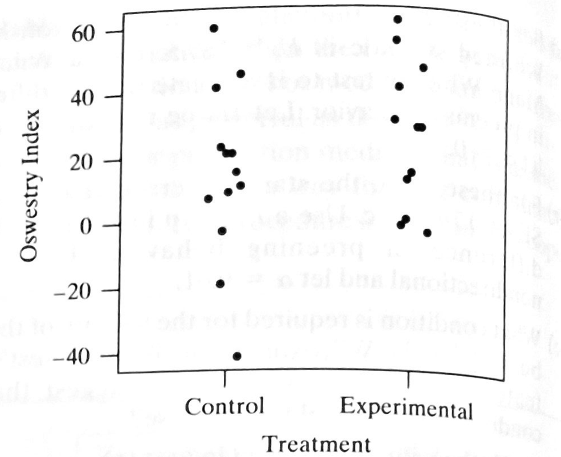
Oswestry Disability Index," which is scored as a percentage, ranging from 0 to 100, with high scores representing high levels of disability. Data are shown in the table.

Experimental	Control
58	18
−4	−6
10	−22
−6	8
52	20
26	12
10	38
38	18
26	42
−2	−44
12	56
44	6
28	4



The Wilcoxon-Mann-Whitney test statistic for these data is 104.5. If the alternative hypothesis is nondirectional, do the data provide evidence of a difference in the two distributions?

Supplementary Exercises 7.S.1–7.S.35

(*Note:* Exercises preceded by an asterisk refer to optional sections.)

Answers to hypothesis testing questions should include a statement of the conclusion in the context of the setting. (See Examples 7.2.4 and 7.2.5.)

7.S.1 For each of the following pairs of samples, compute the standard error of $(\bar{Y}_1 - \bar{Y}_2)$.

(a)

	Sample 1	Sample 2
n	12	13
$\bar{y}$	42	47
s	9.6	10.2

(b)

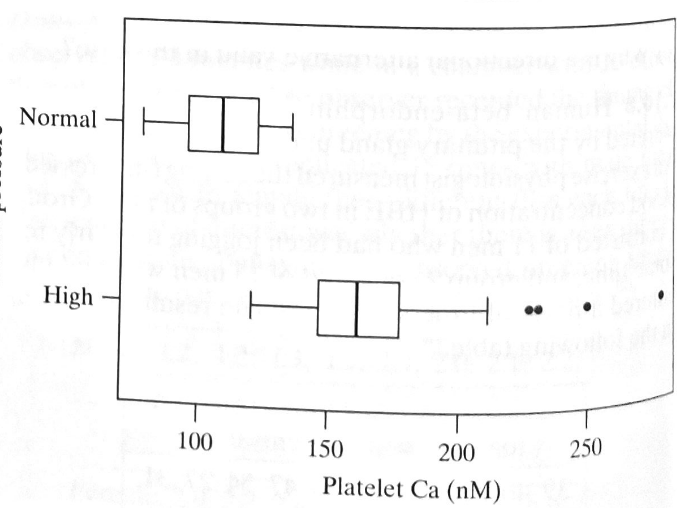
	Sample 1	Sample 2
n	22	19
$\bar{y}$	112	126
s	2.7	1.9

(c)

	Sample 1	Sample 2
n	5	7
$\bar{y}$	14	16
SE	1.2	1.4

7.S.2 To investigate the relationship between intracellular calcium and blood pressure, researchers measured the free calcium concentration in the blood platelets of 38 people with normal blood pressure and 45 people with high blood pressure. The results are given in the table and the distributions are shown in the boxplots.[71] Use the t test to compare the means. Let $\alpha = 0.01$ and let H_A be nondirectional. [*Note:* Formula (6.7.1) yields 67.5 df.]

		Platelet calcium (nM)	
Blood pressure	n	Mean	SD
Normal	38	107.9	16.1
High	45	168.2	31.7



7.S.3 Refer to Exercise 7.S.2. Construct a 95% confidence interval for the difference between the population means.

7.S.4 Refer to Exercise 7.S.2. The boxplot for the high blood pressure group is skewed to the right and includes outliers. Does this mean that the t test is not valid for these data? Why or why not?

