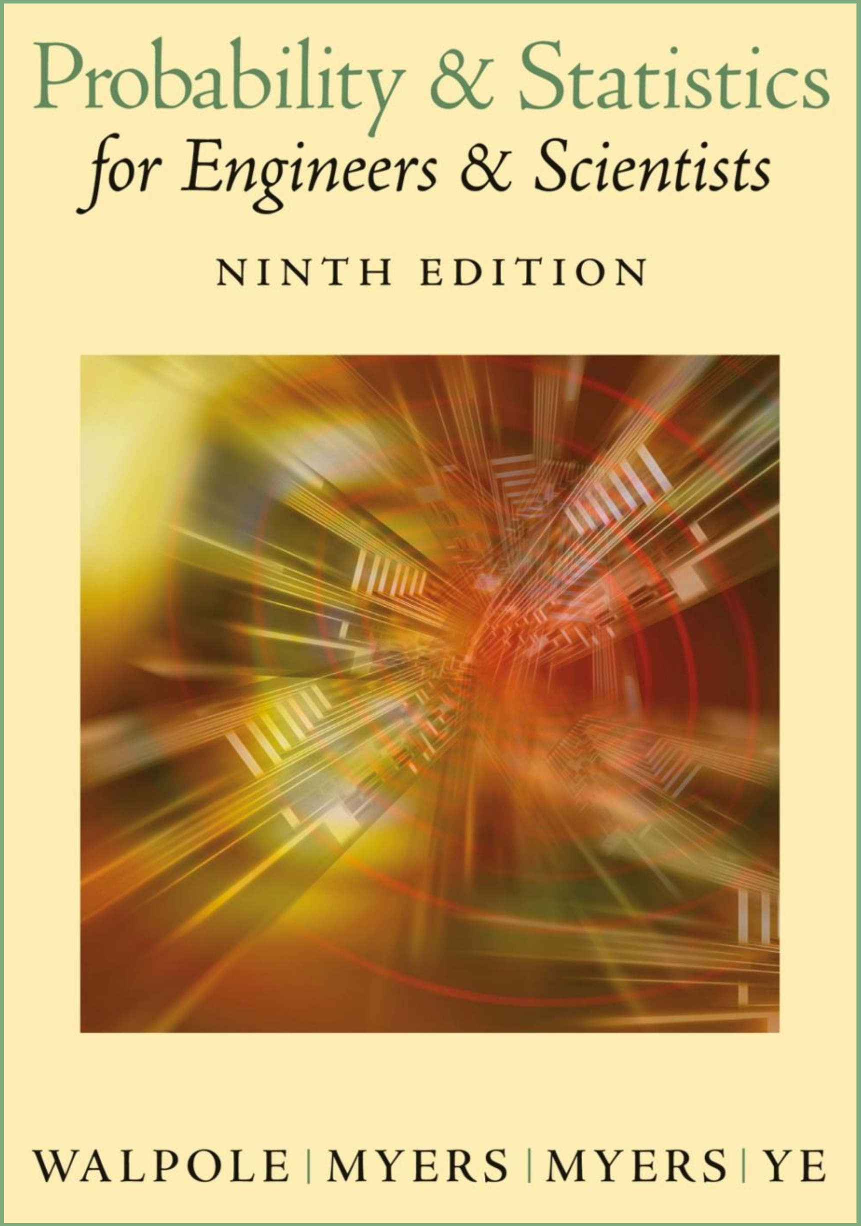
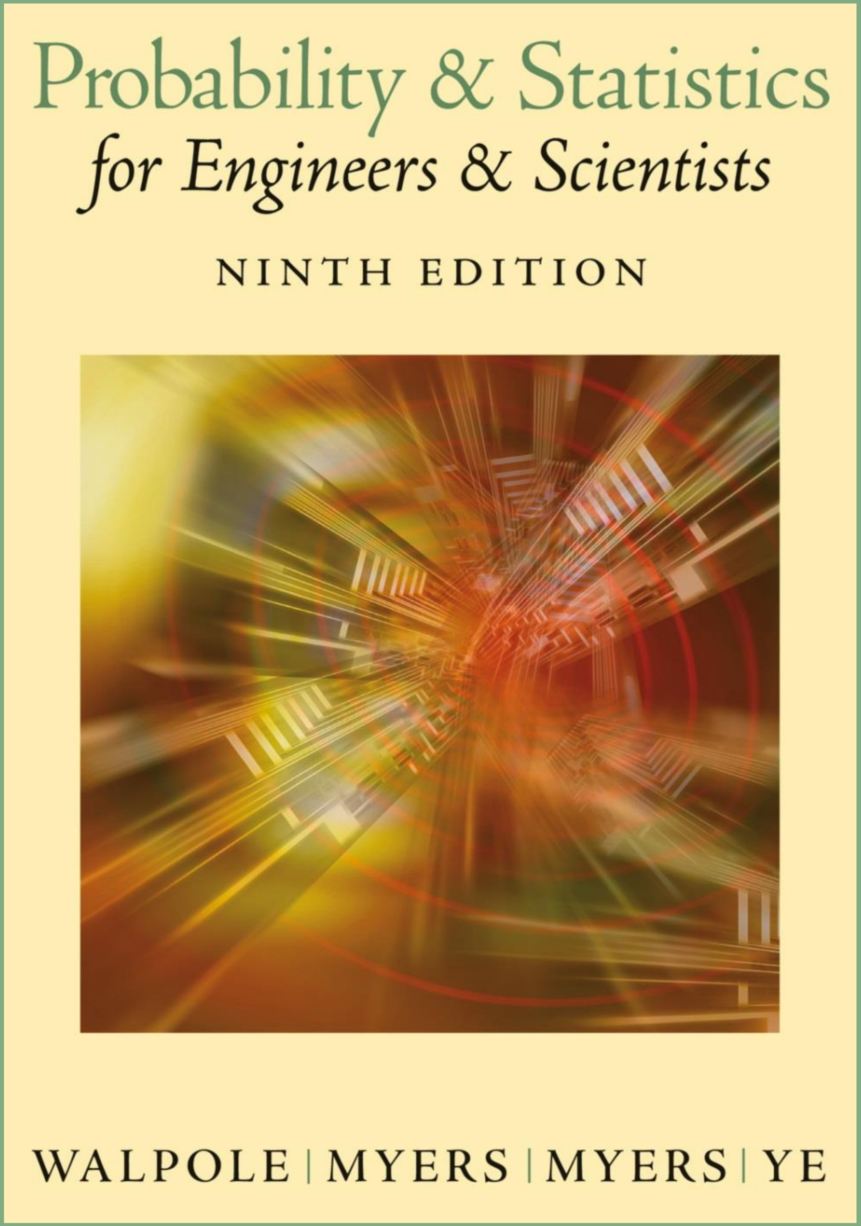
Chapter 10

One- and Two-Sample Tests of Hypotheses



Section 10.1

Statistical Hypotheses: General Concepts



Definition 10.1

A **statistical hypothesis** is an assertion or conjecture concerning one or more populations.

Section 10.2

Testing a Statistical Hypothesis

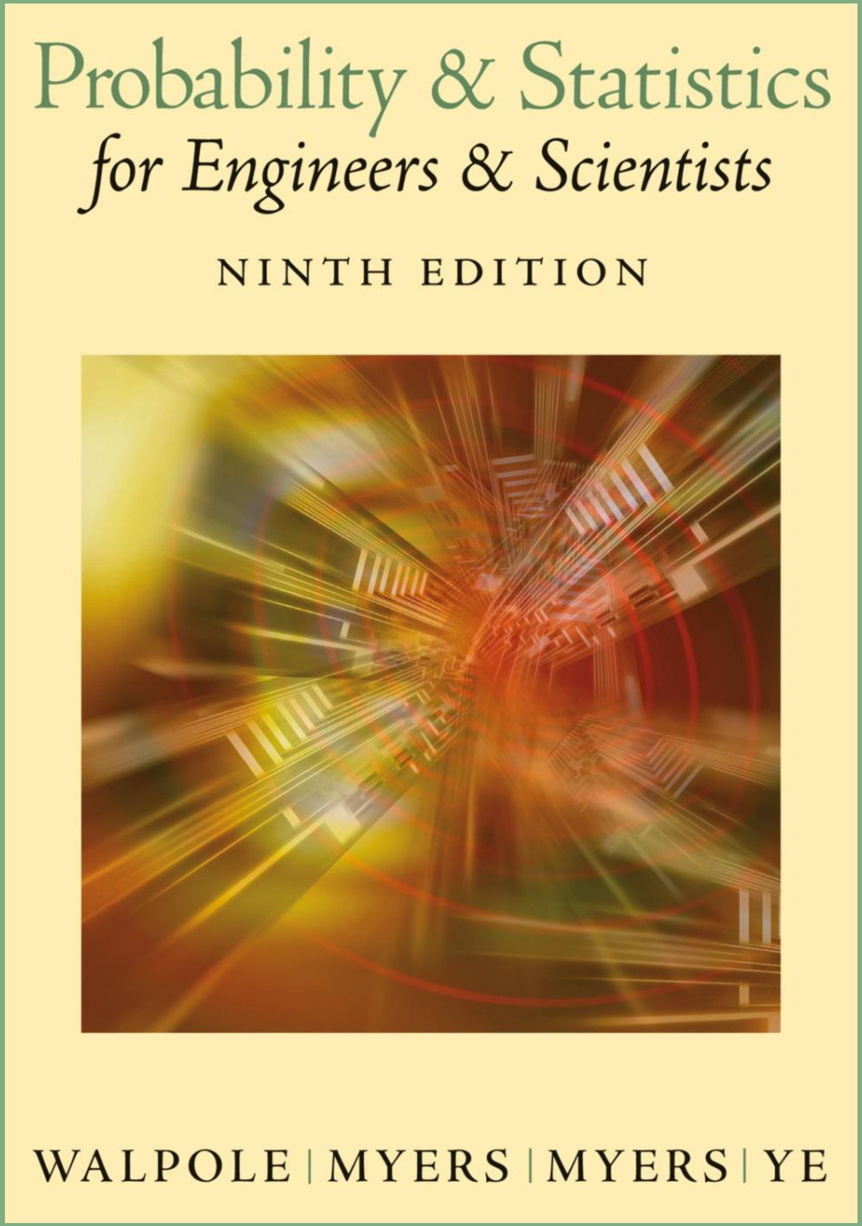
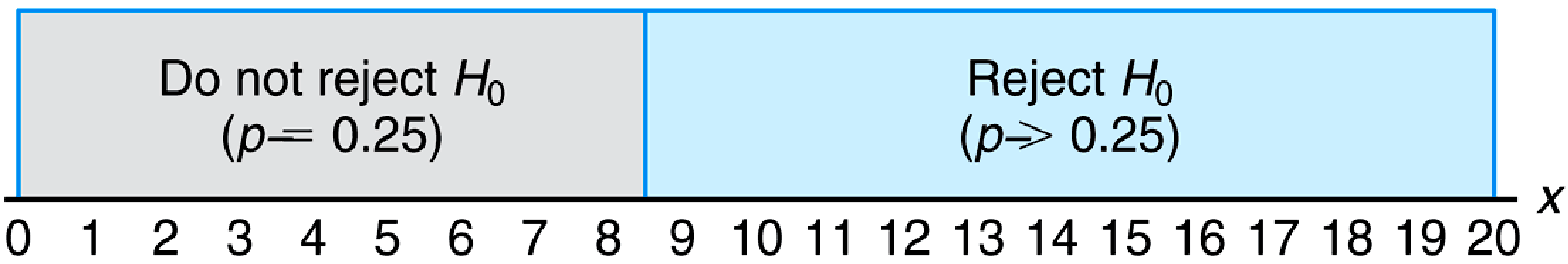


Figure 10.1 Decision criterion for testing $p = 0.25$ versus $p > 0.25$



Definition 10.2

Rejection of the null hypothesis when it is true is called a **type I error**.

Definition 10.3

Nonrejection of the null hypothesis when it is false is called a **type II error**.

Table 10.1 Possible Situations for Testing a Statistical Hypothesis

	H_0 is true	**H_0 is false**
Do not reject H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

