

CBA · NAU

College of Business
Administration
Northern Arizona
University
Box 15066
Flagstaff AZ 86011

**Logical Extremes,
Beta and the Power of
the Test¹**

**Working Paper Series 02-03—
Revised December 2002**

James V. Pinto

Professor of Economics

(928) 523-7356

Fax (928) 523-7331

James.Pinto@nau.edu

Pin Ng

Assistant Professor of Economics

David S. Allen

Associate Professor of Finance

College of Business Administration

Northern Arizona University

Box 15066

Flagstaff, Arizona 86011-5600 USA

¹ The authors wish to thank two anonymous referees for their detailed and useful comments. All remaining errors are the responsibility of the authors.

Logical Extremes, Beta and the Power of the Test

James V. Pinto, Ph.D.,
Pin Ng, Ph.D. and
David S. Allen, Ph.D.

1. INTRODUCTION

It is not an easy nor smooth task to calculate the power of a test for beginning statistics students. Until computer algorithms and simulations were designed to help in this process (Doane, Mathieson and Tracy, 2001), the best one could do was to calculate a single point on the power function representing $1 - \beta$ (Kvanli, Ravur and Guynes, 2000). Yu and Behrens (1994) illustrate the usefulness of simulations in determining the power of a test. It is possible to calculate every point on the power curve, but in-class demonstrations are usually limited to single points due to computational constraints and the repetitive nature of the process. Logical extremes, as detailed below, were usually ignored. The purpose of this note is to demonstrate that the logical extremes are nontrivial and that one of the extremes results in calculations of beta and the power of the test that are inconsistent with the common pedagogy. A detailed history of the nature of hypothesis testing with a section on the power of a test in light of the Fisher approach versus Neyman-Pearson approach can be found in Lehmann (1993). He concluded that the two approaches could be unified by combining the best features of both. This historical debate will not be addressed in this paper.

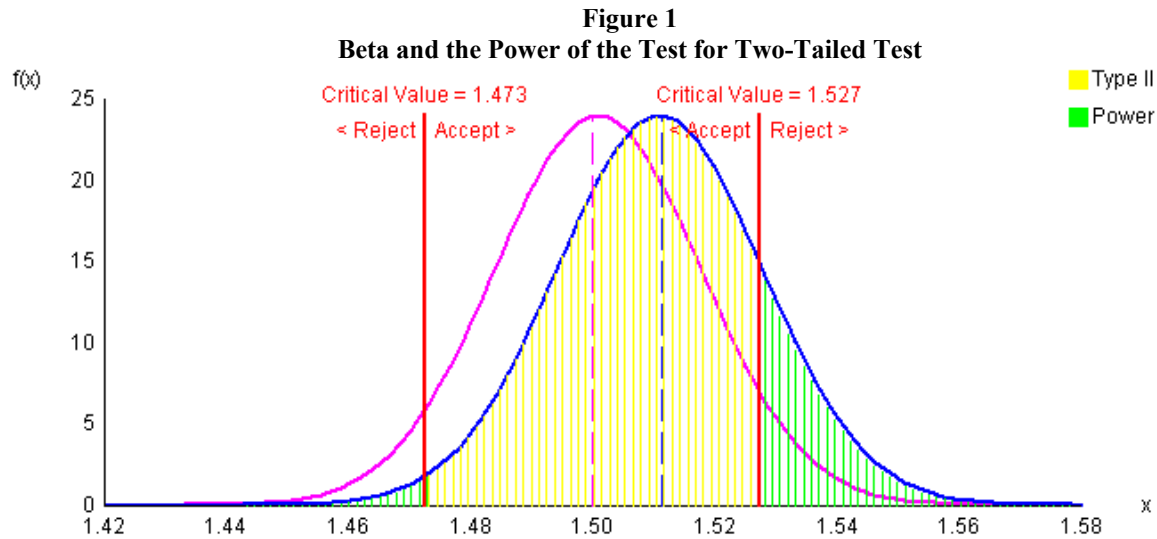
2. CALCULATION PROCESS

We will use the familiar test on the population mean of a normal distribution with a known variance to set up the stage for logical extremes. We assume that the data generating process is $\phi(\mu, \sigma^2)$ where ϕ is the probability density function of a normal distribution with an unknown mean μ and a known variance σ^2 .