7.S.5 In a study of methods of producing sheep's milk for use in cheese manufacture, ewes were randomly allocated to either a mechanical or a manual milking method. The investigator suspected that the mechanical method might irritate the udder and thus produce a higher concentration of somatic cells in the milk. The accompanying data show the average somatic cell count for each animal.[72]

	Somatic count (10^{-3} × cells/ml)	
	Mechanical milking	Manual milking
	2,966	186
	269	107
	59	65
	1,887	126
	3,452	123
	189	164
	93	408
	618	324
	130	548
	2,493	139
n	10	10
Mean	1,215.6	219.0
SD	1,342.9	156.2

(a) Do the data support the investigator's suspicion? Use a t test against a directional alternative at $\alpha = 0.05$. The standard error of $(\bar{Y}_1 - \bar{Y}_2)$ is SE = 427.54 and formula (6.7.1) yields 9.2 df.

(b) Do the data support the investigator's suspicion? Use a Wilcoxon-Mann-Whitney test against a directional alternative at $\alpha = 0.05$. (The value of the Wilcoxon-Mann-Whitney statistic is $U_s = 69$.) Compare with the result of part (a).

(c) What condition is required for the validity of the t test but not for the Wilcoxon-Mann-Whitney test? What features of the data cast doubt on this condition?

(d) Verify the value of U_s given in part (b).

7.S.6 A plant physiologist conducted an experiment to determine whether mechanical stress can retard the growth of soybean plants. Young plants were randomly allocated to two groups of 13 plants each. Plants in one group were mechanically agitated by shaking for 20 minutes twice daily, while plants in the other group were not agitated. After 16 days of growth, the total stem length (cm) of each plant was measured, with the results given in the accompanying table.[73]

Consider a t test to compare the treatments with the alternative hypothesis that stress tends to retard growth.

	Control	Stress
n	13	13
$\bar{y}$	30.59	27.78
s	2.13	1.73

(a) What is the value of the t test statistic for comparing the means?

(b) The P-value for the t test is 0.0006. If $\alpha = 0.01$, what is your conclusion regarding H_0?

7.S.7 Refer to Exercise 7.S.6. Construct a 95% confidence interval for the population mean reduction in stem length. Does the confidence interval indicate whether the effect of stress is "horticulturally important," if "horticulturally important" is defined as a reduction in population mean stem length of at least

(a) 1 cm

(b) 2 cm

(c) 5 cm

7.S.8 Refer to Exercise 7.S.6. The observations (cm), in increasing order, are shown. Compare the treatments using a Wilcoxon-Mann-Whitney test at $\alpha = 0.01$. Let the alternative hypothesis be that stress tends to retard growth.

Control	Stress
25.2	24.7
29.5	25.7
30.1	26.5
30.1	27.0
30.2	27.1
30.2	27.2
30.3	27.3
30.6	27.7
31.1	28.7
31.2	28.9
31.4	29.7
33.5	30.0
34.3	30.6

7.S.9 One measure of the impact of pollution along a river is the diversity of species in the river floodplain. In one study, two rivers, the Black River and the Vermilion River, were compared. Random 50 m × 20 m plots were sampled along each river and the number of species of trees in each plot was recorded. The following table contains the data.[74]

Vermilion River	Black River
9 9 16 13 12	13 10 6 9
13 13 13 8 11	10 7 6 18
9 9 10	6

The Black River was considered to have been polluted quite a bit more than the Vermilion River, and this was expected to lead to lower biodiversity along the Black River. Consider a Wilcoxon-Mann-Whitney test, with $\alpha = 0.10$, of the null hypothesis that the populations from which the two samples were drawn have the same biodiversity (distribution of tree species per plot) versus an appropriate directional alternative. The P-value for the test is 0.078. State your conclusion regarding H_0 in the context of this setting.

7.S.10 A developmental biologist removed the oocytes (developing egg cells) from the ovaries of 24 frogs (*Xenopus laevis*). For each frog the oocyte pH was determined. In addition, each frog was classified according to its response to a certain stimulus with the hormone progesterone. The pH values were as follows[75]:

Positive response:

7.06, 7.18, 7.30, 7.30, 7.31, 7.32, 7.33, 7.34, 7.36, 7.36, 7.40, 7.41, 7.43, 7.48, 7.49, 7.53, 7.55, 7.57

No response:

7.55, 7.70, 7.73, 7.75, 7.75, 7.77

Investigate the relationship of oocyte pH to progesterone response using a Wilcoxon-Mann-Whitney test at $\alpha = 0.05$. Use a nondirectional alternative.

7.S.11 Refer to Exercise 7.S.10. Summary statistics for the pH measurements are given in the following table. We wish to investigate the relationship of oocyte pH to progesterone response using a t test at $\alpha = 0.05$ and a nondirectional alternative.

