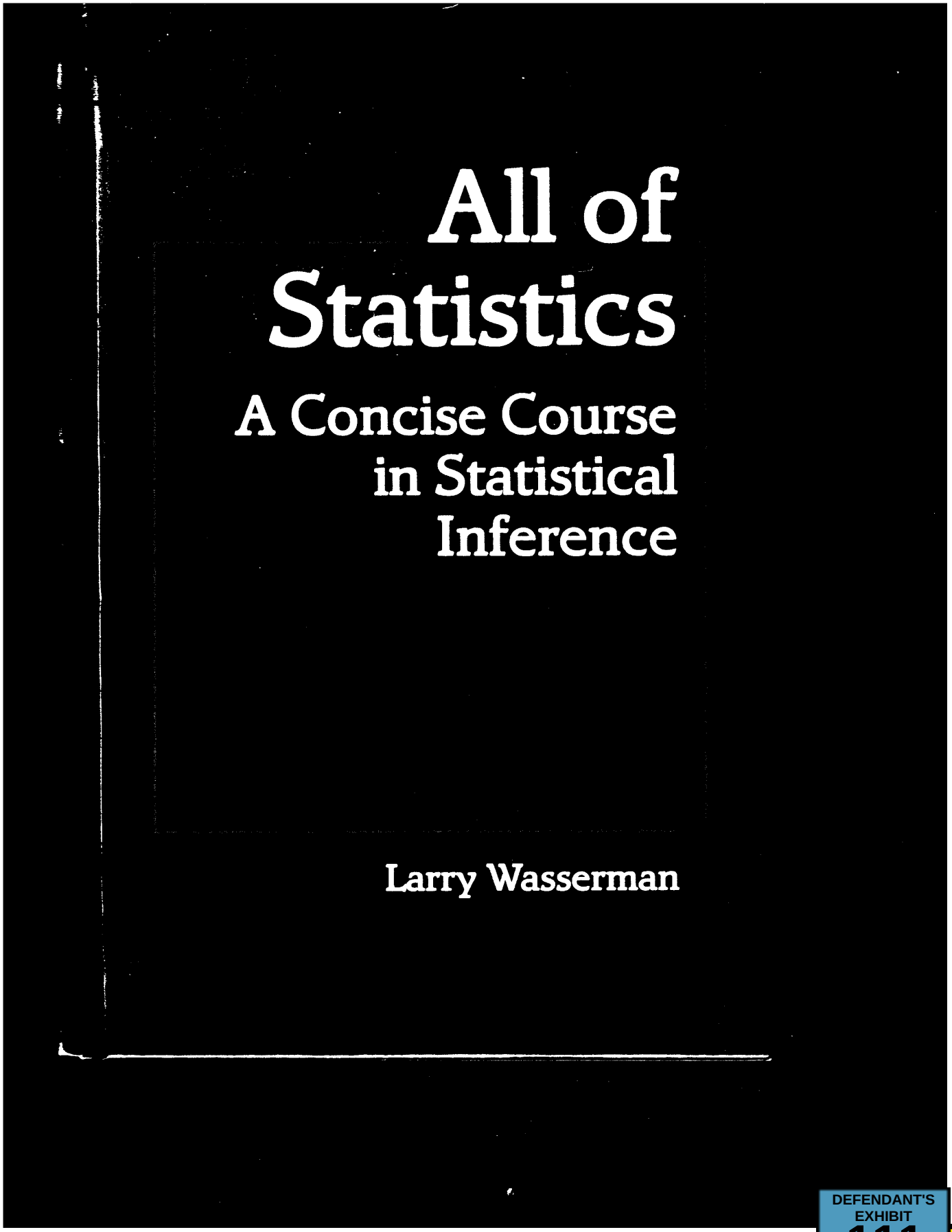
FILED
2024 Sep-25 AM 10:20
U.S. DISTRICT COURT
N.D. OF ALABAMA



DEFENDANT'S EXHIBIT 111
BIT 2

Springer Texts in Statistics

Advisors:
George Casella Stephen Fienberg Ingram Olkin

Social Sciences
tochastic Processes
te Statistics
Applications
Series and Forecasting,

ata in S-Plus
pendence, Interchangeability,

ultivariate, Time Series, and
and Response Surface

e Regression, Second Edition
essions: The Theory of Linear

dels and Statistical Inference
f Repeated Measurements
eriments
ratory Data Analysis

ng. Second Edition

ics
and Data Display: An Intermediate
ed SAS
. Volume I: Regression and

. Volume II: Categorical and

ence. Volume I: Probability,

ence. Volume II: Statistical Inference,

. Second Edition

Formulae
Control of Stochastic Systems

(continued after index)

