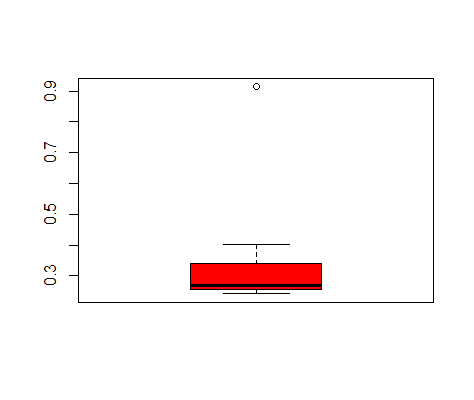
**Topics: Descriptive Statistics and Probability**

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% |

Ans:- µ = 33.27% σ = 0.169454 σ2 = 0.0287



Outlier is 91.36% as per box plot.



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.

IQR = 12-5 = 7, it means data distributed around from 5 to 12 with median 7.

1. What can we say about the skewness of this dataset?

Skewed right means positively skewed.

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

25th percentile or Q1 starts with 2.5 and move towards left and outlier will not be there and it becomes symmetric.

1. 

Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Mode lies in two number 5 and 7.

1. Comment on the skewness of the dataset.

Positively skewed as moves towards right side.

1. 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.

The boxplot showed the median, Q1 and Q3 and outliers but histograms shows the mode or frequency.

1. 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.)

1/2\*5/200 = 1/80

1. 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?

2000

1. Is the venture likely to be successful? Explain

After plotting histogram gradual increase is seen till 2000 but after that drastically decreases.

1. What is the long-term average earning of business ventures of this kind? Explain

1/6

1. What is the good measure of the risk involved in a venture of this kind? Compute this measure

Risk involved is standard deviation which is 1870.

**Topics: Normal distribution, Functions of Random Variables**

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

Ans:- P(X >50) = 1 – P(X<=50) = 1- 0.7323 = 0.2677.

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.

Ans:- True, because probability of former is higher than latter.

P(X>=44) = 0.841 and P(38<=X<=44) = 0.341

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

Ans:- P(X<=30) = 0.912, so number of employees with probability of 0.0912 is 0.0912\*400 = 36.48 which is 36. Hence above statement is 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.

Ans:- As we know that if X ∼ N(µ1, σ1^2 ), and Y ∼ N(µ2, σ2^2 ) are two independent random variables then X + Y ∼ N(µ1 + µ2, σ1^2 + σ2^2 ) , and X − Y ∼ N(µ1 − µ2, σ1^2 + σ2^2 ) .

Similarly if Z = aX + bY , where X and Y are as defined above, i.e Z is linear combination of X and Y , then Z ∼ N(aµ1 + bµ2, a^2σ1^2 + b^2σ2^2 ).

Therefore according to the question

2X1~ N(2 u,4 σ^2) and

X1+X2 ~ N(µ + µ, σ^2 + σ^2 ) ~ N(2 u, 2σ^2 )

2X1-(X1+X2) = N( 4µ,6 σ^2)

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

Ans:- Here we have to work in reverse order as following:

Probability of random value if taken outside of a and b is 1-0.99 = 0.01

Probability from left towards a is -0.005 and from right towards b is +0.005.

Then value of X is given by=

X = Z\*σ + µ , Z value of probability 0.005 from Z table is -2.57

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

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

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
4. Which of the two divisions has a larger probability of making a 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.
2. The sample size of the survey should at least be a fixed percentage of the population size in order to produce representative results.

TRUE

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

FALSE, it should only contain those who responds

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

TRUE

1. *PC Magazine* asked all of its readers to participate in a survey of their satisfaction with different brands of electronics. In the, 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:
2. The population = 9000
3. The parameter of interest = 7.5
4. The sampling frame = 2004 survey
5. The sample size = 225
6. The sampling design = simple random sampling
7. Any potential sources of bias or other problems with the survey or sample = bias is underestimating the population very less sample size is taken.
8. For each of the following statements, indicate whether it is True/False. If false, explain why.
9. 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

1. 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

1. The 95% Confidence-Interval for *μ* only applies if the sample data are nearly normally distributed.

FALSE, when µ applies it follows t-distribution not normal distribution.

1. What are the chances that ?
2. ¼
3. ½
4. ¾
5. 1

Ans:- option A. 1/4

1. 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.
2. If the sample were based on 2,000 users, could Microsoft conclude that Mozilla has a less than 5% share of the market?

Ans:- NO as sample size is very small it is possible that Mozilla may acquire more than 5%. Hence above statement can’t be concluded.

1. 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?

Ans:- Here sample size is fairly large than the previous case so it true to conclude that % share is less than 5.

1. 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?
2. All shipments are between 205 and 295 books.
3. 95% of shipments are between 205 and 295 books.
4. The procedure that produced this interval generates ranges that hold the population mean for 95% of samples.
5. If we get another sample, then we can be 95% sure that the mean of this second sample is between 205 and 295.
6. We can be 95% confident that the range 160 to 340 holds the population mean.

Ans:- option B ,C and D are correct.

1. Which is shorter: a 95% *z*-interval or a 95% *t*-interval for *μ* if we know that σ =s?
2. The z-interval is shorter
3. The t-interval is shorter
4. Both are equal
5. We cannot say

Ans:- option D , we cannot say as t-value also depends on sample size(n)

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.

1. 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)?
2. 600
3. 400
4. 550
5. 1000

Soln:- margin of error = 5%= 0.05

C= 95% hence Z-value is 1.96

For unknown sample size, n = [(Z2 \*0.25)/E2] = [(1.962\*0.25)/0.052] = 600.25.

Therefore minimum number is 600.

1. Suppose we want the above margin of error to be based on a 98% confidence level. What sample size (minimum) must we now use?
2. 1000
3. 757
4. 848
5. 543

Soln:- only Z-value changes to 2.33 for 98% CI so the minimum value of sample size is 848