Solutions to Practice Final; Statistics 371; Spring 2012; Professor Wardrop

1. First we note that in the collapsed table \( \hat{p}_1 = 0.34 \) is smaller than \( \hat{p}_2 = 0.356 \). For Simpson’s Paradox to occur, this inequality must be reversed in both component tables.

   In the first component table, we need
   \[
   \frac{c}{100} < \frac{60}{200} \text{ or } c < 30 \text{ or } c \leq 29.
   \]

   In the second component table, we need
   \[
   \frac{d}{150} < \frac{42}{100} \text{ or } d < 63 \text{ or } d \leq 62.
   \]

   But we also need the component tables to be consistent with the collapsed table; i.e., we need \( c + d = 89 \).

   So, we have three conditions: \( c \leq 9 \), \( d \leq 30 \) and \( c + d = 42 \). Can all three conditions be satisfied? No, because the largest \( c + d \) can be and still keep both reversals is 39. On the exam it would suffice to say: It is impossible for \( c \leq 9 \), \( d \leq 30 \) and \( c + d = 42 \); I will assume you know why. (You won’t have time for long essay answers on the exam.)

2. We note that in the collapsed table \( \hat{p}_1 = 0.362 \) is smaller than \( \hat{p}_2 = 0.42 \). For Simpson’s Paradox to occur, this inequality must be reversed in both component tables.

   In the first component table, we need
   \[
   \frac{c}{35} < \frac{21}{75} \text{ or } c < 9.8 \text{ or } c \leq 9.
   \]

   In the second component table, we need
   \[
   \frac{d}{65} < \frac{26}{55} \text{ or } d < 30.7 \text{ or } d \leq 30.
   \]

   But we also need the component tables to be consistent with the collapsed table; i.e. we need \( c + d = 42 \).

3. This problem presents a twist on what you have seen in that one of the unknown counts is in the collapsed table. (When I first put this on an exam—last semester—many students immediately concluded it was impossible.)

   In A, \( \hat{p}_2 = 0.50 > 0.40 = \hat{p}_1 \). Thus, for SP, we need this same pattern in B and its reversal in the collapsed table. You can solve this problem by focusing on either of these (B or collapsed).

   In B, we need \( \frac{c_B}{150} > \frac{80}{200} \) which reduces to \( c_B > 60 \) or \( c_B \geq 61 \). Thus, for consistency, \( c \geq 75 + 61 = 136 \).

   But \( c \geq 136 \) implies that \( \hat{p}_2 > \hat{p}_1 \); hence, SP is impossible.

4. This problem presents a twist on what you have seen in that one of the unknown counts is in the collapsed table. (When I first put this on an exam—last semester—many students immediately concluded it was impossible.)

   In A, \( \hat{p}_1 = 0.40 > 0.30 = \hat{p}_2 \). Thus, for SP, we need this same pattern in B and its reversal in the collapsed table. You can solve this problem by focusing on either of these (B or collapsed).

   In B, we need \( \frac{c_B}{200} > \frac{90}{300} \) which reduces to \( c_B > 60 \) or \( c_B \geq 61 \). Thus, for consistency, \( c \geq 80 + 61 = 141 \).
But \( c \geq 141 \) implies that \( \hat{p}_2 > \hat{p}_1 \); hence, SP is impossible.

5. I will define population 1 to be females and population 2 to be males. Assume independent random samples from these populations. Based on the data given, the 95% CI for \( p_1 - p_2 \) is \((0.407 - 0.364)\pm \sqrt{0.407(0.593)\frac{450}{450} + 0.364(0.636)\frac{538}{538}} = 0.043 \pm 1.96(0.0311) = 0.043 \pm 0.061 = [-0.018, 0.104].

6. I will define population 1 to be males and population 2 to be females. Assume independent random samples from these populations. Based on the data given, the 95% CI for \( p_1 - p_2 \) is \((0.257 - 0.173)\pm \sqrt{0.257(0.743)\frac{537}{537} + 0.173(0.827)\frac{446}{446}} = 0.084 \pm 1.96(0.0260) = 0.084 \pm 0.051 = [0.033, 0.135].

