# STATISTICS WORKSHEET- 6

1-D

2-A

3-A

4-C

5-B

6-A

7-C

8-B

9-B

10-Histograms and box plots are graphical representations for the frequency of numeric data values. They aim to describe the data and explore the central tendency and variability before using advanced statistical analysis techniques.

- Both histograms and box plots allow to visually assess the central tendency, the amount of variation in the data as well as the presence of gaps, outliers or unusual data points.
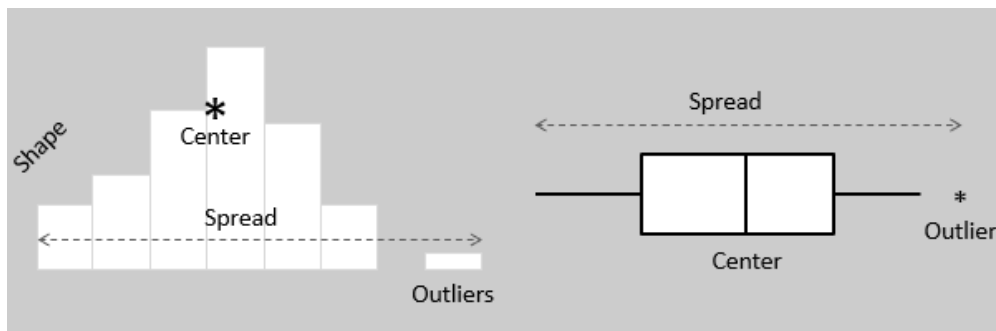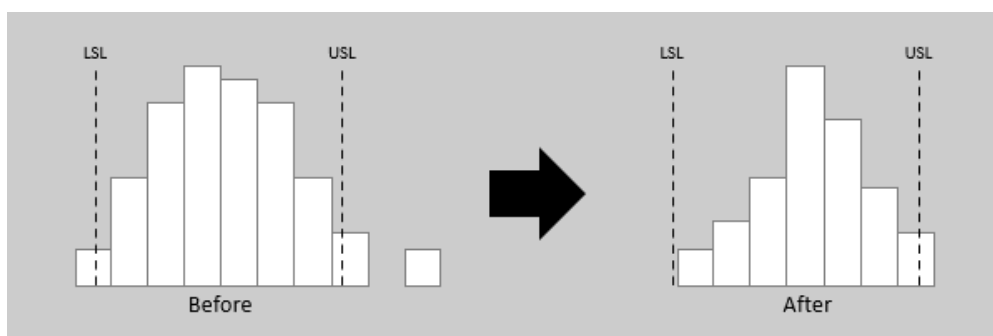

- Both histograms and box plots are used to explore and present the data in an easy and understandable manner. Histograms are preferred to determine the underlying probability distribution of a data. Box plots on the other hand are more useful when comparing between several data sets. They are less detailed than histograms and take up less space.
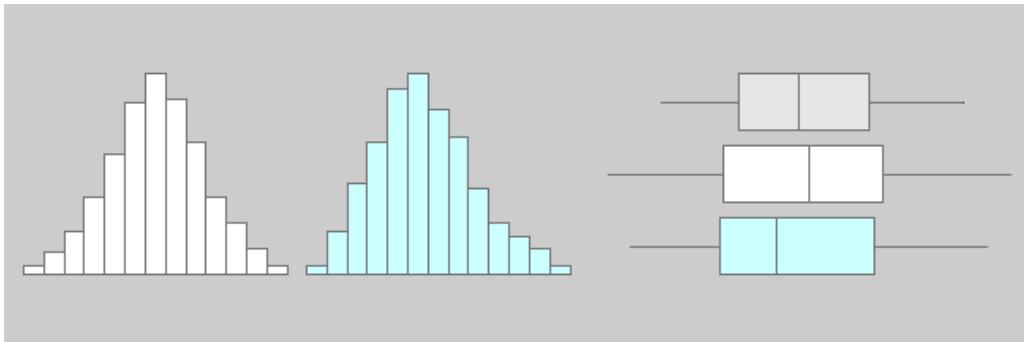
- Although histograms are better in displaying the distribution of data, you can use a box plot to tell if the distribution is symmetric or skewed. In a symmetric distribution, the mean and median are nearly the same, and the two whiskers has almost the same length.
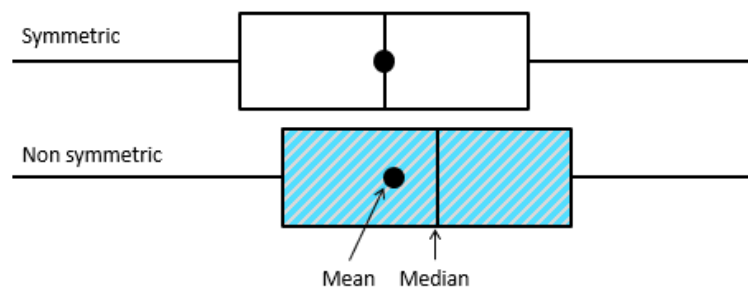
- You can use histograms and box plots to verify whether an improvement has been achieved by exploring the data before and after the improvement initiative. Both tools can be helpful to identify whether variability is within specification limits, whether the process is capable, and whether there is a shift in the process over time.

