Statistical Bulletin
Reliability & Variation Research

DETROIT RESEARCH INSTITUTE
15224 KERCHEVAL AVE. • GROSSE POINTE PARK, MICHIGAN 48230 • (313) 499-1150

Leonard G. Johnson, EDITOR

Vol. 1 January, 1972

Bulletin No. 7 Page 1

ECONOMICS OF STATISTICAL EXPERIMENTAL DESIGN

"It is not too well realized that competence in practically every field of endeavor requires the ability to take in and interpret information, often to a greater extent than the ability to give it out." [A. Rapepert: "Saying What You Mean"; Etc., Summer, 1956.]

Within industrial complexes, and in scientific experimentation, information is coming to us daily, and we must interpret the results. Have you ever heard a couple of fellows arguing, where the first one says to the second one, "But I kept everything else constant and only changed this one factor, and look what happened."? "You're a liar!" says the second one. "I performed the very same type of test by keeping everything else constant and changed only this one factor you mention, and I got the opposite result!" Who is right? Is it possible to have such discrepancies in results?

As a matter of fact, both of these fellows show a lack of understanding concerning the design and analysis of experiments. This situation reminds us of the old story about the farmer who having been paid $100 in cash, counted the first 70 bills and then put all the money into his pocket saying that since the count was right so far he had no reason to suspect the remainder. We hope to shed some new light on this important subject in the next few pages.

Yes, it is possible to get opposite results in experiments conducted the way these fellows did theirs. It's not enough to keep everything else at one constant level, while varying one particular factor. We must try everything else at different constant levels and then in each such new situation vary the factor whose effect we are studying. (We must count all the bills, not just the first 70.)

EXAMPLE

$$Z = (B - 10)\,X + 100$$

B = 20

X	Z	
10	200	Increase in Z
20	300	

B = 8

X	Z	
10	80	Decrease in Z
20	60	

These two fellows who were arguing with one another had kept everything else at a constant level all right, but each one had made his own choice of constants. Changing constants can make responses do an about face.

Wouldn't it be foolish of me to think that I can change the constants in an equation and then expect the equation to show the same directional behaviour as before in response to a change in the value of the variable whose effect I am studying? This is the same foolishness present in the old unscientific method of testing ONE VARIABLE AT A TIME while keeping all other variables at their nominal levels. Even if no other excuse can be found for replacing the old method of experimentation with the STATISTICAL DESIGN OF EXPERIMENTS, we can still say that the old method must be discontinued simply because it is incorrect. But, fortunately the statistical method of experimental design not only is correct, but is also more economical and productive of fruitful results.

STUDY ONE VARIABLE AT A TIME ---- AND LEAD YOURSELF ASTRAY

Two experimenters (Mr. A and Mr. B) are making separate studies of the unknown relation

$$Z = (Y - 10)X + 50.$$

When we say the relation is unknown, we really mean that it is unknown to Messrs. A and B, although it is known to us.

Neither Mr. A nor Mr. B has been trained in the modern statistical design of experiments. Each one intends to study one variable at a time in two steps. These two steps are

STEP 1: Keep FACTOR Y fixed and vary X from a low value to a higher value to observe whether Z increases or decreases with an increase in X.

STEP 2: Keep FACTOR X fixed at its last level in STEP 1 and increase Y from what it was in STEP 1 to observe whether Z increases or decreases with an increase in Y.

MR. A's STRATEGY AND CONCLUSIONS ARE AS FOLLOWS:

STEP 1: Keep FACTOR Y fixed at Y = 8, while measuring Z at X = 4 and at X = 7. From this Mr. A finds that Z decreases from 42 to 36 as X increases from X = 4 to X = 7. From STEP 1 Mr. A concludes

Z is a decreasing function of X. (See FIGURE 1)

STEP 2: Keep FACTOR X fixed at X = 7, while measuring Z at Y = 8 and at Y = 10. From this Mr. A finds that Z increases from 36 to 50 as Y increases from Y = 8 to Y = 10. From STEP 2 Mr. A concludes

Z is an increasing function of Y.

