The Alpha — Beta Wars
Which Risk Are You Willing To Live With?

Annabeth L. Propst
Process Management Institute - Bloomington, MN 55435

Introduction

In situations where a decision must be made based on imperfect or incomplete information, we are faced with the probability of making the wrong decision. In the simplistic case of a dichotomous decision, there are two types of errors. In traditional statistical nomenclature they are called Type I and Type II errors. With a constant information base the likelihood of a Type I error (alpha) is inversely related to the likelihood of a Type II error (beta). Traditionally, we have tried to minimize alpha (producer's risk) and let beta (consumer's risk) fall where it will. Hence the phrase "caveat emptor" . . . let the buyer beware. In today's world of high precision, extended warranties, and product liability suits this may no longer be acceptable the consumer's risk has become the producer's risk. In this paper we will address four points:

1. Definition of alpha and beta (in the traditional sense)
2. Expanded definitions of alpha and beta
3. The cause for concern
4. What we can do to end the war

Definition of Alpha and Beta

In a traditional hypothesis test we are concerned with testing some null hypothesis (Ho). There are two states of nature: either Ho is true or Ho is false. Of course, we know that Ho is probably not exactly true. What we are really concerning with discovering is whether it is so different that we can no longer act as if it were. The two decisions we can make are to 'reject Ho' or 'fail to reject Ho'.

If Ho is true and we fail to reject, or if Ho is false and we reject, then a correct decision has been made. (See Figure I.) If Ho is true but we reject it, we have made a Type I error. The likelihood, or probability, of making this error is defined as alpha. When Ho is false but we fail to reject, we have made a Type II error. This probability is called beta. In a hypothesis test, alpha and beta (given a margin of difference we wish to detect) can be computed fairly easily. In fact, we usually predetermine alpha, pick a reasonable sample size, and calculate beta. Often, the computation of beta is omitted.

In many statistics course the analogy is drawn between the traditional hypothesis test and a jury trial. In a jury trial the dichotomous reality is the innocence or guilt of the defendant. (See Figure 2.) The verdict 'guilty' or 'not guilty'. If a guilty person is found guilty or an innocent person is found not guilty, justice has been served. When an innocent man is found guilty we have made the Type I error of punishing and innocent person. On the other hand, when we find the guilty person to be not guilty, we have made the Type II error of allowing the guilty person go free, roaming the streets to err again — a potential time bomb.

This analogy goes even deeper. In the legal process in this country, the defendant is innocent until proven guilty. The jury, however, never brings back a verdict of 'innocent' just as a statistician never decides that the null hypothesis is true. Just as the statistician is not prepared to say the null hypothesis is true, the jury is not prepared to say that the defendant is innocent — how could they know? Instead they bring back a verdict of not guilty, which means that there is not enough evidence to satisfy them beyond a doubt that the defendant is guilty. That is because we feel that the conse-

