Thinking Like A Statistician

Richard J. Burke, Robert D. Davis and Frank C. Kaminsky
Productivity Sciences Inc. Greenfield, MA 01301

Introduction

In a modern approach to management it is assumed that 80 to 95 percent of all problems are with the system. The role of the manager is to understand the system and to change the system to make it operate with higher quality and, consequently, with higher productivity. The only way to understand the system is to understand the concept of variability and, since statistics is the study of variability, it is absolutely essential to learn statistics.

In order to understand variability one must learn to *think like a statistician*. This is not as difficult as it may sound. Each day everyone is exposed to situations that illustrate the concept of variability.

Random Experiments

The process of thinking like a statistician begins with the translation of a situation of interest into something called a basic random experiment. In this process an outcome of interest from a real world situation must be defined and the following three questions must be asked.

1. Does the outcome change?
2. Is the outcome unpredictable?
3. Is the experiment repeatable under essentially the same conditions?

If the answer is yes to each of the three questions, then the situation has been translated into a basic random experiment. Before delving more deeply into this concept, this manner of thinking is illustrated with several examples.

Example 1:

Situation: ***A worker tests an ignition system.***

Outcome of interest:

Ignition system is good (performs properly; i.e., conforming) or bad (does not perform properly; i.e., nonconforming).

Answers to the three questions:

1. Obviously, the outcome may change from one ignition system to the next.
2. Although the outcome will be one of two possibilities, the result will not be known until the test is complete. It is, therefore, unpredictable.
3. For each ignition system produced or purchased, the experiment may be repeated under essentially the same conditions.

Example 2:

Situation: ***An invoice is examined for errors.***

Outcome of interest:

The number of errors in the invoice.

Answers to the three questions:

1. The outcome may be any number from zero to the maximum number of entries that should have been made. It is presumed that omissions may be counted as errors.
2. The number of errors is unknown until the audit is completed.
3. The experiment can be repeated for each invoice under essentially the same conditions.

Example 3:

Situation: ***A new work period begins.***

Outcome of interest:

The number of people absent.

Answers to the three questions:

1. The outcome may be any number from zero to the total number of employees.
2. The number of people absent will not be known until the work period begins.
3. The experiment may be repeated each work period under essentially the same conditions.

In each of these examples, the number of possible outcomes is countable. Random experiments with a countable set of outcomes are called discrete. In these cases an integer value is normally assigned to the outcome of the experiment. This was an automatic result in Examples 2 and 3. This was not the case in Example 1. A statistician would assign either a zero or a one to the outcome good and then either a one or a zero to the outcome bad. The choice might be arbitrary. If the number of good or conforming items is of primary interest, then a one would be usually assigned to the outcome good.

In general a random experiment is classified as either discrete (if the outcome is countable) or continuous (if the outcome is measurable). This distinction is illustrated in Table 1.

	TYPE OF OUTCOME:	
	DISCRETE:	CONTINUOUS:
PROPERTY:	Countable	Measurable
EXAMPLE VALUES:	0, 1 0, 1, 2, 3, ...	1.1, 2.5, 9.4, ... .0025, .0017, .0021, ...
PHYSICAL CORRESPONDENCE OF EXAMPLE VALUES:	Good, Bad Go, No Go No. Of Nonconformities Etc.	Length Area Volume Time

Classification of Random Experiments
TABLE 1

Continued

Random experiments for which the outcome is continuous are illustrated in examples 4 through 6.

Example 4:

Situation: ***A shipment of shoulder belts that are nominally 72 inches in length arrives and a shoulder belt is selected.***

Outcome of interest:

The actual length of the belt.

Answers to the three questions:

1. The actual length may be any number greater than zero but it will probably fall somewhere in the vicinity of 72 inches; e.g., 73.6", 70.4", 71.9", ...
2. The actual length is unknown until the measurement is taken.
3. The experiment may be repeated for each belt in the shipment of belts and every time a shipment is received under essentially the same conditions.

Example 5:

Situation: ***An employee who lives about 25 miles from work drives to work each day.***

Outcome of interest:

The actual miles driven.

Answers to the three questions:

1. The actual number of miles driven will vary but will probably remain in the vicinity of 25 miles; e.g., 25.2, 26.2, 24.9, ...
2. Until the employee arrives at work the miles driven will not be known.
3. The experiment may be repeated each working day under essentially the same conditions.

For this random experiment the reader should collect some data in their own driving situation.

Example 6:

Situation: ***An item is selected for a quality audit. A perfect item has a quality index of 100%.***

Outcome of interest:

Actual quality index.

Answers to the three questions:

1. Any number that is no larger than 100% is a possible value; e.g., 90.2, 85.7, 98.6, ...
2. Until the audit is completed the quality index is unknown.
3. The experiment may be repeated each time an item is available for an audit under essentially the same conditions.

Based upon the concepts illustrated in the previous six examples, a variety of situations that are encountered every day can be transformed into random experiments. Each experiment should be classified as to whether its outcome is discrete or continuous.

Repeated Random Experiments

The two characteristics of a random experiment which are extremely important to a statistician, and to the eventual concepts of statistical quality control, are that the outcome changes and is therefore *unpredictable*, and that the experiment is *repeatable* under essentially the same conditions.

If either of these characteristics is not present, then the random experiment is of no value. Generally, results from several repetitions of the basic random experiment, rather than the outcome of a single random experiment, form the basis for decisions. This repetition is viewed as follows:

Repeated Random Experiment:

i. Define A Basic Random Experiment, called E.
   The outcome of E becomes *variable 1.*
ii. Repeat the Basic Random Experiment a fixed number of times, say k.
   Call these k experiments E_1, E_2, ... , E_k.
   A variable 1 outcome is produced by each E_1, E_2, ... , E_k.
iii. An overall outcome is computed from the individual outcomes.
   This outcome becomes *variable 2.*

In statistical process control, the second variable ("variable 2") is usually monitored. Examples of variable 2 are the total number of nonconforming items, the average number of nonconforming items, fraction nonconforming, percent nonconforming, the average quality index of the k items, the average length of k shoulder belts, the average number of people absent in k days, the average number of miles driven to work in k days, and the average number of nonconformities in k invoices. This type of repeated random experiment is illustrated in Example 7.

Example 7:

Basic random experiment: ***A worker tests an ignition system.***

Outcome: The ignition system works (variable 1 = 0) or does not work (variable 1 = 1).

Repeat experiment 40 times:

Experiment repetition: E_1 E_2 E_3 ... E_{40}.

Individual outcome: 0 1 0 ... 1.

Possible values for variable 2:

a. Total number of non-working items = outcome of E_1 + outcome of E_2 + outcome of E_3 + ... + outcome of E_{40}
b. Fraction of non-working items = Total number of non-working items divided by 40.

The Language Of Statistics

The collection of measurements obtained from repetitions of the basic random experiment is a set of numbers. This set of numbers is referred to as the data from the composite random experiment. Statisticians use a variety of graphical and statistical methods to extract information from the data.

Reference:

R.J. Burke, R.D. Davis, and F.C. Kaminsky, Statistics And Quality Control For The Workplace, (unpublished manuscript). The book will be available soon through Quality Press.