Chapter 8

MATH 4753

CONTENTS

8.1 The Relationship Between Statistical Tests of Hypotheses and Confidence Intervals

8.2 Elements and Properties of a Statistical Test

8.3 Finding Statistical Tests: Classical Methods

8.4 Choosing the Null and Alternative Hypotheses

8.5 The Observed Significance Level for a Test: *p*-values

8.6 Testing a Population Mean

8.7 Testing the Difference Between Two Population Means: Independent Samples

8.8 Testing the Difference Between Two Population Means: Matched Pairs

8.9 Testing a Population Proportion

8.10 Testing the Difference Between Two Population Proportions

8.11 Testing a Population Variance

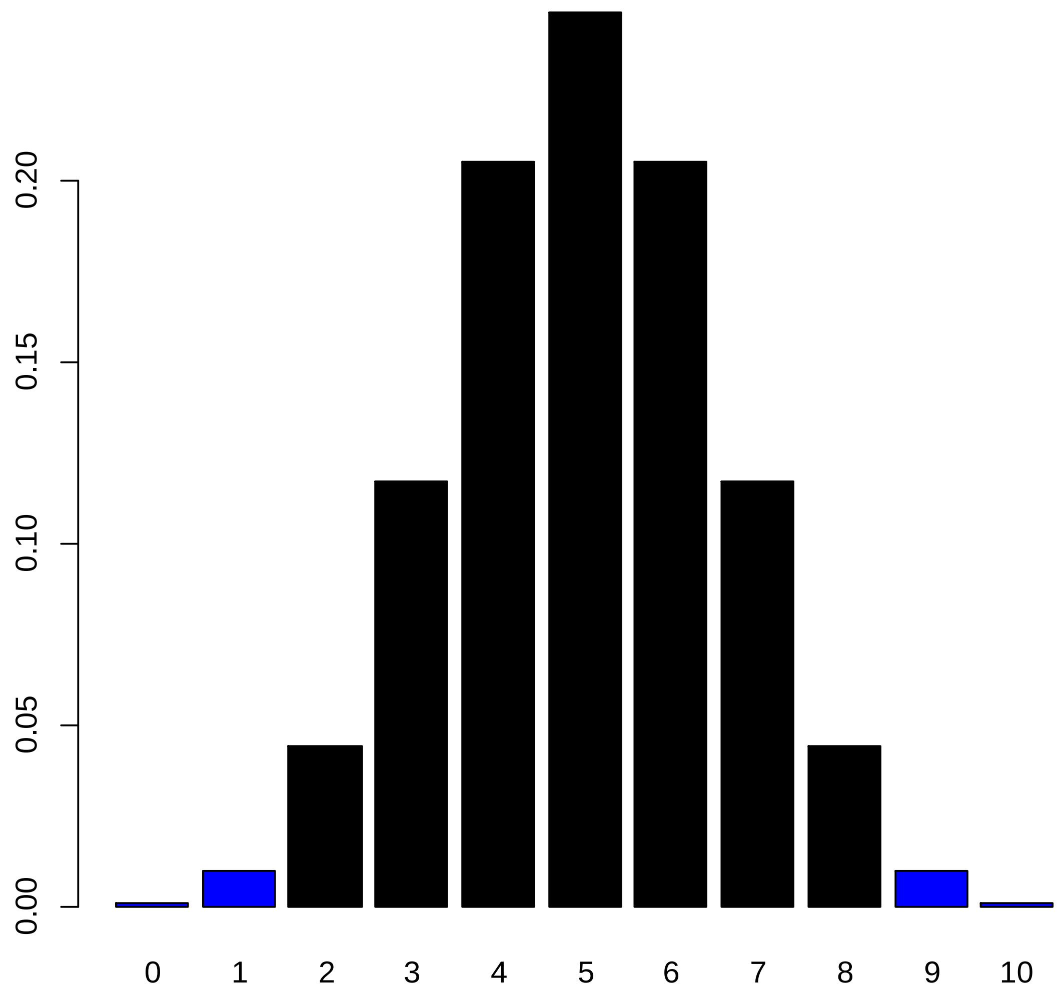
8.12 Testing the Ratio of Two Population Variances

8.13 Alternative Testing Procedures: Bootstrapping and Bayesian Methods (*Optional*)

By the end of this lesson you should know:

- What hypothesis testing is.
- What type 1 and 2 errors are.
- What power is.

Development of Hypothesis testing (R script)



Test errors

- **Type 1 error** happens when you **reject H_o when it is true.**
  - **Probability of type 1 error =α**
- **Type 2 error** happens when you **accept H_o when it is false.**
  - **Probability of type 2 error=β**

The Courtroom

In the start of the procedure, there are two hypotheses H_0: "the defendant is not guilty", and H_1: "the defendant is guilty". The first one is called *null hypothesis*, and is for the time being accepted. The second one is called *alternative (hypothesis)*. It is the hypothesis one hopes to support.

The hypothesis of innocence is only rejected when an error is very unlikely, because one doesn't want to convict an innocent defendant. Such an error is called *error of the first kind* (i.e., the conviction of an innocent person), and the occurrence of this error is controlled to be rare. As a consequence of this asymmetric behaviour, the *error of the second kind* (acquitting a person who committed the crime), is often rather large.

	H_0 is true **Truly not guilty**	**H_1 is true** **Truly guilty**
Accept Null Hypothesis Acquittal	Right decision	Wrong decision Type II Error
Reject Null Hypothesis Conviction	Wrong decision Type I Error	Right decision

A criminal trial can be regarded as either or both of two decision processes: guilty vs not guilty or evidence vs a threshold ("beyond a reasonable doubt"). In one view, the defendant is judged; in the other view the performance of the prosecution (which bears the burden of proof) is judged. A hypothesis test can be regarded as either a judgment of a hypothesis or as a judgment of evidence.

	H_0 is true Truly not guilty	H_1 is true Truly guilty
Accept Null Hypothesis Acquittal	Right decision	Wrong decision Type II Error
Reject Null Hypothesis Conviction	Wrong decision Type I Error	Right decision

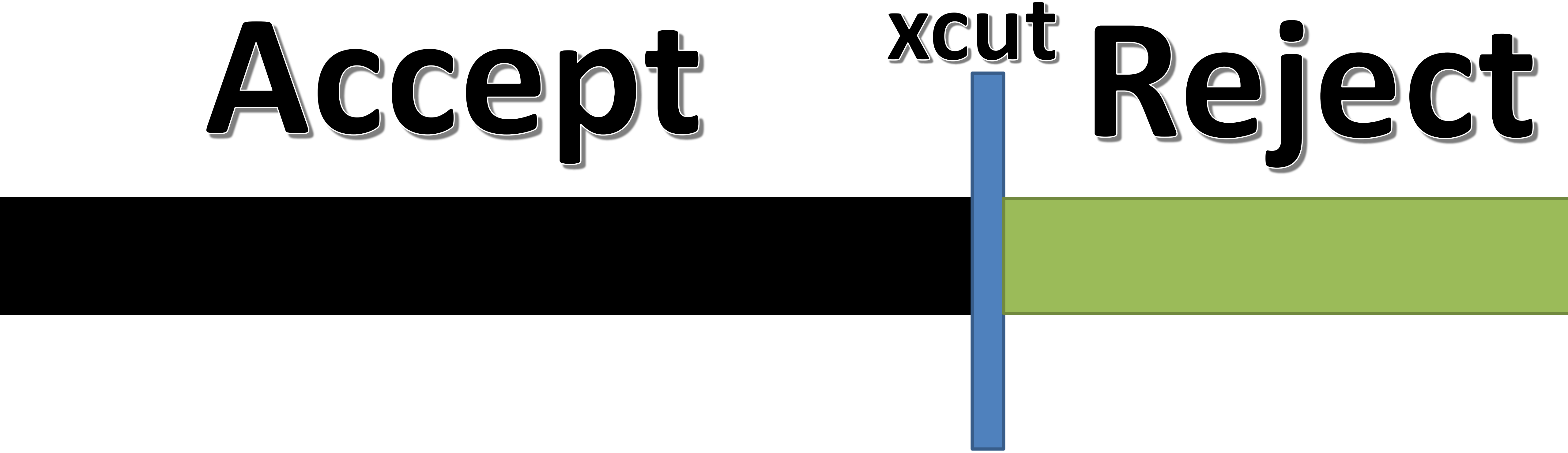
Power

- Probability of **rejecting H_0 when it is false** is $1 - \beta$
- This is called the power of a test.
- **Power is not the probability of an error!**
- **Power=P(Rejecting $H_o|H_o$ is False)=1-P(Accepting $H_o|H_o$ is False)**

What is the importance of power?

