Topics: Descriptive Statistics and Probability

1. **Look at the data given below. Plot the data, find the outliers, and find out** μ, σ, σ^2

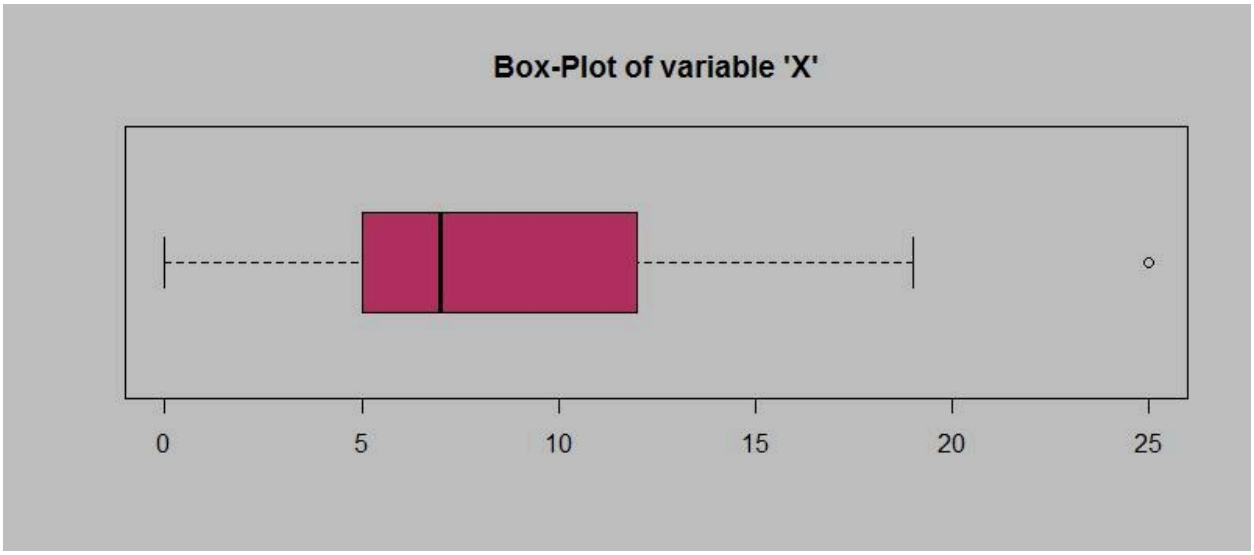
Name of company	Measure X
Allied Signal	24.23%
Bankers Trust	25.53%
General Mills	25.41%
ITT Industries	24.14%
J.P.Morgan & Co.	29.62%
Lehman Brothers	28.25%
Marriott	25.81%
MCI	24.39%
Merrill Lynch	40.26%
Microsoft	32.95%
Morgan Stanley	91.36%
Sun Microsystems	25.99%
Travelers	39.42%
US Airways	26.71%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

Ans: 1) Morgan Stanley is the outlier in the given data.

2) Mean 33.271333
Variance 287.146612
Standard Deviation 16.945401

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

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.

Ans: The IQR is about 7. IQR gives a spread of the 50% data in the box plot.

