1000 | 0.2 |
| 2000 | 0.3 |
| 3000 | 0.1 |

(i) What is the most likely monetary outcome of the business venture?
(ii) Is the venture likely to be successful? Explain
(iii) What is the long-term average earning of business ventures of this kind? Explain
(iv) What is the good measure of the risk involved in a venture of this kind? Compute this measure

I)Max, P= 0.3 for P(2000).So most likely outcome is 2000

II) (P(x>0)=0.6,implies there is a 60%chance that the venture would yield profits or greater than expected returns .P is only 0.2 so the venture is likely to be successful.

III)Weighted average= x*P(x)=800 this means the average expected earnings over a long period of the time would be 800

Iv)P(loss)=P(X=2000)+P(x=1000)0.2 so risk associated with this venture is 20%

## <u>Topics: Normal distribution, Functions of Random Variables</u>

1. The time required for servicing transmissions is normally distributed with $\mu$ = 45 minutes and $\sigma$ = 8 minutes. The service manager plans to have work begin on the transmission of a customer's car 10 minutes after the car is dropped off and the customer is told that the car will be ready within 1 hour from drop-off. What is the probability that the service manager cannot meet his commitment?

   A. 0.3875
   B. 0.2676
   C. 0.5
   D. 0.6987

**Answer: Since work being 10 min after the car is dropped, the time left to complete work 50 mins, Probability that Service Manager cannot meet his commitment=P(X>50)=1-Pr(x<=50)**

**Convert 50 to z-score**

**Standard normal variable Z=(X-μ)/σ=(x-45)/8**

**P(X<=50)=P(Z<=(50-45)/8)=PR(Z<=0.625) =0.73237= 73.237%**

**Probability that service manager will not meet his commmitment is 100-73.237=26.763%=0.2676**

**The answer B.**

2. The current age (in years) of 400 clerical employees at an insurance claims processing center is normally distributed with mean $\mu$ = 38 and Standard deviation $\sigma$=6. For each statement below, please specify True/False. If false, briefly explain why.
   A. More employees at the processing center are older than 44 than between 38 and 44.

      0.3413447460685429

      **True**

      **So, the statement" More employees at the processing center are older th an 44 than between 38 and 44"**

   B. A training program for employees under the age of 30 at the center would be expected to attract about 36 employees.

**True**

**The statement of "training program for employees under the age of 30 at the center would be expected to attract about 36 employees**"

36.484487890347154

3. If $X_1$ ~ $N(\mu, \sigma^2)$ and $X_2$ ~ $N(\mu, \sigma^2)$ are *iid* normal random variables, then what is the difference between 2 $X_1$ and $X_1 + X_2$? Discuss both their distributions and parameters.

   **Answer: Here X1 and X2 are two independent random variables then**

   **X1+X2 ~N(μ+μ,$\sigma^2$ +$\sigma^2$)andX1-X2~N(μ-μ,σ $^2\sigma^2$)**

   **2X1~N(2μ,2$\sigma^2$)**

   **2X1-(X1+X2) = N(2μ,2$\sigma^2$)-N(μ+μ,σ $^2$+$\sigma^2$)**

<u>**Topics: Normal distribution, Functions of Random Variables**</u>

4. The time required for servicing transmissions is normally distributed with $\mu$ = 45 minutes and $\sigma$ = 8 minutes. The service manager plans to have work begin on the transmission of a customer's car 10 minutes after the car is dropped off and the customer is told that the car will be ready within 1 hour from drop-off. What is the probability that the service manager cannot meet his commitment?

   E. 0.3875

F. 0.2676
G. 0.5
H. 0.6987

**Answer: Since work being 10 min after the car is dropped, the time left to complete work 50 mins, Probability that Service Manager cannot meet his commitment=P(X>50)=1-Pr(x<=50)**

**Convert 50 to z-score**

**Standard normal variable Z=(X-μ)/σ=(x-45)/8**

**P(X<=50)=P(Z<=(50-45)/8)=PR(Z<=0.625) =0.73237= 73.237%**

**Probability that service manager will not meet his commmitment is 100-73.237=26.763%=0.2676**

**The answer B.**

5. The current age (in years) of 400 clerical employees at an insurance claims processing center is normally distributed with mean $\mu$ = 38 and Standard deviation $\sigma$=6. For each statement below, please specify True/False. If false, briefly explain why.
   C. More employees at the processing center are older than 44 than between 38 and 44.