MR. B's STRATEGY AND CONCLUSIONS ARE AS FOLLOWS:

STEP 1: Keep FACTOR Y fixed at Y = 12, while measuring Z at X = 4 and at X = 7. From this Mr. B finds that Z increases from 58 to 64 as X increases from X = 4 to X = 7. From STEP 1 Mr. B concludes

Z is an increasing function of X. (See FIGURE 1)

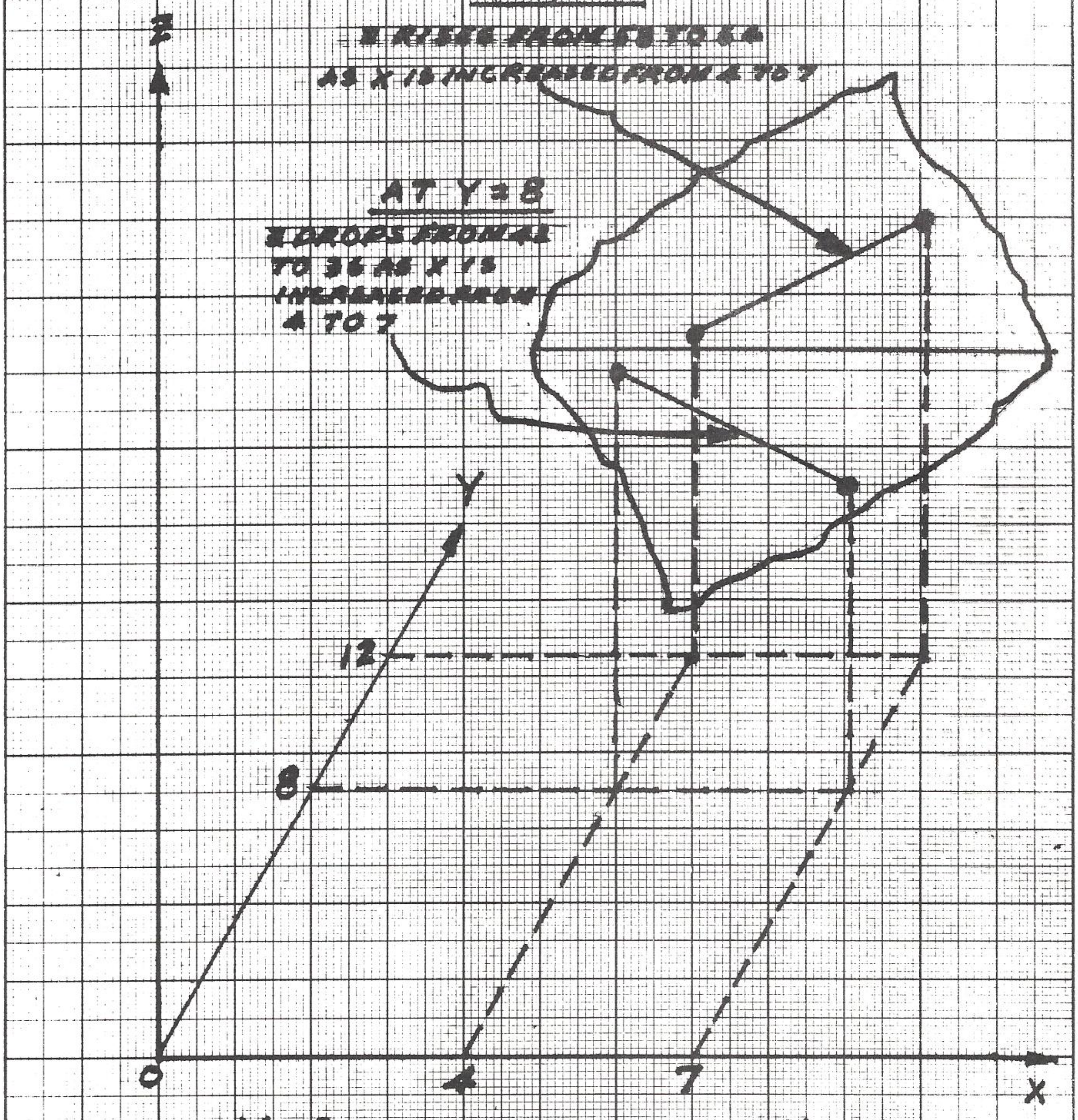
STEP 2: Keep FACTOR X fixed at X = 7, while measuring Z at Y = 12 and at Y = 14. From this Mr. B finds that Z increases from 64 to 78 as Y increases from Y = 12 to Y = 14. From STEP 2 Mr. B concludes

Z is an increasing function of Y.