FIGURE 1

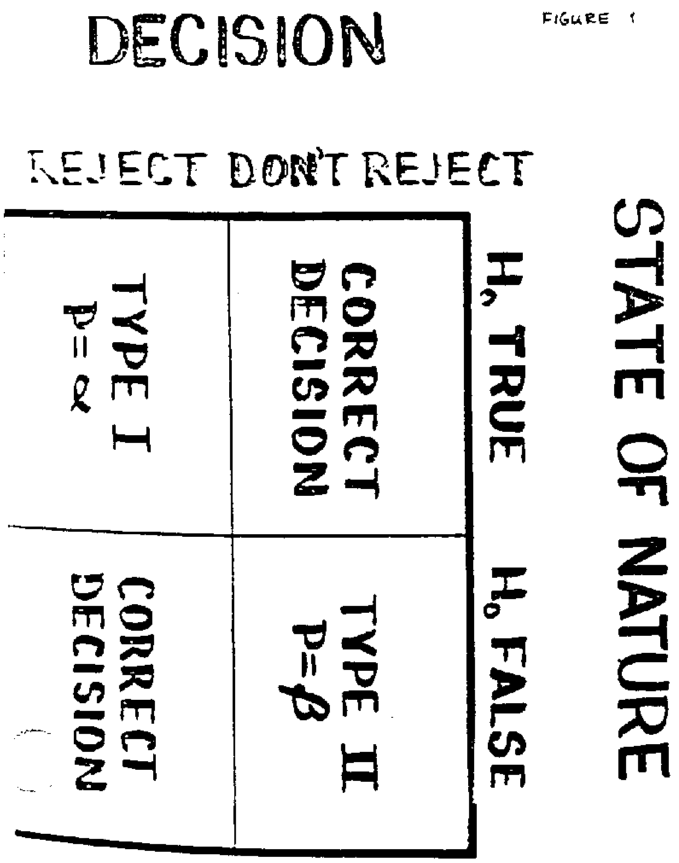
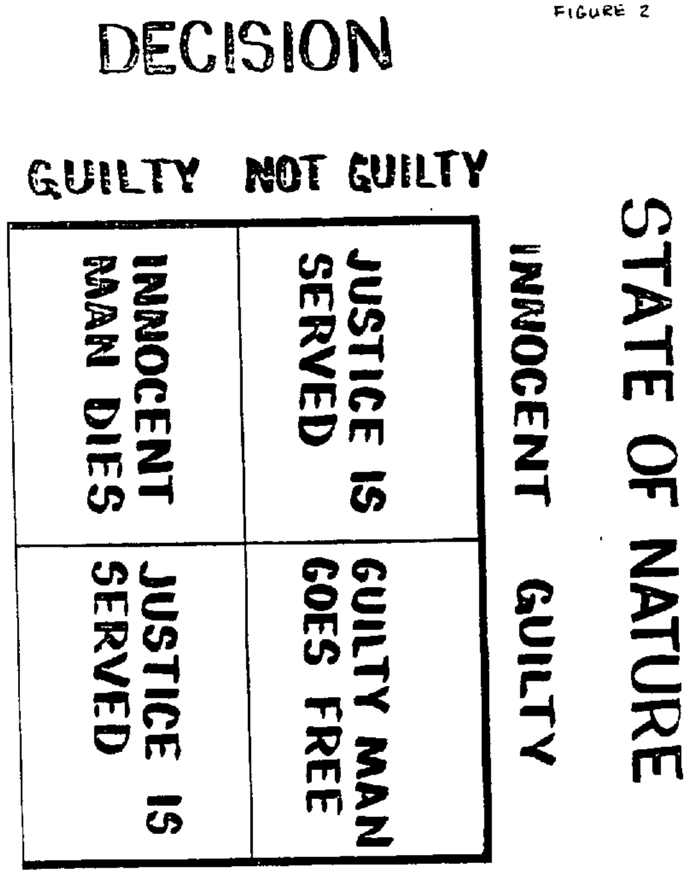


FIGURE 2



Alpha — Beta Wars

(Continued from page 7)

quences of convicting an innocent person are much more serious than the consequences of allowing a guilty person to go free. This results in a tendency to allow guilty people to go free. The same is true of the traditional hypothesis test. Alpha is predetermined to be small. Therefore beta for reasonable sample sizes could be large.

In addition, once a jury has failed to convict the defendant, he cannot be retried for the same offense. A defendant judged 'guilty' can appeal. In a statistical hypothesis test we get a clue when we make a Type I error — we reject something we thought might be true. But when we make a Type II error, we have no clue — the Fat, Dumb, and Happy Syndrome.

The analogy between the hypothesis test and the jury trial ends when we try to calculate alpha and beta. In the hypothesis test situation it is fairly easy to determine the probabilities of error. In the jury trial it is not possible to calculate these values, even though we can say that alpha is small compared to beta.