   `0.3413447460685429`

   **True**

   **So, the statement" More employees at the processing center are older th an 44 than between 38 and 44"**

   D. A training program for employees under the age of 30 at the center would be expected to attract about 36 employees.

**True**

**The statement of "training program for employees under the age of 30 at the center would be expected to attract about 36 employees**"

`36.484487890347154`

6. If $X_1 \sim N(\mu, \sigma^2)$ and $X_2 \sim N(\mu, \sigma^2)$ are *iid* normal random variables, then what is the difference between 2 $X_1$ and $X_1 + X_2$? Discuss both their distributions and parameters.

   **Answer: Here X1 and X2 are two independent random variables then**

   **X1+X2 ~N(μ+μ,σ$^2$ +σ$^2$)andX1-X2~N(μ-μ,σ $^2$σ$^2$)**

   **2X1~N(2μ,2σ$^2$)**

   **2X1-(X1+X2) = N(2μ,2σ$^2$)-N(μ+μ,σ $^2$+σ$^2$)**

## <u>Topics: Normal distribution, Functions of Random Variables</u>

7. The time required for servicing transmissions is normally distributed with $\mu$ = 45 minutes and $\sigma$ = 8 minutes. The service manager plans to have work begin on the transmission of a customer's car 10

minutes after the car is dropped off and the customer is told that the car will be ready within 1 hour from drop-off. What is the probability that the service manager cannot meet his commitment?

I. 0.3875
J. 0.2676
K. 0.5
L. 0.6987

**Answer: Since work being 10 min after the car is dropped, the time left to complete work 50 mins, Probability that Service Manager cannot meet his commitment=P(X>50)=1-Pr(x<=50)**

**Convert 50 to z-score**

**Standard normal variable Z=(X-μ)/σ=(x-45)/8**

**P(X<=50)=P(Z<=(50-45)/8)=PR(Z<=0.625) =0.73237= 73.237%**

**Probability that service manager will not meet his commmitment is 100-73.237=26.763%=0.2676**

**The answer B.**

8. The current age (in years) of 400 clerical employees at an insurance claims processing center is normally distributed with mean $\mu = 38$ and Standard deviation $\sigma=6$. For each statement below, please specify True/False. If false, briefly explain why.

   E. More employees at the processing center are older than 44 than between 38 and 44.

   `0.3413447460685429`

   **True**

   **So, the statement" More employees at the processing center are older th an 44 than between 38 and 44"**

   F. A training program for employees under the age of 30 at the center would be expected to attract about 36 employees.

**True**

**The statement of "training program for employees under the age of 30 at the center would be expected to attract about 36 employees**"

`36.484487890347154`

9. If $X_1 \sim N(\mu, \sigma^2)$ and $X_2 \sim N(\mu, \sigma^2)$ are *iid* normal random variables, then what is the difference between 2 $X_1$ and $X_1 + X_2$? Discuss both their distributions and parameters.

   **Answer: Here X1 and X2 are two independent random variables then**

   **X1+X2 ~N(μ+μ,$\sigma^2$ +$\sigma^2$)andX1-X2~N(μ-μ,σ $^2\sigma^2$)**

   **2X1~N(2μ,2$\sigma^2$)**

   **2X1-(X1+X2) = N(2μ,2$\sigma^2$)-N(μ+μ,σ $^2$+$\sigma^2$)**

10. Let X ~ N(100, $20^2$). Find two values, *a* and *b*, symmetric about the mean, such that the probability of the random variable taking a value between them is 0.99.

   A. 90.5, 105.9
   B. 80.2, 119.8
   C. 22, 78
   D. 48.5, 151.5
   E. 90.1, 109.9

**Solution: The probability of getting value between a &b is 0.99**

**So, the probability of getting value outside a &b is 1-0.99=0.01**

**The probability towards left of a =-0.01/2=-0.05**

**The probability towards right of b= 0.01/2=0.05**

**By finding Standard Normal Variable(z), need to calculate X.**

**$Z*\sigma+\mu=x$**

**-(-2.57)*20+100=151.4**

**(-2.57)*20+100=48.6**

**D is the correct answer**

11. Consider a company that has two different divisions. The annual profits from the two divisions are independent and have distributions $Profit_1$ ~ N (5, $3^2$) and $Profit_2$ ~ N(7, $4^2$) respectively. Both the profits are in $ Million. Answer the following questions about the total profit of the company in Rupees. Assume that $1 = Rs. 45
   A. Specify a Rupee range (centered on the mean) such that it contains 95% probability for the annual profit of the company.
   B. Specify the $5^{th}$ percentile of profit (in Rupees) for the company

Which of the two divisions has a larger probability of making a loss in a given year?