Figure 10.2 Probability of a type I error

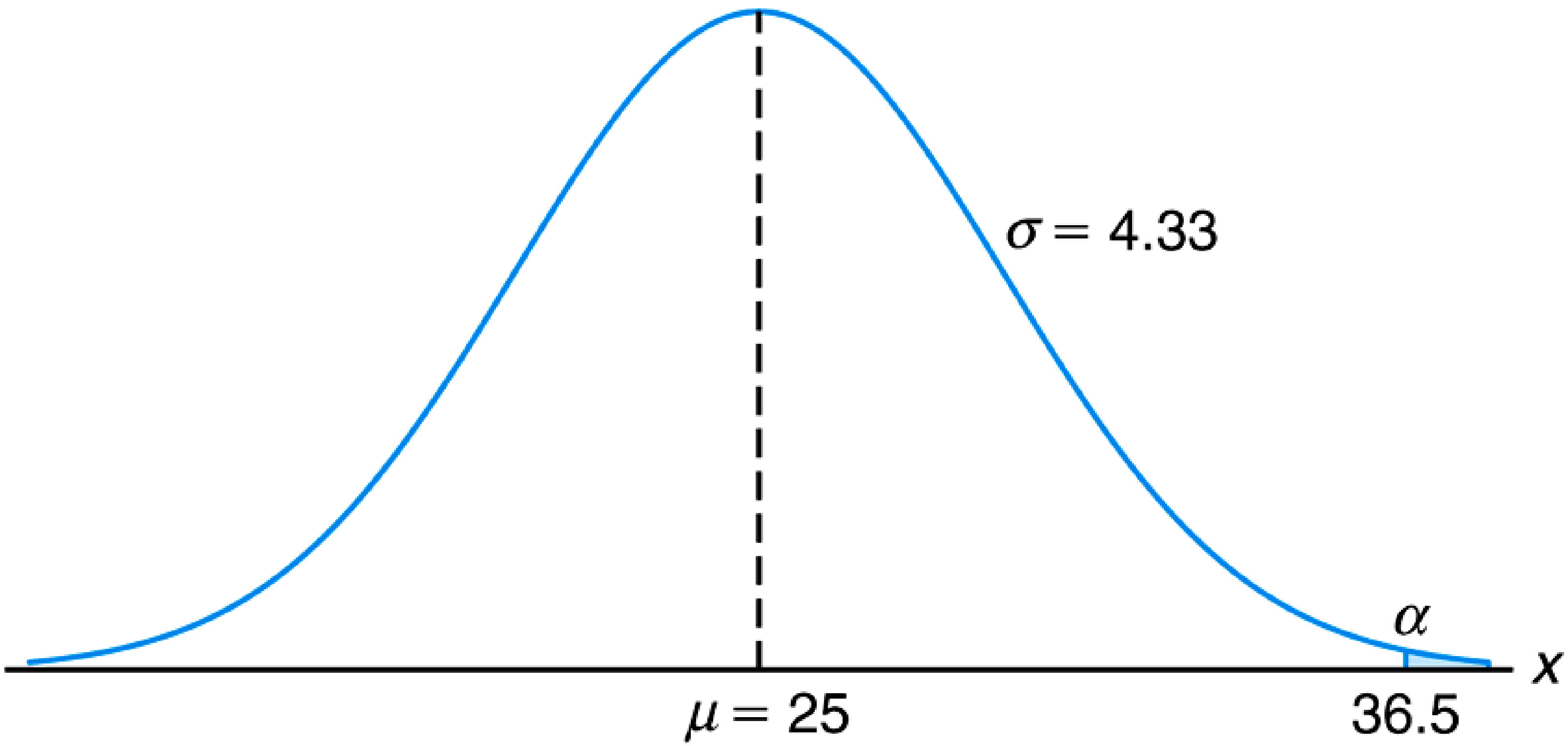


Figure 10.3 Probability of a type II error

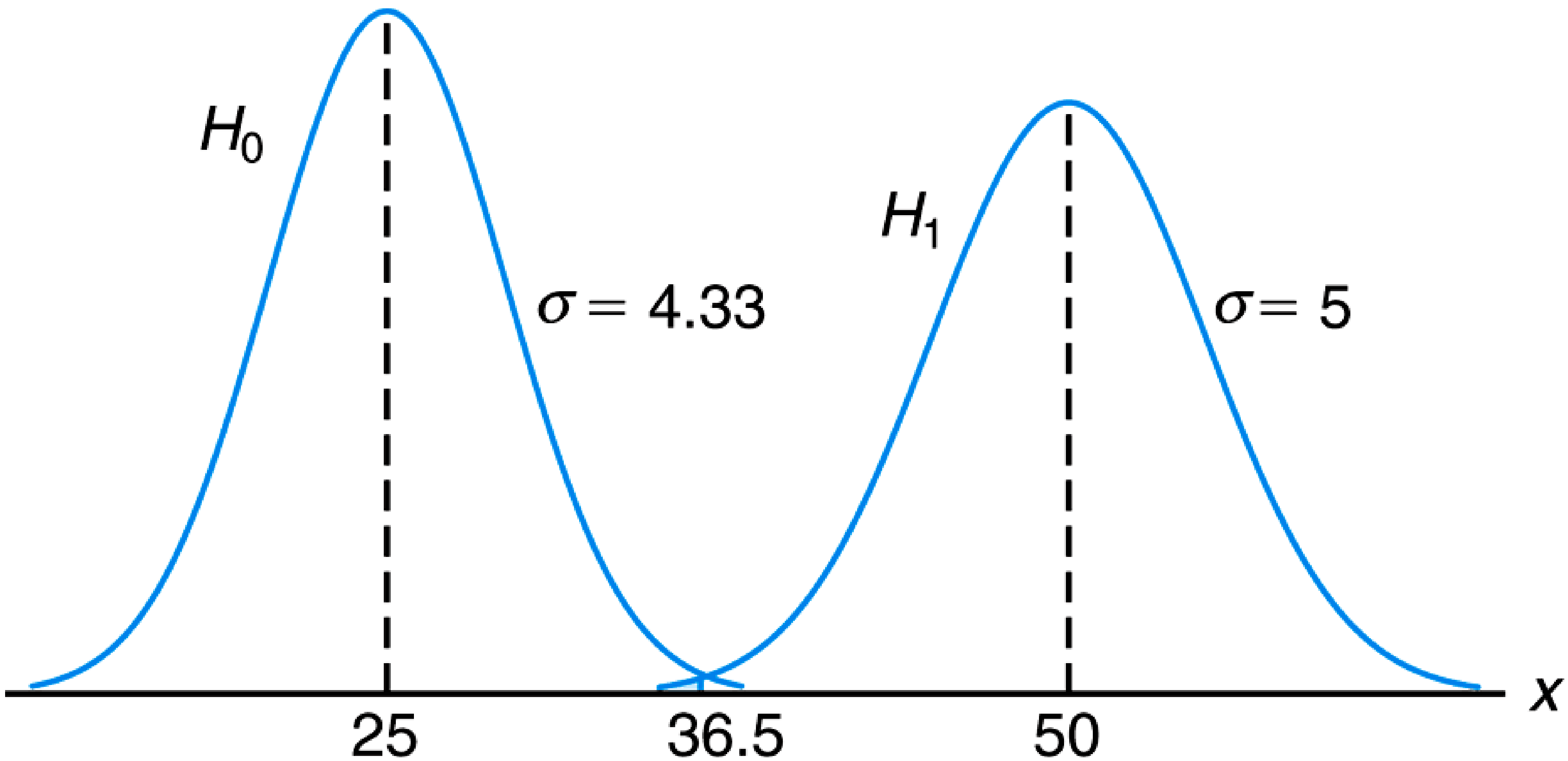


Figure 10.4 Critical region (in blue)

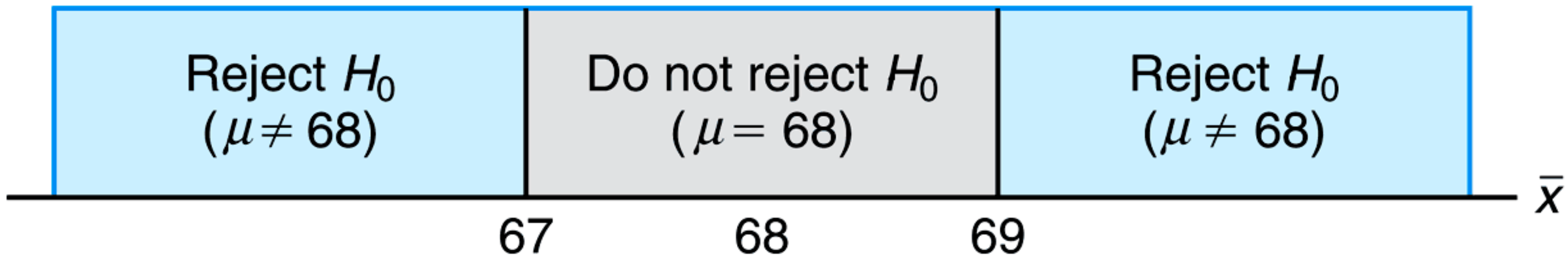


Figure 10.5 Critical Region for testing $\mu = 68$ versus $\mu \neq 68$

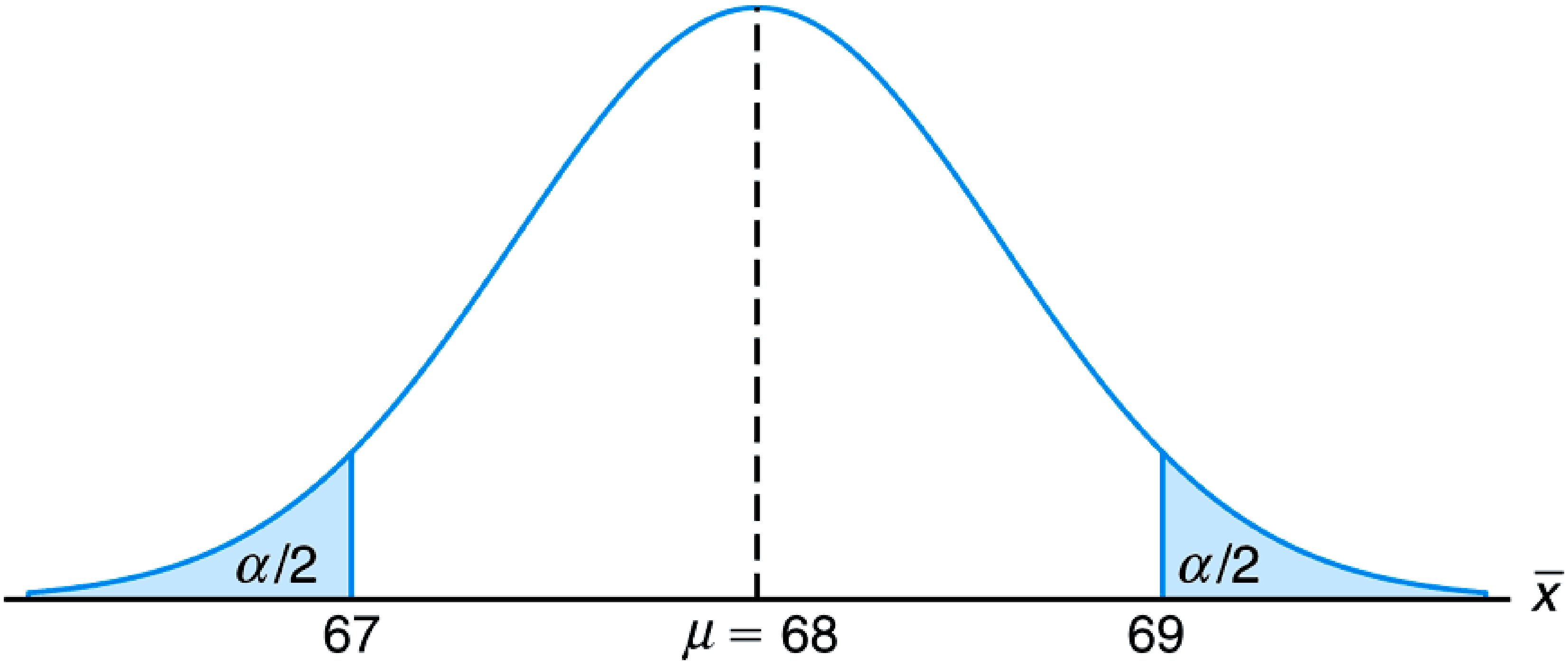


Figure 10.6 Probability of type II error for testing $\mu = 68$ versus $\mu = 70$

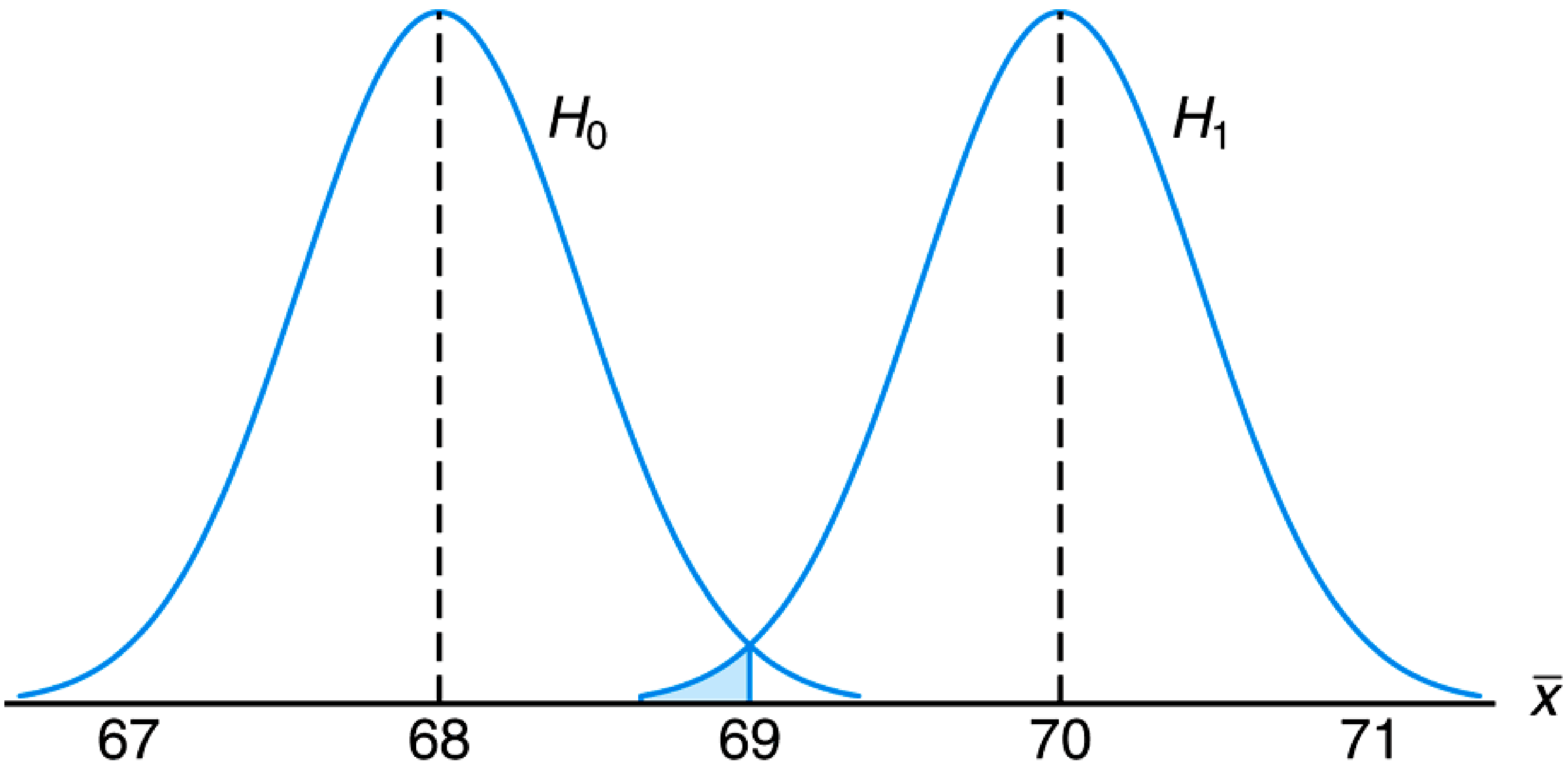
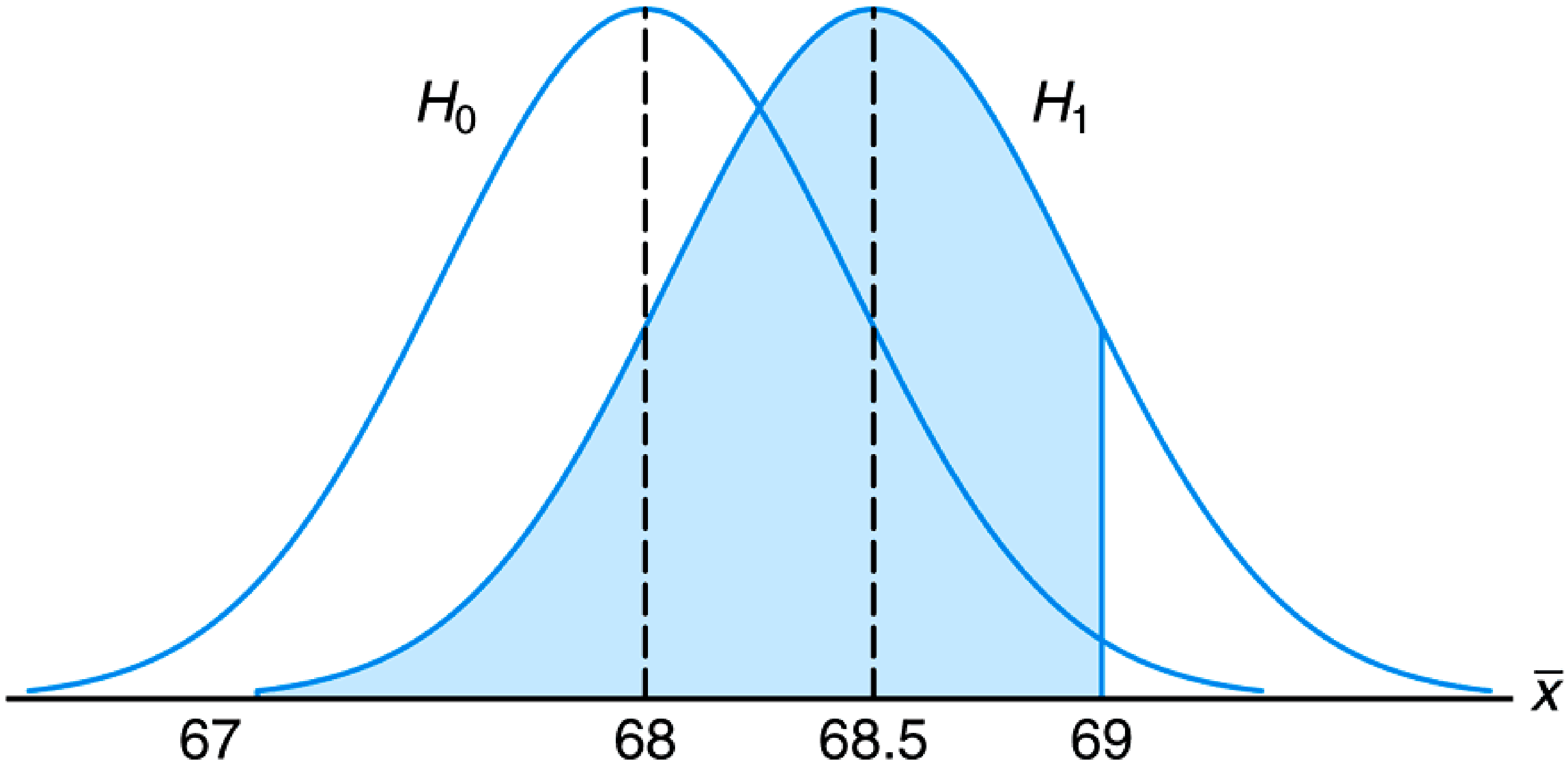


Figure 10.7 Type II error for testing $\mu = 68$ versus $\mu = 68.5$



Definition 10.4

The **power** of a test is the probability of rejecting H_0 given that a specific alternative is true.