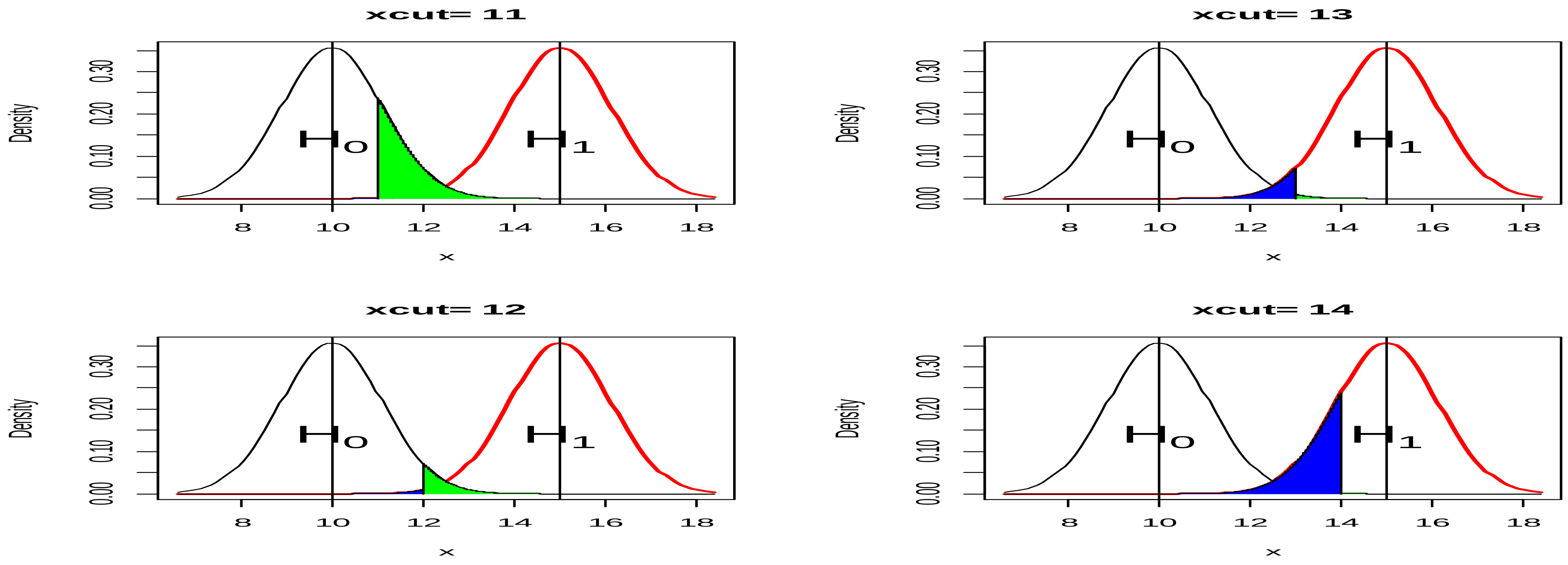
- The higher the value of the power the more we can detect departures from H_o.

Cut off and Acceptance, Rejection regions



Investigating POWER

$H_0: \mu = 10$
$H_1: \mu = 15$



Can you dig it?

- A) Yes
- B) No

Elements of a Statistical Test

1. **Null hypothesis, H_0**, about one or more population parameters
2. **Alternative hypothesis, H_a**, that we will accept if we decide to reject the null hypothesis
3. **Test statistic**, computed from sample data
4. **Rejection region**, indicating the values of the test statistic that will imply rejection of the null hypothesis
5. **Conclusion,** the decision made on whether to accept or reject the null hypothesis

Environmental Protection Agency

Confidence intervals and hypothesis tests are related and can be used to make decisions about parameters. For example, suppose an investigator for the Environmental Protection Agency (EPA) wants to determine whether the mean level μ of a certain type of pollutant released into the atmosphere by a chemical company meets the EPA guidelines. If 3 parts per million is the upper limit allowed by the EPA, the investigator would want to use sample data (daily pollution measurements) to decide whether the company is violating the law, i.e., to decide whether $\mu > 3$. If, say, a 99% confidence interval for μ contained only numbers greater than 3, then the EPA would be confident that the mean exceeds the established limit.

Test

- $H_0 : \mu = 3$
- $H_1 : \mu > 3$

TABLE 8.1 Conclusions and Consequences for the EPA's Test of Hypothesis

	True State of Nature	
EPA Decision	Company Not in Violation (H_0 true)	Company in Violation (H_a true)
Company in Violation (Reject H_0)	Type I error	Correct decision
Company Not in Violation (Accept H_0)	Correct decision	Type II error

Definition 8.1

Rejecting the null hypothesis if it is true is a **Type I error**. The probability of making a Type I error is denoted by the symbol α.

Definition 8.2

Accepting the null hypothesis if it is false is a **Type II error**. The probability of making a Type II error is denoted by the symbol β.

Definition 8.3

The **power** of a statistical test, $(1 - \beta)$, is the probability of rejecting the null hypothesis H_0 when, in fact, H_0 is false.

Power calculation

Example 8.4

Computing the Power of a Test

Refer to the test of hypothesis in Example 8.1. Find the power of the test if in fact $p = .3$.

Solution

From Definition 8.3, the power of the test is the probability $(1 - \beta)$. The probability of making a Type II error, i.e., failing to reject H_0: $p = .2$, if in fact $p = .3$, will be larger than the value of β calculated in Example 8.3 because $p = .3$ is much closer to the hypothesized value of $p = .2$. Thus,

$$\beta = P(Y \leq 3 \mid p = .3) = \sum_{y=0}^{3} p(y) \qquad \text{for } n = 10 \text{ and } p = .3$$

The value of this partial sum, given in Table 2 of Appendix B for a binomial random variable with $n = 10$ and $p = .3$, is .650. Therefore, the probability that we will fail to reject H_0: $p = .2$ if in fact $p = .3$ is $\beta = .650$ and the power of the test is $(1 - \beta) = (1 - .650) = .350$. You can see that the closer the actual value of p is to the hypothesized null value, the more unlikely it is that we will reject H_0: $p = .2$.