DISCUSSION OF THE OUTCOME

We see that the two experimenters arrive at opposite conclusions from STEP 1, but agree in their conclusions from STEP 2. How come? Who is right? The answer is that each one is right for the EXPERIMENTAL PATH he chose. What was overlooked by each one was the fact that there is an interaction between X and Y, and that this interaction came into play within the range of Y values (Y = 8 to Y = 12) which were chosen in STEP 1 by the two experimenters. The problem is of such a nature that one is led astray by checking on variations of Z vs. X while keeping Y fixed at one level only. The nature of the function changes with the fixed level of Y chosen. For this reason a properly designed experiment with a sufficient number of fixed levels of X and Y is needed to tell the whole story.

THE RELATION: $Z=(Y-10)X+60$



Y=8	
X	Z
4	42
7	36

Y=12	
X	Z
4	58
7	64

FIGURE 1

A common complaint in modern industry is the drop in productivity. Workers are less productive per labor dollar. Budgets are less productive per budget dollar. Investments in new facilities are less productive per investment dollar. We can cite example after example to show that productivity has been on the decline during the past two decades in American industry. The profit squeeze has become so critical that every effort should now be made to increase productivity in all phases of industry. A very important part of the whole industrial operation of a corporation is the expense incurred from the experimentation performed to uncover facts for research, development, decision making, policy making, materials selection, design validation, future plans, etc. Consequently, experimentation is also one of those operations of an industry which must be made more productive, if such an industry is to remain viable. The purpose of this issue of our bulletin is to indicate how more productive experimentation can be achieved through the proper statistical design of experiments.

THE UNIVERSAL GOAL OF ALL EXPERIMENTATION

Whether we admit it or not, all experimentation has a basic type of goal which can be described in general terms. Once we are able to describe the purpose of experimentation in general terms it becomes quite clear why some forms of experimentation are not very productive, while other forms are highly productive. When we stop and think about the subject seriously we will have to admit that the following general definition of an experiment is a valid concept:

DEFINITION OF AN EXPERIMENT:

An EXPERIMENT is an attempt to construct a usable "MAP" of an actual "TERRITORY" (or "TOPOGRAPHY") imbedded in one or more measurable "DIMENSIONS".

WHAT A MAP IS SUPPOSED TO DO

In order to be usable a MAP must be a reasonable representation of the "TERRITORY" it is supposed to APPROXIMATE. There are rough maps and fine maps, as well as certain grades in between. With a fine map we can do a good job of interpolating, or predicting, what the situation is between any two adjacent points. However, one thing a map will not permit is extrapolation into "territory" it does not cover. Because extrapolation into new "territory" is not possible, it is important that we decide in advance of each experiment how much "territory" will be subjected to interpolation when we come to use the constructed map resulting from the experiment. The "DIMENSIONS" of a "MAP" are the FACTORS affecting the outcome (or RESPONSE VARIABLE) in an experiment. In essence, a "MAP" permits interpolation in accordance with its fineness, but never extrapolation.

A SIMPLE EXAMPLE OF A MAP AND ITS EXPERIMENT

To take a simple example of a map, let us consider a daily WEATHER MAP of the United States (FIGURE 2).

