STATISTICAL BULLETIN
Reliability & Variation Research

LEONARD G. JOHNSON
EDITOR

DETROIT RESEARCH INSTITUTE
P.O. Box 36504 • Grosse Pointe, MI 48236 • (313) 886-8435

WANG H. YEE
DIRECTOR

Volume 23
Bulletin 8

February, 1994
Page 1

THE SAMPLE SIZE THEORY OF EVIDENCE (BOTH FROM SINGLE SAMPLES AND SEQUENTIAL TESTS)

INTRODUCTION

Evidence is a factor which is involved in any kind of decision making that intelligent human beings daily must face. It comes into play in courts of justice, as well as in all scientific projects. We all must agree that it is only right that enough evidence be obtained in any decision making process before we ever conclude that our decision is a fair and valid one. In this bulletin we present the principles involved in the theory of evidence based on adequate statistical sampling, together with the gain and loss basis for coming up with what level of evidence or odds in favor of being correct is needed to avoid social injustice in courts of law, or financial failure in any consumer product business where adequate product assurance is demanded, both by customer and seller.

The discussion of evidence is centered around an argument between two individuals, named Mr. Gambler (Mr. G) and Mr. Evidence Expert (Mr. EE). They get into their argument over a new pet design which Mr. G is desirous of getting accepted and ready to be sold to his company's customers. In the ensuing discussion it becomes very clear that Mr. G is unaware of the proper level of evidence needed for his product's reliability, simply because he failed to establish that evidence level from gains and losses, and the number of test specimens needed in order to assure profitability.

A Discussion Between Mr. Gambler and Mr. Evidence Expert

"What in the world do you call evidence?" yelled Mr. Gambler (Mr. G) as he was arguing with Mr. Evidence Expert (Mr. EE). The topic of discussion was a proposed design which Mr. Gambler was banking on to bring his company new record sales. Mr. Evidence Expert patiently waited until Mr. G cooled down enough to listen. Then Mr. EE eased into the discussion by telling Mr. G that his favorite design had not as yet demonstrated enough evidence to be accepted as one which could live up to the guaranteed B-1% service life of 730 days (2 years), which means that no more than 1% of the items are allowed to fail in 2 years. "You see", said Mr. EE, "If we don't pass the required B-1% life of 730 days, we stand to lose 10 million dollars, while our profits from selling a complying design are only half a million dollars." . " This means", continued Mr. EE, "That just in order to break even, we must have 20 to 1 odds in our favor, i.e., the ratio (10 Million/Half Million).". "But, we don't just want to break even. If we want our profits from complying lots to be **twice** the losses from non-complying lots, then we need 40 to1 odds in favor of compliance with the B-1% life of 730 days." . "So, you see, Mr. G, your favorite design, which showed a B-1% life of 905 days from a test sample of 10 specimens with a Weibull slope of 2.5, has only odds of 9:1 in favor of complying with the B-1% goal of 730 days." . "I thought there was plenty of evidence that my design would comply with a B-1% life of 730 days", said Mr. G. "But," Mr. EE continued, "The evidence you have demonstrated is only equal to the natural logarithm of the 9:1 odds, i.e., $\ln(9) = 2.19722$ units of evidence, whereas, the required odds of 40:1 for twice as many dollars gained as lost, requires an evidence equal to the natural logarithm of 40, i.e., $\ln(40) = 3.68888$ units of evidence." "So, you see, Mr. G, you are still lacking $3.68888 - 2.19722 = 1.49166$ units of evidence." "What, then, does that mean as far as our efforts are concerned?", asked Mr. G. "It means", replied Mr. EE, "That your test sample was too small to yield the needed evidence." "How large should the test sample have been to yield the desired evidence?" asked Mr. G.