Section 10.3

The Use of *P*-Values for Decision Making in Testing Hypotheses

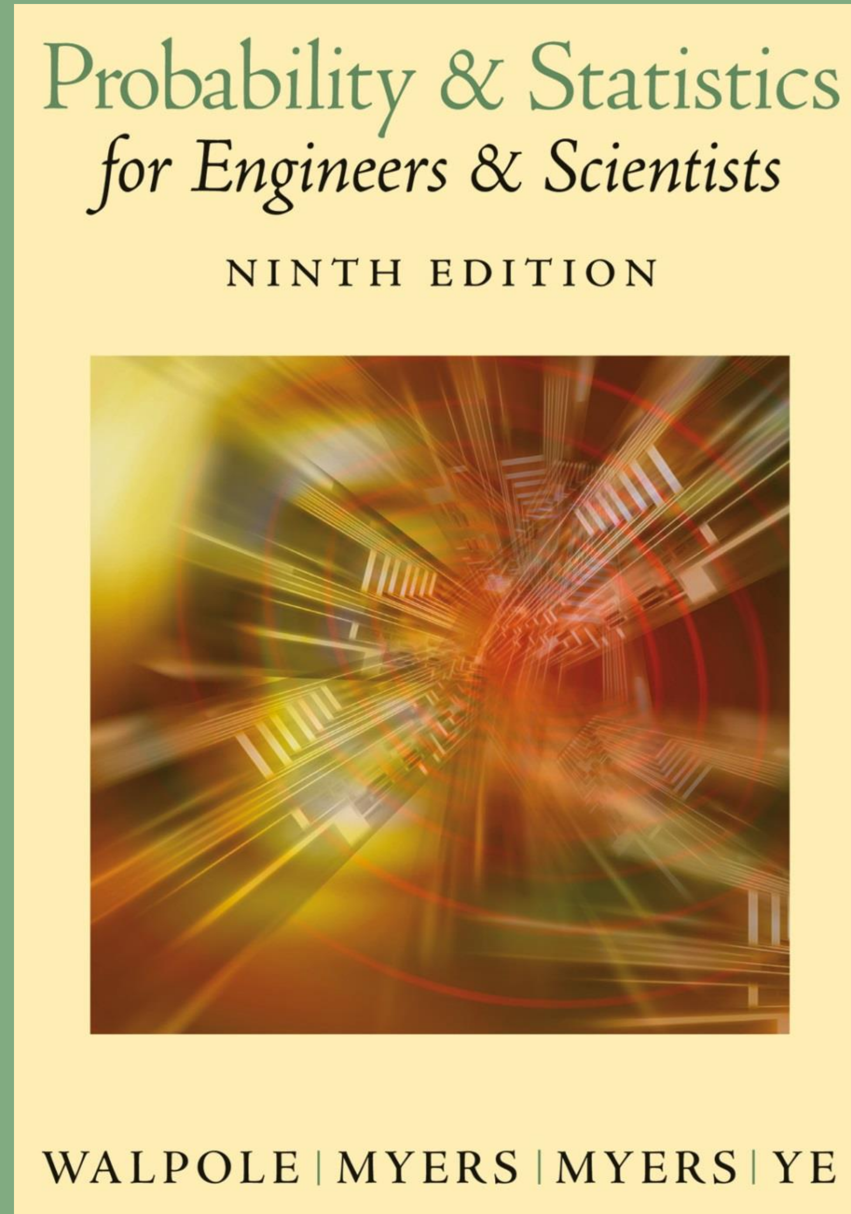
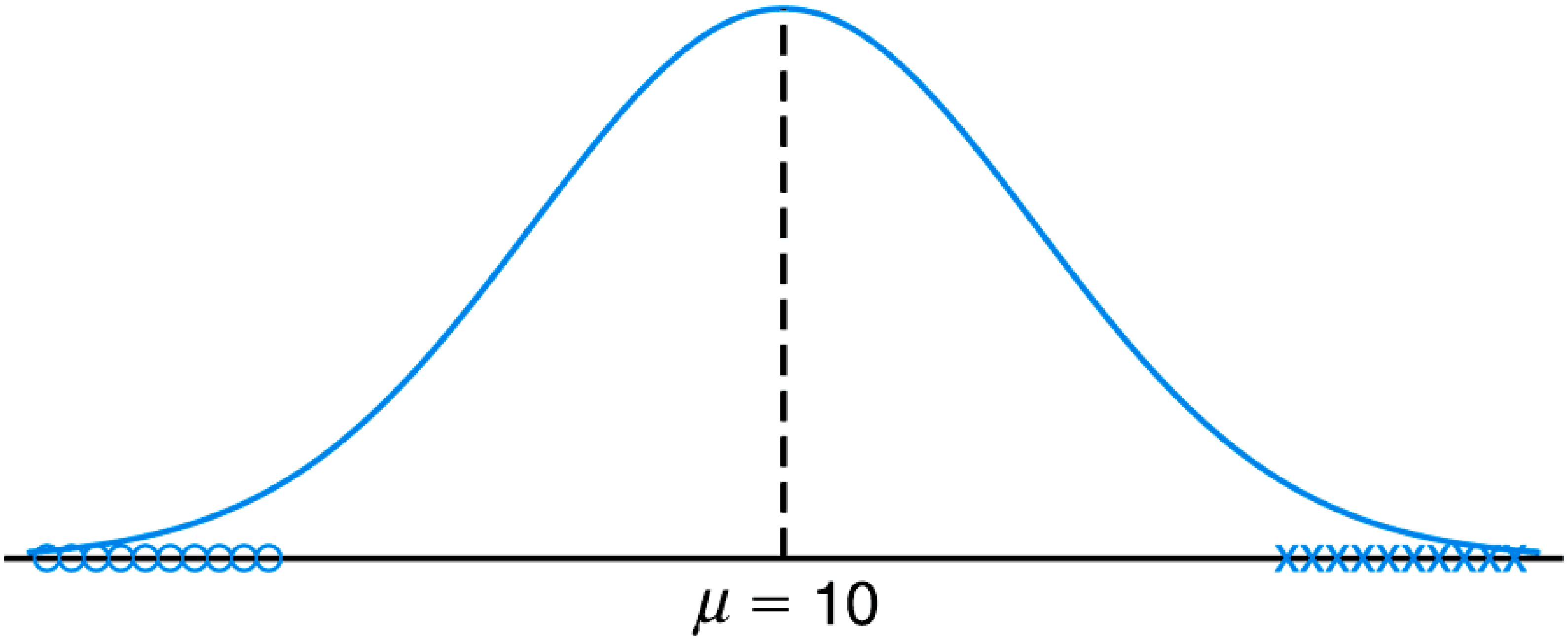


Figure 10.8 Data that are likely generated from populations having two different means



Definition 10.5

A ***P*-value** is the lowest level (of significance) at which the observed value of the test statistic is significant.