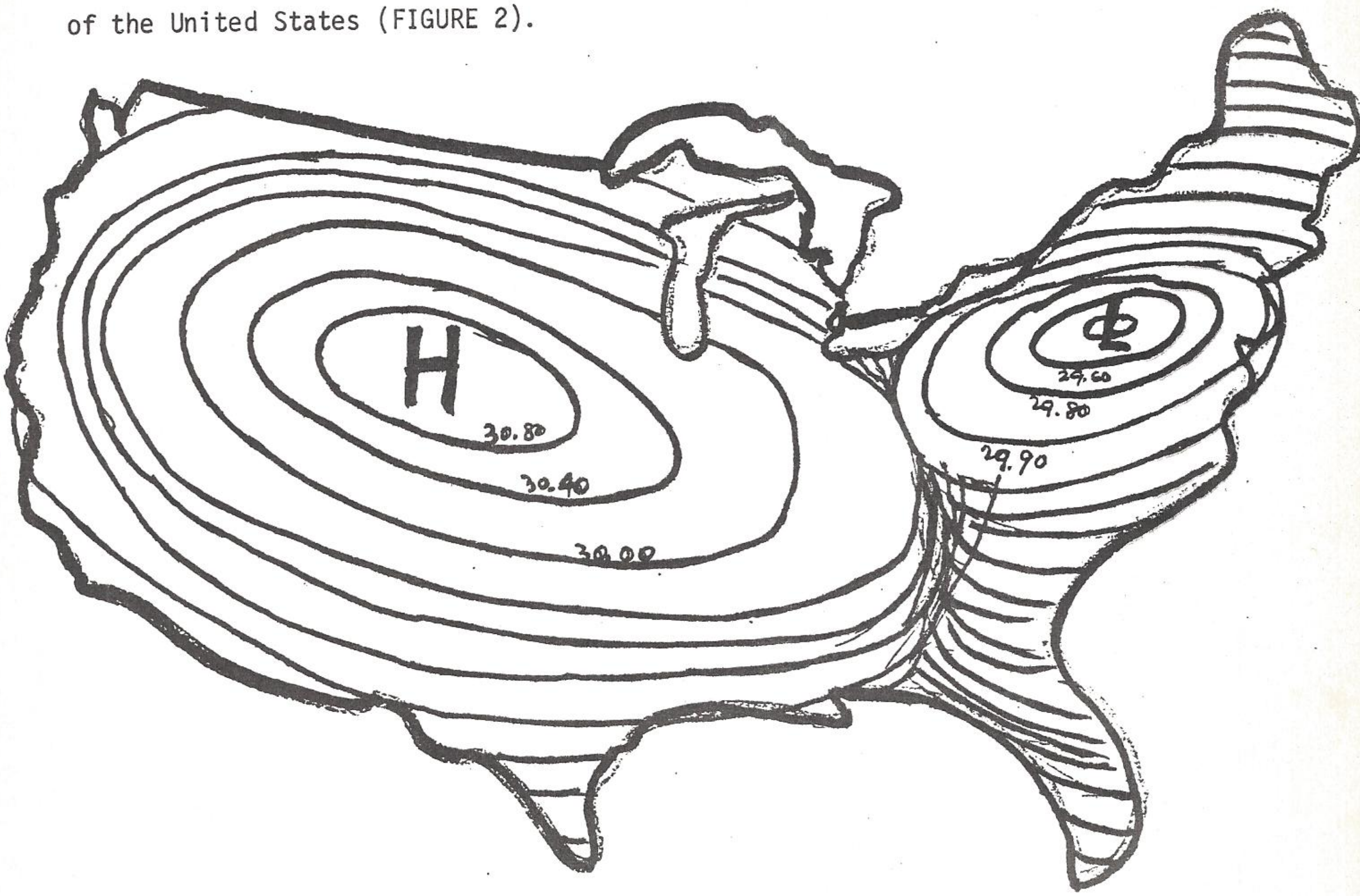


FIGURE 2

Shown in the Weather Map of FIGURE 2 are a HIGH PRESSURE AREA and a LOW PRESSURE AREA. The numbered ovals are called ISOBARS (lines of equal BAROMETRIC PRESSURE [inches of mercury]).