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

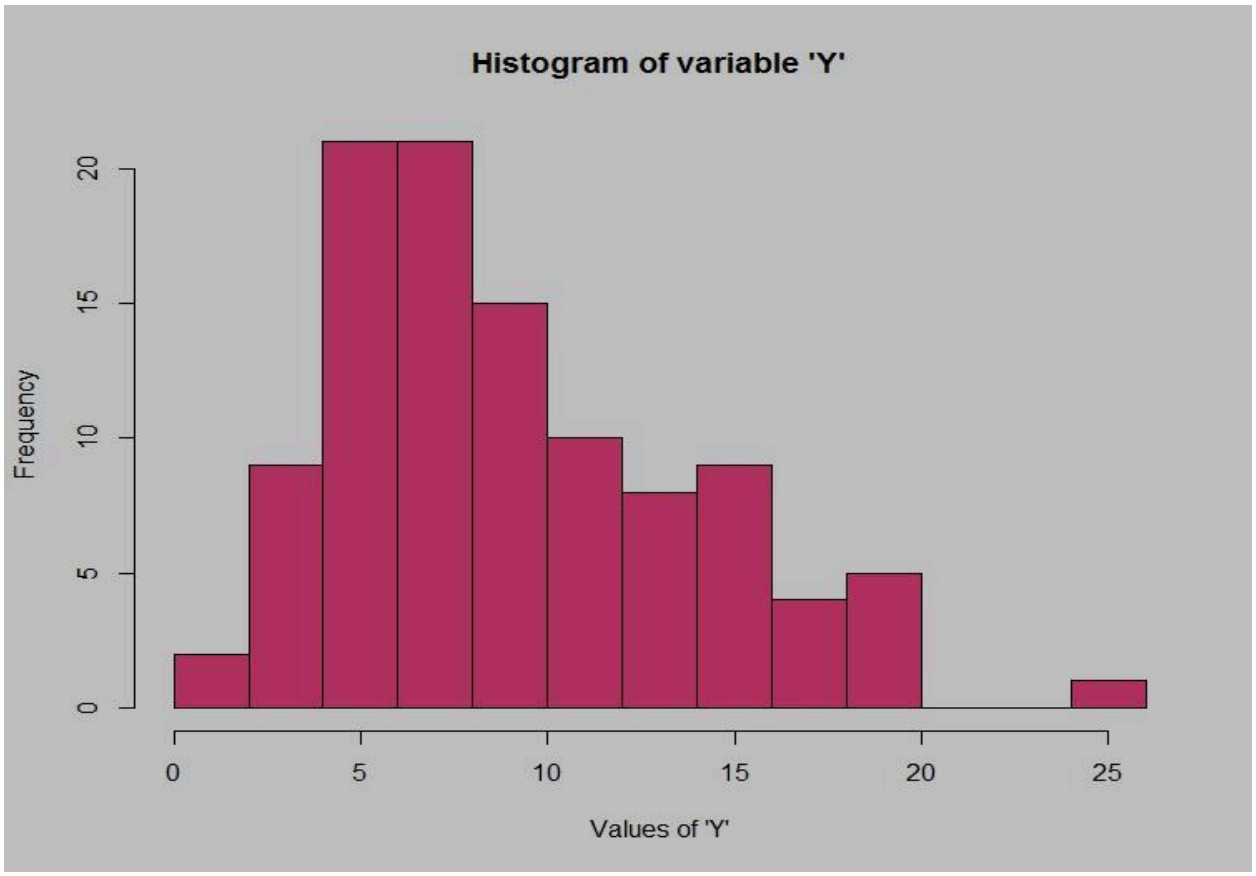
Ans: The Boxplot is positively skewed.

(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?

Ans: 1) There will be no outliers.

2) Median will not change but the IQR will change.

3.



Answer the following three questions based on the histogram above.

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

Ans: The mode will lie between 4 to 8 approximately.

(ii) Comment on the skewness of the dataset.

Ans: The data is positively skewed i.e left tailed.

(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.

Ans: Both provide outliers. The BoxPlot gives the median clearly and the histogram provides visualization.

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.)**

Ans: p=1/200. q= 1-p= 199/200.

To calculate at least one of five i.e.

1- no calls reach the wrong number

$= 1-P(0)$

$= 1- {}^{5}C_{0}(1/200)^{0}(199/200)^{5-0}$

$= 0.02475.$

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

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

Ans: 2000 is the most likely monetary outcome as it has the highest probability of 0.3.

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

Ans: Yes venture is likely to be successful as the profit has the highest probability of about 80%.

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

Ans: The long-term average investment is about 800$. The total returns x.P(x) is 800$.

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

Ans: The risk involved can be calculated by Variance, if Variance is high risk is also high. So the total variance is 216000. So we can conclude that the high amount risk is involved in this venture.

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?**
   - **A. 0.3875**
   - **B. 0.2676**
   - **C. 0.5**
   - **D. 0.6987**

Ans: We have to find the probability that the service manager cannot meet his commitment i.e probability of work done on car greater than 50(as work begins after 10 min, so 60-10=50).

$P(X>50)= 1-P(X<50)$

$= 1-0.73$

$= 0.26$

2. **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.**
   - **A. More employees at the processing center are older than 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.**

Ans: Both the statements are true.

a) There are about 63 employees older than 44 and 137 employees between 38 and 44. So the statement is true.

b) The training program under the age of 30 has approximately 36 employees. So this statement is also true.

3. **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.**

Ans: Distributions:

a) $2\,X_1$

Mean= 2μ

Varaince = $4\sigma^2$

b) $X_1 + X_2$

Mean= 2μ

Variance = $2\sigma^2$

Both a) and b) have same mean but variance differ. The Varaince of a) is twice than the b).

4. **Let $X \sim N(100, 20^2)$. 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

Ans: D. Mean=100

Variance=20

a= Mean-Z(Variance)

b=Mean+Z(Variance)

As the confidenceinterval is 99% so value of z is 2.576

hence a= 100-2.576(20) = 48.5

b= 100+2.576(20) = 151.5

5. **Consider a company that has two different divisions. The annual profits from the two divisions are independent and have distributions $\text{Profit}_1 \sim N(5, 3^2)$ and $\text{Profit}_2 \sim N(7, 4^2)$ 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

A. Specify a Rupee range (centered on the mean) such that it contains a 95% probability for the annual profit of the company.

Ans. As both the divisions are independent, the sum of independent normal random variables results in a normal random variable with mean equal to the sum of the individual means and variance equal to the sum of the individual variances.