	Positive response	No response
n	18	6
$\bar{y}$	7.373	7.708
s	0.129	0.081

(a) Write the null hypothesis and the alternative hypothesis in symbols.

(b) Here is computer output for the t test. Explain what the P-value means in the context of this study. [*Note:* 3.042e-06 means 3.042×10^{-6} which is 0.000003042.]

```
t = -7.46, df = 14.003, p-value = 3.042e-06
alternative hypothesis: true difference in
means is not equal to 0
```

(c) If $\alpha = 0.05$, what is your conclusion regarding H_0?

7.S.12 A proposed new diet for beef cattle is less expensive than the standard diet. The proponents of the new diet have conducted a comparative study in which one group of cattle was fed the new diet and another group was fed the standard. They found that the mean weight gains in the two groups were not statistically significantly different at the 5% significance level, and they stated that this finding supported the claim that the new cheaper diet was as good (for weight gain) as the standard diet. Criticize this statement.

***7.S.13** Refer to Exercise 7.S.12. Suppose you discover that the study used 25 animals on each of the two diets, and that the coefficient of variation of weight gain under the conditions of the study was about 20%. Using this additional information, write an expanded criticism of the proponents' claim, indicating how likely such a study would be to detect a 10% deficiency in weight gain on the cheaper diet (using a two-tailed test at the 5% significance level).

7.S.14 In a study of hearing loss, endolymphatic sac tumors (ELSTs) were discovered in 13 patients. These 13 patients had a total of 15 tumors (i.e., most patients had a single tumor, but two of the patients had 2 tumors each). Ten of the tumors were associated with the loss of functional hearing in an ear, but for 5 of the ears with tumors the patient had no hearing loss.[76] A natural question is whether hearing loss is more likely with large tumors than with small tumors. Thus, the sizes of the tumors were measured. Suppose that the sample means and standard deviations were given and that a comparison of average tumor size (hearing loss versus no hearing loss) was being considered.

(a) Explain why a t test to compare average tumor size is not appropriate here.

(b) If the raw data were given, could a Wilcoxon-Mann-Whitney test be used?

7.S.15 (Computer exercise) In an investigation of the possible influence of dietary chromium on diabetic symptoms, 14 rats were fed a low-chromium diet and 10 were fed a normal diet. One response variable was activity of the liver enzyme GITH, which was measured using a radioactively labeled molecule. The accompanying table shows the results, expressed as thousands of counts per minute per gram of liver.[77] Use a t test to compare the diets at $\alpha = 0.05$. Use a nondirectional alternative. [*Note:* Formula (6.7.1) yields 21.9 df.]

Low-chromium diet		Normal diet	
42.3	52.8	53.1	53.6
51.5	51.3	50.7	47.8
53.7	58.5	55.8	61.8
48.0	55.4	55.1	52.6
56.0	38.3	47.5	53.7
55.7	54.1		
54.8	52.1		

7.S.16 (Computer exercise) Refer to Exercise 7.S.15. Use a Wilcoxon-Mann-Whitney test to compare the diets at $\alpha = 0.05$. Use a nondirectional alternative.

7.S.17 (Computer exercise) Refer to Exercise 7.S.15.

(a) Construct a 95% confidence interval for the difference in population means.

(b) Suppose the investigators believe that the effect of the low-chromium diet is "unimportant" if it shifts mean GITH activity by less than 15%—that is, if the population mean difference is less than about 8000 cpm/gm. According to the confidence interval of part (a), do the data support the conclusion that the difference is "unimportant"?

(c) How would you answer the question in part (b) if the criterion were 4000 rather than 8000 cpm/gm?

7.S.18 (Computer exercise) In a study of the lizard *Sceloporus occidentalis*, researchers examined field-caught lizards for infection by the malarial parasite *Plasmodium*. To help assess the ecological impact of malarial infection, the researchers tested 15 infected and 15 noninfected lizards for stamina, as indicated by the distance each animal could run in 2 minutes. The distances (meters) are shown in the table.[78]

Infected animals		Uninfected animals	
16.4	36.7	22.2	18.4
29.4	28.7	34.8	27.5
37.1	30.2	42.1	45.5
23.0	21.8	32.9	34.0
24.1	37.1	26.4	45.5
24.5	20.3	30.6	24.5
16.4	28.3	32.9	28.7
29.1		37.5	

Do the data provide evidence that the infection is associated with decreased stamina? Investigate this question using

(a) a t test.