Larry Wasserman

All of Statistics

A Concise Course in Statistical Inference

With 95 Figures

Springer

Larry Wasserman
Department of Statistics
Carnegie Mellon University
Baker Hall 228A
Pittsburgh, PA 15213-3890
USA
larry@stat.cmu.edu

Editorial Board

George Casella
Department of Statistics
University of Florida
Gainesville, FL 32611-8545
USA

Stephen Fienberg
Department of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213-3890
USA

Ingram Olkin
Department of Statistics
Stanford University
Stanford, CA 94305
USA

Library of Congress Cataloging-in-Publication Data
Wasserman, Larry A. (Larry Alan), 1959–
All of statistics: a concise course in statistical inference / Larry a. Wasserman.
p. cm. — (Springer texts in statistics)
Includes bibliographical references and index.
ISBN 0-387-40272-1 (alk. paper)
1. Mathematical statistics. I. Title. II. Series.
QA276.12.W37 2003
519.5—dc21 2003062209

ISBN 0-387-40272-1 Printed on acid-free paper.

© 2004 Springer Science+Business Media, Inc.
All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, Inc., 233 Spring St., New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.
The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed in the United States of America. (MVY)

9 8 7 6 5 4 3 (Corrected second printing, 2005)

springeronline.com

6.12 Definition. *An estimator is* **asymptotically Normal** *if*

$$\frac{\widehat{\theta}_n - \theta}{\mathsf{se}} \rightsquigarrow N(0,1). \tag{6.8}$$

6.3.2 Confidence Sets

A $1-\alpha$ **confidence interval** for a parameter θ is an interval $C_n = (a,b)$ where $a = a(X_1,\ldots,X_n)$ and $b = b(X_1,\ldots,X_n)$ are functions of the data such that

$$\mathbb{P}_\theta(\theta \in C_n) \geq 1-\alpha, \quad \text{for all } \theta \in \Theta. \tag{6.9}$$

In words, (a,b) traps θ with probability $1-\alpha$. We call $1-\alpha$ the **coverage** of the confidence interval.

Warning! C_n is random and θ is fixed.

Commonly, people use 95 percent confidence intervals, which corresponds to choosing $\alpha = 0.05$. If θ is a vector then we use a **confidence set** (such as a sphere or an ellipse) instead of an interval.

Warning! There is much confusion about how to interpret a confidence interval. A confidence interval is not a probability statement about θ since θ is a fixed quantity, not a random variable. Some texts interpret confidence intervals as follows: if I repeat the experiment over and over, the interval will contain the parameter 95 percent of the time. This is correct but useless since we rarely repeat the same experiment over and over. A better interpretation is this:

> On day 1, you collect data and construct a 95 percent confidence interval for a parameter θ_1. On day 2, you collect new data and construct a 95 percent confidence interval for an unrelated parameter θ_2. On day 3, you collect new data and construct a 95 percent confidence interval for an unrelated parameter θ_3. You continue this way constructing confidence intervals for a sequence of unrelated parameters $\theta_1, \theta_2, \ldots$. Then 95 percent of your intervals will trap the true parameter value. There is no need to introduce the idea of repeating the same experiment over and over.

6.13 Example. Every day, newspapers report opinion polls. For example, they might say that "83 percent of the population favor arming pilots with guns." Usually, you will see a statement like "this poll is accurate to within 4 points

...of the time." They ...
...for the true but unk...
...with guns. If you form ...
...your life, 95 percent ...
...is true even though you ...
...question) every day. ■

...14 Example. The fact that ... about θ is confusing. Consi... Let θ be a fixed, known ... variables such that $\mathbb{P}(X_i = ...) = ... - X_i$ and suppose that you ... "confidence interval" which ...

$$C = \begin{cases} \{Y_1 - 1 \\ \{Y_1 - 1 \end{cases}$$

... can check that, no matter wh... 75 percent confidence interval ... set $Y_1 = 15$ and $Y_2 = 17$. Then ... wever, we are certain that $\theta = ...$ ...ement about θ you would pr... ...hing wrong with saying that ... not a probability statement ...