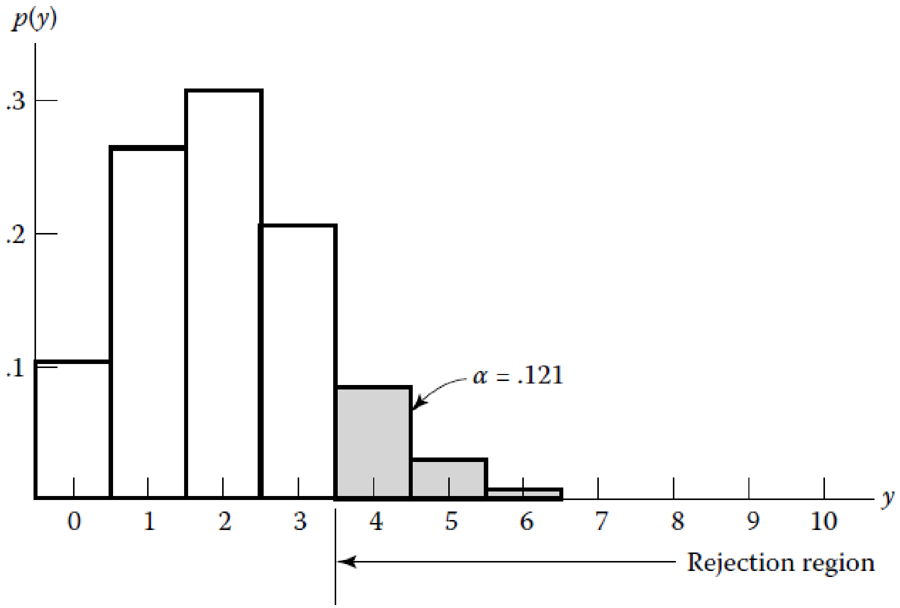
Example 8.1

Elements of a Statistical Test: Proportion of Software Purchasers

A manufacturer of notebook computers believes that it can sell a particular software package to more than 20% of the buyers of its computers. Ten prospective purchasers of the notebook computer were randomly selected and questioned about their interest in the software package. Of these, four indicated that they planned to buy the package. Does this sample provide sufficient evidence to indicate that more than 20% of the computer purchasers will buy the software package?

Solution

Let p be the true proportion of all prospective notebook computer buyers who will purchase the software package. Since we want to show that $p > .2$, we choose H_a: $p > .2$ for the alternative hypothesis and H_0: $p = .2$ for the null hypothesis. We will use the binomial random variable Y, the number of prospective purchasers in the sample who plan to buy the software, as the test statistic and will reject H_0: $p = .2$ if Y is large. A graph of $p(y)$ for $n = 10$ and $p = .2$ is shown in Figure 8.1.



Large values of Y will support the alternative hypothesis, H_a: $p > .2$, but what values of Y should we include in the rejection region? Suppose that we select values of $Y \geq 4$ as the rejection region. Then the elements of the test are

H_0: $p = .2$

H_a: $p > .2$

Test statistic: $Y = y$

Rejection region: $y \geq 4$

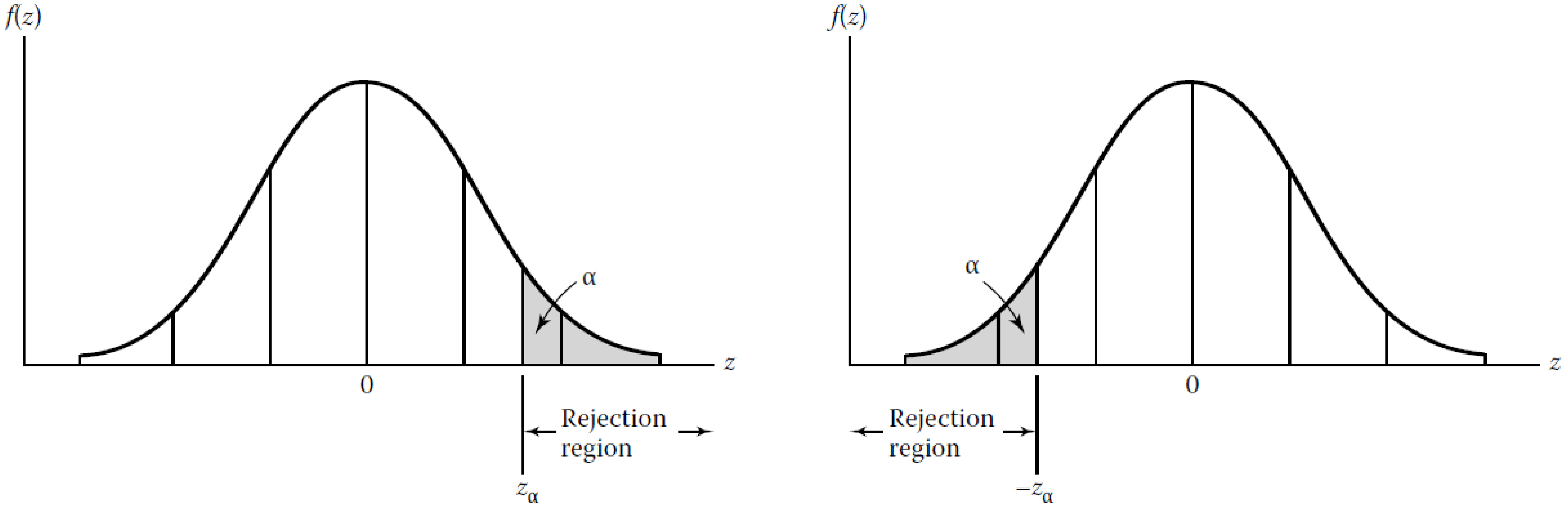
To conduct the test, we note that the observed value of Y, $y = 4$, falls in the rejection region. Thus, for this test procedure, we reject the null hypothesis, H_0: $p = .2$, and conclude that the manufacturer is correct, i.e., $p > .2$.

8.3 Finding Statistical Tests: Classical Methods

To find a statistical test about one or more population parameters, we must (1) find a suitable test statistic and (2) specify a rejection region. Classical statisticians use a method proposed by R. A. Fisher for finding a reasonable test statistic for testing a hypothesis. For example, suppose we want to test a hypothesis about the sole parameter θ of a probability function $p(y)$ or density function $f(y)$, and let L represent the likelihood function of the sample. Then to test the null hypothesis, H_0: $\theta = \theta_0$, Fisher's **likelihood ratio test statistic** is

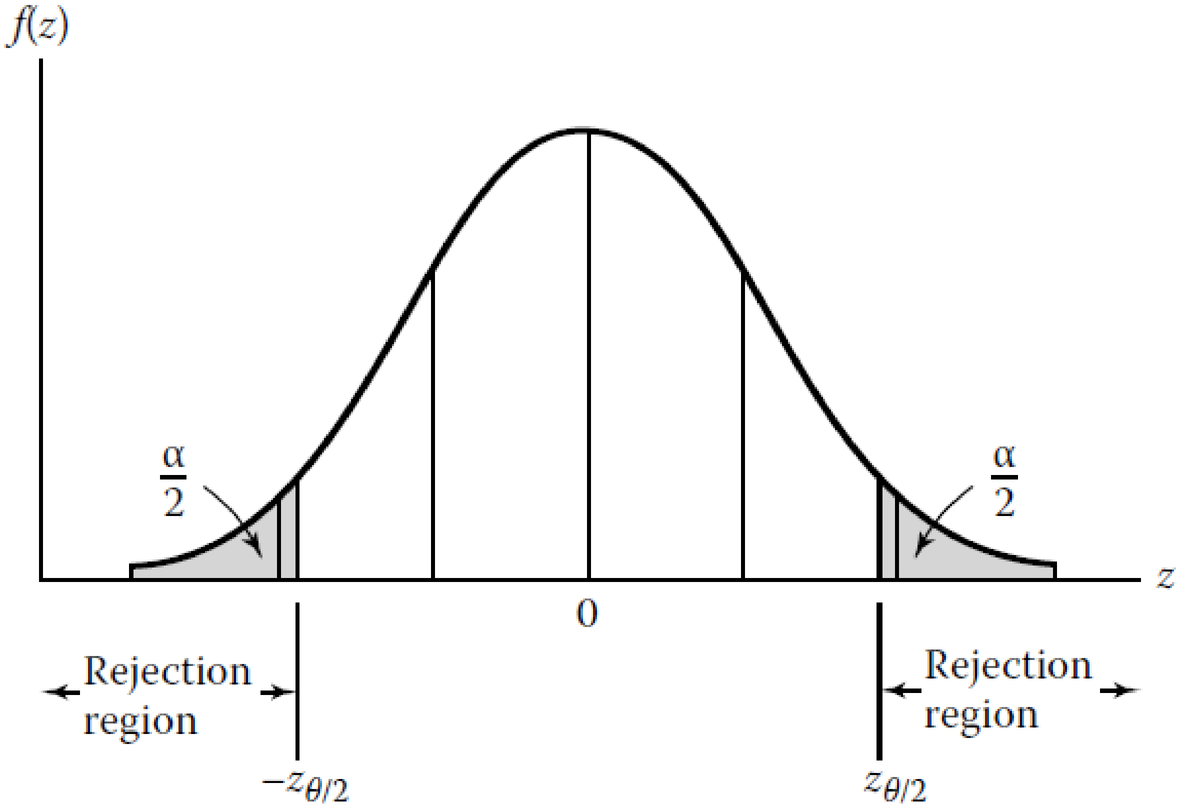
$$\lambda = \frac{\text{Likelihood assuming } \theta = \theta_0}{\text{Likelihood assuming } \theta = \hat{\theta}} = \frac{L(\theta_0)}{L(\hat{\theta})}$$