The accurate construction of a Weather Map such as in Figure 2 requires weather stations at many different locations (combinations of longitude and latitude) in the United States. In conducting this experiment of measuring barometric pressures with the purpose of constructing such a weather map everyone will readily admit that it would be utterly foolish to move NORTH and SOUTH at the middle of the United States (i.e., at the average longitude of the nation), and just measure the barometric pressure at a certain number of latitudes along this NORTH-SOUTH center line, and then, similarly, to take the EAST-WEST center line (ave. latitude of the U.S.) and just measure the barometric pressure along a certain number of longitudes on this EAST-WEST center line. Oddly enough, however, this is exactly the type of unproductive experimentation which has been done in the past whenever we were testing ONE VARIABLE AT A TIME while keeping all other variables constant at their nominal values. We cannot estimate the BAROMETRIC PRESSURE in MAINE or in the STATE OF WASHINGTON just from data on two center lines through the United States ----- that would be extrapolation. This is what L. Seder* calls the GUINEA-PIG FALLACY.

* L. Seder: Technique of Experimenting in the Factory MECHANICAL ENGINEERING, July 1948, pp. 593-598.

WHY STATISTICAL DESIGNS ARE THE MOST PRODUCTIVE

A statistically designed experiment has certain desirable features. These are

(1) It adequately covers the "territory" we want to map.

(2) It is balanced, i.e., it considers each level of each factor in its combinations, thus avoiding the need for extrapolations.
(The weather map example along the center lines of the U.S. was not balanced, and thus gave no information about the four corners of the U.S.)

(3) It is more economical because it produces usable maps without the added delay of making additional spot measurements after the experiment.

In summary, a statistically designed experiment is actually the most logical layout for a planned "MAP" of an experimental "TERRITORY". Everyone must admit that this is the most productive and most economical way of experimenting. To increase productivity let us design productive experiments!

ECONOMICS OF STATISTICAL EXPERIMENTAL DESIGN

A correctly designed experiment for a given situation is one which yields exactly the needed amount of information for a correct decision ---- no more and no less.

Let I_o = The Needed Amount of Information For a Statistical Decision.

Let I_- denote INSUFFICIENT INFORMATION

Let I_+ denote EXCESS INFORMATION

Associated with the difference $(I_o - I_-)$ is a loss due to lack of needed information for a correct decision. Call it L_1.

Furthermore, associated with the difference $(I_+ - I_o)$ is a waste due to excessive testing for extra information. Call it L_W.

L_1, due to lack of engough information to make a correct statistical decision involves all the losses incurred from an incorrect decision.

Let P_1 = PROBABILITY of an incorrect decision under I_-.

The expected dollar loss due to I_- is then $E_1 = L_1 P_1$.

L_W, due to wasted information involves all the extra testing time, material, and labor over and above what would be needed for the required information I_o.

Thus, $L_W = NC_s + Ht$

This type of loss is a sure thing (i.e., with PROBABILITY = 1) whenever excess information is gathered from an experiment.

N = Number of Extra Test Specimens

C_s = Cost Per Test Specimen

t = Extra Hours of Testing

H = Hourly Labor Rate for Testing

P_1 becomes 1 for $I_- = 0$.

P_1 becomes 0 for $I_- = I_o$.

Let us take $P_1 = 1 - \frac{I_-}{I_o}$ as an example

Then $E_1 = L_1 \left(1 - \frac{I_-}{I_o}\right)$.

We can now construct the following loss table:

	I_- INSUFFICIENT INFORMATION	I_+ EXCESS INFORMATION
Correct Decision	$\left(\text{Decision Probability} = \frac{I_-}{I_0}\right)$ LOSS = 0	$\left(\text{Decision Probability} = 1\right)$ LOSS = L_W
Incorrect Decision	$\left(\text{Decision Probability} = 1 - \frac{I_-}{I_0}\right)$ LOSS = L_1	$\left(\text{Decision Probability} = 0\right)$ LOSS = L_W
	EXPECTED LOSS $= E_1 = L_1\left(1 - \frac{I_-}{I_0}\right)$	EXPECTED LOSS $= L_W$

In the long run, when experiments are designed haphazardly, it will turn out that the probabilities of insufficient information and excess information are both equal to .5. Hence, after a total of T haphazard experiments, we can say that T/2 will have insufficient information and T/2 will have excess information.