Section 10.4

Single Sample: Tests Concerning a Single Mean

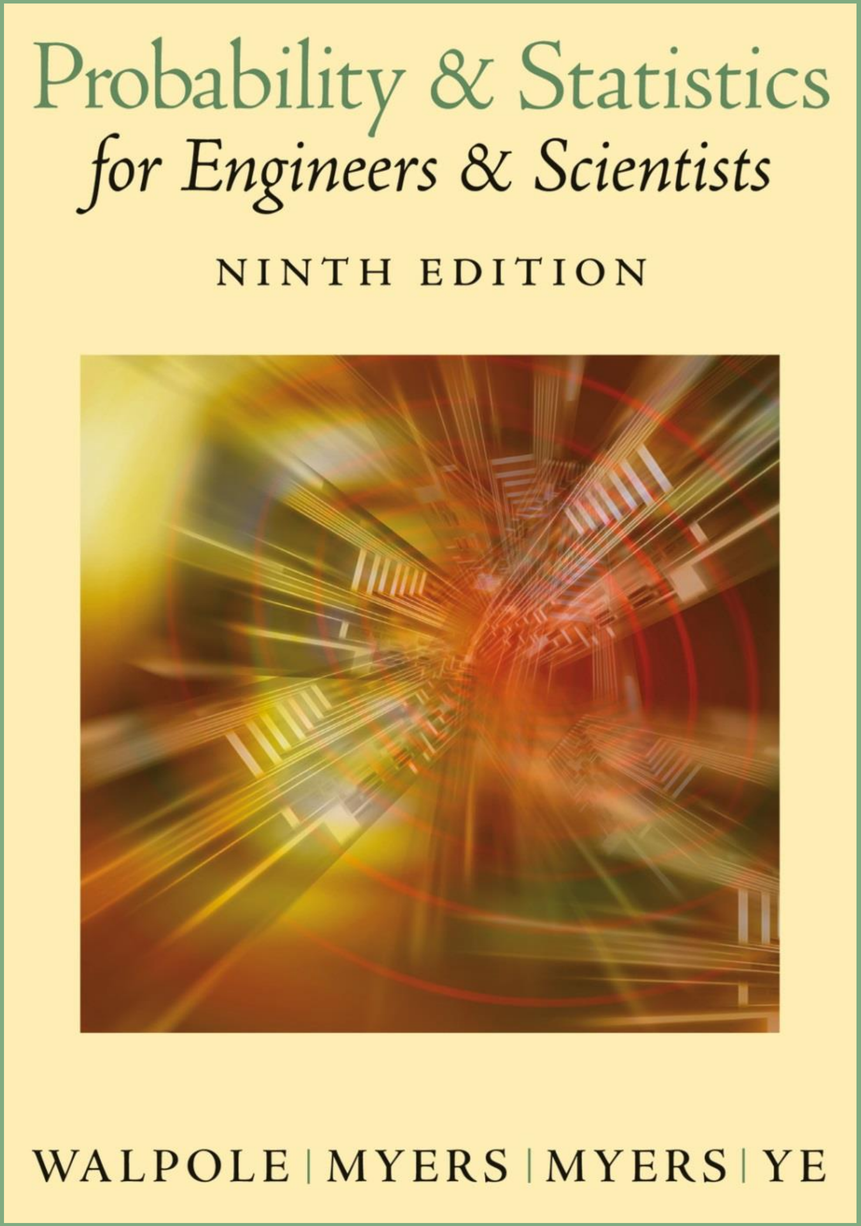


Figure 10.9 Critical region for the alternative hypothesis $\mu \neq \mu_o$

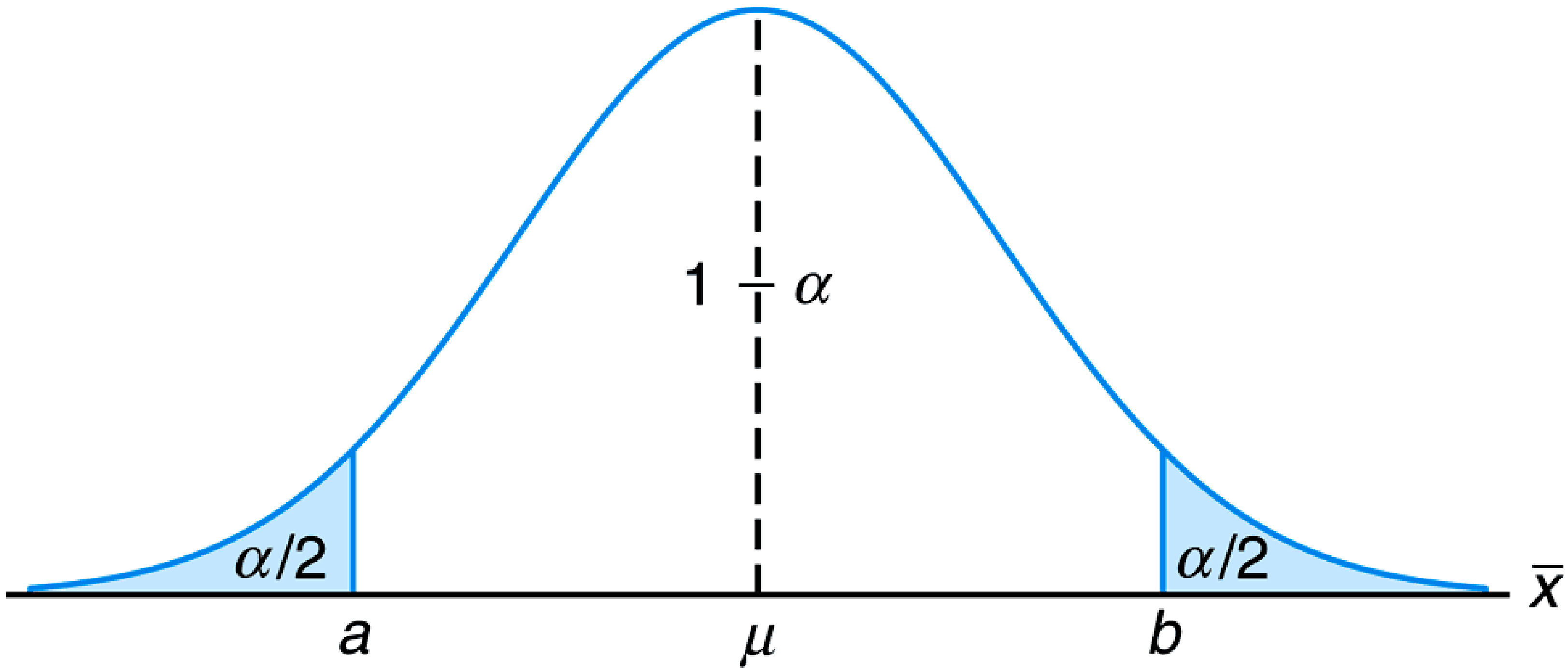


Figure 10.10 P-value for Example 10.3

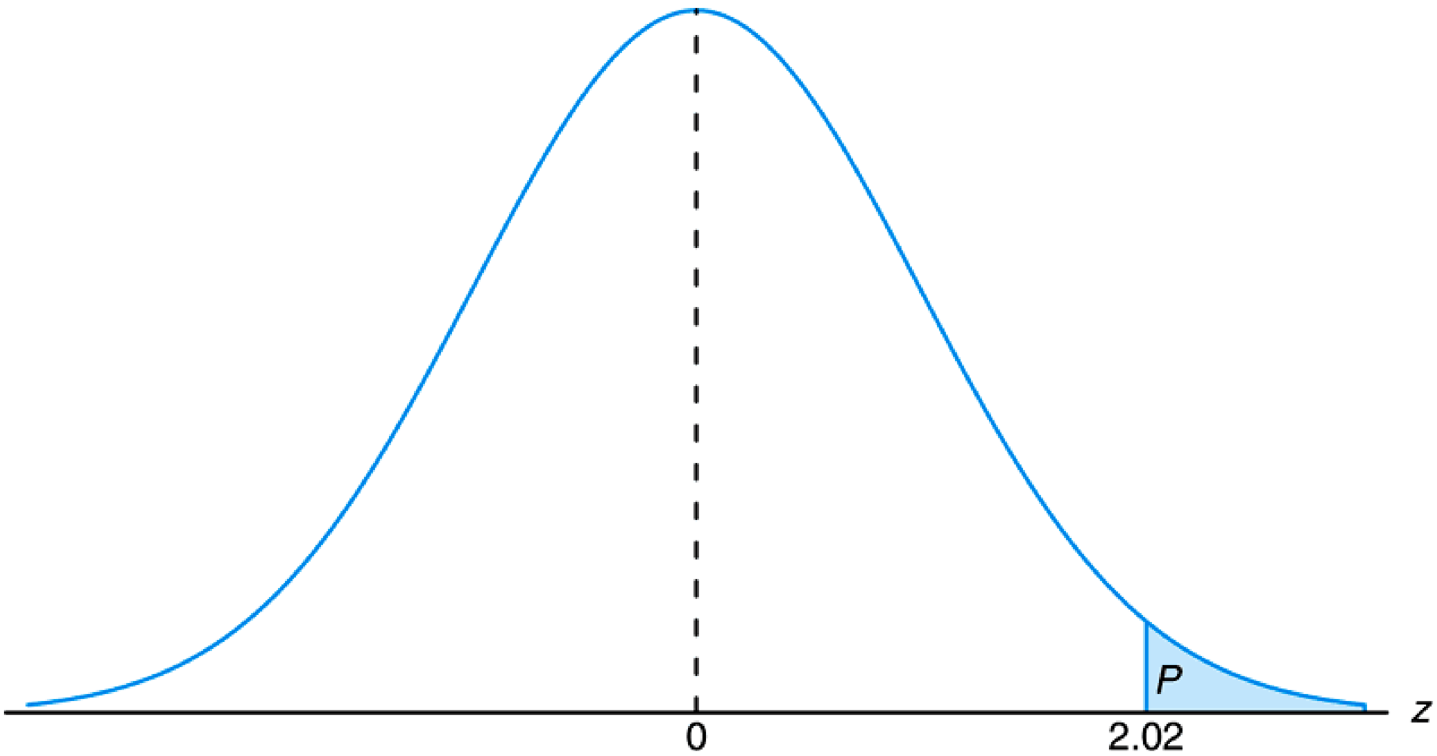


Figure 10.11 P-value for Example 10.4

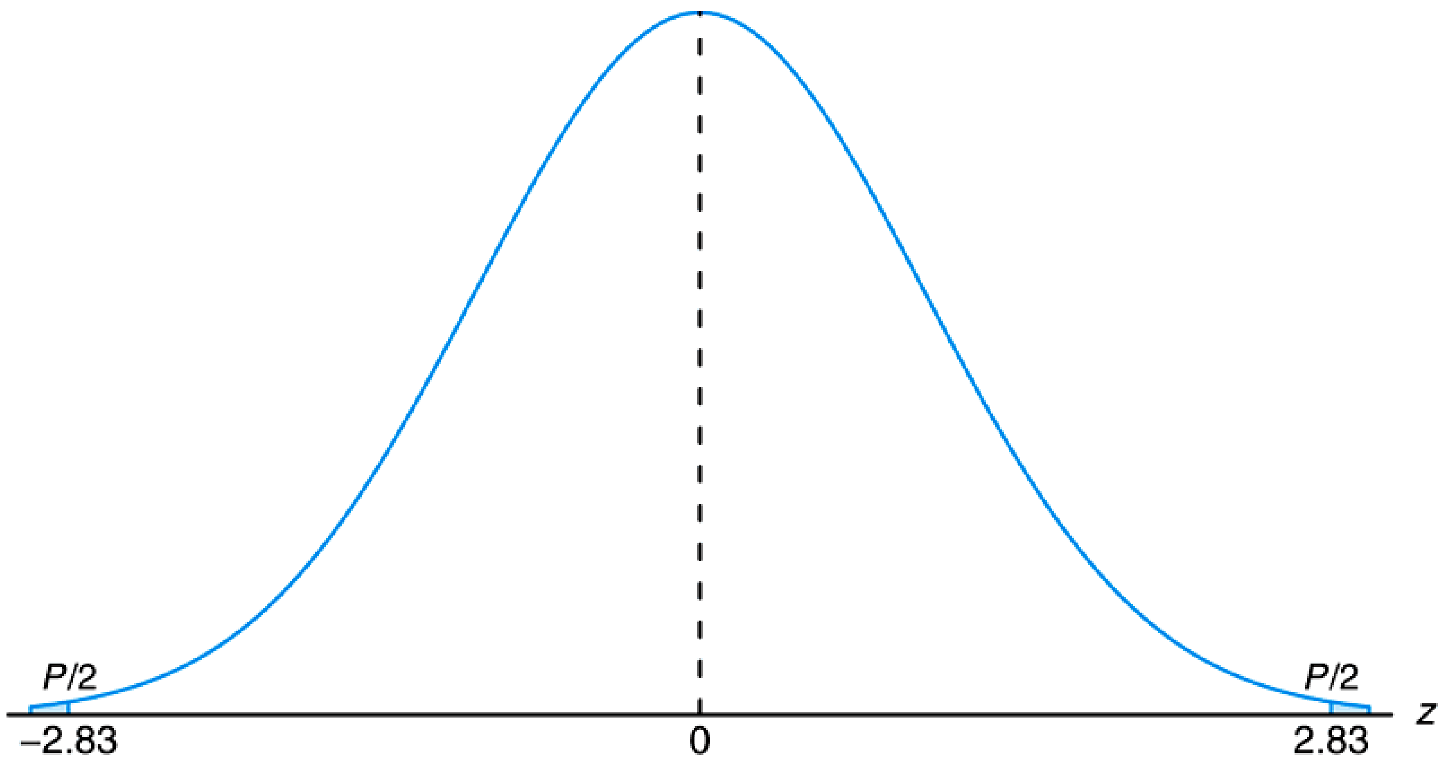


Figure 10.12 MINITAB printout for one sample *t*-test for pH meter

```
pH-meter
  7.07   7.00   7.10   6.97   7.00   7.03   7.01   7.01   6.98   7.08
MTB > Onet 'pH-meter'; SUBC>   Test 7.

One-Sample T: pH-meter Test of mu = 7 vs not = 7
Variable  N    Mean   StDev  SE Mean       95% CI            T      P
pH-meter 10 7.02500 0.04403  0.01392 (6.99350, 7.05650) 1.80 0.106
```

Section 10.5

Two Samples: Tests on Two Means

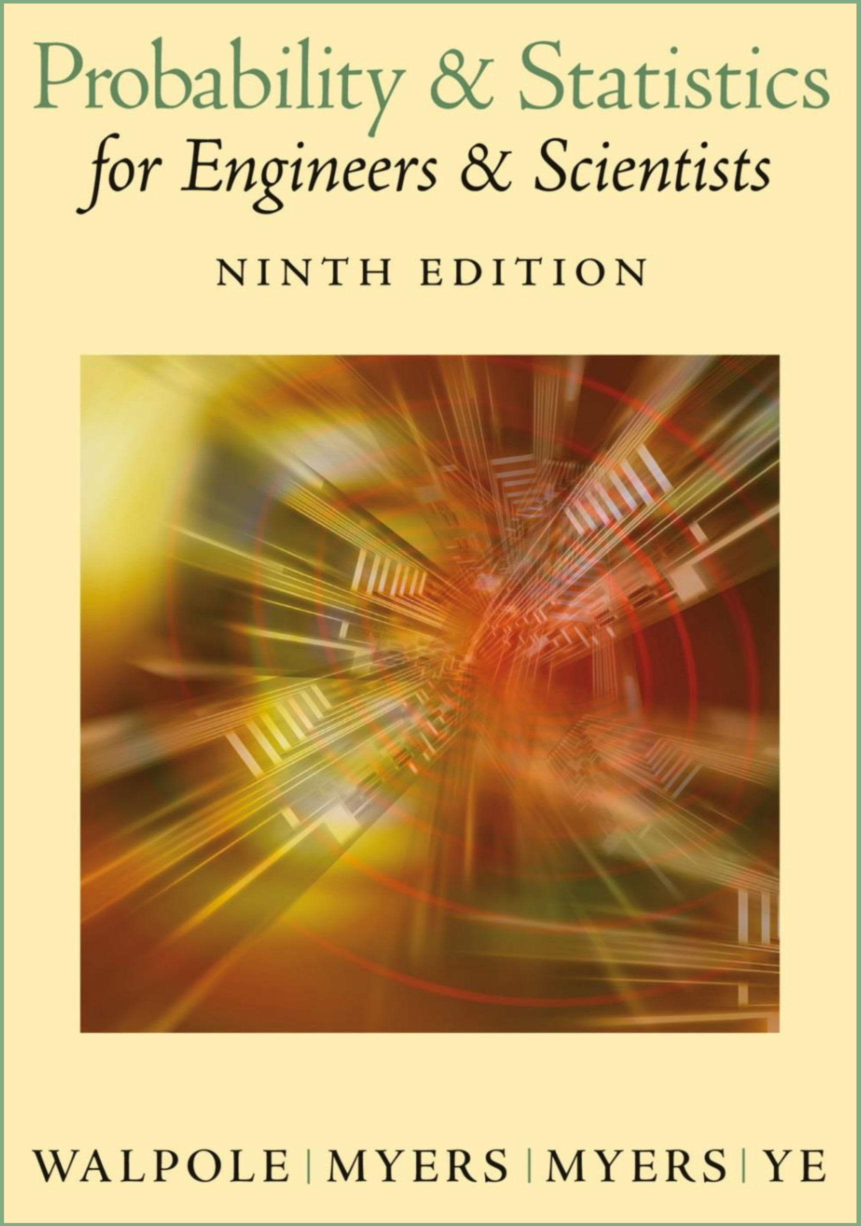


Table 10.2 Data for Case Study 10.1

Deer	Androgen (ng/mL) At Time of Injection	Androgen (ng/mL) 30 Minutes after Injection	d_i
1	2.76	7.02	4.26
2	5.18	3.10	−2.08
3	2.68	5.44	2.76
4	3.05	3.99	0.94
5	4.10	5.21	1.11
6	7.05	10.26	3.21
7	6.60	13.91	7.31
8	4.79	18.53	13.74
9	7.39	7.91	0.52
10	7.30	4.85	−2.45
11	11.78	11.10	−0.68
12	3.90	3.74	−0.16
13	26.00	94.03	68.03
14	67.48	94.03	26.55
15	17.04	41.70	24.66

Figure 10.13 SAS printout of paired *t*-test for data of Case Study 10.1

Analysis Variable : Diff

N	Mean	Std Error	t Value	Pr > |t|
15	9.8480000	4.7698699	2.06	0.0580

Section 10.6

Choice of Sample Size for Testing Means

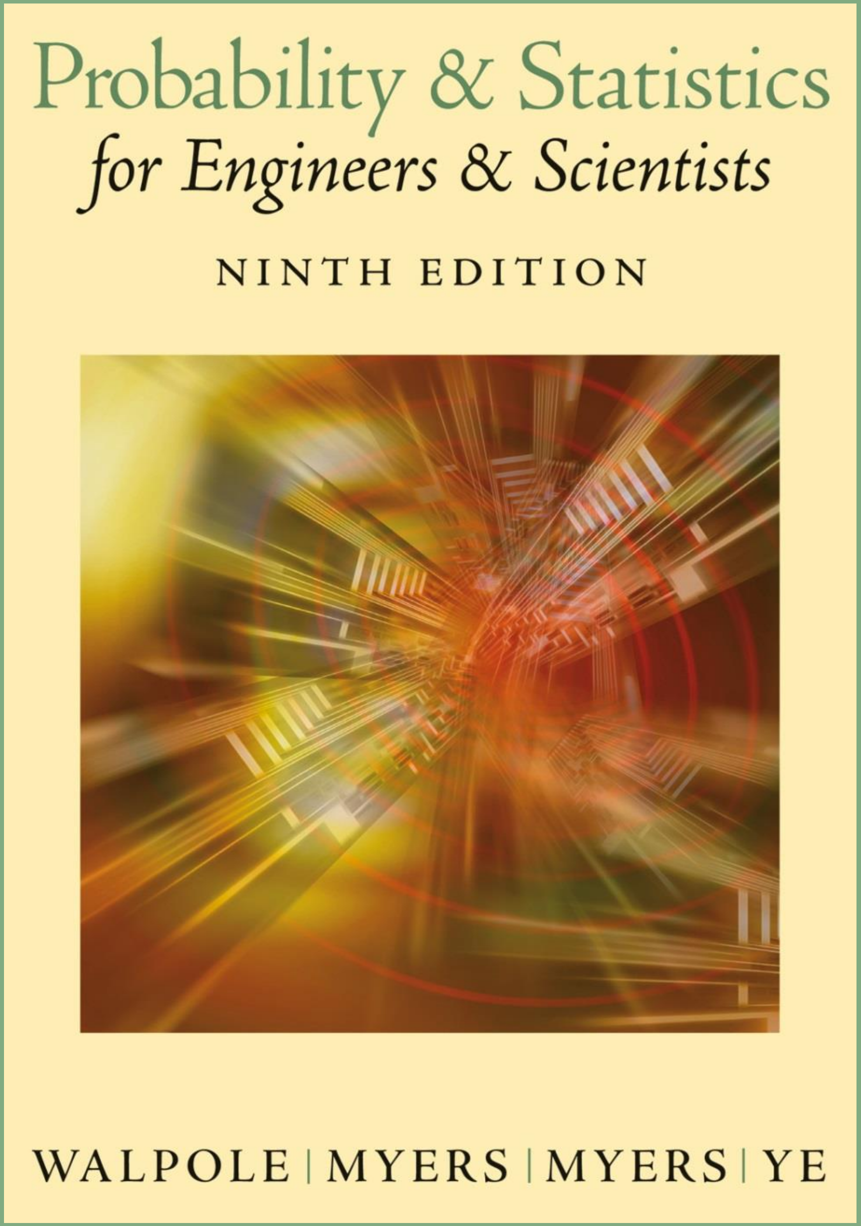


Table 10.3 Tests Concerning Means

H_0	Value of Test Statistic	H_1	Critical Region
$\mu = \mu_0$	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$; σ known	$\mu < \mu_0$ $\mu > \mu_0$ $\mu \neq \mu_0$	$z < -z_\alpha$ $z > z_\alpha$ $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$
$\mu = \mu_0$	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$; $v = n - 1$, σ unknown	$\mu < \mu_0$ $\mu > \mu_0$ $\mu \neq \mu_0$	$t < -t_\alpha$ $t > t_\alpha$ $t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$
$\mu_1 - \mu_2 = d_0$	$z = \dfrac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$; σ_1 and σ_2 known	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$z < -z_\alpha$ $z > z_\alpha$ $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$
$\mu_1 - \mu_2 = d_0$	$t = \dfrac{(\bar{x}_1 - \bar{x}_2) - d_0}{s_p\sqrt{1/n_1 + 1/n_2}}$; $v = n_1 + n_2 - 2$, $\sigma_1 = \sigma_2$ but unknown, $s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$t < -t_\alpha$ $t > t_\alpha$ $t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$
$\mu_1 - \mu_2 = d_0$	$t' = \dfrac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$; $v = \dfrac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$, $\sigma_1 \neq \sigma_2$ and unknown	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$t' < -t_\alpha$ $t' > t_\alpha$ $t' < -t_{\alpha/2}$ or $t' > t_{\alpha/2}$
$\mu_D = d_0$ paired observations	$t = \dfrac{\bar{d} - d_0}{s_d/\sqrt{n}}$; $v = n - 1$	$\mu_D < d_0$ $\mu_D > d_0$ $\mu_D \neq d_0$	$t < -t_\alpha$ $t > t_\alpha$ $t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$

