**ASSIGNMENT-2**

1. Look at the data given below. Plot the data, find the outliers and find out

|  |  |
| --- | --- |
| **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% |

A1)

OUTLIERS- MORGAN STANLEY-91.36%

|  |  |  |
| --- | --- | --- |
| MEAN |  | 0.332713333 |
| STDEVIATION |  | 0.169454009 |
| VARIANCE |  | 0.028714661 |

IQ

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

1. What is inter-quartile range of this dataset? (please approximate the numbers) In one line, explain what this value implies.
2. What can we say about the skewness of this dataset?
3. If it was found that the data point with the value 25 is actually 2.5, how would the new box-plot be affected?

**A2)**

(I) IQR=Q3-Q1

12-5 IQR

=7

IQR RANGES FROM 5 TO 12

this represents the range which contains 50% of the data points

(II)

IT IS RIGHT SKEWED

(III)

IF VALUE 25 DOES NOT EXIT THEN OUTLIER 25 DATA POINT DOES NOT EXIT

3)



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?
2. Comment on the skewness of the dataset.
3. 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.

A3)

(i)MODE LIES BETWEEN 4 &8

(ii)

MOSTLY RIGHT SKEWED

(iii)

BOX PLOT TELLS US MAX VALUE ,MIN ,Q1,Q3 AND MEDIAN BUT NO MODE

HISTOGRAM TELLS US MODE BUT NOT OUTLIERS

TWO PLOTS COMPLEMENT EACH OTHER IT IS VERY USEFUL

4.

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.)

A4)

One wrong number out of 200

Probability of wrong number: P(WN) = 1/200 = 0.005

Probability of not wrong number: 1 - P(WN) =1- 1/200 = 0.995

Probability of at least one out of five is a wrong number= 1 – Probability that all five calls are not wrong numbers

= 1 – (1 – P(WN))^5= 1 – (1- 0.005)^5= 1 – 0.975= 0.024= 2.5%

5.

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 |

1. What is the most likely monetary outcome of the business venture?
2. Is the venture likely to be successful? Explain
3. What is the long-term average earning of business ventures of this kind? Explain
4. What is the good measure of the risk involved in a venture of this kind? Compute this measure

A5)

(i) What is the most likely monetary outcome of the business venture?

ANS: Max. P = 0.3 for P(2000). So most likely outcome is 2000

(ii) Is the venture likely to be successful? Explain

ANS: P(x>0) = 0.6, implies there is a 60% chance that the venture would yield profits or greater

than expected returns. P(Incurring losses) is only 0.2. So the venture is likely to be successful.

(iii) What is the long-term average earning of business ventures of this kind? Explain

ANS: Weighted average = x\*P(x) = 800. This means the average expected earnings over a long

period of time would be 800(including all losses and gains over the period o f time)

(iv) What is the good measure of the risk involved in a venture of this kind? Compute this

measure

ANS: P(loss) = P(x= -2000)+P(x=-1000)=0.2. So the risk associated with this venture is 20%.

**SET-2**

1. The time required for servicing transmissions is normally distributed with *μ* = 45 minutes and *σ* = 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?
2. 0.3875
3. 0.2676
4. 0.5
5. 0.6987

A1)

We have a normal distribution with MEW = 45 and SIGMA = 8.0.

Let X be the amount of time it takes to complete the repair on a customer's car.

To finish in one hour you must have X <= 50 so the first question is to find

Pr(X > 50).

Pr(X > 50) = 1 - Pr(X <= 50).

Z = (X - mu)/sigma = (X - 45)/8.0

Pr(X ≤ 50) = Pr(Z ≤ (50 - 45)/8.0) = Pr(Z ≤ 0.625)=73.4%

PROBABILITY THAT THE SERVICE MANAGER WILL NOT MEET HIS DEMANDS

= 100-73.4 = 26.6% or 0.2676

1. The current age (in years) of 400 clerical employees at an insurance claims processing center is normally distributed with mean *μ* = 38 and Standard deviation *σ* =6. For each statement below, please specify True/False. If false, briefly explain why.
2. More employees at the processing center are older than 44 than between 38 and 44.
3. A training program for employees under the age of 30 at the center would be expected to attract about 36 employees.