The TOTAL EXPECTED LOSS per experiment will then be

$$\overline{E}_T = 1/2 \cdot \overline{L}_1 \cdot (1 - \bar{\rho}) + 1/2 \cdot \overline{L}_W$$ [* See Footnote]

where $\bar{\rho}$ is the average insufficiency ratio $\left(\frac{I_-}{I_0}\right)$ on these experiments which turn out to have insufficient information.

Furthermore, $\overline{L}_1$ = Ave. loss due to each insufficient experiment which leads to an incorrect decision,

and $\overline{L}_W$ = Ave. loss due to excess time, equipment, material, and labor in these experiments which yield excess information.

* This formula can be generalized to read

$$\overline{E}_T = (p_-) \; \overline{L}_1 (1 - \bar{\rho}) + (p_+) \; \overline{L}_W$$

where p_- = Tendency (or Probability) of insufficient experimentation,

and p_+ = Tendency (or Probability) of excessive experimentation. $(p_- + p_+) = 1$

AN ACTUAL EXAMPLE

ACCEPTANCE TESTING OF A NEW SPECIAL TYPE OF BEARING MATERIAL

CORRECTLY DESIGNING THE TEST

It is claimed that the new bearing material triples the B_{10} life when compared to AIR MELT material. The correct way to design a test for accepting or rejecting the new bearing material vs. AIR MELT material is to use the NOMOGRAPH of FIGURE 3. We make use of the fact discovered in our past experience that the Weibull Slope of the ball bearings is 1.3, and take the B_{10} Life Ratio to be 3 (as claimed). Then for 90% confidence the NOMOGRAPH tells us to test 11 bearings made of the new material versus 11 bearings made of the AIR MELT material (as control). This would then require us to test a total of 22 bearings for (11 - 1) (11 - 1) = 100 TOTAL DEGREES OF FREEDOM.

INCORRECTLY DESIGNING THE TEST

Somebody who is ever-zealous might claim that we should test enough bearings to estimate the TRUE MEAN LIFE within $\pm$ 10% for the 3-SIGMA LIMITS on the mean life. For a Weibull Slope of 1.3 and a mean life of 1000 hours, the sigma of the population would be 775 hours. Hence, a sample of size N would yield a sigma for the MEAN LIFE equal to $\frac{775}{\sqrt{N}}$.

Now, what is being required in this test design is that $3 \times \frac{775}{\sqrt{N}}$ be only 10% of the 1000 hours of mean life.

Thus, $3 \times \frac{775}{\sqrt{N}} = 100$, or $\sqrt{N} = \frac{2325}{100} = 23.25$