Figure 10.14 Testing $\mu = \mu_o$ versus $\mu = \mu_o + \delta$

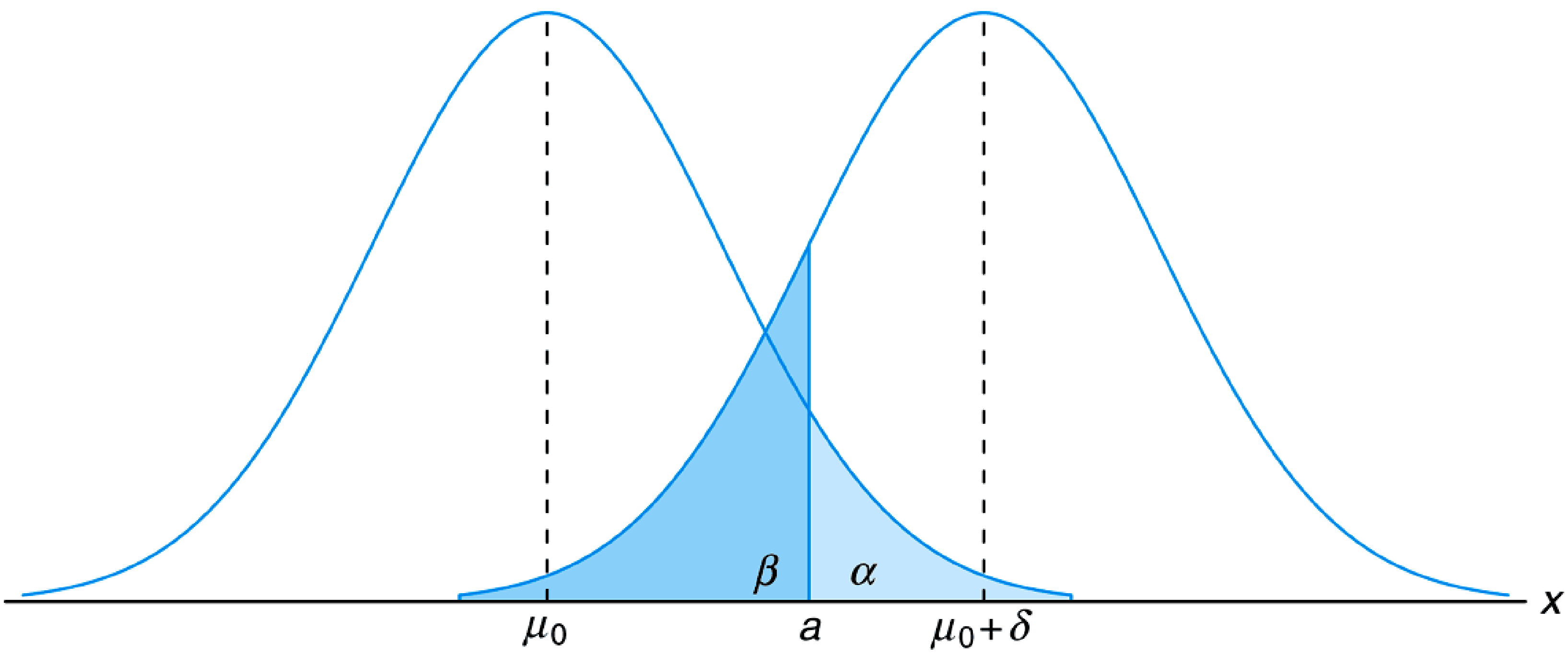
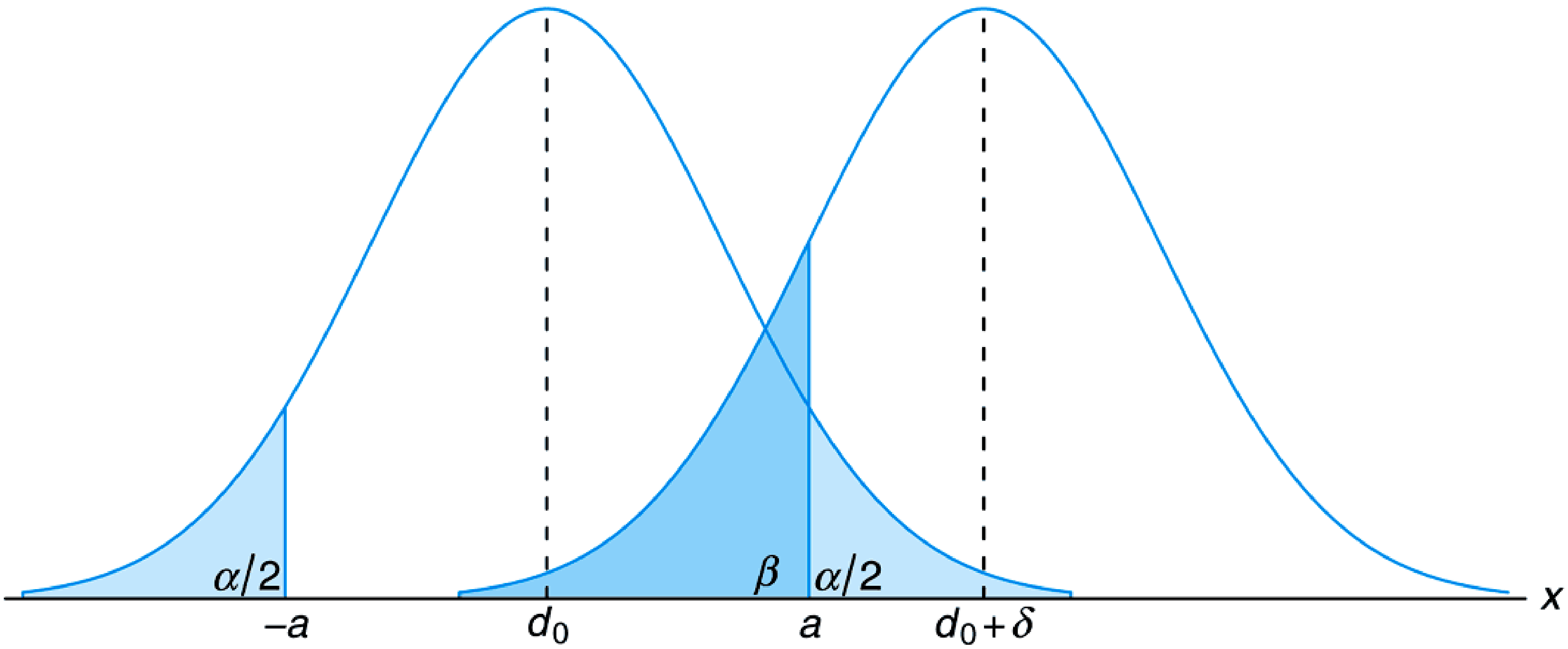


Figure 10.15 Testing $\mu_1 - \mu_2 = d_o$ versus $\mu_1 - \mu_2 = d_o + \delta$



Section 10.7

Graphical Methods for Comparing Means

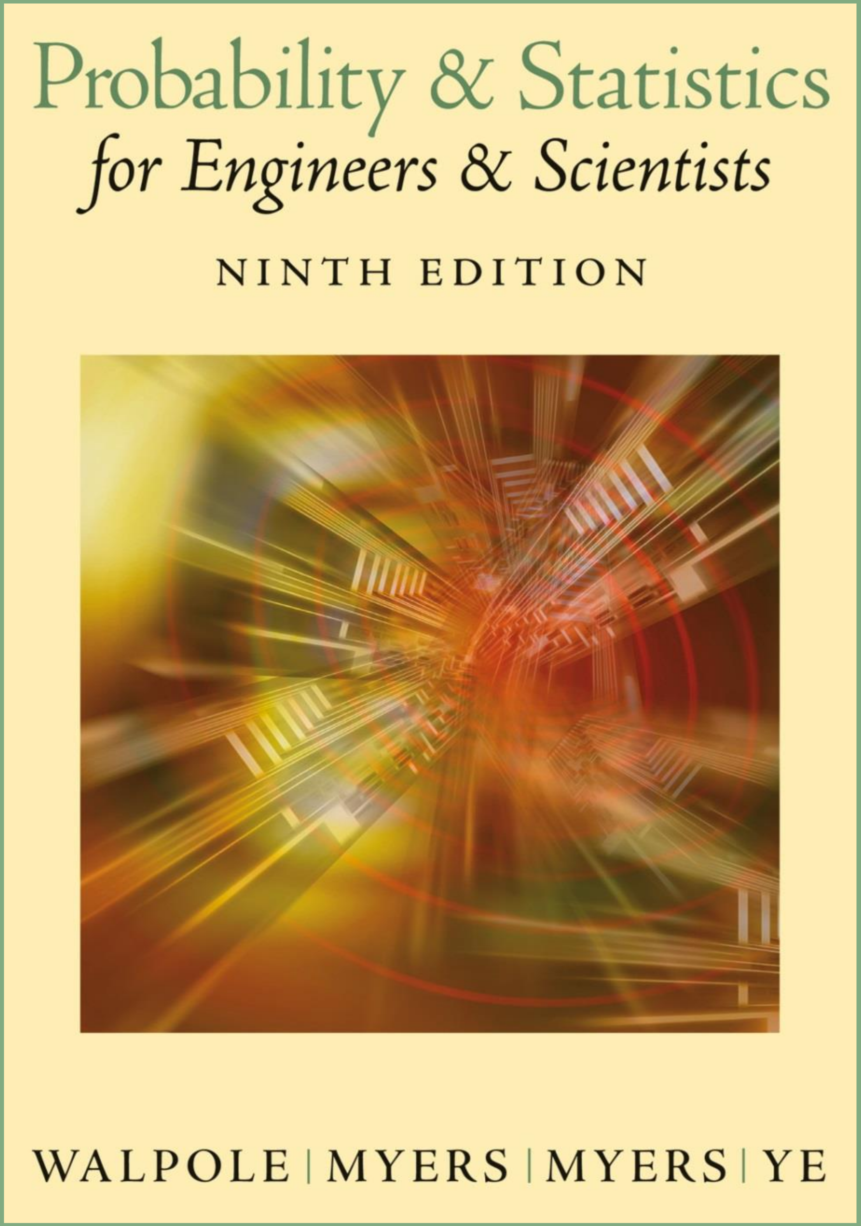


Figure 10.16 Two box-and whisker plots of plasma ascorbic acid in smokers and nonsmokers