where $\hat{\theta}$ is the maximum likelihood estimator of θ. Fisher reasoned that if θ differs from θ_0, then the value of the likelihood L when $\theta = \hat{\theta}$ will be larger than when $\theta = \theta_0$. Thus, the rejection region for the test contains values of λ that are small—say, smaller than some value λ_R.



a. One-tailed test;
H_a: $\theta > \theta_0$

b. One-tailed test;
H_a: $\theta < \theta_0$



c. Two-tailed test;
H_{θ}: $\theta \neq \theta_0$

FIGURE 8.3

Rejection regions for one- and two-tailed tests

A Large-Sample Test Based on the Standard Normal z Test Statistic

One-Tailed Test	*Two-Tailed Test*
$H_0\colon\ \theta = \theta_0$	$H_0\colon\ \theta = \theta_0$
$H_a\colon\ \theta > \theta_0 \quad (\text{or } H_a\colon\ \theta < \theta_0)$	$H_a\colon\ \theta \neq \theta_0$
Test statistic: $Z = \dfrac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$	*Test statistic:* $Z = \dfrac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$
Rejection region: $Z > z_\alpha \quad (\text{or } Z < -z_\alpha)$	*Rejection region:* $\lvert Z \rvert > z_{\alpha/2}$
where $P(Z > z_\alpha) = \alpha$	where $P(Z > z_{\alpha/2}) = \alpha/2$

Example 8.6

Calculating β for the Traveling Trucks Test

Refer to the one-tailed test for μ, Example 8.5. If the mean number μ of heavy freight trucks traveling a particular 25-mile stretch of interstate highway is in fact 78 per hour, what is the probability that the test procedure of Example 8.5 would fail to detect it? That is, what is the probability β that we would fail to reject H_0: $\mu = 72$ in this one-tailed test if μ is actually equal to 78?

Solution

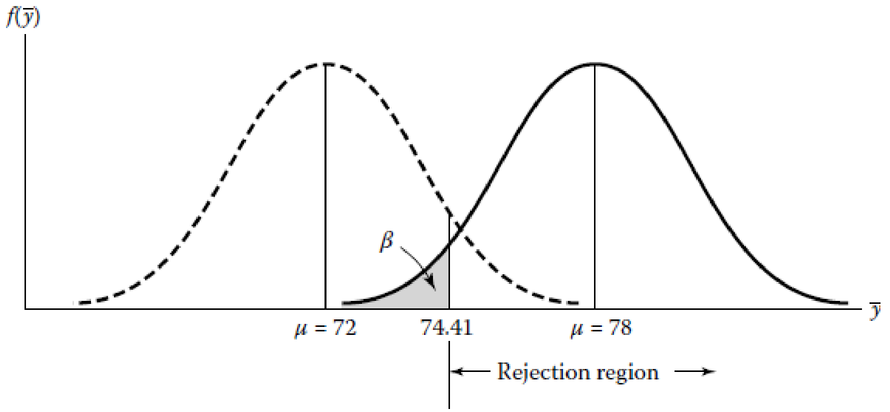
To calculate β for the large-sample Z test, we need to specify the rejection region in terms of the point estimator $\hat{\theta}$, where, for this example, $\hat{\theta} = \bar{y}$. From Figure 8.4, you can see that the rejection region consists of values of $Z \geq 1.28$. To determine the value of $\bar{y}$ corresponding to $z = 1.28$, we substitute into the equation

$$Z = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}} \approx \frac{\bar{y} - \mu_0}{s/\sqrt{n}} \quad \text{or} \quad 1.28 = \frac{\bar{y} - 72}{13.3/\sqrt{50}}$$

Solving for $\bar{y}$, we obtain $\bar{y} = 74.41$. Therefore, the rejection region for the test is $Z \geq 1.28$ or, equivalently, $\bar{y} \geq 74.41$.

The dotted curve in Figure 8.5 is the sampling distribution for $\bar{y}$ if H_0: $\mu = 72$ is true. This curve was used to locate the rejection region for $\bar{y}$ (and, equivalently, z), i.e., values of $\bar{y}$ contradictory to H_0: $\mu = 72$. The solid curve is the sampling distribution for $\bar{y}$ if $\mu = 78$. Since we want to find β if H_0 is in fact false and $\mu = 78$, we want to find the probability that $\bar{y}$ does not fall in the rejection region if $\mu = 78$. This

FIGURE 8.5
The probability β of making a Type II error if $\mu = 78$ in Example 8.6



Calculating β for a Large-Sample Z Test

Consider a large-sample test of H_0: $\theta = \theta_0$ at significance level α. The value of β for a specific value of the alternative $\theta = \theta_a$ is calculated as follows:

Upper-tailed test: $\beta = P\left(Z < \dfrac{\hat{\theta}_0 - \theta_a}{\sigma_{\hat{\theta}}}\right)$

where $\hat{\theta}_0 = \theta_0 + z_\alpha \sigma_{\hat{\theta}}$ is the value of the estimator corresponding to the border of the rejection region

Lower-tailed test: $\beta = P\left(Z > \dfrac{\hat{\theta}_0 - \theta_a}{\sigma_{\hat{\theta}}}\right)$

where $\hat{\theta}_0 = \theta_0 - z_\alpha \sigma_{\hat{\theta}}$ is the value of the estimator corresponding to the border of the rejection region

Two-tailed test: $\beta = P\left(\dfrac{\hat{\theta}_{0,\mathrm{L}} - \theta_a}{\sigma_{\hat{\theta}}} < Z < \dfrac{\hat{\theta}_{0,\mathrm{U}} - \theta_a}{\sigma_{\hat{\theta}}}\right)$

where $\hat{\theta}_{0,\mathrm{U}} = \theta_0 + z_\alpha \sigma_{\hat{\theta}}$ and $\hat{\theta}_{0,\mathrm{L}} = \theta_0 - z_\alpha \sigma_{\hat{\theta}}$ are the values of the estimator corresponding to the borders of the rejection region

Example 8.7

Choosing H_0 and H_a for Testing the Mean Diameter of Bearings

A metal lathe is checked periodically by quality control inspectors to determine whether it is producing machine bearings with a mean diameter of .5 inch. If the mean diameter of the bearings is larger or smaller than .5 inch, then the process is out of control and needs to be adjusted. Formulate the null and alternative hypotheses that could be used to test whether the bearing production process is out of control.