A2)

|  |
| --- |
|  |
|  | a)Probabilty of employees greater than age of 44= Pr(X>44) |
|  | Pr(X > 44) = 1 - Pr(X ≤ 44). |
|  | Z = (X -MEW )/ SIGMA = (X - 38)/6 |
|  | Thus the question can be answered by using the normal table to find |
|  | Pr(X ≤ 44) = Pr(Z ≤ (44 - 38)/6) = Pr(Z ≤ 1)=84.1345% |
|  | Probabilty that the employee will be greater than age of 44 = 100-84.1345=15.86% |
|  | So the probability of number of employees between 38-44 years of age = Pr(X<44)-0.5=84.1345-0.5= 34.1345% |
|  | Therefore the statement that “More employees at the processing center are older than 44 than between 38 and 44” is TRUE. |
|  | b) Probabilty of employees less than age of 30 = Pr(X<30). |
|  | Z = (X - MEW )/SIGMA = (30 - 38)/6 |
|  | Thus the question can be answered by using the normal table to find |
|  | Pr(X ≤ 30) = Pr(Z ≤ (30 - 38)/6) = Pr(Z ≤ -1.333)=9.12% |
|  | So the number of employees with probability 0.912 of them being under age 30 = 0.0912\*400=36.48( or 36 employees). |
|  | Therefore the statement B of the question is also TRUE. |

1. If *X1* ~ *N*(μ, σ2) and *X*2 ~ *N*(μ, σ2) 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.

A3)

LET X1=1,2,3,4,5 X2=5,4,3,2,1

2X1=2,4,6,8,10

X1+X2=6,6,6,6,6

MEAN WILL BE SAME BUT VARIANCE DIFFERERS

1. Let X ~ N(100, 202). 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.
2. 90.5, 105.9
3. 80.2, 119.8
4. 22, 78
5. 48.5, 151.5
6. 90.1, 109.9

A4)

MEAN=100

VARIANCE=20

XBAR+-Z99%(VARIANCE)

=100+2.58(20) , 100-2.58(20)

=151.5,48.5

D IS CORRECT

1. Consider a company that has two different divisions. The annual profits from the two divisions are independent and have distributions Profit1 ~ N(5, 32) and Profit2 ~ N(7, 42) 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
2. Specify a Rupee range (centered on the mean) such that it contains 95% probability for the annual profit of the company.
3. 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

A5)

X1 MEAN=5,VARIANCE=3

X2 MEAN=7 VARIANCE=4

IN RUPEES 45\*(5+7)=540MILLION RUPEES

ST DEVIANCE=45\*SQRT(9+16)=225

(540,225^2)

A)

Z(95)=1.96

XBAR+-Z95%(SIGMA)

540+-1.96(225)=540+-441

( 99 ,981)

B)

Z(90%)=1.64

540+-1.64(225)

540+-369

(909,171)

171 IS 5TH PERCENTILE

WHICH DIVISION IS UNDER LOSS(X=0)

X-MEAN/STD

0-5/3

=-1.667

0-7/4

=-1.75

LEAST VALUE-1.75

LOSS MORE IN SECOND DIVISION

**SET-3**

1. For each of the following statements, indicate whether it is True/False. If false, explain why.
2. The sample size of the survey should at least be a fixed percentage of the population size in order to produce representative results.
3. The sampling frame is a list of every item that appears in a survey sample, including those that did not respond to questions.
4. Larger surveys convey a more accurate impression of the population than smaller surveys.

A1)

1. False(including those things which will respond to our questions)
2. TRUE
3. *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:
4. The population
5. The parameter of interest
6. The sampling frame
7. The sample size
8. The sampling design
9. Any potential sources of bias or other problems with the survey or sample

A2)

The population

Ans) p=x/n=225/9000=0.025

The parameter of interest

Ans) sample size,average,scale

The sampling frame

Ans)9000

The sample size

Ans 225

The sampling design

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

**SET-4**

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?



**A1)**

1. C
2. D
3. A
4. 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.

A2)

I)false

As sampling distribution will be normal when distribution in data in population is normal and sampling distribution can be normal even if population is normal if sample size is big or fairly large

II)

Standard error=sigma/sqrt(n)=5/sqrt(25)=5/5=1

So 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?
2. 1.25%
3. 2.5%
4. 10.55%
5. 21.1%
6. 50%

A3)

Mew=50

Sigma=40

N=100

P(45<=x<=55)=

Stderror=40/sqrt(100)

=4

Z=x-mew/sigma=45-50/40=-5/4=-1.25

Z value=55-50/40=5/4=1.25

p(-1.25<=x<=1.25)=0.8944-0.1056=0.788

1-0.788

=0.211

21.1%

D is correct

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

A4)

P(95%)=1.96

-1.96 to +1.96

Z=x-mew/sigmaof x

1.96=55-50/sigma

Sigma=5/1.96

=2.55

Sigma=s/sqrt(n)

2.55=40/sqrt(n)

N=40^2/2.55^2

=246 approx

D is correct

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

A5)

Mean(Mew) =720

Sigma=120

Scores between 650-790

According to centrel limit theorem

Population is normal then sample distribution will also normal

Average of Mean will across sample will also be same

D is correct

Average of mean across several sample will be 720