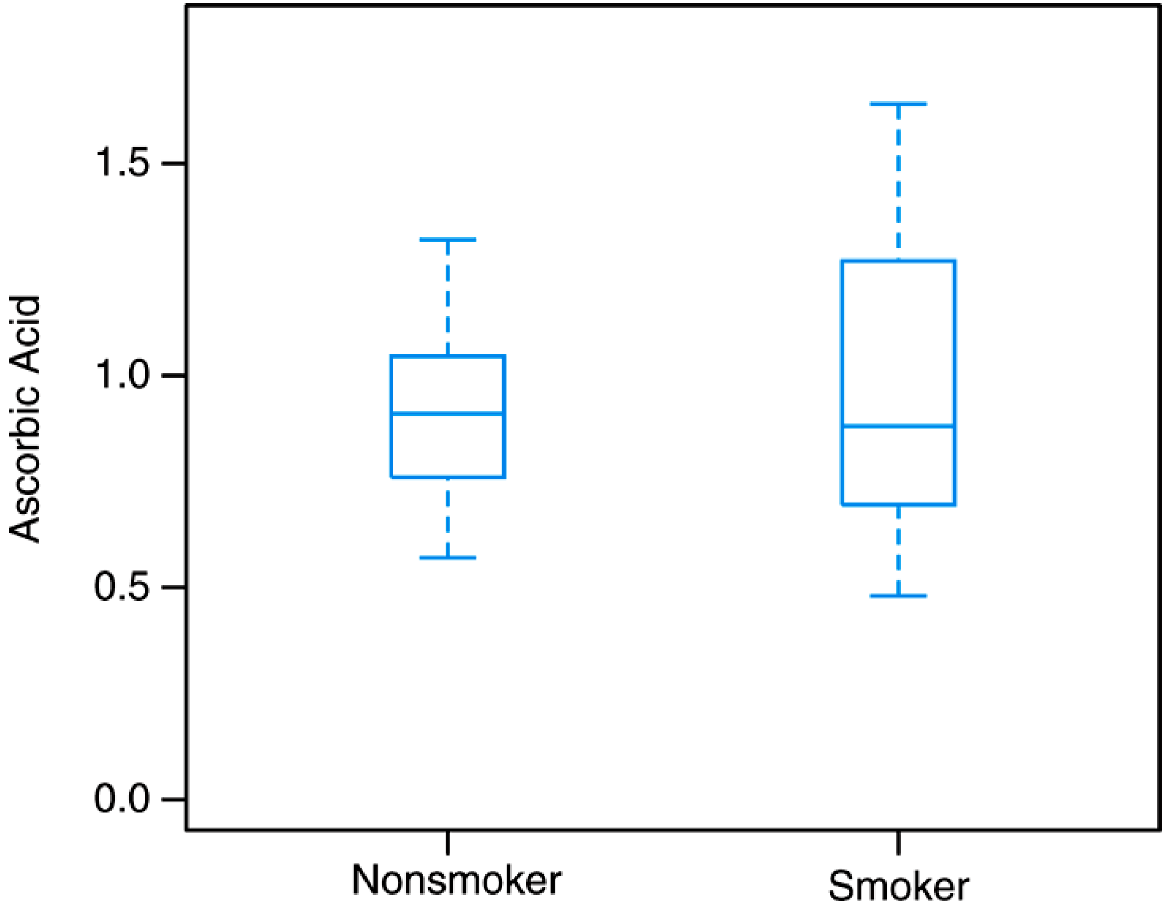


Figure 10.17 Two box-and-whisker plots of seedling data

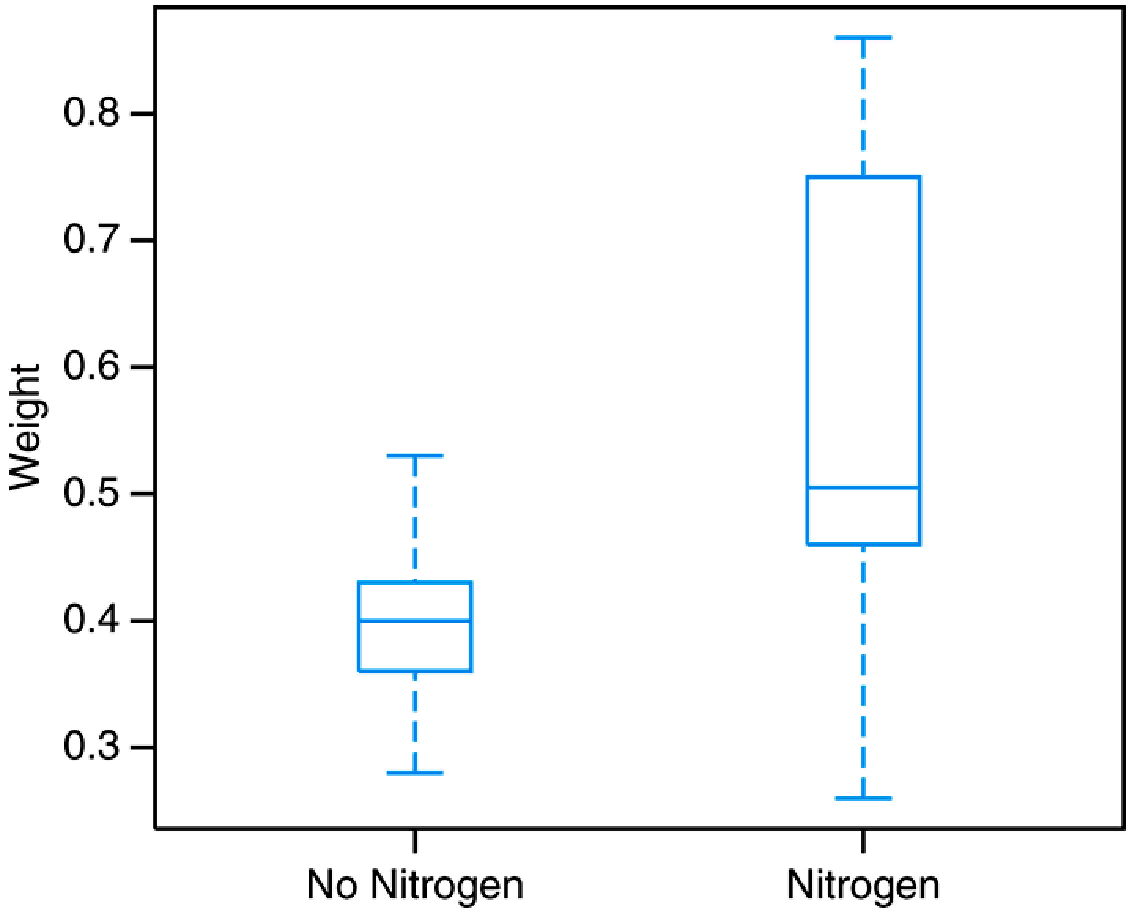


Figure 10.18 SAS printout for two-sample *t*-test

TTEST Procedure

Variable Weight

Mineral	N	Mean	Std Dev	Std Err
No nitrogen	10	0.3990	0.0728	0.0230
Nitrogen	10	0.5650	0.1867	0.0591

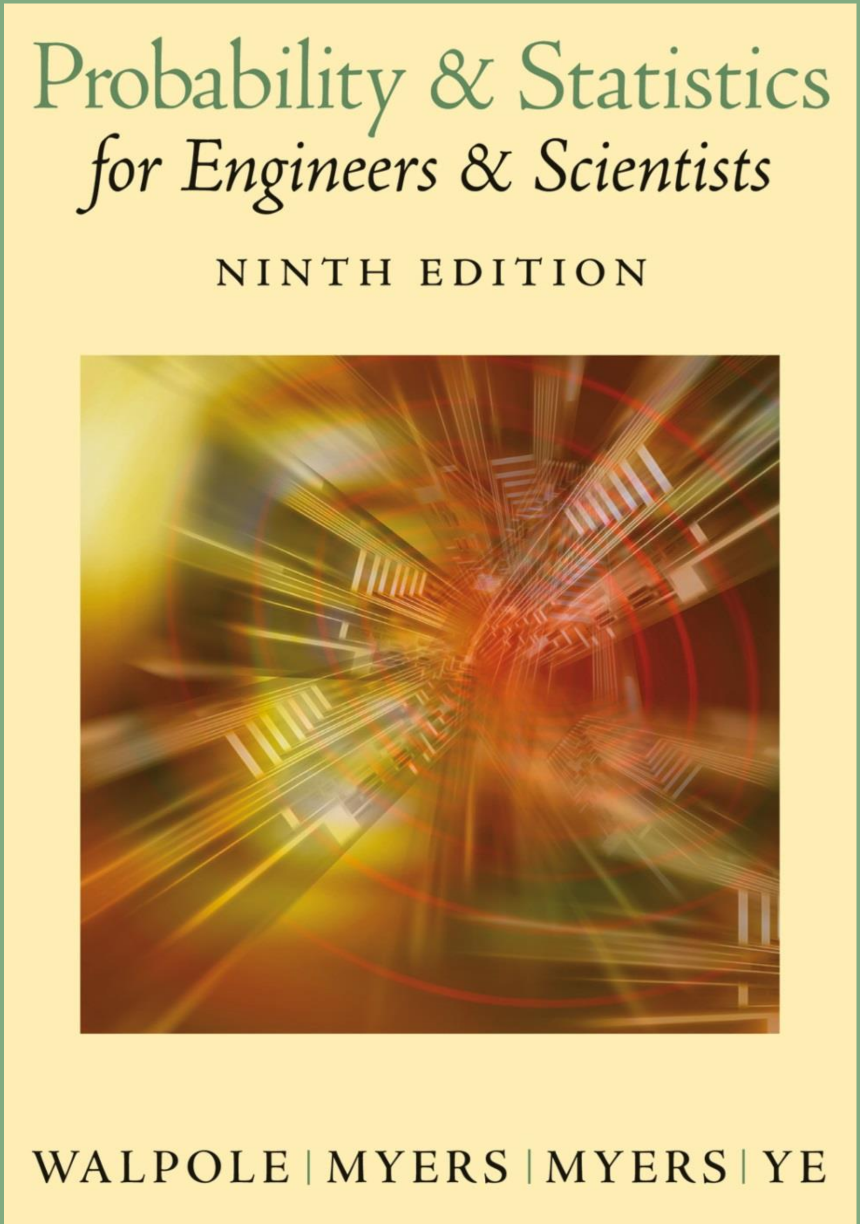
Variances	DF	t Value	Pr > |t|
Equal	18	2.62	0.0174
Unequal	11.7	2.62	0.0229

