STATISTICS

1.0 CONTINUOUS DISTRIBUTIONS

Continuous distributions are formed because everything in the world that can be measured varies to some degree. Measurements are like snowflakes and fingerprints, no two are exactly alike. The degree of variation will depend on the precision of the measuring instrument used. The more precise the instrument, the more variation will be detected. A distribution, when displayed graphically, shows the variation with respect to a central value.

Everything that can be measured forms some type of distribution that contains the following characteristics:

Measures of central tendency:

- Arithmetic mean or average
- Median
- Mode

Measures of spread or dispersion from the center:

- Range
- Variance
- Standard deviation

Shapes of distributions:

- Symmetrical - normal
- Symmetrical - not normal
- Skewed right or left
- More than one peak

2.0 MEASURES OF CENTRAL TENDENCY

Measures of central tendency are values that represent the center of the distribution.

2.1 Arithmetic Mean or Average

The arithmetic mean or average of sample data is denoted by $\bar{x}$. The mean or average of an entire population or universe is denoted by μ. The value of $\bar{x}$ may always be used as an estimate of μ.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{(x_1 + x_2 + x_3 + \ldots + x_n)}{n}$$

The symbol Σ stands for "sum of."

Five parts are measured and the following data are obtained:

2.6'', 2.2'', 2.4'', 2.3'', 2.5''

$$\bar{x} = \frac{\sum x_i}{n} = \frac{(2.6 + 2.2 + 2.4 + .2.3 + 2.5)}{5} = 2.4$$

2.2 Median

The median is the middle value of the data points.

To find the median, the data must be rank ordered in either ascending or descending order.

2.2, 2.3, 2.4, 2.5, 2.6
The Median is 2.4

For an even number of data points, the median is the average of the two middle points.

2.3 Mode

The mode is the value that occurs most frequently.

The data 2.6, 2.2, 2.4, 2.3, 2.5 do not contain a mode because no value occurs more than any other.

The following data are taken from another product:

6, 8, 13, 13, 20
The Mode is 13

3.0 MEASURES OF SPREAD OR DISPERSION FROM THE CENTER

How much can data points vary from a center or central value and still be considered reasonable variation? The question can be answered by calculating what is considered to be the natural spread of the data values.

3.1 Range

The calculation of the range provides a simple method of obtaining the spread or dispersion of a set of data. The range is the difference between the highest and lowest number in the set and is denoted by the letter r. The range and average are points plotted on control charts (a subject covered in a subsequent chapter). For the data set 2.6, 2.2, 2.4, 2.3, and 2.5, the high value is 2.6 and the low value is 2.2.

$$\text{Range} = r = (2.6 - 2.2) = .4$$

3.2 Variance

The variance is the mean squared deviation from the average in a set of data. It is used to determine the standard deviation, which is an indicator of the spread or dispersion of a data set.

$$\text{Variance} = \sigma^2 = \text{Sigma Squared} = \frac{\Sigma(x_i - \bar{x})^2}{n}$$

3.3 Standard deviation

The standard deviation is the square root of the variance. It is also known as the root-mean-square deviation because it is the square root of the mean of the squared deviations.

$$\text{Standard Deviation} = \sigma = \text{Sigma} = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}}$$

The average and standard deviation together can provide a great deal of information about a process or product. These two statistics are very powerful values used to make inferences about the entire population based on sample data.

When an inference is made about a population from sample data, (n - 1) is used instead of n in the denominator of the variance formula. The term (n - 1) is defined as degrees of freedom. When (n - 1) is used, the calculated value is called the unbiased estimator of the true variance and is usually denoted by s^2. When the standard deviation is obtained from the unbiased estimator of the variance it is denoted by s or σ'.

If a sample is taken and the average and standard deviation are not used to make inferences about the entire population, then the sample is considered to be the population and the standard deviation is indicated by σ. The symbol μ is used to denote the population average and $\bar{x}$ is used to denote the sample average. The value of $\bar{x}$ may always be used as an estimate of μ.

3.4 Variance and Standard Deviation Formulas

The following terminology and formulas will be used for the variance and associated standard deviation:

- Variance and standard deviation using all data values of a finite population:

$$\text{Variance} = \sigma^2 = \frac{\Sigma(x_i - \mu)^2}{N}$$

$$\text{Standard Deviation} = \sqrt{\text{Variance}} = \sigma$$

- Variance and standard deviation using a subset (sample) of an infinite (very large) population:

$$\text{Variance} = s^2 = (\sigma')^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1}$$

$$\text{Standard Deviation} = \sqrt{\text{Variance}} = s \text{ or } \sigma'$$

This is called the unbiased estimator of the population variance σ^2.

- Variance and standard deviation using a subset (sample) of a finite population:

$$\text{Variance} = s^2 = (\sigma')^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1} \bullet \frac{(N-1)}{N}$$

$$\text{Standard Deviation} = \sqrt{\text{Variance}} = s \text{ or } \sigma'$$

This is also called the unbiased estimate of the population variance σ^2.

- The standard deviation for a distribution of averages is called the standard error.

$$\text{Standard Error} = \sigma_{\bar{x}} = \sqrt{\frac{\Sigma(\bar{x}_i - \bar{\bar{x}})^2}{N_s}} \text{ or } \frac{s}{\sqrt{n}}$$

N_s is the number of samples and n is the sample size.

Example 1

Compute the variance and standard deviation for the data: 2.6" 2.2", 2.4", 2.3", 2.5". Assume that the data is the entire population.

$$\sigma^2 = \frac{\Sigma(x_i - \mu)^2}{N} \quad \text{where } \mu = 2.4$$

$(2.6 - 2.4)^2 =$	$(.2)^2 =$	.04
$(2.2 - 2.4)^2 =$	$(-.2)^2 =$	.04
$(2.4 - 2.4)^2 =$	$(0)^2 =$	0
$(2.3 - 2.4)^2 =$	$(-.1)^2 =$	.01
$(2.5 - 2.4)^2 =$	$(.1)^2 =$	.01
	Total =	.10

Therefore σ^2 = .10/5 = .02