Expanded Definitions of Alpha and Beta

The preceding discussion of alpha as the probability of a Type I error and beta as the probability of a Type II error should be familiar to us. Let us agree to continue to use this definition of alpha and beta as the probabilities of Type I and Type II errors, but redefine the errors in the context of some other dichotomous realities.

For instance, in the case of a control chart, we could view the reality relative to process change. Either the process has changed appreciably or it has not. The decision could be 'take action based on an out of control signal' of 'do not take action based on the absence of an out of control signal'. (See Figure 3.) If the process had changed and we take corrective action based on a signal (assuming such action is truly corrective) we have made a correct decision, allowing us to control the process and minimize variability. If the process has not changed and we fail to take action due to the absence of a signal, we have again made a correct decision. This results in a consistent process, producing the minimum variability possible for that process.

If the process has not changed, but we receive a signal resulting in unnecessary action, we have made a Type I error. The result is an over-controlled process with increased variability. The Type II error occurs when the process has changed appreciably and, because we received no signal, we failed to take action. This process is under-controlled and also has increased variability. In the case of the st dard Shewhart control chart we can cc pute approximate values of alpha and beta for specific process shifts, assuming a normal distribution. If we add supplementary runs rules, the computation becomes unnice. However, it is still true to say that alpha is small compared to beta.

We can extend this definition to other cases where it is no longer possible to calculate even approximate values for alpha and beta. Suppose the reality is the existence of a single common cause system. Of course, if a single common cause system does not exist, we say special causes are present. The two decisions are 'work in the process' and work on the process'. If special causes are present and we work in the process, we have made a correct decision, optimizing the existing process. (See Figure 4.) Similarly, if we have a single common cause system and we work on the process to reduce those common causes, we have made a correct decision, and process improvement occurs.

On the other hand, if we are working in the process to eliminate special causes when we, in fact, have a single common cause system, we have made a Type I er

(Continued on page 9)

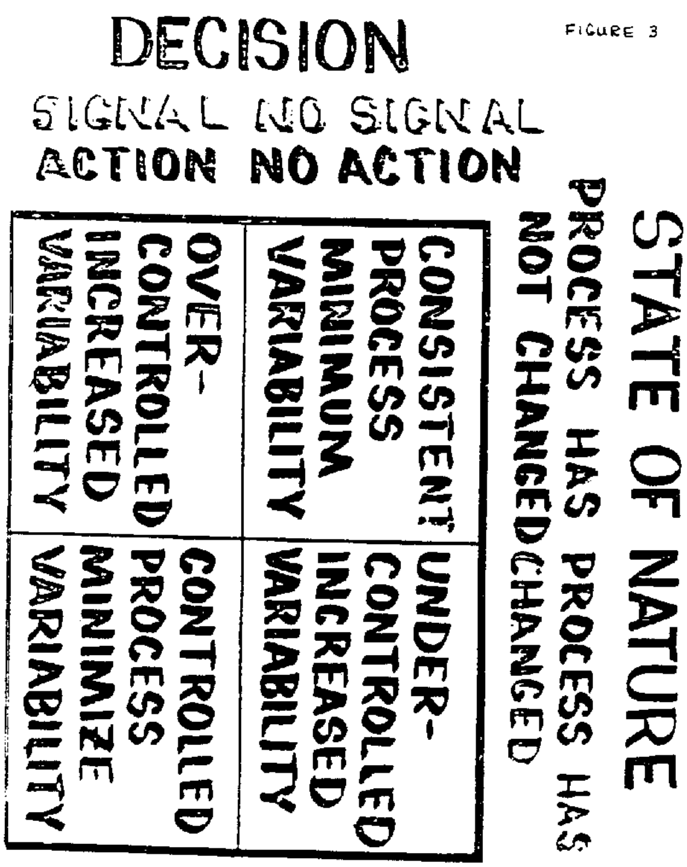
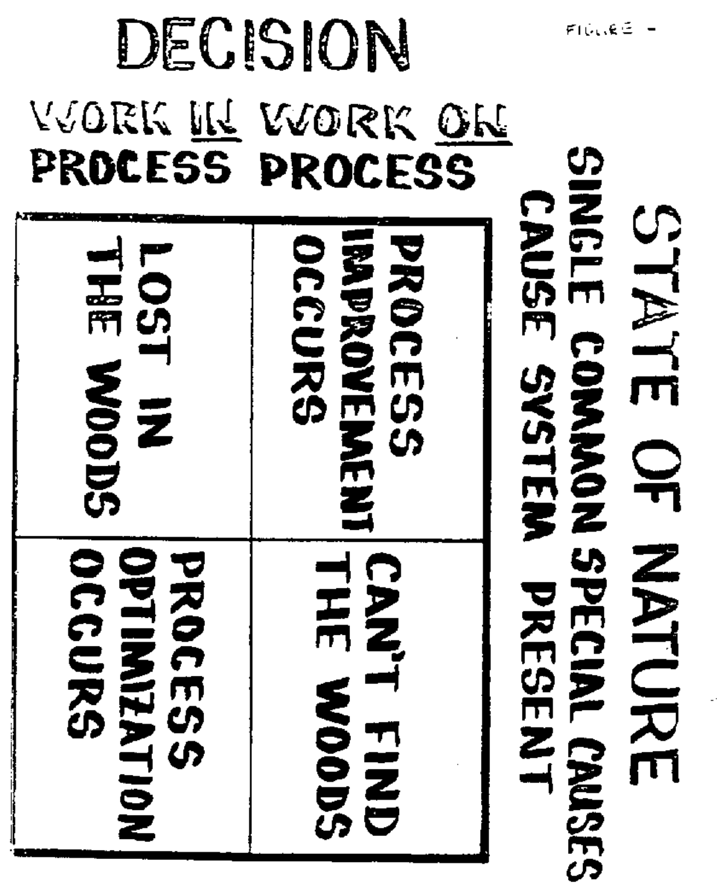