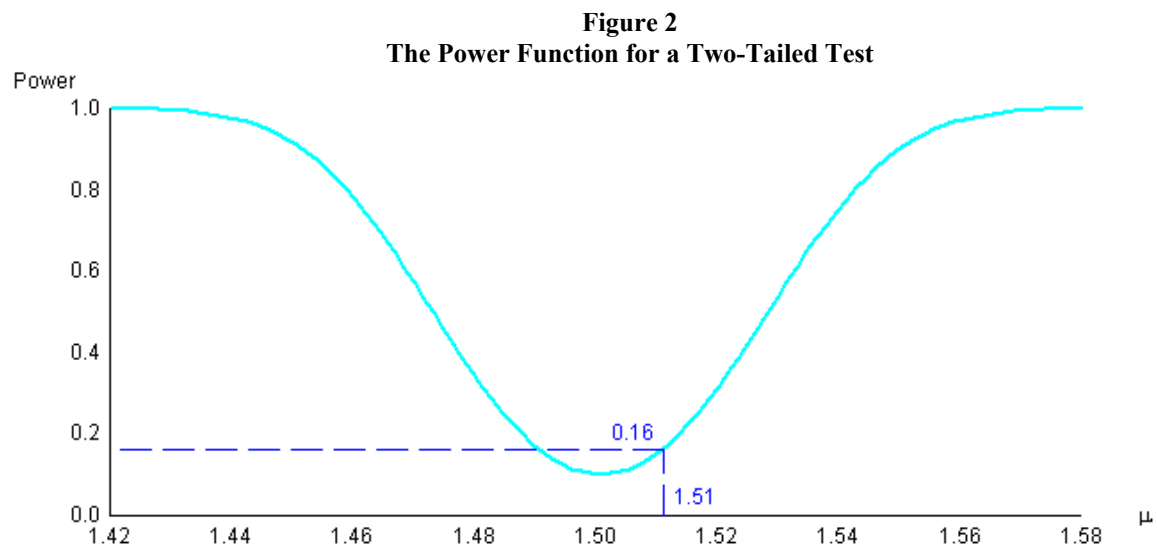
We first consider testing the simple null hypothesis $H_0 : \mu = \mu_0 = 1.5$ against a two-sided composite alternative $H_1 : \mu \neq \mu_0$ assuming $\sigma = 0.1$ and $n = 36$. In our example, under the null hypothesis, the sampling

distribution of the sample mean is normal, centered at $\mu_0 = 1.5$, with a standard error of $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.1}{\sqrt{36}}$.

This is illustrated by the normal distribution on the left in Figure 1 which was generated using Visual Statistics (Doane, Mathieson and Tracy, 2001). The size of the z test (the probability of committing a type I error), α , is chosen to be 10%. Figure 1 also presents a second normal distribution on the right, that is the sampling distribution of the sample mean at the specific value of $\mu = 1.51$ when the null hypothesis is false. The shaded area under this sampling distribution in the middle acceptance region measures the probability of committing a type II error while the shaded areas in the rejection regions provide the power of the test, $1 - \beta$.



The power function, defined as the probability of rejecting the null given a specific value of μ , for the z test is given by $\pi(\mu) = P(\bar{X} > 1.527 \cup \bar{X} < 1.473 | \mu) = \Phi\left(\frac{1.473 - \mu}{\frac{0.1}{\sqrt{36}}}\right) + 1 - \Phi\left(\frac{1.527 - \mu}{\frac{0.1}{\sqrt{36}}}\right)$, where Φ is the cumulative distribution function of the standard normal distribution. This yields $1 - \beta = 0.16$ on the power function at this single point of $\mu = 1.51$. The entire power function is shown in Figure 2 and resembles an upside down normal distribution. It should be noted that the power function in this example approaches one asymptotically as we move away from 1.5 in either direction.