Test the Equality of Variances

Variable	Num DF	Den DF	F Value	Pr > F
Weight	9	9	6.58	0.0098

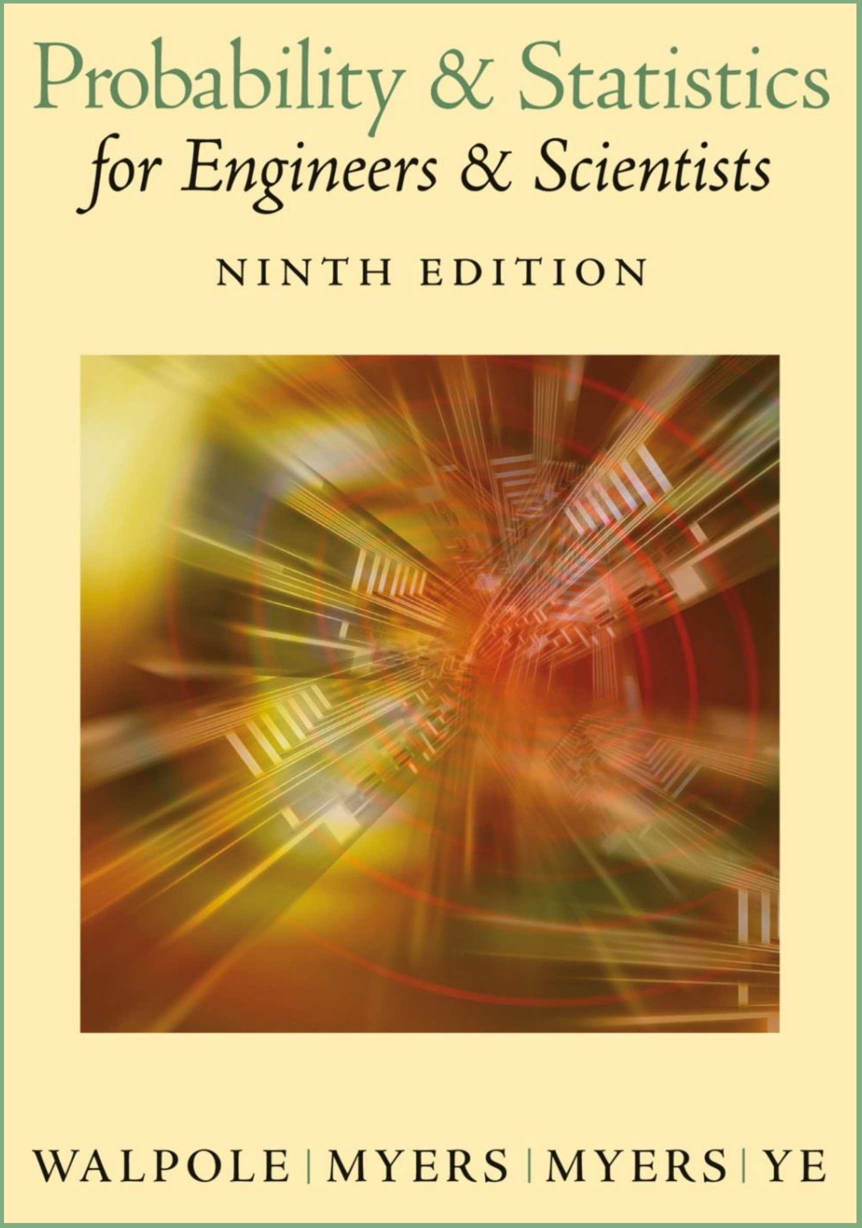
Section 10.8

One Sample: Test on a Single Proportion



Section 10.9

Two Samples: Tests on Two Proportions



Section 10.10

One- and Two-Sample Tests Concerning Variances

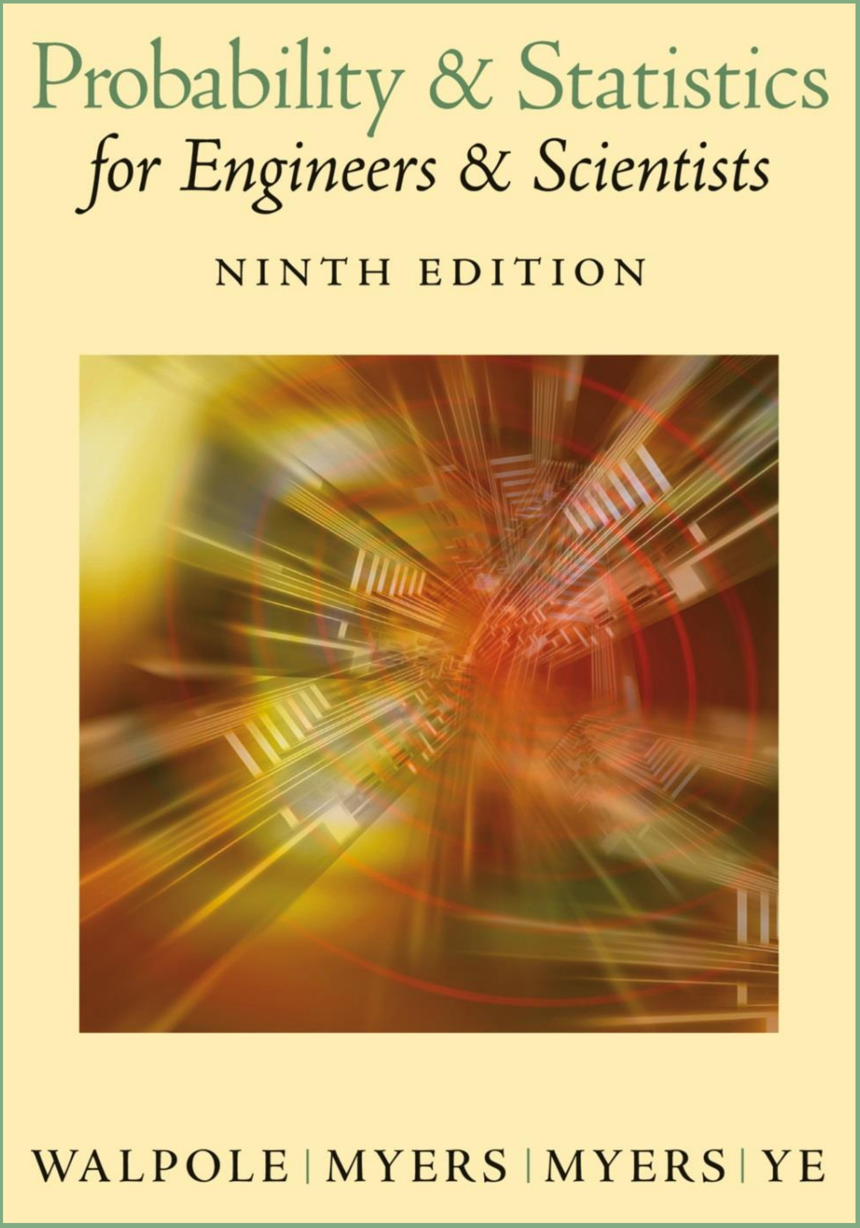


Figure 10.19 Critical region for the alternative hypothesis $\sigma > 0.9$

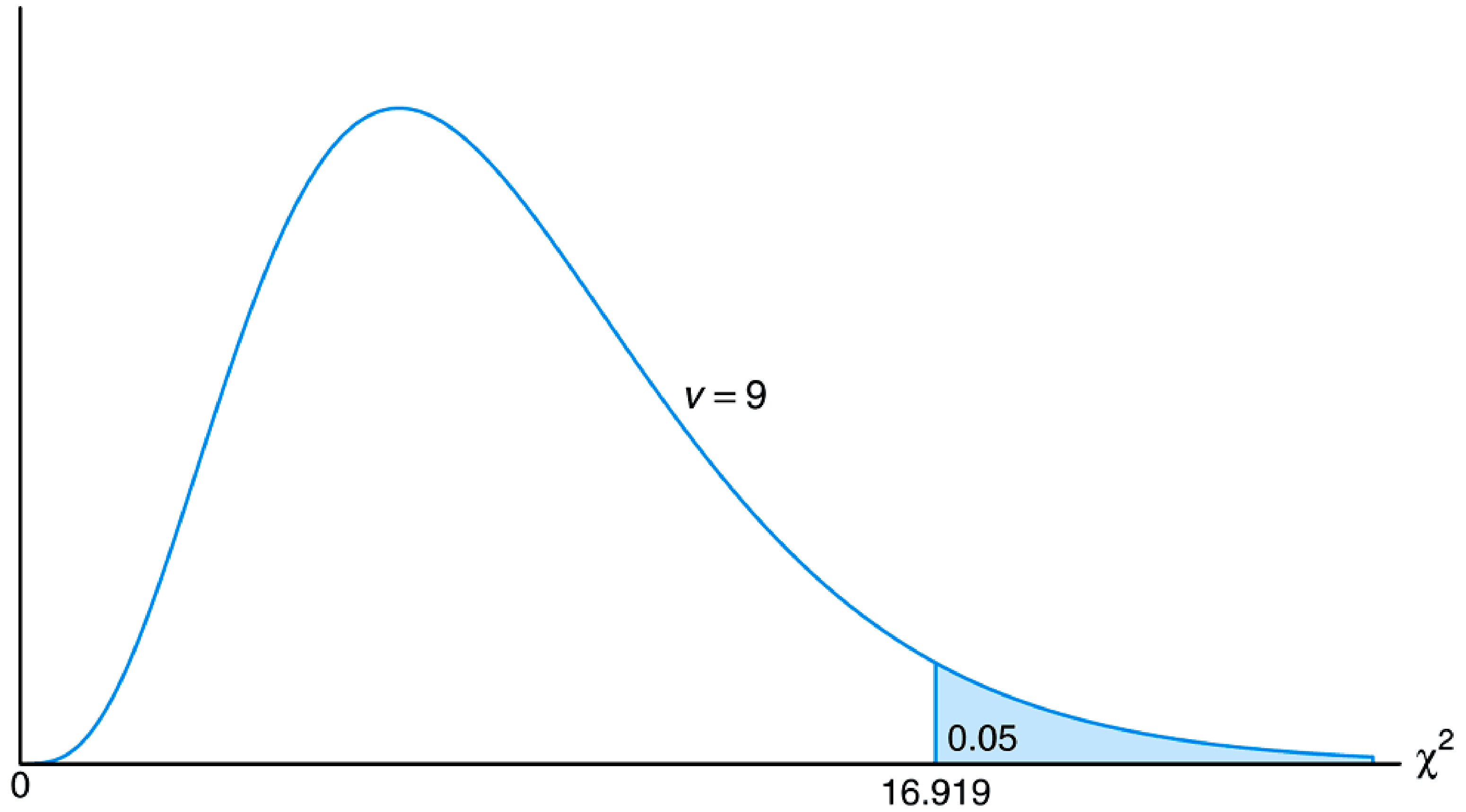
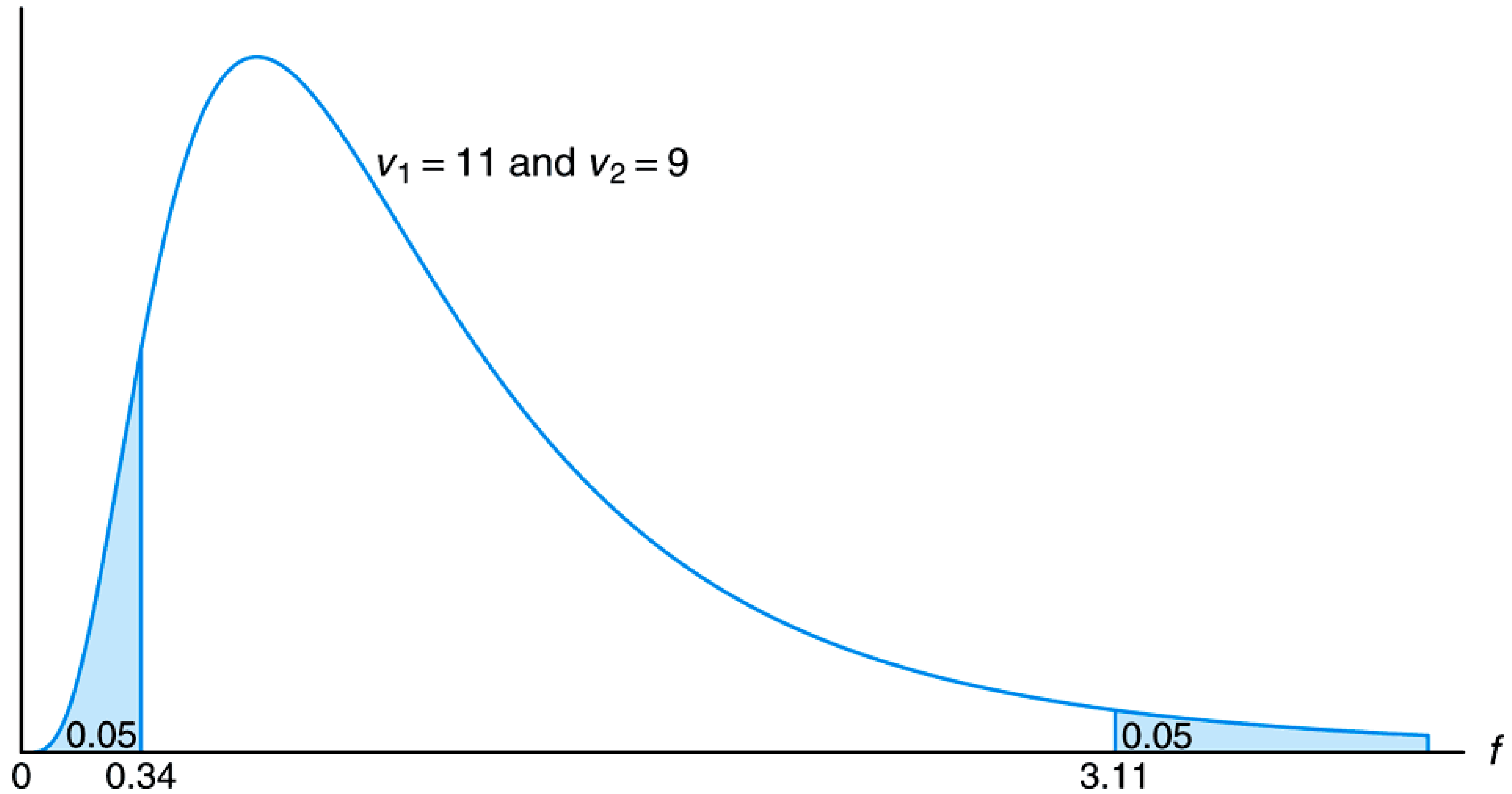


Figure 10.20 Critical region for the alternative hypothesis $\sigma_1^2 \neq \sigma_2^2$



Section 10.11

Goodness-of-Fit Test

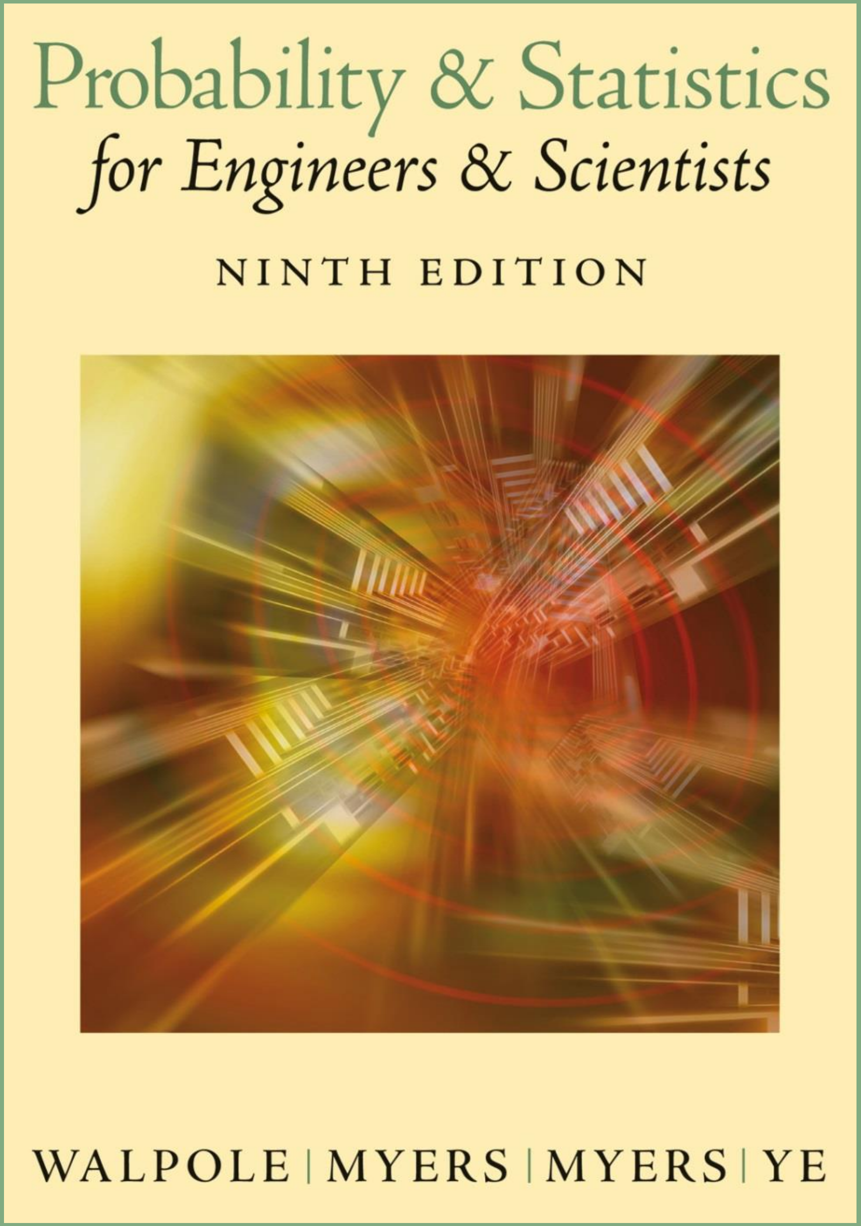


Table 10.4 Observed and Expected Frequencies of 120 Tosses of a Die

Face:	**1**	**2**	**3**	**4**	**5**	**6**
Observed	20	22	17	18	19	24
Expected	20	20	20	20	20	20

Table 10.5 Observed and Expected Frequencies of Battery Lives, Assuming Normality

Class Boundaries	o_i		e_i	
1.45–1.95	2	7	0.5	8.5
1.95–2.45	1		2.1	
2.45–2.95	4		5.9	
2.95–3.45	15		10.3	
3.45–3.95	10		10.7	
3.95–4.45	5	8	7.0	10.5
4.45–4.95	3		3.5	

Section 10.12

Test for Independence (Categorical Data)

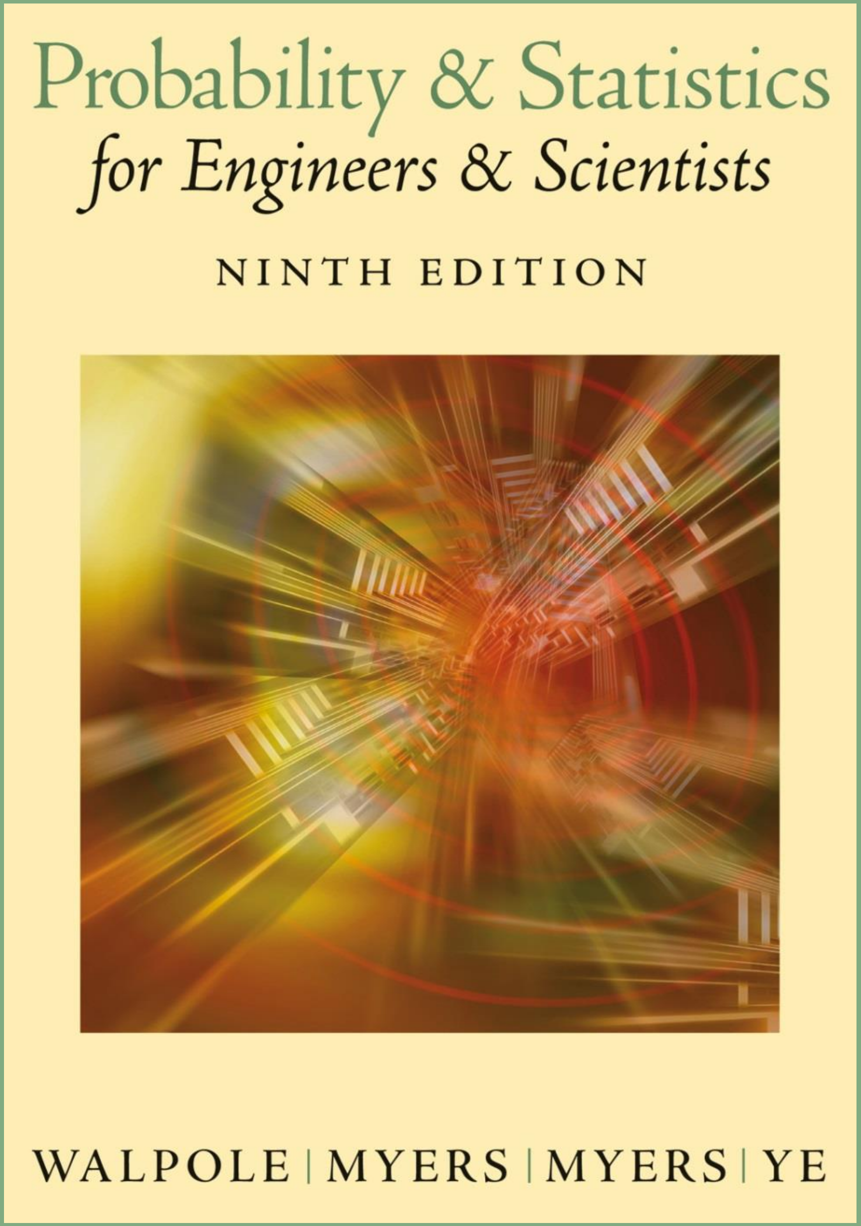


Table 10.6 2 × 3 Contingency Table

	Income Level			
Tax Reform	**Low**	**Medium**	**High**	**Total**
For	182	213	203	598
Against	154	138	110	402
Total	336	351	313	1000

Table 10.7 Observes and Expected Frequencies

	Income Level			
Tax Reform	**Low**	**Medium**	**High**	**Total**
For	182 (200.9)	213 (209.9)	203 (187.2)	598
Against	154 (135.1)	138 (141.1)	110 (125.8)	402
Total	336	351	313	1000

Section 10.13

Test for Homogeneity

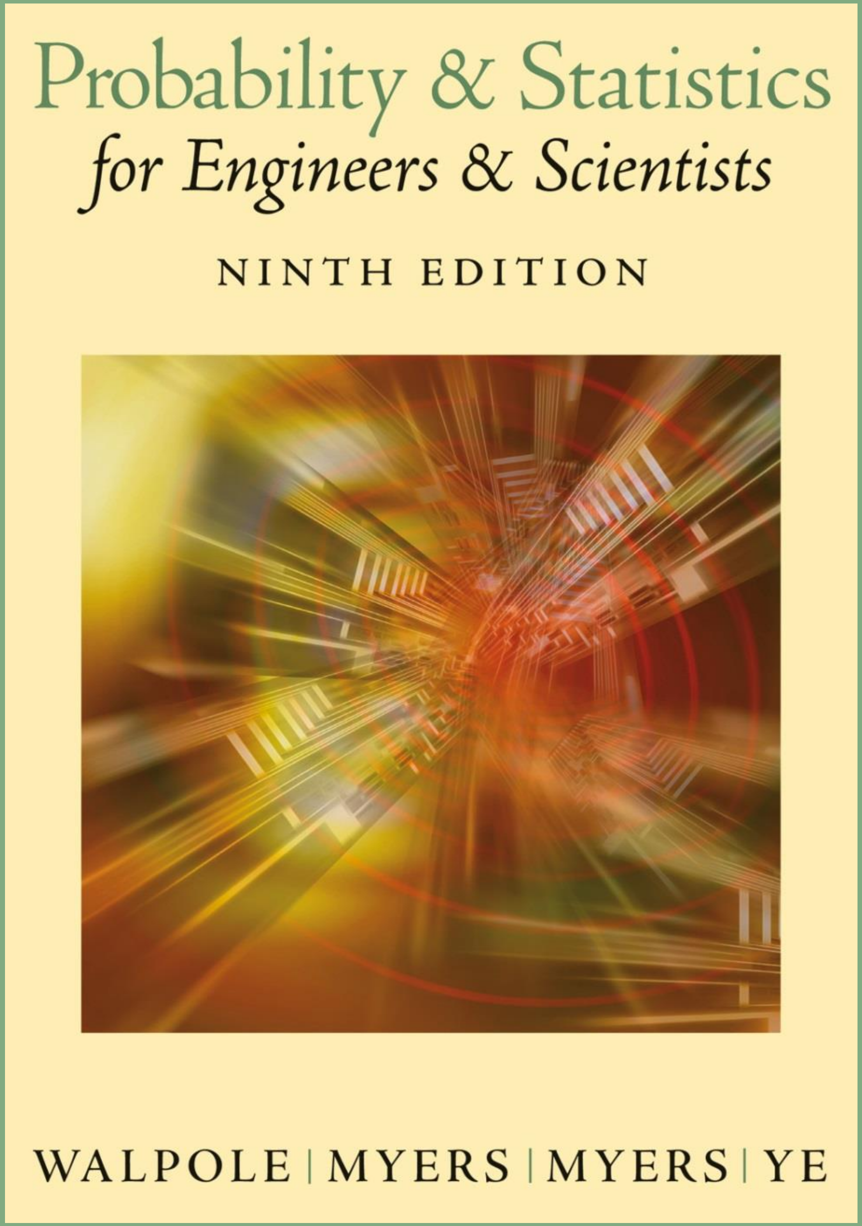


Table 10.8 Observed Frequencies

	Political Affiliation			
Abortion Law	**Democrat**	**Republican**	**Independent**	**Total**
For	82	70	62	214
Against	93	62	67	222
Undecided	25	18	21	64
Total	200	150	150	500