- Both histograms and box plots are ideal to represent moderate to large amount of data. They may not accurately display the distribution shape if the data size is too small. In practice, a sample size of at least 30 data values would be sufficient for both tools.

N = 40   N = 14

N = 40

N = 14
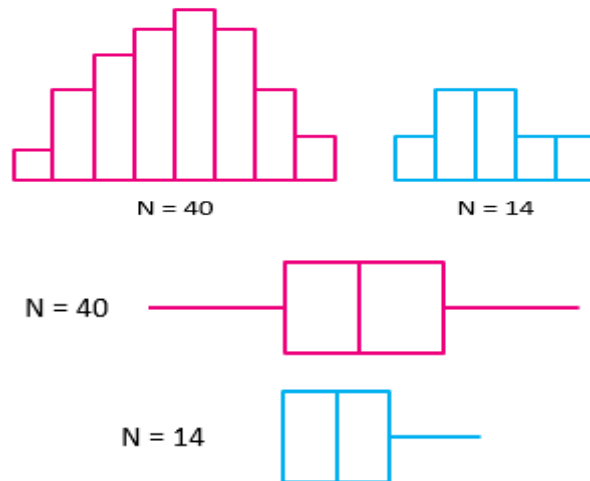
11-Selection of metrices is based on multiple factors as shared below.

- Define your primary objective
  Before you even begin to sift through the various metrics and statistics available to you, it is essential that your company's governing objectives have been clearly established.

- Choose your metric(s) - determine cause and effect
  Once a clear, overarching objective has been established, most marketing companies look to major metrics to determine their success—factors such as the generation of sales and leads.

  But these metrics aren't the only indicator of a company's success. Less easily quantifiable factors such as customer satisfaction and brand loyalty also play a significant role in the ability to achieve overall marketing objectives, especially in the long term.

- Create relevant activities
  Digital technology has made it easier than ever to track the engagement of various types of marketing materials, be they a video, article, or even a podcast.

  By taking advantage of the analytical tools provided by various online platforms, a marketing agency could easily find that for one client, video content is the chief driver of engagement, while for another, lengthy, informative blogs are the most effective type of content. Within these types of content, further factors could also impact engagement (such as the length of a video or the number of list items contained in a blog post).

- Evaluate periodically

  Of course, the ever-changing nature of marketing (and the business world as a whole), ensures that the measures you use to link activities with your primary goals must be constantly re-evaluated. The metrics and statistics that drive value for your clients can change over time, especially as new technologies emerge and target demographics shift.

12-Statistical significance is a determination that a relationship between two or more variables is caused by something other than chance. Statistical significance is used to provide evidence concerning the plausibility of the null hypothesis, which hypothesizes that there is nothing more than random chance at work in the data. Statistical hypothesis testing is used to determine whether the result of a data set is statistically significant. Generally, a p-value of 5% or lower is considered statistically significant.

Statistical significance is a determination about the null hypothesis, which suggests that the results are due to chance alone. A data set provides statistical significance when the p-value is sufficiently small.

When the p-value is large, then the results in the data are explainable by chance alone, and the data are deemed consistent with (while not proving) the null hypothesis

13-Exponential distributions do not have a log-normal distribution or a Gaussian distribution. In fact, any type of data that is categorical will not have these distributions as well.

The exponential distribution is often concerned with the amount of time until some specific event occurs. For example, the amount of time (beginning now) until an earthquake occurs has an exponential distribution. Other examples include the length of time, in minutes, of long-distance business telephone calls, and the amount of time, in months, a car battery lasts. It can be shown, too, that the value of the change that you have in your pocket or purse approximately follows an exponential distribution.

14-Mean is simply another term for "Average." It takes all of the numbers in the dataset, adds them together, and divides them by the total number of entries. Median, on the other hand, is the 50% point in the data, regardless of the rest of the data

Example of where the median is a better measure than the mean.

Let's say you run a customer satisfaction survey with a sample of 9 and rate their overall satisfaction scores on a scale of 1 to 10. You get an average of 5.22. You know that in general, you tend to retain customers with a score over 3, so you're satisfied,

because this indicates that you're still above where you want to be. But then, suddenly, you lose 6 of those 9 customers. You go back to look at your data, and you find these scores:

1, 3, 3, 3, 3, 5, 9, 10, 10

The median of this group is a 3, indicating that at least half of your customers or more were unhappy. The scores became lopsided because of the unexpected 10's, and you missed out on an important part of your data – the midpoint that indicated that as many as half of your customers or more were dissatisfied with your company.

Median can play a major role in things like income level research as well, because a few millionaires may make it look like the socio-economic status of your sample is higher than it really is.

Whenever a graph falls on a normal distribution, using the mean is a good choice. But if your data has extreme scores (such as the difference between a millionaire and someone making 30,000 a year), you will need to look at median, because you'll find a much more representative number for your sample.

15-The likelihood is a quantity proportional to the probability that, from a population having a particular value of θ, a sample having the observed value x_0, should be obtained. Likelihood, being the outcome of a likelihood function thus defined, describes the plausibility, under a certain statistical model (the null hypothesis in hypothesis testing), of a certain parameter value after observing a particular outcome.