For the one-tailed test, we consider the composite null hypothesis $H_0 : \mu \leq \mu_0$ versus the composite alternative $H_1 : \mu > \mu_0$. The sampling distributions of the sample mean under the null of $\mu = \mu_0 = 1.5$ and the specific alternative of $\mu = 1.51$ are illustrated, respectively, by the left and right distributions in Figure 3. The

power function is given by $\pi(\mu) = P(\bar{X} > 1.521 | \mu) = 1 - \Phi\left(\frac{1.521 - \mu}{\frac{0.1}{\sqrt{36}}}\right)$. The power of the test at

$\mu = 1.51$ is 0.25 and the whole power function is illustrated in Figure 4. It should be noted that the power function in this example approaches one and zero asymptotically as we move away from 1.5.

Figure 3
Beta and the Power of the Test for One-Tail Test

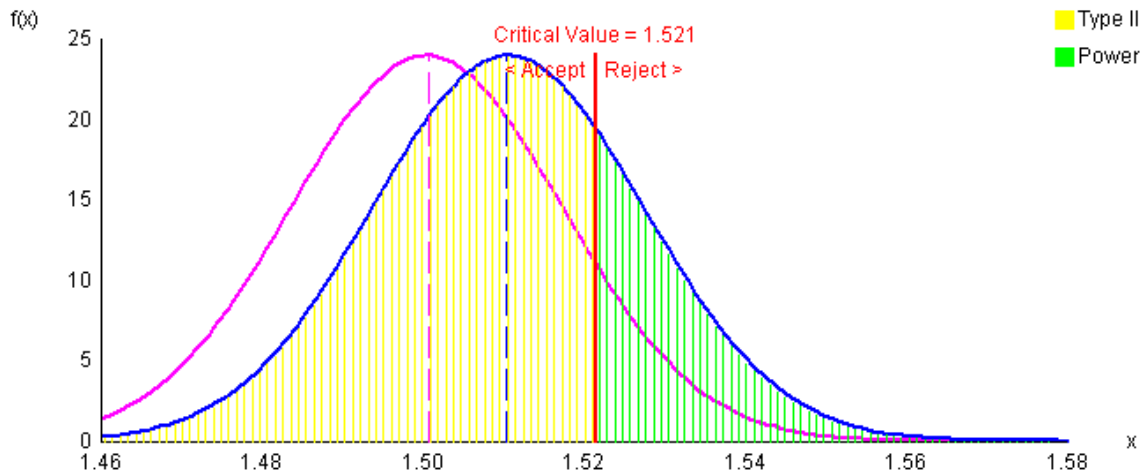
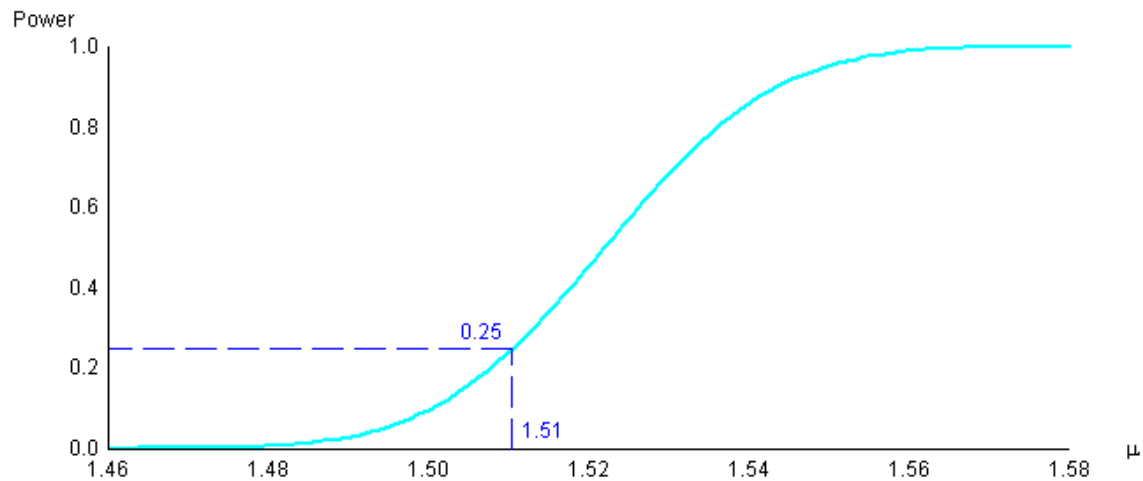