Solution

The hypotheses must be stated in terms of a population parameter. Thus, we define

$$\mu = \text{true mean diameter (in inches) of all bearings produced by the lathe}$$

If either $\mu > .5$ or $\mu < .5$, then the metal lathe's production process is out of control. Since we wish to be able to detect either possibility, the null and alternative hypotheses would be

H_0: $\mu = .5$ (i.e., the process is in control)
H_a: $\mu \neq .5$ (i.e., the process is out of control)

Definition 8.4

The **observed significance level**, or ***p*-value**, for a specific statistical test is the probability (assuming H_0 is true) of observing a value of the test statistic that is at least as contradictory to the null hypothesis, and supportive of the alternative hypothesis, as the one computed from the sample data.

Large-Sample ($n \geq 30$) Test of Hypothesis About a Population Mean μ

One-Tailed Test	*Two-Tailed Test*
H_0: $\mu = \mu_0$	H_0: $\mu = \mu_0$
H_a: $\mu > \mu_0$ (or H_a: $\mu < \mu_0$)	H_a: $\mu \neq \mu_0$
Test statistic: $Z = \dfrac{\bar{y} - \mu_0}{\sigma_{\bar{y}}} \approx \dfrac{\bar{y} - \mu_0}{s/\sqrt{n}}$	*Test statistic:* $Z = \dfrac{\bar{y} - \mu_0}{\sigma_{\bar{y}}} \approx \dfrac{\bar{y} - \mu_0}{s/\sqrt{n}}$
Rejection region: $Z > z_\alpha$ (or $Z < -z_\alpha$)	*Rejection region:* $\lvert Z \rvert > z_{\alpha/2}$
p-value $= P(Z > z_c)$ [or, $P(Z < z_c)$]	p-value $= 2P(Z > \lvert z_c \rvert)$

where $P(Z > z_\alpha) = \alpha$, $P(Z > z_{\alpha/2}) = \alpha/2$, μ_0 is our symbol for the particular numerical value specified for μ in the null hypothesis, and z_c is the computed value of the test statistic.

Assumptions: None (since the central limit theorem guarantees that $\bar{y}$ is approximately normal regardless of the distribution of the sampled population)

Small-Sample Test of Hypothesis About a Population Mean μ

One-Tailed Test

H_0: $\mu = \mu_0$

H_a: $\mu > \mu_0$ (or H_a: $\mu < \mu_0$)

Two-Tailed Test

H_0: $\mu = \mu_0$

H_a: $\mu \neq \mu_0$

$$\textit{Test statistic: } T = \frac{\bar{y} - \mu_0}{s/\sqrt{n}}$$

One-tailed:

Rejection region: $T > t_\alpha$ (or $T < -t_\alpha$)

p-value $= P(T \geq t_c)$ [or, $P(T \leq t_c)$]

Two-tailed:

Rejection region: $|T| > t_{\alpha/2}$

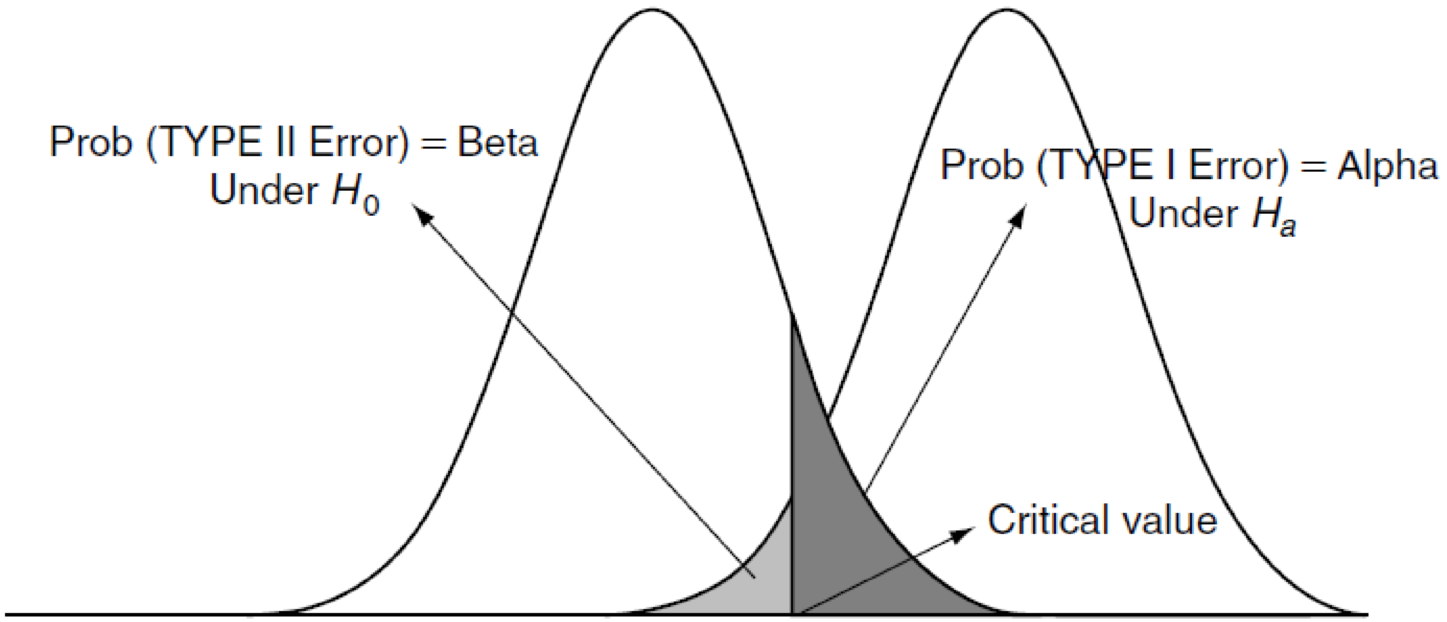
p-value $= 2P(T \geq |t_c|)$

where the distribution of t is based on $(n - 1)$ degrees of freedom; $P(T > t_\alpha) = \alpha$; $P(T > t_{\alpha/2}) = \alpha/2$, and t_c is the computed value of the test-statistic.

Assumption: The relative frequency distribution of the population from which the sample was selected is approximately normal.

Warning: If the data depart greatly from normality, this small-sample test may lead to erroneous inferences. In this case, use the nonparametric sign test that is discussed in Section 15.2.

Hypothesis Testing: Errors



8.5 *Defective power meters.* A manufacturer of power meters, which are used to regulate energy thresholds of a data-communications system, claims that when its production process is operating correctly, only 10% of the power meters will be defective. A vendor has just received a shipment of 25 power meters from the manufacturer. Suppose the vendor wants to test H_0: $p = .10$ against H_a: $p > .10$, where p is the true proportion of power meters that are defective. Use $Y \geq 6$ as the rejection region.

a. Determine the value of α for this test procedure.
b. Find β if in fact $p = .2$. What is the power of the test for this value of p?
c. Find β if in fact $p = .4$. What is the power of the test for this value of p?