(b) a Wilcoxon-Mann-Whitney test.

Let H_A be directional and $\alpha = 0.05$.

7.S.19 In a study of the effect of amphetamine on water consumption, a pharmacologist injected four rats with amphetamine and four with saline as controls. She measured the amount of water each rat consumed in 24 hours. The following are the results, expressed as ml water per kg body weight[79]:

Amphetamine	Control
118.4	122.9
124.4	162.1
169.4	184.1
105.3	154.9

(a) Use a t test to compare the treatments at $\alpha = 0.10$. Let the alternative hypothesis be that amphetamine tends to suppress water consumption.

(b) Use a Wilcoxon-Mann-Whitney test to compare the treatments at $\alpha = 0.10$, with the directional alternative that amphetamine tends to suppress water consumption.

(c) Why is it important that some of the rats received saline injections as a control? That is, why didn't the researchers simply compare rats receiving amphetamine injections to rats receiving no injection?

7.S.20 Researchers studied subjects who had pneumonia and classified them as being in one of two groups: those who were given medical therapy that is consistent with American Thoracic Society (ATS) guidelines and those who were given medical therapy that is inconsistent with ATS guidelines. Subjects in the "consistent" group were generally able to return to work sooner than were subjects in the "inconsistent" group. A Wilcoxon-Mann-Whitney test was applied to the data; the P-value for the test was 0.04.[80] For each of the following, say whether the statement is true or false and say why.

(a) There is a 4% chance that the "consistent" and "inconsistent" population distributions really are the same.

(b) If the "consistent" and "inconsistent" population distributions really are the same, then a difference between the two samples at least as large as the difference that these researchers observed would only happen 4% of the time.

(c) If a new study were done that compared the "consistent" and "inconsistent" populations, there is a 4% probability that H_0 would be rejected again.

7.S.21 A student recorded the number of calories in each of 56 entrees—28 vegetarian and 28 nonvegetarian—served at a college dining hall.[81] The following table summarizes the data. Graphs of the data (not given here) show that both distributions are reasonably symmetric and bell shaped. A 95% confidence interval for $\mu_1 - \mu_2$ is $(-27, 85)$. For each of the following, say whether the statement is true or false and say why.

	n	Mean	SD
Vegetarian	28	351	119
Nonvegetarian	28	322	87

(a) 95% of the data are between −27 and 85 calories.

(b) We are 95% confident that $\mu_1 - \mu_2$ is between −27 and 85 calories.

(c) 95% of the time $\bar{Y}_1 - \bar{Y}_2$ will be between −27 and 85 calories.

(d) 95% of the vegetarian entrees have between 27 fewer calories and 85 more calories than the average nonvegetarian entree.

7.S.22 Refer to Exercise 7.S.21. True or false (and say why): 95% of the time, when conducting a study of this size, the difference in sample means $(\bar{Y}_1 - \bar{Y}_2)$ will be within approximately $\frac{(85 - (-27))}{2} = 56$ calories of the difference in population means $(\mu_1 - \mu_2)$.

7.S.23 (Computer exercise) Lianas are woody vines that grow in tropical forests. Researchers measured liana abundance (stems/ha) in several plots in the central Amazon region of Brazil. The plots were classified into two types: plots that were near the edge of the forest (less than 100 meters from the edge) or plots far from the edge of the forest. The raw data are given and are summarized in the table.[82]

	n	Mean	SD
Near	34	438	125
Far	34	368	114

Near			Far		
639	601	600	470	339	384
605	581	555	309	395	393
535	531	466	236	252	407
437	423	380	241	215	427
376	362	350	320	228	445
349	346	337	325	267	451
320	317	310	352	294	493
285	271	265	275	356	502
250	450	441	181	418	540
436	432	420	250	425	590
419	407		266	495	
702	676		338	648	

(a) Make normal probability plots of the data to confirm that the distributions are mildly skewed.

(b) Conduct a *t* test to compare the two types of plots at $\alpha = 0.05$. Use a nondirectional alternative.

(c) Apply a logarithm transformation to the data and repeat parts (a) and (b).

(d) Compare the *t* tests from parts (b) and (c). What do these results indicate about the effect on a *t* test of mild skewness when the sample sizes are fairly large?