Figure 4
The Power Function for a One-Tail Test



3. THE LOGICAL EXTREMES

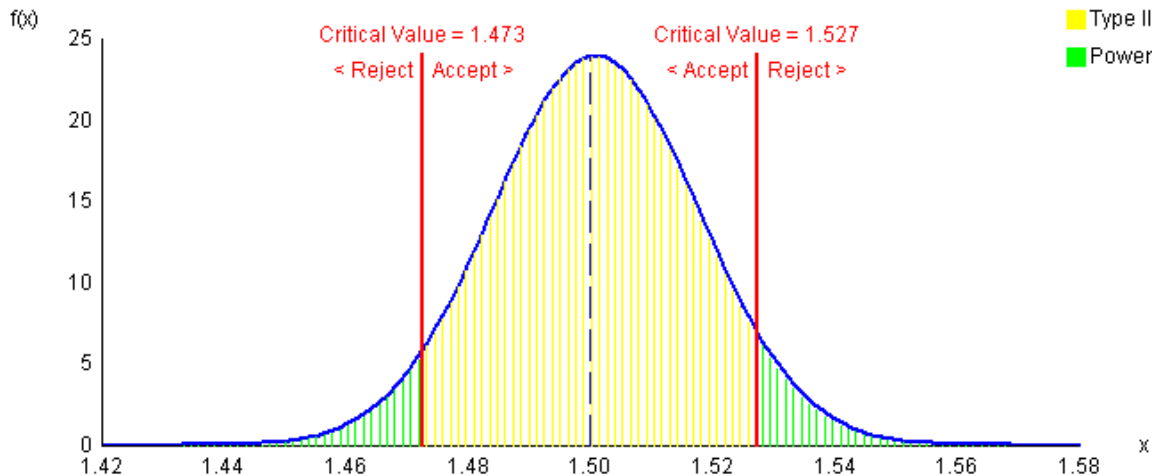
The first logical extreme occurs when the two sampling distributions in both Figure 1 and 3 pull apart to a point where there is very little overlap between them. In this case the power of the test approaches 1 (as the two distributions in Figure 1 and Figure 3 separate), and the β approaches zero. This can easily be seen in Figure 2 when μ moves away from $\mu_0 = 1.5$ in both directions and also in Figure 4 when μ moves away from $\mu_0 = 1.5$ to the right.

The second logical extreme occurs when μ approaches $\mu_0 = 1.5$. One subtle point that has not usually been made explicitly in the definition of the power function is the fact that the power function yields the power of a test only when the null hypothesis is false. Under the null hypothesis, the power function provides the probability of committing a type I error instead of the power of the test. As presented in Figure 2, the power function yields the 10% size at $\mu = \mu_0 = 1.5$ according to the accepted pedagogy (Freund and Walpole, 1987).