In Chapter 11 we will discuss Ba... ...e a random variable and we ... ...rticular, we will make statement... ... data, is 95 percent." However, ...-belief probabilities. These Bay... ...rameter 95 percent of the time.

...15 Example. In the coin flippi... ... $= \log(2/\alpha)/(2n)$. From Hoeffdi...

$$\mathbb{P}(p \in ...$$

... every p. Hence, C_n is a $1-\alpha$ ...

As mentioned earlier, point est... ...ribution, meaning that equation ... ...se we can construct (approxima...

...ptotically **Normal** *if*

... (6.8)

...meter θ is an interval $C_n = (a, b)$ ... X_i are functions of the data

... for all $\theta \in \Theta$. (6.9)

... We call $1-\alpha$ the **coverage** of ... intervals, which corresponds ... use a **confidence set** (such as ...

... how to interpret a confidence ... bability statement about θ since ... Some texts interpret confidence ... over and over, the interval will ... This is correct but useless since ... over. A better interpretation

...struct a 95 percent confidence ... you collect new data and con... for an unrelated parameter θ_2, ... construct a 95 percent confi...meter ... You continue this way ... a sequence of unrelated param... intervals will trap the true ... introduce the idea of repeating

... opinion polls. For example, they ... favor arming pilots with guns." ... is accurate to within 4 points ... percent of the time." They are saying that 83 ± 4 is a 95 percent confidence interval for the true but unknown proportion p of people who favor arming pilots with guns. If you form a confidence interval this way every day for the rest of your life, 95 percent of your intervals will contain the true parameter. This is true even though you are estimating a different quantity (a different poll question) every day. ■

6.14 Example. The fact that a confidence interval is not a probability statement about θ is confusing. Consider this example from Berger and Wolpert (1984). Let θ be a fixed, known real number and let X_1, X_2 be independent random variables such that $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = 1/2$. Now define $Y_i = \theta + X_i$ and suppose that you only observe Y_1 and Y_2. Define the following "confidence interval" which actually only contains one point:

$$C = \begin{cases} \{Y_1 - 1\} & \text{if } Y_1 = Y_2 \\ \{(Y_1 + Y_2)/2\} & \text{if } Y_1 \neq Y_2. \end{cases}$$

You can check that, no matter what θ is, we have $\mathbb{P}_\theta(\theta \in C) = 3/4$ so this is a 75 percent confidence interval. Suppose we now do the experiment and get $Y_1 = 15$ and $Y_2 = 17$. Then our 75 percent confidence interval is $\{16\}$. However, we are certain that $\theta = 16$. If you wanted to make a probability statement about θ you would probably say that $\mathbb{P}(\theta \in C | Y_1, Y_2) = 1$. There is nothing wrong with saying that $\{16\}$ is a 75 percent confidence interval. But it is not a probability statement about θ. ■

In Chapter 11 we will discuss Bayesian methods in which we treat θ as if it were a random variable and we do make probability statements about θ. In particular, we will make statements like "the probability that θ is in C_n, given the data, is 95 percent." However, these Bayesian intervals refer to degree-of-belief probabilities. These Bayesian intervals will not, in general, trap the parameter 95 percent of the time.

6.15 Example. In the coin flipping setting, let $C_n = (\widehat{p}_n - \epsilon_n, \widehat{p}_n + \epsilon_n)$ where $\epsilon_n = \sqrt{\log(2/\alpha)/(2n)}$. From Hoeffding's inequality (4.4) it follows that

$$\mathbb{P}(p \in C_n) \geq 1 - \alpha$$

for every p. Hence, C_n is a $1-\alpha$ confidence interval. ■

As mentioned earlier, point estimators often have a limiting Normal distribution, meaning that equation (6.8) holds, that is, $\widehat{\theta}_n \approx N(\theta, \widehat{\mathsf{se}}^2)$. In this case we can construct (approximate) confidence intervals as follows.