7.S.24 Androstenedione (andro) is a steroid that is thought by some athletes to increase strength. Researchers investigated this claim by giving andro to one group of men and a placebo to a control group of men. One of the variables measured in the experiment was the increase in "lat pulldown" strength (in pounds) of each subject after 4 weeks. (A lat pulldown is a type of weightlifting exercise.) The raw data are given below and are summarized in the table.[83]

	n	Mean	SD
Andro	9	14.4	13.3
Control	10	20.0	12.5

Control				Andro			
30	10	10	30	0	10	0	10
40	20	30	20	10	40	20	10
10	0			30			

(a) Conduct a *t* test to compare the two groups at $\alpha = 0.10$. Use a nondirectional alternative. [*Note:* Formula (6.7.1) yields 16.5 df.]

(b) Prior to the study it was expected that andro would increase strength, which means that a directional alternative might have been used. Redo the analysis in part (a) using the appropriate directional alternative.

7.S.25 The following is a sample of computer output from a study.[84] Describe the problem and the conclusion, based on the computer output.

```
Y = number of drinks in the previous 7 days

  Two-sample T for treatment vs. control:

                       n      Mean       SD
    Treatment        244     13.62    12.39
    Control          238     16.86    13.49

   95% CI for mu1 – mu2:(–5.56, – 0.92)

   T-test mu1 = mu2 (vs  < ):
   T = –2.74 P = .0031 DF = 474.3
```

7.S.26 In a controversial study to determine the effectiveness of AZT, a group of HIV-positive pregnant women were randomly assigned to get either AZT or a placebo. Some of the babies born to these women were HIV-positive, while others were not.[85]

(a) What is the explanatory variable?

(b) What is the response variable?

(c) What are the experimental units?

7.S.27 Patients suffering from acute respiratory failure were randomly assigned to either be placed in a prone (face down) position or a supine (face up) position. In the prone group, 21 out of 152 patients died. In the supine group, 25 out of 152 patients died.[86]

(a) What is the explanatory variable?

(b) What is the response variable?

(c) What are the experimental units?

7.S.28 A study of postmenopausal women on hormone replacement therapy (HRT) reported that they had a reduced heart attack rate, but had even greater reductions in death from homicide and accidents—two causes of death that cannot be linked to HRT. It seems that the women on HRT differ from others in many other aspects of their lives—for instance, they exercise more; they also tend to be wealthier and to be better educated.[87] Use the language of statistics to discuss what these data say about the relationships among HRT, heart attack risk, and variables such as exercise, wealth, and education. Use a schematic diagram similar to Figure 7.4.1 or Figure 7.4.2 to support your explanation.

7.S.29 Alice did a two-sample t test of the hypotheses $H_0: \mu_1 = \mu_2$ versus $H_A: \mu_1 \neq \mu_2$, using samples sizes of $n_1 = n_2 = 15$. The P-value for the test was 0.13, and α was 0.10. It happened that $\bar{y}_1$ was less than $\bar{y}_2$. Unbeknownst to Alice, Linda was interested in the same data. However, Linda had reason to believe, based on an earlier study of which Alice was not aware, that either $\mu_1 = \mu_2$ or else $\mu_1 < \mu_2$. Thus, Linda did a test of the hypotheses $H_0: \mu_1 = \mu_2$ versus $H_A: \mu_1 < \mu_2$. So, for Linda's test

(i) the P-value would still be 0.13 and H_0 would not be rejected if $\alpha = 0.10$.

(ii) the P-value would still be 0.13 and H_0 would be rejected if $\alpha = 0.10$.

(iii) the P-value would be less than 0.13 and H_0 would not be rejected if $\alpha = 0.10$.

(iv) the P-value would be less than 0.13 and H_0 would be rejected if $\alpha = 0.10$.

(v) the P-value would be greater than 0.13 and H_0 would not be rejected if $\alpha = 0.10$.

(vi) the P-value would be greater than 0.13 and H_0 would be rejected if $\alpha = 0.10$.

Choose one of these and briefly explain why your choice is true.

7.S.30 Jayden did a two-sample t test of the hypotheses $H_0: \mu_1 = \mu_2$ versus $H_A: \mu_1 \neq \mu_2$, using samples sizes of $n_1 = n_2 = 12$. The P-value for the test was 0.12, and α was 0.10. Suppose that everything except the sample sizes were the same (i.e., $\bar{y}_1$ was the same, $\bar{y}_2$ was the same, etc.), but the sample sizes had been $n_1 = n_2 = 7$. Then

(i) the P-value would still be 0.12 and α would be less than 0.10.