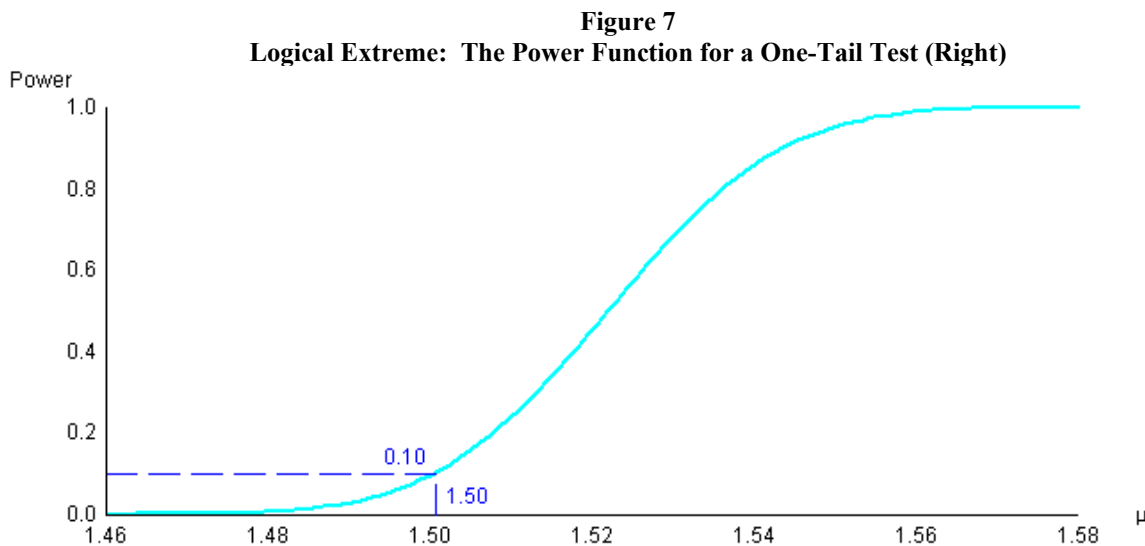
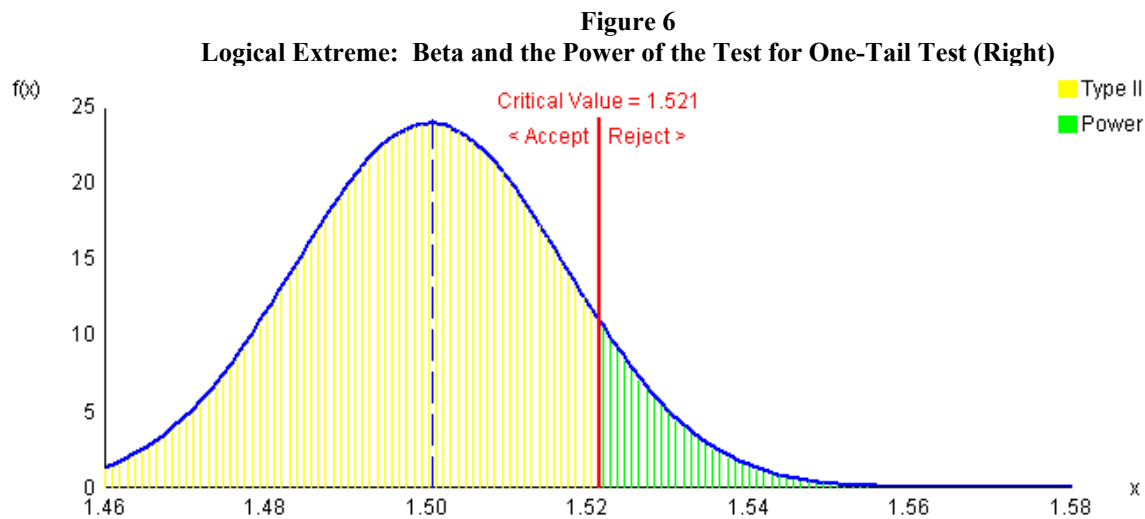
This is where this logical extreme breaks down. In this case, it is incorrect to say that “the power of the test is 10% or the probability of committing a type II error is 90%” since the power is defined as “the probability of rejecting a false null hypothesis” and a type II error is defined to be “the failure to reject the false null hypothesis.” But there is no false null hypothesis when the null hypothesis is true at $\mu = \mu_0 = 1.5$. β should be undefined rather than having a maximum value of 90%. Thus, the contention that the power of the test is at a minimum of 10% under the null cannot be supported because there is no false null hypothesis when the null hypothesis is true at $\mu = \mu_0 = 1.5$. Both the power and the probability of committing a type II error are undefined under this logical extreme.