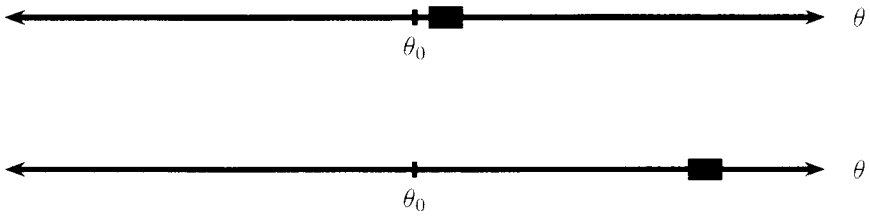


FIGURE 10.2. Scientific significance versus statistical significance. A level α test rejects $H_0 : \theta = \theta_0$ if and only if the $1 - \alpha$ confidence interval does not include θ_0. Here are two different confidence intervals. Both exclude θ_0 so in both cases the test would reject H_0. But in the first case, the estimated value of θ is close to θ_0 so the finding is probably of little scientific or practical value. In the second case, the estimated value of θ is far from θ_0 so the finding is of scientific value. This shows two things. First, statistical significance does not imply that a finding is of scientific importance. Second, confidence intervals are often more informative than tests.

10.2 p-values

Reporting "reject H_0" or "retain H_0" is not very informative. Instead, we could ask, for every α, whether the test rejects at that level. Generally, if the test rejects at level α it will also reject at level $\alpha' > \alpha$. Hence, there is a smallest α at which the test rejects and we call this number the p-value. See Figure 10.3.

10.11 Definition. *Suppose that for every $\alpha \in (0, 1)$ we have a size α test with rejection region R_α. Then.*

$$\text{p-value} = \inf\left\{\alpha : T(X^n) \in R_\alpha\right\}.$$

That is, the p-value is the smallest level at which we can reject H_0.

Informally, the p-value is a measure of the evidence against H_0: the smaller the p-value, the stronger the evidence against H_0. Typically, researchers use the following evidence scale:

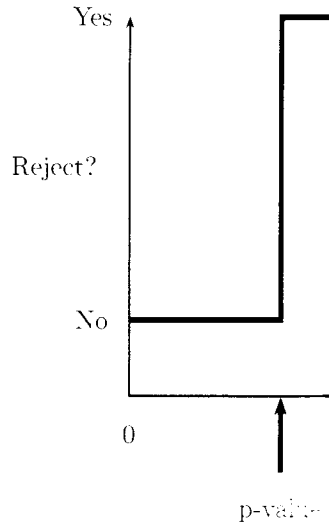


FIGURE 10.3. p-values explained. For
at level α? The p-value is the smalles
against H_0 is strong, the p-value wil

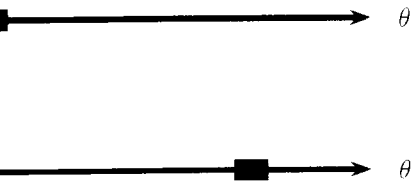
p-value	evidence
$< .01$	very
$.01 - .05$	strong
$.05 - .10$	weak
$> .1$	little

Warning! A large p-value is n
p-value can occur for two reasons:
has low power.

Warning! Do not confuse the p-v
not the probability that the null

The following result explains how

[2]We discuss quantities like $\mathbb{P}(H_0|\text{Data}$



statistical significance. A level α test
confidence interval does not include
Both exclude θ_0 so in both cases the
estimated value of θ is close to θ_0 so
practical value. In the second case, the
finding is of scientific value. This shows
not imply that a finding is of scientific
often more informative than tests.

is not very informative. Instead, we
reports at that level. Generally, if the
at level $\alpha' > \alpha$. Hence, there is a
call this number the p-value. See

$\in (0, 1)$ *we have a size* α *test*

$X^n \in R_\alpha \Big\}$.

which we can reject H_0.

the evidence against H_0: the smaller
against H_0. Typically, researchers use

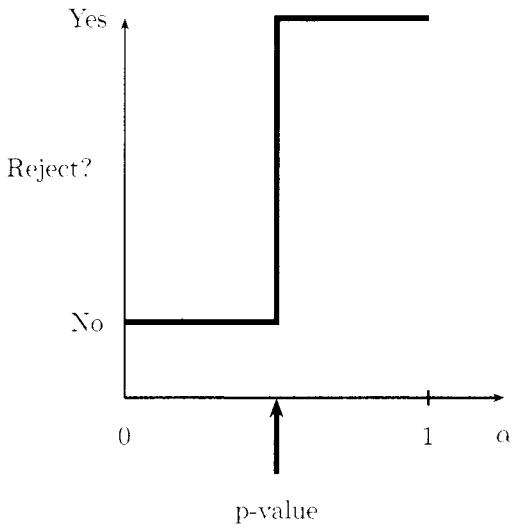


FIGURE 10.3. p-values explained. For each α we can ask: does our test reject H_0 at level α? The p-value is the smallest α at which we do reject H_0. If the evidence against H_0 is strong, the p-value will be small.