$\therefore N = (23.25)^2 = 540$ bearings per material

CONCLUSION

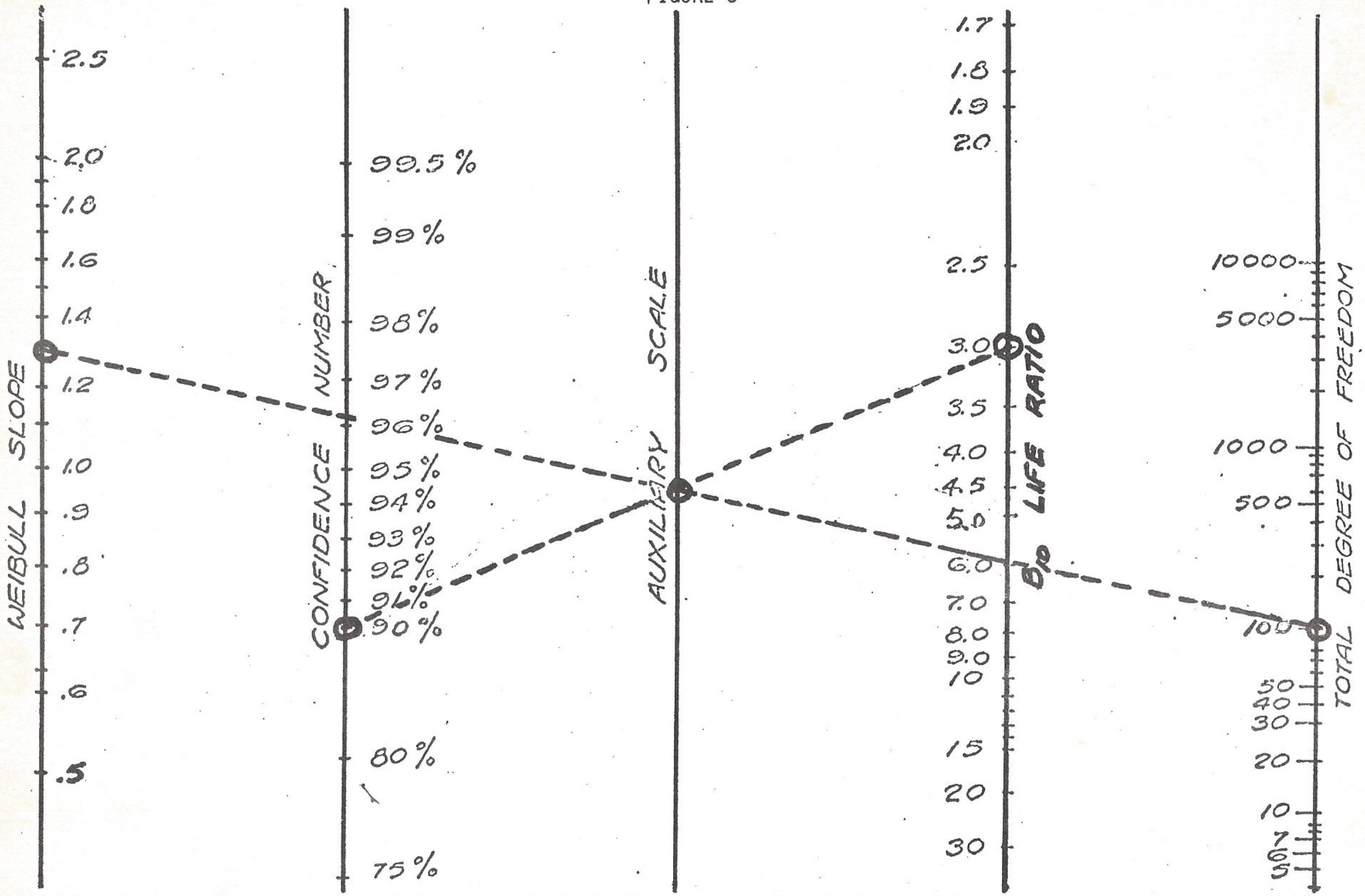
It can be seen that the INCORRECT TEST DESIGN (based on the incorrect criterion of closely estimating the true mean life) would require the testing of a total of 1080 bearings as compared to 22 bearings in the CORRECT TEST DESIGN.

Obviously the INCORRECT TEST DESIGN is a big waste of time and money, for it is an attempt to seek much more information than is really needed for the decision to adopt the new bearing material in place of the AIR MELT material.

Going in the other direction, toward insufficient information, somebody who is not familiar with statistical behavior might just test one new type bearing versus exactly one AIR MELT bearing and reject the new because he happened to select a low life new type bearing and a long life AIR MELT bearing. As a consequence, his decision to reject the new type would deprive the manufacturer of all the good publicity and extra sales generated by the new bearing material. THE MORAL OF THIS WHOLE LESSON IS: DESIGN YOUR EXPERIMENTS TO YIELD THE RIGHT AMOUNT OF INFORMATION FOR YOUR TEST OBJECTIVE.

CONFIDENCE NUMBER NOMOC ?H AT B-10 LEVEL

FIGURE 3



1. Connect total degrees of freedom with Weibull slope and locate intersection point on auxiliary scale.
2. Connect life ratio with intersection point and continue to intercept on confidence number.
3. For unequal Weibull slopes perform operation for each slope and average the confidence numbers so obtained.

In above example - Total Degrees of Freedom = 100, Weibull Slope = 1.3, Life Ratio = 3, Confidence No. = 90%.