7. (a) It cannot be a frequency histogram because the height is not an integer.
(b) The relative frequency of the interval is 0.20; thus, the frequency is \( 0.20(500) = 100 \).
(c) The relative frequency of the interval is its area: \( 2(0.20) = 0.40 \). Thus, its frequency is \( 0.40(500) = 200 \).

8. (a) The frequency of the interval is 5.
(b) It cannot be a relative frequency histogram because a relative frequency cannot exceed one.
(c) The relative frequency of the interval is its area: \( 5(0.10) = 0.50 \). Thus, its frequency is \( 0.50(800) = 400 \).

9. (a) Begin with the following table.

<table>
<thead>
<tr>
<th>Class Int.</th>
<th>( w )</th>
<th>Fr.</th>
<th>R. Fr.</th>
<th>Dens.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00–2.00</td>
<td>2</td>
<td>20</td>
<td>0.40</td>
<td>0.20</td>
</tr>
<tr>
<td>2.00–6.00</td>
<td>4</td>
<td>18</td>
<td>0.36</td>
<td>0.09</td>
</tr>
<tr>
<td>6.00–12.00</td>
<td>6</td>
<td>12</td>
<td>0.24</td>
<td>0.04</td>
</tr>
</tbody>
</table>

(b) Because \( n = 50 \) is even, the median is the average of 2.72 and 2.94; i.e. it is 2.83.
(c) The first quartile is the 25th percentile, so we calculate \( 0.25(50) = 12.5 \). We round this up to 13 and the first quartile is 1.32.
(d) We calculate \( 0.42(50) = 21 \). Thus, the 42nd percentile is the mean of the numbers in positions 21 and 22: \( \frac{2.26 + 2.40}{2} = 2.33 \).
(e) The total of the 50 numbers is \( 50(3.76) = 188.00 \). The total of the new data set of 49 numbers is \( 188.00 - 11.86 = 176.14 \). Thus, its mean is \( 176.14/49 = 3.5947 \).
(f) After the two deletions, \( n = 48 \). Thus, \( 0.25(48) = 12 \). The first quartile is the mean of the numbers in positions 12 and 13 (in the new listing): \( \frac{1.32 + 1.38}{2} = 1.35 \).
10. (a) Begin with the following table.

<table>
<thead>
<tr>
<th>Class Int.</th>
<th>w</th>
<th>Fr.</th>
<th>R. Fr.</th>
<th>Dens.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00–0.50</td>
<td>0.5</td>
<td>16</td>
<td>0.40</td>
<td>0.80</td>
</tr>
<tr>
<td>0.50–1.50</td>
<td>1.0</td>
<td>18</td>
<td>0.45</td>
<td>0.45</td>
</tr>
<tr>
<td>1.50–3.00</td>
<td>1.5</td>
<td>6</td>
<td>0.15</td>
<td>0.10</td>
</tr>
</tbody>
</table>

(b) Because \( n = 40 \) is even, the median is the average of 0.69 and 0.73; i.e. it is 0.71.

(c) We calculate \( 0.85(40) = 34 \). Thus, the 85th percentile is the mean of the numbers in positions 34 and 35:

\[
\frac{1.49 + 1.51}{2} = 1.50
\]

(d) The total of the 40 numbers is \( 40(0.800) = 32.00 \). The total of the new data set of 38 numbers is

\[
32.00 - 2.72 - 0.04 = 29.24.
\]

Thus, its mean is \( 29.24/38 = 0.7695 \). Also, \( 0.75(38) = 28.5 \), which we round up to 29. The 75th percentile is 1.18.