8.5 a. $\alpha =$ probability of rejecting H_0 when H_0 is true

$$= P(Y \geq 6 \text{ if } p = 0.1) = 1 - P(Y \leq 5) = 1 - \sum_{y=0}^{5} p(y) = 1 - 0.9666 = 0.0334$$

Note: $p(y)$ is found using Table 2, Appendix B, with $n = 25$ and $p = 0.1$

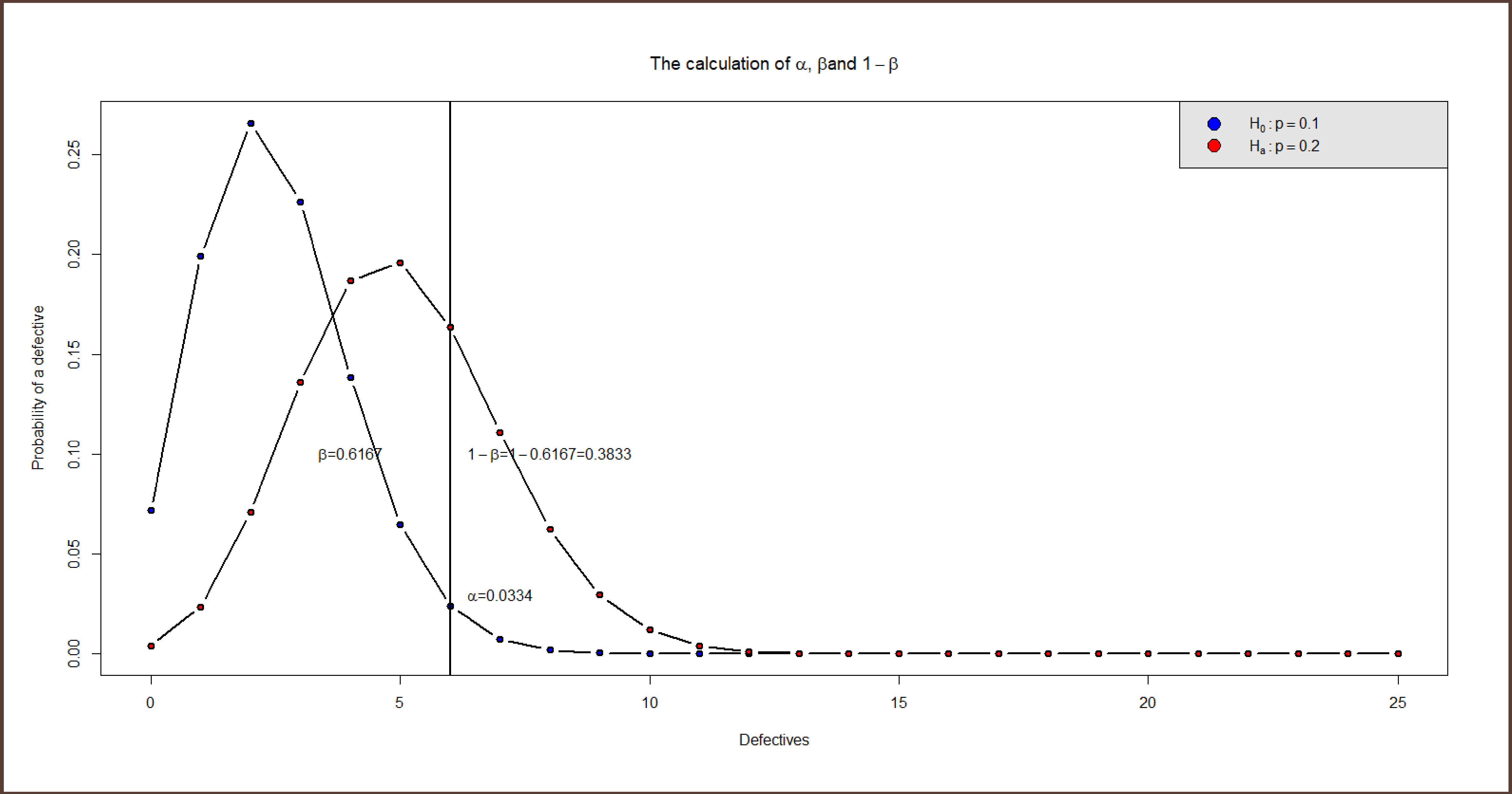
b. $\beta =$ probability of accepting H_0 when H_0 is false

$$= P(Y \leq 5 \text{ if } p = 0.2) = \sum_{y=0}^{5} p(y) = 0.6167$$

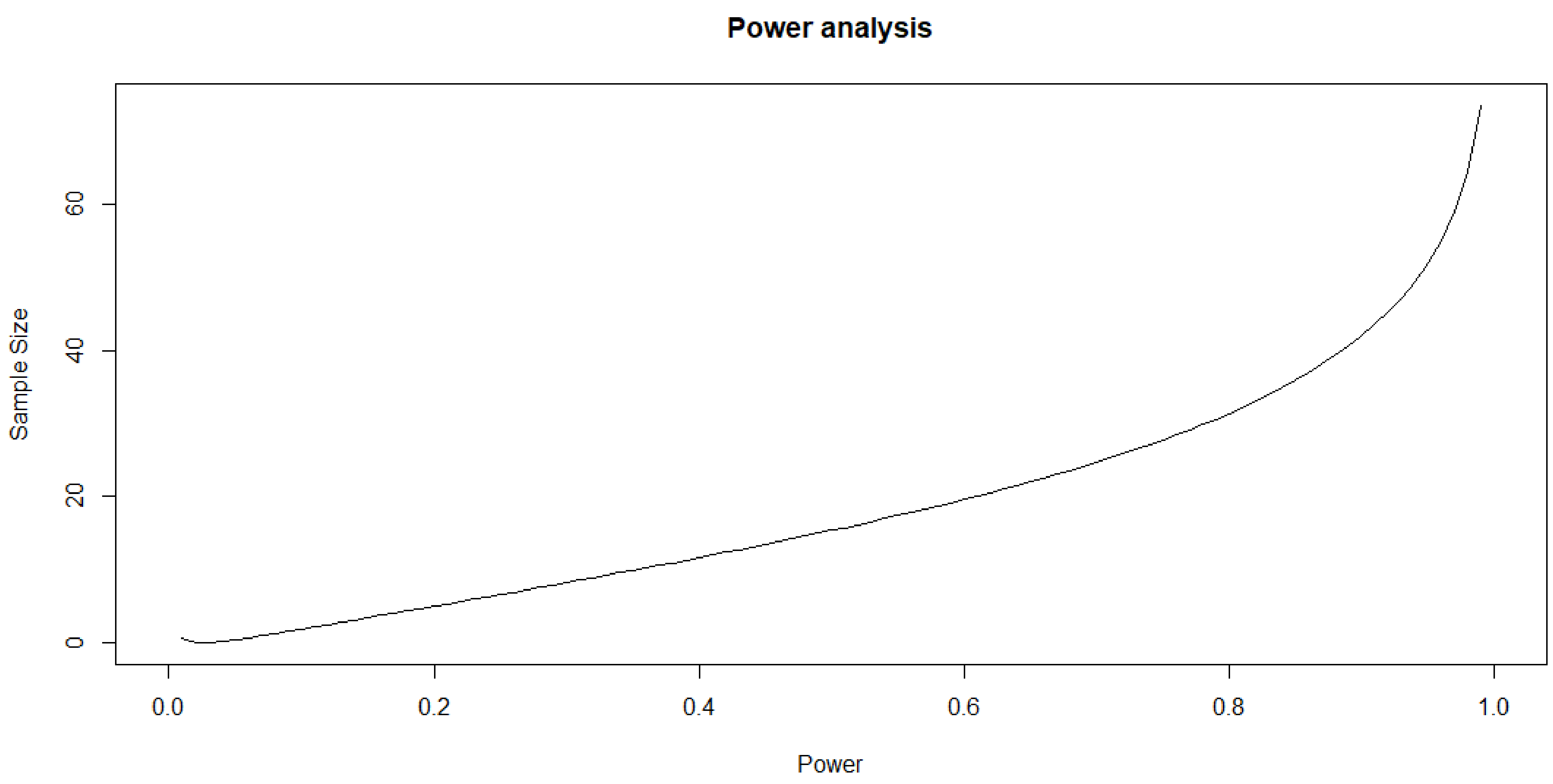
Note: $p(y)$ is found using Table 2, Appendix B, with $n = 25$ and $p = 0.2$

The power of the test $= 1 - \beta = 1 - 0.6167 = 0.3833$.