**Ans: Profit1+profit2~N(5+7,3 $^2$+$4^2$)= Profit~N(12,5)**

**A) Range = 12-19.6×5,12+ 1.96 ×5**

**=$2.2, $22.8**

**=Rs.99, Rs.1026**

**B) P(Z<=(p-12)/5)=0.05**

**p-12/5= -1.644**

**p=12 – 8.22 = $3. 78 = rs.170.1**

**C) When profit is less than 0 then loss**

**p-12/5= -1.644**

**p= 12-8.22 - $3.78 = Rs.170.1**

**C)When profit is less 0 then loss**

**D) The first division of company, thus have large probability of making Loss in a given year.**

**Topics: Confidence Intervals**

1. For each of the following statements, indicate whether it is True/False. If false, explain why.

    I. The sample size of the survey should at least be a fixed percentage of the population size in order to produce representative results.

**Answer : False**

**Reason: A sample size of 30 is considered large enough, but that may or may not be adequate**

    II. The sampling frame is a list of every item that appears in a survey sample, including those that did not respond to questions.

**Answer: True**

**Reason: The population is generic and the sampling frame is a specific list of all items in the population.**

    III. Larger surveys convey a more accurate impression of the population than smaller surveys.

**Answer: True**

2. *PC Magazine* asked all of its readers to participate in a survey of their satisfaction with different brands of electronics. In the 2004 survey, which was included in an issue of the magazine that year, more than 9000 readers rated the products on a scale from 1 to 10. The magazine reported that the average rating assigned by 225 readers to a Kodak compact digital camera was 7.5. For this product, identify the following:

12. Let X ~ N(100, $20^2$). Find two values, *a* and *b*, symmetric about the mean, such that the probability of the random variable taking a value between them is 0.99.

F. 90.5, 105.9
G. 80.2, 119.8
H. 22, 78
I. 48.5, 151.5
J. 90.1, 109.9

**Solution: The probability of getting value between a &b is 0.99**
**So, the probability of getting value outside a &b is 1-0.99=0.01**
**The probability towards left of a =-0.01/2=-0.05**
**The probability towards right of b= 0.01/2=0.05**
**By finding Standard Normal Variable(z), need to calculate X.**
**Z*σ+μ=x**
**-(-2.57)*20+100=151.4**
**(-2.57)*20+100=48.6**
**D is the correct answer**

13. Consider a company that has two different divisions. The annual profits from the two divisions are independent and have distributions $Profit_1 \sim N(5, 3^2)$ and $Profit_2 \sim N(7, 4^2)$ respectively. Both the profits are in $ Million. Answer the following questions about the total profit of the company in Rupees. Assume that $1 = Rs. 45
    C. Specify a Rupee range (centered on the mean) such that it contains 95% probability for the annual profit of the company.
    D. Specify the 5th percentile of profit (in Rupees) for the company

Which of the two divisions has a larger probability of making a loss in a given year?

**Ans: Profit1+profit2~N(5+7,3 $^2$+4$^2$)= Profit~N(12,5)**

**B) Range = 12-19.6×5,12+ 1.96 ×5**

**=$2.2, $22.8**

**=Rs.99, Rs.1026**

**B) P(Z<=(p-12)/5)=0.05**

**p-12/5= -1.644**

**p=12 – 8.22 = $3. 78 = rs.170.1**

**C) When profit is less than 0 then loss**

**p-12/5= -1.644**

**p= 12-8.22 - $3.78 = Rs.170.1**

**C)When profit is less 0 then loss**

**D) The first division of company, thus have large probability of making Loss in a given year.**

**Topics: Confidence Intervals**

3. For each of the following statements, indicate whether it is True/False. If false, explain why.

IV. The sample size of the survey should at least be a fixed percentage of the population size in order to produce representative results.

**Answer : False**

**Reason: A sample size of 30 is considered large enough, but that may or may not be adequate**

V. The sampling frame is a list of every item that appears in a survey sample, including those that did not respond to questions.