(ii) the P-value would still be 0.12 and α would be greater than 0.10.

(iii) the P-value would be less than 0.12 and α would still be 0.10.

(iv) the P-value would be less than 0.12 and α would be less than 0.10.

(v) the P-value would be less than 0.12 and α would be greater than 0.10

(vi) the P-value would be greater than 0.12 and α would still be 0.10.

(vii) the P-value would be greater than 0.12 and α would be less than 0.10

(viii) the P-value would be greater than 0.12 and α would be greater than 0.10

Choose one of these and explain why your choice is true.

7.S.31 Omar collects data from two populations. Sahar uses those data to construct a (two-sided) 90% confidence interval for $\mu_1 - \mu_2$ and gets $(-12.5, -3.7)$. At the same time, Miguel uses the data to test $H_0: \mu_1 = \mu_2$ versus $H_A: \mu_1 < \mu_2$ with $\alpha = 0.05$. Can we tell from the information if Miguel rejects H_0? Explain your reasoning.

7.S.32 A group of 477 healthy children, aged 5–11 years, were randomly divided into two groups. Children in the "sugar-free" group were given one can (250 mL) of an artificially sweetened beverage each day for 18 months. Children in the "sugar-containing" group were given one can each day for 18 months of a beverage sweetened with sugar. The researchers developed the two drinks specifically for this experiment and made them taste and look as similar as possible.

The researchers recorded body mass index (BMI, which is weight [kg] divided by the square of height [m].) In order to control for age and sex effects, the response variable was change in BMI z score; that is, the number of SDs that a BMI differed from the national average for the child's age and sex. The table below shows the means and standard deviations of change in BMI z score for the two groups.[88] Consider a t test with a directional alternative to investigate the research hypothesis that change in BMI is lower for the sugar-free condition.

	Sugar-free	Sugar-containing
Mean	0.02	0.15
SD	0.41	0.42
n	225	252

(a) Write the null hypothesis and the alternative hypothesis in symbols.

(b) The P-value for the t test is 0.0003. Using $\alpha = 0.01$, state your conclusion regarding the hypotheses in the context of this problem.

7.S.33 Adults who broke their wrists were randomly divided into two groups. The "intervention" group was given exercises to do under the direction of a physiotherapist. The "control" group did not receive such intervention. The primary variable of interest was improvement in wrist extension, measured in degrees with a device called a goniometer, after 6 weeks.[62] Consider a t test, with a directional alternative, to investigate the research hypothesis that the intervention improves wrist extension.

	Intervention	Control
Mean	25.6	20.5
SD	17.5	9.0
n	27	20

(a) Write the null hypothesis and the alternative hypothesis in symbols.

(b) Here is computer output for the t test. Explain what the P-value means in the context of this study.

```
t = 1.29, df = 40.8, p-value = 0.1022
alternative hypothesis: true difference in
means is greater than 0
```

(c) If $\alpha = 0.10$, what is your conclusion regarding H_0?

(d) If the research hypothesis were nondirectional, how would your answer to part (c) change, if at all?

7.S.34 Parents of newborn babies were divided into three groups as part of an experiment to see if the age at which a baby first walks can be affected by special exercises. For simplicity we will call the groups Grp1, Grp2, and Grp3.[89] This computer output compares Grp1 to Grp2:

```
t = 1.28, df = 9.35, p-value = 0.2301
```

This computer output compares Grp1 to Grp3:

```
t = 3.04, df = 8.66, p-value = 0.01453
```

It seems that we are being told that $\mu_1 = \mu_2$ and that $\mu_1 \neq \mu_3$. This might lead us to think that $\mu_2 \neq \mu_3$. However, when we conduct a test to compare Grp2 to Grp3 we get a large P-value:

```
t = 1.1, df = 7.65, p-value = 0.3042
```

All three t tests used a nondirectional alternative. Discuss how it is possible that the three tests, taken together, seem to say that $\mu_1 = \mu_2$ and $\mu_2 = \mu_3$ but $\mu_1 \neq \mu_3$.

7.S.35 Patients with pleural infection were randomly assigned to be given either an active medical treatment ($n = 52$) or a placebo ($n = 55$).[90] One outcome that was measured on each patient was duration of hospital stay. Consider conducting a t test of the null hypothesis that average hospital stay is the same for treatment and for control. Assuming that there are 57 degrees of freedom for the test, what is the distribution of the test statistic if the null hypothesis is true?