Use R



The calculation of α, βand 1 – β
$H_0 : p = 0.1$
$H_a : p = 0.2$
β=0.6167
1 – β=1 – 0.6167=0.3833
α=0.0334
Probability of a defective
Defectives
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
20
25



Power analysis
Sample Size
0
20
40
60
0.0
0.2
0.4
0.6
0.8
1.0
Power

Population of Wisconsin lakes has $\mu = 15$

8.28 a. Using MINITAB, the descriptive statistics are:

Descriptive Statistics: DOC

Variable	N	Mean	StDev	Variance	Minimum	Q1	Median	Q3	Maximum
DOC	25	14.52	12.96	168.06	2.40	4.15	13.20	19.10	56.90

To determine if the mean DOC value differs from 15, we test:

$$H_0: \mu = 15$$
$$H_a: \mu \neq 15$$

```
> qt(1-0.1/2, 25-1)
[1] 1.710882
```

The test statistic is $t = \dfrac{\bar{y} - \mu_0}{s/\sqrt{n}} = \dfrac{14.52 - 15}{12.96/\sqrt{25}} = -0.185$.

The rejection region requires $\alpha/2 = 0.10/2 = 0.05$ in each tail of the t distribution. From Table 7, Appendix B, with $df = n - 1 = 25 - 1 = 24$, $t_{0.05} = 1.711$. The rejection region is $t < -1.711$ or $t > 1.711$.

Since the observed value of the test statistic does not fall in the rejection region $(t = -0.185 \not< -1.711)$, H_0 is not rejected. There is insufficient evidence to indicate the mean DOC value differs from 15 at $\alpha = 0.10$.

Probability the test will detect a mean that differs from 15 gm/m^3 if $\mu_a = 14$

b. We must find the rejection region in part a in terms of $\overline{y}$. We know

$$t = \frac{\overline{y} - \mu_0}{s/\sqrt{n}} \Rightarrow t\left(\frac{s}{\sqrt{n}}\right) = \overline{y} - \mu_0 \Rightarrow \overline{y} = \mu_0 + t\left(\frac{s}{\sqrt{n}}\right)$$

For $t = -1.711$, $\overline{y} = \mu_0 + t\left(\frac{s}{\sqrt{n}}\right) = 15 - 1.711\left(\frac{12.96}{\sqrt{25}}\right) = 10.565$

For $t = 1.711$, $\overline{y} = \mu_0 + t\left(\frac{s}{\sqrt{n}}\right) = 15 + 1.711\left(\frac{12.96}{\sqrt{25}}\right) = 19.435$

Thus, we would reject H_0 if $\overline{y} < 10.565$ or $\overline{y} > 19.435$.

We want to find

$$P(\overline{y} < 10.565 \mid \mu_a = 14) + P(\overline{y} > 19.435 \mid \mu_a = 14)$$

$$= P\left(t < \frac{10.565 - 14}{\frac{12.96}{\sqrt{25}}}\right) + P\left(t > \frac{19.435 - 14}{\frac{12.96}{\sqrt{25}}}\right) = P(t < -1.33) + P(t > 2.10)$$

$$= 0.0980 + 0.0232 = 0.1212$$

(using a computer package with 24 degrees of freedom)

Taken from Montgomery and Runger App. Stat pg 311 Fourth edition

9-2.2 TYPE II ERROR AND CHOICE OF SAMPLE SIZE

In testing hypotheses, the analyst directly selects the type I error probability. However, the probability of type II error β depends on the choice of sample size. In this section, we will show how to calculate the probability of type II error β. We will also show how to select the sample size to obtain a specified value of β.

Finding the Probability of Type II Error β

Consider the two-sided hypotheses

$$H_0: \mu = \mu_0 \qquad H_1: \mu \neq \mu_0$$

Suppose that the null hypothesis is false and that the true value of the mean is $\mu = \mu_0 + \delta$, say, where $\delta > 0$. The test statistic Z_0 is

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{X} - (\mu_0 + \delta)}{\sigma/\sqrt{n}} + \frac{\delta\sqrt{n}}{\sigma}$$

Therefore, the distribution of Z_0 when H_1 is true is

$$Z_0 \sim N\left(\frac{\delta\sqrt{n}}{\sigma}, 1\right) \tag{9-19}$$

The distribution of the test statistic Z_0 under both the null hypothesis H_0 and the alternate hypothesis H_1 is shown in Fig. 9-9. From examining this figure, we note that if H_1 is true, a type II error will be made only if $-z_{\alpha/2} \leq Z_0 \leq z_{\alpha/2}$ where $Z_0 \sim N(\delta\sqrt{n}/\sigma, 1)$. That is, the probability of the type II error β is the probability that Z_0 falls between $-z_{\alpha/2}$ and $z_{\alpha/2}$ *given that H_1 is true*. This probability is shown as the shaded portion of Fig. 9-12. Expressed mathematically, this probability is

Probability of a Type II Error for a Two-Sided Test on the Mean, Variance Known

$$\beta = \Phi\left(z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) - \Phi\left(-z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) \tag{9-20}$$

where $\Phi(z)$ denotes the probability to the left of z in the standard normal distribution. Note that Equation 9-20 was obtained by evaluating the probability that Z_0 falls in the interval

Taken from Montgomery and Runger App. Stat pg 311 Fourth edition

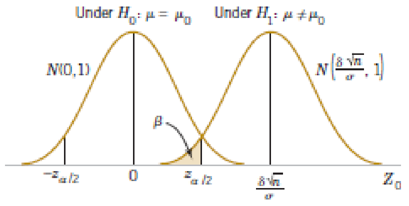


FIGURE 9-12 The distribution of Z_0 under H_0 and H_1.

$\left[-z_{\alpha/2}, z_{\alpha/2}\right]$ when H_1 is true. Furthermore, note that Equation 9-20 also holds if $\delta < 0$, because of the symmetry of the normal distribution. It is also possible to derive an equation similar to Equation 9-20 for a one-sided alternative hypothesis.

Sample Size Formulas

One may easily obtain formulas that determine the appropriate sample size to obtain a particular value of β for a given Δ and α. For the two-sided alternative hypothesis, we know from Equation 9-20 that

$$\beta = \Phi\left(z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) - \Phi\left(-z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right)$$

or, if $\delta > 0$,

$$\beta \simeq \Phi\left(z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) \qquad (9\text{-}21)$$

because $\Phi\left(-z_{\alpha/2} - \delta\sqrt{n}/\sigma\right) \simeq 0$ when δ is positive. Let z_β be the 100β upper percentile of the standard normal distribution. Then, $\beta = \Phi\left(-z_\beta\right)$. From Equation 9-21,

$$-z_\beta \simeq z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}$$

or

Sample Size for a Two-Sided Test on the Mean, Variance Known

$$n \simeq \frac{(z_{\alpha/2} + z_\beta)^2 \sigma^2}{\delta^2} \quad \text{where} \quad \delta = \mu - \mu_0 \qquad (9\text{-}22)$$

If n is not an integer, the convention is to round the sample size up to the next integer. This approximation is good when $\Phi(-z_{\alpha/2} - \delta\sqrt{n}/\sigma)$ is small compared to β. For either of the one-sided alternative hypotheses, the sample size required to produce a specified type II error with probability β given δ and α is

Sample Size for a One-Sided Test on the Mean, Variance Known

$$n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{\delta^2} \quad \text{where} \quad \delta = \mu - \mu_0 \qquad (9\text{-}23)$$

8.53 *Modeling transport of gases.* In *AIChE Journal* (Jan. 2005), chemical engineers published a new method for modeling multicomponent transport of gases. Twelve gas mixtures consisting of neon, argon, and helium were prepared at different ratios and at different temperatures. The viscosity of each mixture (10^{-5}Pa $\cdot$ s) was measured experimentally and was calculated with the new model. The results are shown in the table below. The chemical engineers concluded that there is "an excellent agreement between our new calculation and experiments." Do you agree? Your answer should include a discussion of practical versus statistical significance.