FIGURE 3



ror. We are lost in the woods without a map, a path, or a compass. We will walk until we bump into a tree or stumble over a root and then change directions. If special causes are present but we are working on the process to reduce common causes of variability, we have made a Type II error. We are not lost in the woods . . . we can't even find the woods. It's a good bet we don't even know they exist.

We could get even more general with our analogy and discuss whether the status is quo or the status is not quo. Possible decisions would be 'do something' or 'do nothing'. If the status is quo and we do nothing, we have made a correct decision, staying home by the fire where it is safe and warm. If the status is not quo and we do something, we have at least begun the journey, even though it may not be where we want to go.

The Type I error occurs when we decide to do something even though the unknown reality is status quo. We rocked the boat (which might not be all bad). Doing nothing when the status is not quo results in the Type II error of missing the boat altogether. This business of drawing analogies could go on forever, but let's not. The real question we should be asking at this point is: Why do we care? Funny you should ask.

The Cause for Concern

We should be concerned about Type II errors because the consequences of making them can be severe. Let's look briefly at the consequences of some well-known Type II errors.

- Thalidomide was introduced in the late forties to combat morning sickness in pregnant women. Unfortunately, it also caused serious physical deformities in the fetus.
- Ford Motor Company paid millions of dollars in lawsuits to families of those who died as the result of Pinto gas tanks bursting into flame following rear-end collisions.
- The company that produced the IUD Dalcon Shield was forced to go out of business due to the claims of women with serious complications (including death) resulting from the use of the shield.
- In January, 1986, the Space Shuttle Challenger exploded only minutes after take-off due to the failure of an O-ring. Six astronauts and a school teacher were killed.
- In August, 1987, a Northwest Airlines flight crashed during take-off from Detroit Metro airport, apparently because the flaps were down and the buzzer designed to warn of this condition failed to go off.