**Answer: True**

**Reason: The population is generic and the sampling frame is a specific list of all items in the population.**

VI. Larger surveys convey a more accurate impression of the population than smaller surveys.

**Answer: True**

4. *PC Magazine* asked all of its readers to participate in a survey of their satisfaction with different brands of electronics. In the 2004 survey, which was included in an issue of the magazine that year, more than 9000 readers rated the products on a scale from 1 to 10. The magazine reported that the average rating assigned by 225 readers to a Kodak compact digital camera was 7.5. For this product, identify the following:

A. The population

**Ans : All the readers of PC manager**

B. The sampling frame

**Ans : Readers that rated the products(around 9000)**

C. The sample size

**Ans :225**

D. The sampling design

E. Any potential sources of bias or other problems with the survey or sample

**Ans: Selection of the readers, selection of the issue which will contain the survey**

5. For each of the following statements, indicate whether it is True/False. If false, explain why.

I. If the 95% confidence interval for the average purchase of customers at a department store is \$50 to \$110, then \$100 is a plausible value for the population mean at this level of confidence.

**True**

**Reason**: **The 95% confidence interval for the average purchase of customers at a department store is 550 is \$110.Which means that there is a 95%chance that the population mean will fall between \$50 and \$110 .Hence ,as \$100 falls between \$50 and \$110,it is a value for the population means at this confidence level**.

II. If the 95% confidence interval for the number of moviegoers who purchase concessions is 30% to 45%, this means that fewer than half of all moviegoers purchase concessions..

**True**

**Reason: The 95%confidence interval for the number of moviegoers who purchase concession is 30%to 45%,this means that there is 95%chance that only 30 to 45 % of moviegoers purchase**

III. The 95% Confidence-Interval for $\mu$ only applies if the sample data are nearly normally distributed.

**True**

6. What are the chances that $\overline{X} > \mu$?

   A. ¼
   B. ½
   C. ¾
   D. 1

**Answer : 1**

7. In January 2005, a company that monitors Internet traffic (WebSideStory) reported that its sampling revealed that the Mozilla Firefox browser launched in 2004 had grabbed a 4.6% share of the market.

   I. If the sample were based on 2,000 users, could Microsoft conclude that Mozilla has a less than 5% share of the market?
   **NO**

   II. WebSideStory claims that its sample includes all the daily Internet users. If that's the case, then can Microsoft conclude that Mozilla has a less than 5% share of the market?
   **YES**

8. A book publisher monitors the size of shipments of its textbooks to university bookstores. For a sample of texts used at various schools, the 95% confidence interval for the size of the shipment was 250 ± 45 books. Which, if any, of the following interpretations of this interval are correct?

   A. All shipments are between 205 and 295 books.

INCORRECT

   B. 95% of shipments are between 205 and 295 books.

CORRECT

C. The procedure that produced this interval generates ranges that hold the population mean for 95% of samples.

CORRECT

D. If we get another sample, then we can be 95% sure that the mean of this second sample is between 205 and 295.

CORRECT

E. We can be 95% confident that the range 160 to 340 holds the population mean.

INCORRECT

9. Which is shorter: a 95% *z*-interval or a 95% *t*-interval for $\mu$ if we know that $\sigma = s$?

   A. The z-interval is shorter
   B. The t-interval is shorter
   C. Both are equal
   D. We cannot say

**A. The z-interval is shorter**

Questions 8 and 9 are based on the following: To prepare a report on the economy, analysts need to estimate the percentage of businesses that plan to hire additional employees in the next 60 days.

10. How many randomly selected employers (minimum number) must we contact in order to guarantee a margin of error of no more than 4% (at 95% confidence)?

   A. 600
   B. 400
   C. 550
   D. 1000

**Ans .z=1-α/2 =1-0.05/2=0.025**

**Z=1.96**

**M=0.04**

**M=z√0.5(1-0.5)n/n**

**0.04=1.96√0.5(1-0.5)/n**

**0.04√n=1.96 ×0.5**

**$\sqrt{n}$=1.96×0.05/0.4**

**$\sqrt{n}$=0.060025**

**So we select randomly 600 employees approximately**.

11. Suppose we want the above margin of error to be based on a 98% confidence level. What sample size (minimum) must we now use?

    A. 1000
    B. 757
    C. 848
    D. 543

**Ans. z=1-σ/2 =1-0.02/2=0.99**

**Z=2.327**

**M=0.05**

**M=z$\sqrt{0.5(1-0.5)/n}$**

**0.05=2.327$\sqrt{0.5(1-0.50/n}$**

**0.05$\sqrt{n}$=2.327×0.05**

**$\sqrt{n}$=2.327×0.5/0.05**

**$\sqrt{n}$=23.27**

**Squaring on both side**

**N=541.49**

**So we now use sample size 541**

(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.