p-value	evidence
$< .01$	very strong evidence against H_0
$.01 - .05$	strong evidence against H_0
$.05 - .10$	weak evidence against H_0
$> .1$	little or no evidence against H_0

Warning! A large p-value is not strong evidence in favor of H_0. A large p-value can occur for two reasons: (i) H_0 is true or (ii) H_0 is false but the test has low power.

Warning! Do not confuse the p-value with $\mathbb{P}(H_0|\text{Data})$. [2] **The p-value is not the probability that the null hypothesis is true.**

The following result explains how to compute the p-value.

[2] We discuss quantities like $\mathbb{P}(H_0|\text{Data})$ in the chapter on Bayesian inference.

10.12 Theorem. *Suppose that the size α test is of the form*

$$\text{reject } H_0 \text{ if and only if } T(X^n) \geq c_\alpha.$$

Then,

$$\text{p-value} = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(T(X^n) \geq T(x^n))$$

where x^n is the observed value of X^n. If $\Theta_0 = \{\theta_0\}$ then

$$\text{p-value} = \mathbb{P}_{\theta_0}(T(X^n) \geq T(x^n)).$$

We can express Theorem 10.12 as follows:

> The p-value is the probability (under H_0) of observing a value of the test statistic the same as or more extreme than what was actually observed.

10.13 Theorem. *Let $w = (\widehat{\theta} - \theta_0)/\widehat{\mathsf{se}}$ denote the observed value of the Wald statistic W. The p-value is given by*

$$\text{p} - \text{value} = \mathbb{P}_{\theta_0}(|W| > |w|) \approx \mathbb{P}(|Z| > |w|) = 2\Phi(-|w|) \tag{10.7}$$

where $Z \sim N(0,1)$.

To understand this last theorem, look at Figure 10.4.
Here is an important property of p-values.

10.14 Theorem. *If the test statistic has a continuous distribution, then under $H_0 : \theta = \theta_0$, the p-value has a Uniform (0,1) distribution. Therefore, if we reject H_0 when the p-value is less than α, the probability of a type I error is α.*

In other words, if H_0 is true, the p-value is like a random draw from a Unif$(0,1)$ distribution. If H_1 is true, the distribution of the p-value will tend to concentrate closer to 0.

10.15 Example. Recall the cholesterol data from Example 7.15. To test if the means are different we compute

$$W = \frac{\widehat{\delta} - 0}{\widehat{\mathsf{se}}} = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}} = \frac{216.2 - 195.3}{\sqrt{5^2 + 2.4^2}} = 3.78.$$

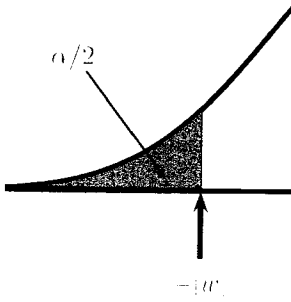


FIGURE 10.4. The p-value is
the p-value for the Wald te
ndary of the rejection region
$= (\widehat{\theta} - \theta_0)/\widehat{\mathsf{se}}$. This implies
$\sim N(0,1)$.

compute the p-value, let Z
riable. Then,

$$\text{p-value} = \mathbb{P}(|Z| >$$

hich is very strong evidence
ns are different, let $\widehat{\nu}_1$ and

$$W = \frac{\widehat{\nu}_1 -}{\widehat{\mathsf{se}}}$$

ere the standard error is

$$\text{p-value} = \mathbb{P}(|Z|$$

hich is strong evidence agains

10.3 The χ^2 Distribu

Before proceeding we need to
ependent, standard Normal
χ^2 distribution with k degree
nsity of V is

$$f($$

> 0. It can be shown that
quantile $\chi^2_{k,\alpha} = F^{-1}(1-\alpha$