Table 10.9 Observed and Expected Frequencies

Abortion Law	Political Affiliation: Democrat	Republican	Independent	Total
For	82 (85.6)	70 (64.2)	62 (64.2)	214
Against	93 (88.8)	62 (66.6)	67 (66.6)	222
Undecided	25 (25.6)	18 (19.2)	21 (19.2)	64
Total	200	150	150	500

Table 10.10 *k* Independent Binomial Samples

Sample:	**1**	**2**	$\cdots$	$\boldsymbol{k}$
Successes	x_1	x_2	$\cdots$	x_k
Failures	$n_1 - x_1$	$n_2 - x_2$	$\cdots$	$n_k - x_k$

Table 10.11 Data for Example 10.15

Shift:	Day	Evening	Night
Defectives	45	55	70
Nondefectives	905	890	870

Table 10.12 Observed and Expected Frequencies

Shift:	**Day**	**Evening**	**Night**	**Total**
Defectives	45 (57.0)	55 (56.7)	70 (56.3)	170
Nondefectives	905 (893.0)	890 (888.3)	870 (883.7)	2665
Total	950	945	940	2835

Section 10.14

Two-Sample Case Study

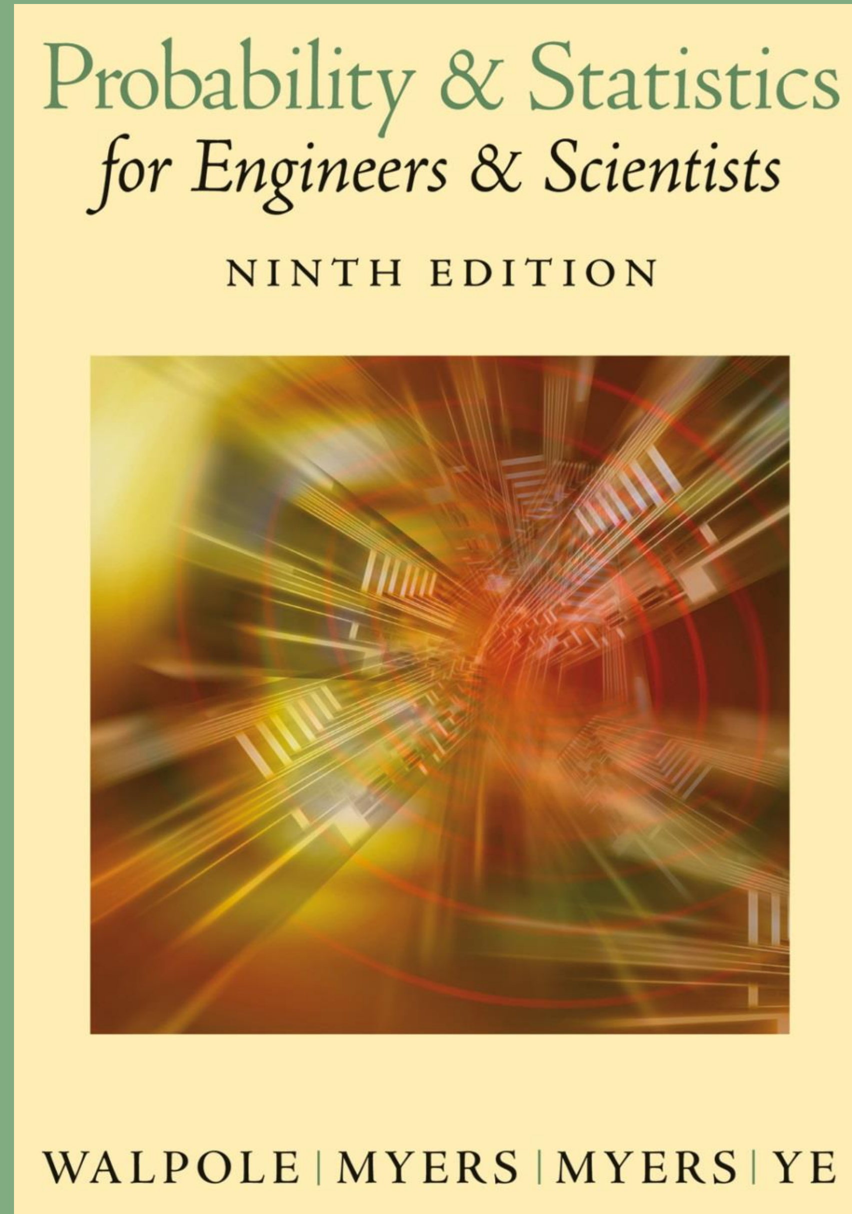


Table 10.13 Data for Two-Sample Case Study

Alloy *A*			Alloy *B*		
88	82	87	75	81	80
79	85	90	77	78	81
84	88	83	86	78	77
89	80	81	84	82	78
81	85		80	80	
83	87		78	76	
82	80		83	85	
79	78		76	79	

Figure 10.21 Normal quantile-quantile plot of data for alloy *A*

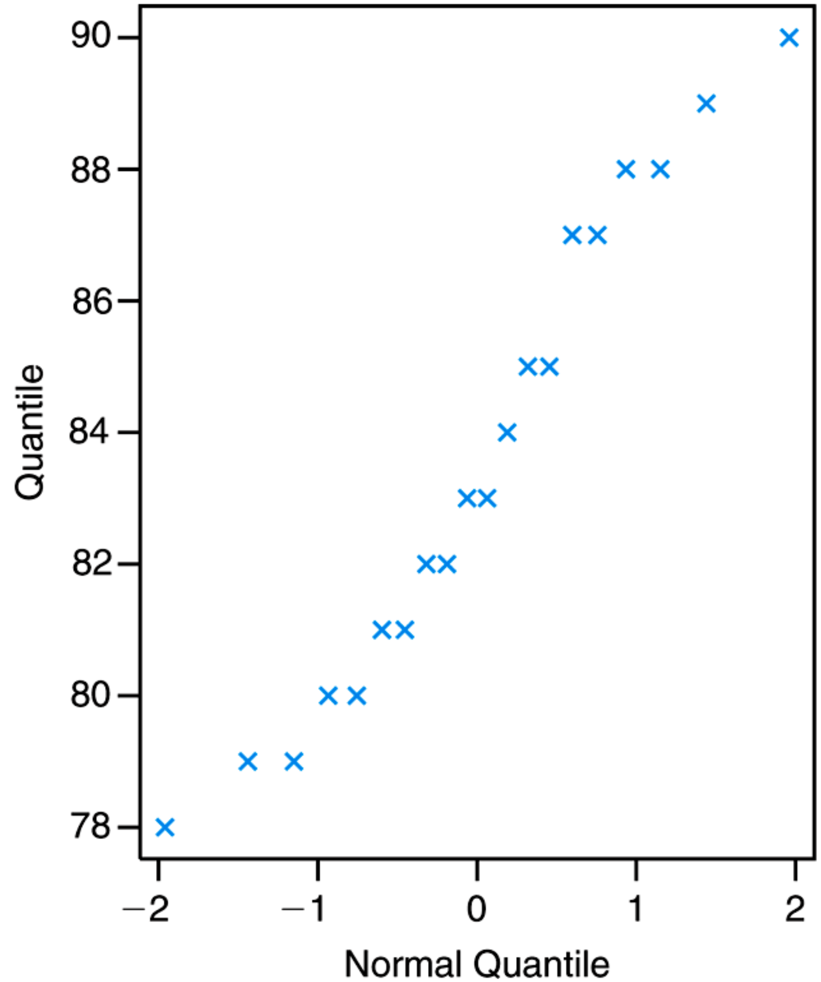


Figure 10.22 Normal quantile-quantile plot of data for alloy *B*

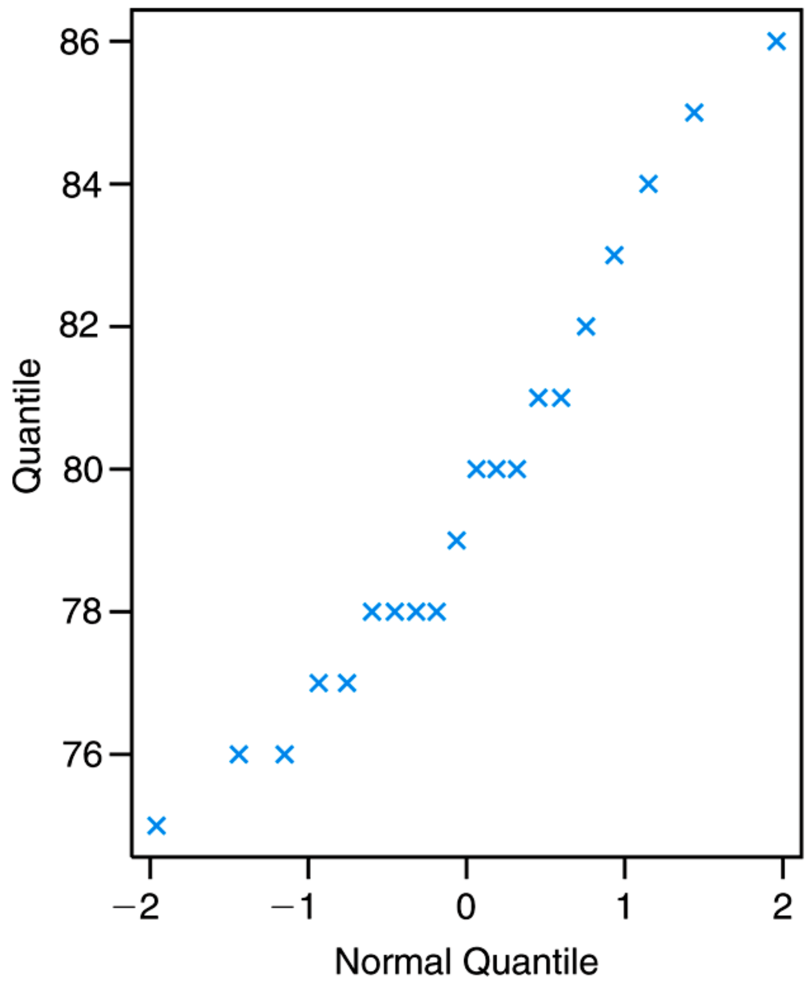


Figure 10.23 Box-and-whisker plots for both alloys

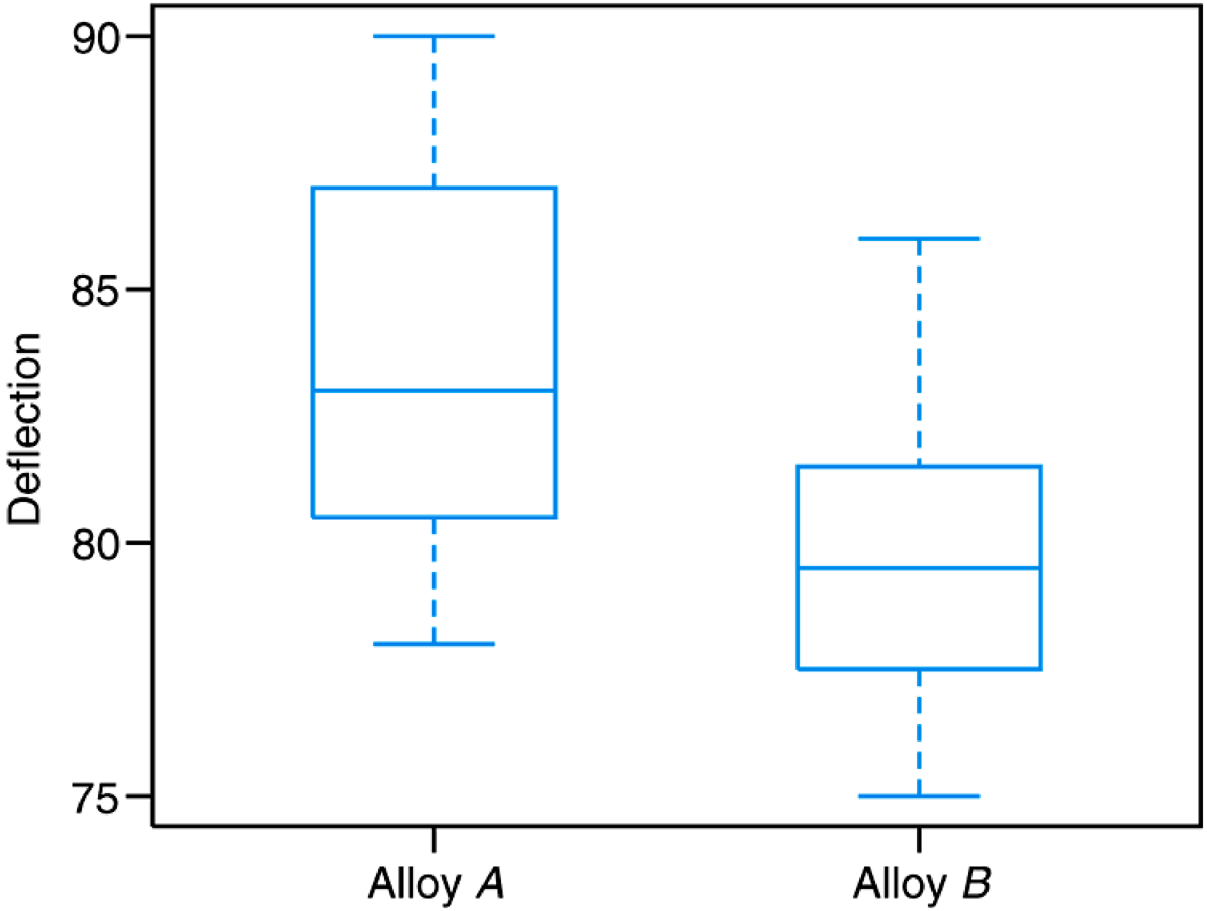


Figure 10.24 Annotated *SAS* printout for alloy data

The TTEST Procedure

Alloy	N	Mean	Std Dev	Std Err
Alloy A	20	83.55	3.6631	0.8191
Alloy B	20	79.7	3.0967	0.6924

Variances	DF	t Value	Pr > |t|
Equal	38	3.59	0.0009
Unequal	37	3.59	0.0010

Equality of Variances

Num DF	Den DF	F Value	Pr > F
19	19	1.40	0.4709

Section 10.15

Potential Misconceptions and Hazards; Relationship to Material in Other Chapters