Unfortunately, this is usually overlooked. For example Figure 5 is generated for $\mu = \mu_0 = 1.5$ using Visual Statistics which labels the area under the “accept” region as “Type II” error and the area under the “Reject” regions as “Power.” This misconception is also apparent in the few current textbooks which attempt to address the power function. In Berenson, Levine and Krehbiel (2002, pp. CD9-6) they state that “Interestingly, if the population mean is actually 368 grams, the power of the test is equal to α , the level of significance (which is 0.05 in this example), because the null hypothesis is actually true.” In Newbold, Carlson and Thorne (2003, p. 357), they make a similar statement that includes, “The power of the test at μ_0 equals α , ...”

Figure 5
Logical Extreme: Beta and the Power of the Test for Two-Tailed Test



Similarly, the sampling distribution and the power function for the one-tailed test are presented in Figure 6 and 7.



The power function in Figure 7 provides the probability of rejecting the null hypothesis over the whole range of values of μ . It can be used to obtain the power of the test only when $\mu > 1.5$. The size of the test is defined as the supremum of the power function over $\mu \leq 1.5$. This yields a size of 10% according to accepted pedagogy (Tamhane and Dunlop, 2000, and Mood, Graybill and Boes, 1974.) The power and β remain undefined under the null when $\mu \leq 1.5$.

4. CONCLUSIONS

There is a potential misuse of the power function under the logical extreme when the null hypothesis is true. The power function is defined to measure the probability of rejecting the null given any value of the parameter being tested. It can be used to obtain the power and the β values only under the alternative hypothesis. When the null is true, the power function can be used to obtain the size of the test. The power and the probability of committing a type II error are, however, undefined and, hence, the power function should not be used to obtain these values.

REFERENCES

- Berenson, M. L., Levine, D. M. and Krehbiel, T. C. (2002) *Basic Business Statistics, 8th edition*, Prentice Hall, New Jersey.
- Doan, D. P., Mathieson, K., and Tracy, R. L. (2001) *Visual Statistics 2.0*, McGraw-Hill/Irwin, New York, N. Y.
- Freund, J. E. and Walpole, R. E. (1987) *Mathematical Statistics*, Prentice-Hall, Inc. Englewoods Cliffs, N. J.
- Kvanli, A. H., Pavur, R. J. and Guynes, C. S. (2000) *Introduction to Business Statistics: A Computer Integrated, Data Analysis Approach, 5th edition*, South-Western College Publishing, Cincinnati, Ohio.
- Lehmann, E. L., “The Fisher, Neyman-Pearson Theories of Testing Hypothesis: One Theory or Two?” *Journal of the American Statistical Association*, vol. 88, no. 424, December 1993, pp.1242-49.
- Mood, A. M., Graybill, F. A. and Boes, D. C. (1974) *Introduction to the Theory of Statistics, 3rd edition*, McGraw-Hill, New York.
- Newbold, P., Carlson, W. L., and Thorne, B. (2003) *Statistics for Business and Economics, 5th edition*, Prentice Hall, Inc., Upper Saddle, NJ.
- Tamhane, A. C. and Dunlop D. D. (2000) *Statistics and Data Analysis from Elementary to Intermediate*, Prentice-Hall, Inc., Upper Saddle, NJ.
- Yu, C. H. and Behrens, J. T., “Identification of Misconceptions in Learning Statistical Power With Dynamic Graphics As a Remedial Tool,” *ASA Proceedings of the Section on Statistical Education*, 1994, pp. 242-6.