So, Total for profits

Mean = 5+7= 12 *45= 540 million rupees

Variance = 9+15=25

Std Deviation = 5*45= 225 million rupees

For 95% probability, z= 1.960

Range= 99 to 981 million rupees

540 -1.96(225) =99

540 +1.96(225) = 981

B. Specify the 5th percentile of profit (in Rupees) for the company

Ans. for 5th percentile, z= 1.675

Profit = 540-1.675(225) = 169.875 million rupees

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

Ans. As both the divisions have profit mean greater than 0 so neither divisions have the probability of mking 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 to produce representative results. **True**

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

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

    A. The population
    Ans: **9000**

    B. The parameter of interest
    Ans: **average rating, sample**

    C. The sampling frame
    Ans: **9000**

    D. The sample size
    Ans: **225**

    E. The sampling design
    Ans: **kodak compact digital camera**

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

Ans: **As the population and sampling frame are same their are no sources of bias**

3. 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.
   Ans: **True**

   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.
   Ans: **True**

   III. The 95% confidence interval for μ only applies if the sample data are nearly normally distributed.
   Ans: **True**

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

   A. ¼
   B. ½
   C. ¾
   D. 1

   Ans: **D**

5. 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**

6. 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**

7. Which is shorter: a 95% *z*-interval or a 95% *t*-interval for μ if we know that σ =s?

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

   Ans: **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.

8. 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: **A. We would need to contact at least 600 randomly selected employers to guarantee a margin of error of no more than 4% at a 95% confidence level.**

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

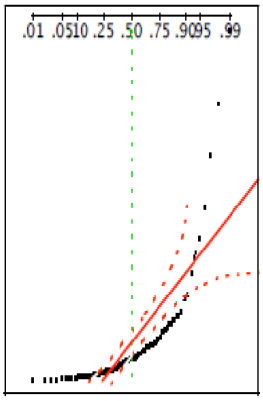
A. 1000
B. 757
C. 848
D. 543

Ans: **B. We would need to contact at least 848 randomly selected employers to guarantee a margin of error of no more than 4% at a 98% confidence level.**

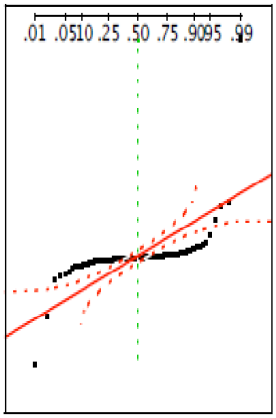
CBA: Practice Problem Set 2
Topics: Sampling Distributions and Central Limit Theorem

1. Examine the following normal Quantile plots carefully. Which of these plots indicates that the data …
   I. Are nearly normal? **C**
   II. Have a bimodal distribution? (One way to recognize a bimodal shape is a "gap" in the spacing of adjacent data values.) **B, D**
   III. Are skewed (i.e. not symmetric)? **A, B, D**
   IV. Have outliers on both sides of the center? **A**

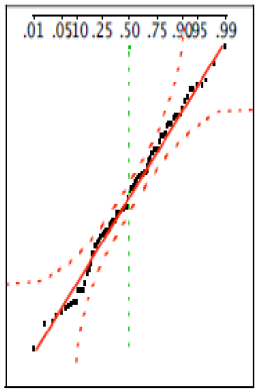
A



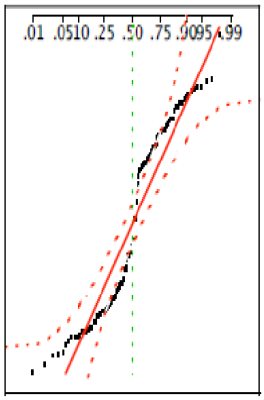
B



C



D



2. 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 $\mu = 22$ lbs. and $\sigma = 5$ lbs.

(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.

Ans. **False. Because it is not only condition for the model to be normally distributed. If the weights are not normal but their sample size is large, the distribution will also be normal.**

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

Ans. **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?

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

Ans. **Calculating z scores**

a) \$45= 45-50/(40/10)= -1.25

b) \$55 = 55-50/(40/10) = 1.25

from the standard normal distribution table,

Probability for -1.25 is 0.1056

1.25 is 0.8944

The values outside probability of 45&55\$ then investating will not happen, so

Probabililty= 1-(0.8644-0.1056) =0.2116

The probability that in any given week, there will be an investigation is 21.6%.

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 of 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. **t_value= 45-50/(40/√n)**

t_value for 95% is 1.96

so,

n= (1.96*8)*(1.96*8)

= 245.86=250

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?

    A. The standard deviation of the scores within any sample will be 120. **False**
    B. The standard deviation of the mean of across several samples will be 120. **False**
    C. The mean score in any sample will be 720. **True**
    D. The average of the mean across several samples will be 720. **True**
    E. The standard deviation of the mean across several samples will be 0.60. **True**