**The standard error of the daily average SE(x) True**

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

**A sampling distribution is a probability distribution that describes how statics, such as the mean, varies True**

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?

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

**Answer: n=100,σ= 40,μ = 50**

**P(no. investigation)P(45<X<55)**

**P(investigation)1-P(45<X<55)**

**Z-score at X=45**

**Z=x-μ /σ/)√n**

**=45 – 50/40/)√100**

**=-1.25**

**Z-score at x=55**

**Z=x-μ /σ/)√n**

**Z-score at X=55**

**Z =x-μ/σ/√n**

**=55-50/40/)√100**

**1.25**

**P(45<X<55)**

**Stats. Norm .cdf (1.25)-stats .norm . cdf**

2. 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 =For 5% probability z value has to be 1.96**

**Z=x-μ/√n σ/**

**1.96=5X σ√n /40**

**√n=15.68**

**N=245.86**

3. 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= Data is distribution with leptokurtic kurtosis mean more information at the center and lesser**
**Information at the tail .This mean there is higher change that average of mean of a aspirant that randomly chosen will be 720 that fall in between 650 and 790 at the center**.

# <u>CBA: Practice Problem Set 2</u>

Questions referred to from *Aczel A., Sounderpandian J., Complete Business Statistics (7ed.)*

Assignment

# 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?

**Ans : C**

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

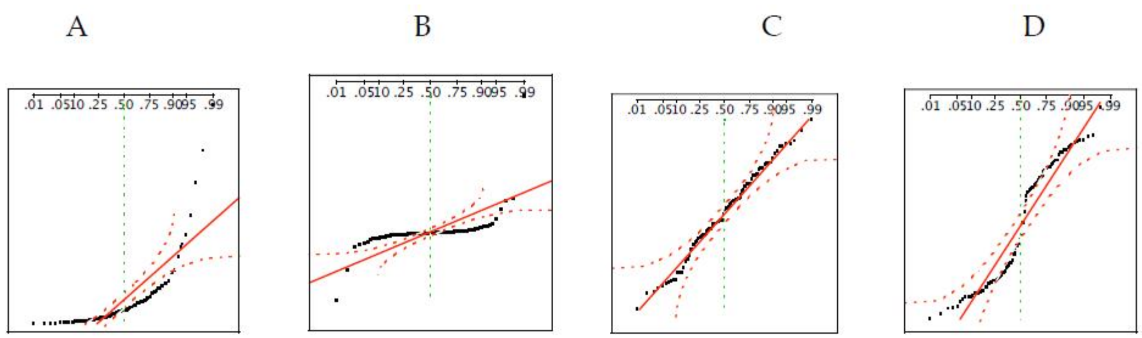
**Ans: B**

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

**Ans: A, C, D**

   IV. Have outliers on both sides of the center?

**Ans: A**




2)For each of the following statements, indicate whether it is <u>True/False</u>. 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.

2 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.

**The standard error of the daily average SE(x) True**

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

**A sampling distribution is a probability distribution that describes how statics, such as the mean, varies 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?

   F. 1.25%
   G. 2.5%
   H. 10.55%
   I. 21.1%
   J. 50%

**Answer: n=100,σ= 40,μ = 50**

**P(no. investigation)P(45<X<55)**

**P(investigation)1-P(45<X<55)**

**Z-score at X=45**

**Z=x-μ /σ/)√n**

**=45 – 50/40/)√100**

**=-1.25**

**Z-score at x=55**

**Z=x-μ /σ/)√n**

**Z-score at X=55**

**Z =x-μ/σ/√n**

**=55-50/40/)√100**

**1.25**

**P(45<X<55)**

**Stats. Norm .cdf (1.25)-stats .norm . cdf**

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.

F. 144
G. 150
H. 196
I. 250
J. Not enough information

**Ans =For 5% probability z value has to be 1.96**

**Z=x-μ/√n σ/**

**1.96=5X σ√n /40**

**√n=15.68**

**N=245.86**

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?

F. The standard deviation of the scores within any sample will be 120.
G. The standard deviation of the mean of across several samples will be 120.
H. The mean score in any sample will be 720.
I. The average of the mean across several samples will be 720.
J. The standard deviation of the mean across several samples will be 0.60

**Ans= Data is distribution with leptokurtic kurtosis mean more information at the center and lesser**
**Information at the tail .This mean there is higher change that average of mean of a aspirant that randomly chosen will be 720 that fall in between 650 and 790 at the center**.