VISCOSITY

	Viscosity Measurements			Viscosity Measurements	
Mixture	Experimental	New Method	Mixture	Experimental	New Method
1	2.740	2.736	7	2.886	2.910
2	2.569	2.575	8	2.957	2.965
3	2.411	2.432	9	3.790	3.792
4	2.504	2.512	10	3.574	3.582
5	3.237	3.233	11	3.415	3.439
6	3.044	3.050	12	3.470	3.476

Source: Kerkhof, P., and Geboers, M. "Toward a unified theory of isotropic molecular transport phenomena." *AIChE Journal*, Vol. 51, No. 1, January 2005 (Table 2).

Intro2MLR - master - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins Intro2MLR

Source

Console | R Markdown | Jobs

D:/RPACKAGES/Intro2MLR/

```
> t.test(vis$`Visc-Exp`,vis$`Visc-New`,paired = TRUE, mu = 0)

        Paired t-test

data:  vis$`Visc-Exp` and vis$`Visc-New`
t = -3.1629, df = 11, p-value = 0.009033
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.01483898 -0.00266102
sample estimates:
mean of the differences
               -0.00875

>
```

Page 426 MS

Bootstrap Hypothesis Tests

Recall that the bootstrap is a Monte Carlo method that involves *resampling*—that is, taking repeated samples of size n (with replacement) from the original sample data set. The bootstrap testing procedure uses resampling to find an approximation for the observed significance level (p-value) of the test. The steps required to obtain the bootstrap p-value estimate for a test on a population mean are listed in the box.

Bootstrap p-Value for Testing a Population Mean, H_0: $\mu = \mu_0$

Let $y_1, y_2, y_3, \ldots, y_n$ represent a random sample of size n from a population with mean $E(Y) = \mu$.

Step 1 Calculate the value of the test statistic for the sample: $t_c = (\bar{y} - \mu_0)/(s/\sqrt{n})$ where $\bar{y}$ is the sample mean and s is the sample standard deviation.

Step 2 Select j, where j is the number of times you will resample. (Usually, j is a very large number, say, $j = 1{,}000$ or $j = 3{,}000$.)

Step 3 Transform each of the sample y values as follows: $x_i = y_i - \bar{y} + \mu_0$. That is, take each sample y value, subtract the sample mean, then add μ_0. (This step will generate sample values with a mean equal to the hypothesized mean in H_0.)

Step 4 Randomly sample, with replacement, n values of X from the transformed sample data set $x_1, x_2, x_3, \ldots, x_n$.

Step 5 Repeat step 4 a total of j times.

Step 6 For each bootstrap sample, compute the test statistic: $t_j = (\bar{x}_j - \mu_0)/(s_j/\sqrt{n})$, where $\bar{x}_j$ and s_j are the mean and standard deviation, respectively, of bootstrap sample j.

Step 7 Find the bootstrap estimated p-value—called the **achieved significance level (ASL)**—as follows:

Upper-tailed test (H_a: $\mu > \mu_0$): ASL $=$ (Number of times $t_j > t_c$)$/j$

Lower-tailed test (H_a: $\mu < \mu_0$): ASL $=$ (Number of times $t_j < t_c$)$/j$

Two-tailed test (H_a: $\mu \neq \mu_0$):

$$\text{ASL} = \frac{(\text{Number of times } t_j > |t_c|) + (\text{Number of times } t_j < -|t_c|)}{j}$$

Bootstrap p-Value for Testing Equality of Population Means, H_0: $(\mu_1 - \mu_2) = 0$

Let $\bar{y}_1$ and s_1 represent the mean and standard deviation of a random sample of size n_1 from a population with mean μ_1. Let $\bar{y}_2$ and s_2 represent the mean and standard deviation of a random sample of size n_2 from a population with mean μ_2.

Step 1 Calculate the value of the test statistic for the sample,

$$t_c = \frac{(\bar{y}_1 - \bar{y}_2)}{\sqrt{(s_1^2/n_1) + (s_2^2/n_2)}}$$

Step 2 Select j, where j is the number of times you will resample.

Step 3 Find the mean $\bar{\bar{y}}$ of the combined samples, then transform each of the sample values as follows:

$$\text{Sample 1: } x_i = y_i - \bar{y}_1 + \bar{\bar{y}} \qquad \text{Sample 2: } x_i = y_i - \bar{y}_2 + \bar{\bar{y}}$$

(That is, take each sample value, subtract its sample mean, then add $\bar{\bar{y}}$.)

Step 4 Randomly sample, with replacement, n_1 transformed values from the first sample. Randomly sample, with replacement, n_2 transformed values from the second sample.

Step 5 Repeat step 4 a total of j times.

Step 6 For each bootstrap sample, compute the test statistic:

$$t_j = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{(s_1^2/n_1) + (s_2^2/n_2)}}$$

where $\bar{x}_1$ and s_1 are the mean and standard deviation, respectively, of bootstrap sample j for sample 1, and $\bar{x}_2$ and s_2 are the mean and standard deviation, respectively, of bootstrap sample j for sample 2.

Step 7 Find the bootstrap estimated p-value—called the **achieved significance level (ASL)**—as follows:

Upper-tailed test (H_a: $\mu_1 - \mu_2 > 0$): ASL = (Number of times $t_j > t_c$)/j

Lower-tailed test (H_a: $\mu_1 - \mu_2 < 0$): ASL = (Number of times $t_j < t_c$)/j

Two-tailed test (H_a: $\mu_1 - \mu_2 \neq 0$):

$$\text{ASL} = \frac{(\text{Number of times } t_j > |t_c|) + (\text{Number of times } t_j < -|t_c|)}{j}$$

Bayesian Testing Procedures

Let $y_1, y_2, y_3, \ldots, y_n$ represent a random sample of size n selected from a population with unknown population parameter θ. The Bayesian approach to testing a hypothesis about θ considers θ as a random variable with a known *prior distribution*, $h(\theta)$. As with interval estimation, we need to find the *posterior distribution*, $g(\theta \mid y_1, y_2, y_3, \ldots, y_n)$. As shown in optional Section 7.14, the posterior distribution is

$$g(\theta \mid y_1, y_2, y_3, \ldots, y_n) = \frac{f(y_1, y_2, y_3, \ldots, y_n \mid \theta) \cdot h(\theta)}{f(y_1, y_2, y_3, \ldots, y_n)}$$

where $f(y_1, y_2, y_3, \ldots, y_n) = \int f(y_1, y_2, y_3, \ldots, y_n \mid \theta) \cdot h(\theta)\, d\theta$

Suppose you want to test H_0: $\theta \leq \theta_0$ versus H_a: $\theta > \theta_0$. The simplest Bayesian test uses the posterior distribution $g(\theta \mid y_1, y_2, y_3, \ldots, y_n)$ to find the following conditional probabilities:

$$P(\theta \leq \theta_0 \mid y_1, y_2, y_3, \ldots, y_n) \quad \text{and} \quad P(\theta > \theta_0 \mid y_1, y_2, y_3, \ldots, y_n)$$

In other words, the posterior distribution is used to find the likelihoods of H_0 and H_a occurring. A simple rule is to accept the hypothesis that is associated with the largest conditional probability. That is,

Accept H_0 if $\quad P(\theta \leq \theta_0 \mid y_1, y_2, y_3, \ldots, y_n) \geq P(\theta > \theta_0 \mid y_1, y_2, y_3, \ldots, y_n)$

Reject H_0 (i.e., *Accept* H_a if $\quad P(\theta \leq \theta_0 \mid y_1, y_2, y_3, \ldots, y_n) < P(\theta > \theta_0 \mid y_1, y_2, y_3, \ldots, y_n)$

We illustrate the Bayesian testing method in the next example.

See BBD for Bayesian Example

$$P(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$$