"Well", replied Mr. EE, "Evidence increases as the **Square Root** of the sample size, so in order to increase evidence from 2.19722 to 3.68888, the square root of sample size must be multiplied by the factor 3.68888/2.19722 = 1.67889. Since the square root of 10 is 3.16228, we must multiply it by 1.67889 to obtain the value (3.16228)(1.67889) = 5.30912 for the needed square root of sample size." "Thus, the test you ran should have had a sample size equal to $(5.30912)^2$ = 28.19, or 29 to the next integer." . "This is assuming that one data set of 29 items would still show a B-1% life of 905 days and a Weibull slope of 2.5. But, Mr. G, you don't have to feel bad that you tested only 10 items to begin with. You can increase the evidence to the desired level by taking a second sample with 1.49166 additional units of evidence on top of what that first sample of 10 gave. Assuming a test B-1% life of 905 days and a Weibull slope of 2.5, the second sample size should have a square root equal to (1.49166/2.19722)10 = 2.14682. " " This means that you should test a second sample of size equal to $(2.14682)^2$, i.e., 4.6, or 5 to the next integer." .

"In other words, by first testing 10 items with B-1% life = 905 days and Weibull slope = 2.5, and then another sample of 5 with the sample B-1% life of 905 days and Weibull slope of 2.5, you would have the required evidence for gaining twice as much as you would lose (in the long run)." . "This type of testing", added Mr. EE, "is known as **Sequential Testing**, and can be a considerable saving in total number of specimens needed in compliance testing to a required life goal. In this case one single sample of 29 items can be replaced by two samples of size 10 and 5 respectively, thus making for a total of only 15 items. Thus, in this example, sequential testing reduced that the total number of items tested to about half of what the single large test of 29 items would have involved for the same required total evidence."

The General Theory of Sample Size Evidence Involved in the Previous Discussion

Basic Theorem # 1: The evidence of compliance to a service life goal which can be demonstrated from a sample is directly proportional to the square root of the sample size.

Basic Theorem # 2: When samples are tested in sequence for compliance to a service life goal, the total evidence is equal to the sum of the evidences from the separate test samples.

As a corollary to the above basic theorems, we can state that the way to reduce the total number of items tested is to test several small samples in sequence instead of testing one large sample all by itself.

Conclusion

From the discussion we presented between Mr. G and Mr. EE, we arrive at two important conclusions, as follows:

Conclusion # 1: It is not enough to have some positive amount of evidence of compliance to a service life goal, if this amount of evidence does not guarantee the profitability we want in spite of possible losses due to any cases of non-compliance. Then, to increase the evidence, we test additional samples for additional evidence to add onto what has already been obtained from the initial test. We continue such sequential testing of further samples until we have accumulated the evidence level dictated by the desired profitability ratio.

Conclusion # 2: It is absolutely necessary to obey **Rule # 1** in the design of a reliability testing program. This rule states that we cannot hope to design a valid life test on a design until we first determine the dollars lost due to non-compliance, as well as the dollars gained when in compliance.

Appendix of Special Questions

Question # 1: What did those 9:1 odds come from for that first test of 10 specimens?

Answer: The 9:1 odds were obtained by using the DRI's **CARS** software "GOALCNF" program described in Bulletin #3 of Volume 23 (July 1993). In that program we are asked for

(1)	The Quantile Level	(Q = .01)
(2)	The B-Q Life Goal	(730)
(3)	The Sample B-Q Life	(903)
(4)	The Sample Weibull Slope	(2.5)
(5)	The Sample Size	(10)

With these inputs, the "GOALCNF" program gives a confidence of 90% for the compliance to the B-1% goal of 730 hours. So, 90% confidence (i.e., C = .9) yields **Odds** of .9/(1 - .9) = .9/.1 = 9 to 1 .

Question # 2: What happens when the B-Q life of the sample is less than the B-Q Goal Life?

Answer: The "GOALCNF" program gives a confidence **less than** 50%, which is equivalent to a **negative evidence**. For Example, for a confidence C = .4 (40%), the evidence becomes ln[.4/(1-.4)] = ln(.4/.6) = ln(.66667) = -.40547.

Question # 3: What common error did Mr. G commit when he came to the false conclusion that the first test sample of ten items validated the design's acceptability?

Answer: Mr. G failed to take **Gains and Losses** into account as the factors which determine the **Odds Required** in favor of compliance to the **Service Life Goal**. Instead Mr. G just thought that 90% confidence (9:1 odds) looked good enough for acceptance, whereas the **Needed Odds** were 40:1 for the desired profitability when **Gains and Losses** were taken into consideration.