We can think of others. One difficulty in addressing Type II errors relates to the difficulty of knowing ahead of time what the costs (or consequences) will be. They are unknown and unknowable . . until it is too late. Who knows what it costs when a defective product gets into the hands of a customer. That is a Type II error. We often find that the costs are quite large. The consequences may be fatal . . . to the customer, the company, and even innocent passersby.

Unfortunately, there are costs associated with both kinds of errors. Given a constant information base (sample size) we can reduce the probability of one error only by increasing the probability of the other. We can reduce both by increasing our information base, but there are costs associated with this also — including the loss of timeliness. Thus we have the alpha-beta wars, the continual struggle between minimizing Type I and Type II errors without incurring prohibitive information costs.

What We Can Do To End The War

If we had perfect knowledge, of course, there would be no errors and no war. In the absence of perfect knowledge we must do something. If consulting the oracle at Delphi is not a useful option (which it isn't) we must develop some alternatives.

First we must call on knowledge of the process. George Box talks about filtering statistical results through engineering knowledge. Dr. Deming tells us that experience teaches nothing in the absence of theory — and often knowledge of the process is as good a theory to start with as any.

This theory is used as the plan in the first trip around the Shewhart cycle, often called the PDCA cycle. We begin our trip around the cycle with a plan or theory (P = Plan). Then we try out the plan or test the theory (D = Do). We collect data, naturally, to check the effectiveness of the plan or the validity of the theory (C = Check). The analysis of the data will tell us whether we have a good plan or a strengthened belief in the theory, or if we wish to modify or discard it (A = Analyze, Act, Adapt, Abandon). This brings us back to P, and around the cycle we go again.

Another approach would be to design out the opportunities for error. If we reduce the number of ways a process can change, we reduce the opportunity for a Type II error. If the probability of producing a defective lot is decreased, the window of opportunity for falsely accepting such a defective lot becomes narrower.

A last possibility is to focus on process improvement. By improving the process we reduce common causes of variability. This reduces the absolute amount of the change that is liable to go undetected. This often reduces the consequences (costs) of making a Type II error.

Note that alpha and beta are conditional probabilities, conditional on the unknown state of nature that actually exists. The relationship between the conditional probabilities and the amount of information is fixed. Only using process knowledge might decrease that conditional probability. Using the PDCA cycle ensures that knowledge is constantly being gained by the combination of theory and experience. We are constantly reassessing the situation rather than being bound by some archival decision. Designing out opportunities for error reduces the probability of an undesirable state of nature. Although the conditional probability of making a wrong decision given this reality is unchanged, the absolute (or marginal) probability is reduced. If we were successful at preventing forest fires, we would no longer have to worry about detecting them. Improving the process may or may not change the conditional probability. It certainly reduces the probability of an undesirable reality, and thus the marginal probability. It often reduces the consequence of making a Type II error.

Not The End

I don't have the answers — I wish I did. I'm just asking some tough questions. If we could accumulate the total knowledge of quality professionals everywhere, we would find better answers. They will be different for every problem, but the thought processes and the techniques used should be similar. We must work together to fight the alpha-beta wars, and we must win in order to survive. We need to know what others are doing. We need to share ideas, success stories, and horror stories. Thus, while this is the end of my paper, it is not The End. It is The Beginning — the beginning of an exciting era of cooperation and collaboration between quality professionals.

Note: This is a written version of a paper presented at the Fall Technical Conference in October 1987.

32nd Technical Conference

(Continued from page 5)

Sheraton Meadowland Hotel, East Rutherford, New Jersey. Co-sponsored by Chemical and Process Industries Division and Statistics Division of the American Society for Quality Control, and the Section on Physical and Engineering Sciences of the American Statistical Association. Hosted by the North Jersey Section of the ASQC. For more information contact: Laura Stauffer (609) 655-7084 or Teri Suzuki (609) 655-7815, Sensory and Statistical Sciences, General Foods, Prospect Plains Road, Cranbury, NJ 08512.