Using Sampling Trees to Assist in Sample Plan Development

Bill Ross
Sigma Science Inc.

Abstract

Multiple organizations are applying analytical techniques to improve performance. These techniques include the application of various sampling methodologies. The intent of sampling is to correlate variation in outputs (Y's) to variation in inputs (x's) to gain understanding of potential causal relationships commonly represented by the equation: Y=*f*(x) Once the relationships are exposed and understood, the sources can be managed, optimized, designed out, or their effects mitigated. It has been observed most training in the use of these techniques is solely on data analysis rather than on planning for the acquisition of data. This paper will focus on the importance of planning and the use of sampling trees as a graphical aid to assist in that planning.

Introduction

The concepts underlying Sigma Science deal with the fact process and product variation is known to be a strong factor affecting manufacturing lead times, product and process costs, process yields, product quality, performance, reliability and ultimately customer satisfaction. A critical part of the work is to identify and measure variation in outputs with the intent of discovering the causes of that variation and to develop efficient, operational means to control or reduce that variation. Sampling is an efficient means of investigating a large number of potential sources simultaneously.

Dr. W. Edwards Deming was the first to differentiate between *analytical* and *enumerative* sampling plans. An enumerative sampling plan is one where conclusions will be made regarding a well-defined, existing population. In an analytical study there is a desire to draw inferences, beyond the data in-hand, about the future. The ultimate intent of an analytical study is to understand and manage the causal structure to predict and improve future performance. While enumerative sampling plans have a role in industry, the most applicable and appropriate sampling is or is intended to be analytical. Designing sampling plans inclusive of all sources of variation is likely inefficient. In most cases designing sequential sampling strategies *representative* of known or suspected sources of variation is the most efficient and effective means of understanding relationships between a large number of variables. The sampling trees graphically display how the potential causal structure components (x's) are being partitioned and compared.