11. (a) The CI is

\[
3.76 \pm 1.96(3.27/\sqrt{40}) = 3.76 \pm 1.96(0.4624) = 3.76 \pm 0.91 = [2.85, 4.67].
\]

(b) The observed value of the test statistic is

\[
t = \frac{3.76 - 2.00}{0.4624} = 1.76/
\]

(c) Calculate

\[
\frac{51}{2} - \frac{1.96 \sqrt{50}}{2} = 25.5 - 6.93 = 18.57,
\]

which we round down to 18. The CI is \([1.56, 3.94]\).

12. (a) The CI is

\[
0.800 \pm 1.96(0.635/\sqrt{40}) = 0.800 \pm 1.96(0.1004) = 0.800 \pm 0.197 = [0.603, 0.997].
\]

(b) The observed value of the test statistic is

\[
t = \frac{0.800 - 1.00}{0.1004} = -0.20/0.1004 = -1.992.
\]

(c) Calculate

\[
\frac{41}{2} - \frac{1.96 \sqrt{40}}{2} = 20.5 - 6.20 = 14.30,
\]

which we round down to 14. The CI is \([0.35, 0.98]\).

13. (a) We don’t need to sort all 15 observations; we need to find the fourth smallest and fourth largest. The CI is \([46.7, 74.1]\).

(b) The 95% CI is

\[
60.51 \pm 2.145(16.47/\sqrt{15}) = 60.51 \pm 9.12 = [51.39, 69.63].
\]

14. (a) We don’t need to sort all 17 observations; we need to find the fourth smallest and fourth largest. The CI is \([113, 145]\).
(b) The 95% CI is
\[ 128.12 \pm 2.120(18.93/\sqrt{17}) = \]
\[ 128.12 \pm 9.73 = [118.39, 137.85]. \]

15. With \( df = n - 1 = 9 \), \( t = 2.262 \). Thus, we have
\[ 2.40 = 2.262s/\sqrt{10}. \]
Solving for \( s \), we get \( s = 3.355 \). With the correct \( df = n - 1 = 19 \), \( t = 2.093 \). Thus, the half width should be
\[ 2.093(3.355)/\sqrt{20} = 1.57. \]
Thus, Tom’s CI is \( \bar{x} \pm 1.57 \).

16. With \( df = n - 1 = 11 \), \( t = 2.201 \). Thus, we have
\[ 2.58 = 2.201s/\sqrt{12}. \]
Solving for \( s \), we get \( s = 4.0606 \). With the correct \( df = n - 1 = 17 \), \( t = 2.110 \). Thus, the half width should be
\[ 2.110(4.0606)/\sqrt{18} = 2.02. \]
Thus, Tom’s CI is \( \bar{x} \pm 2.02 \).

17. (a) To solve a problem like this, you must save consideration of the correct CI’s to the end. Too small means that the upper bound, \( U \) is smaller than \( \mu \). This must be the smallest of the \( U \)’s, which is 25. Thus 25 < \( \mu \). Too large means that the lower bound, \( L \), is larger than \( \mu \). Thus 29 > \( \mu \). The remaining two CI’s must be the correct ones; this tells us that 26 \( \leq \mu \leq 30 \) and 23 \( \leq \mu \leq 47 \). Using the logic of ‘and,’ we see that
\[ 26 \leq \mu < 29 \]
is the answer. BTW, on the exam, I also would give full credit to 26 < \( \mu \leq 29 \), 26 \( \leq \mu \leq 29 \) or 26 < \( \mu \leq 29 \); i.e. in these ‘logic’ questions, I don’t care about the endpoints of the intervals in your answers. You must, however, realize that—even for a count population—a mean can be a real number; i.e. it is not restricted to integers. Thus, the answer \( \mu = 26, 27 \) or 28, would lose substantial credit.

(b) It helps to note that the first two CI’s do not overlap; thus, they cannot be simultaneously correct. So, three correct means either 1, 3 and 4 are correct or 2, 3 and 4 are correct. CI’s 1 and 4 do not overlap, so the only possibility is for 2, 3 and 4 to be correct. Using the logic of ‘and,’ we get
\[ 29 \leq \mu \leq 30. \]

18. (a) Two upper bounds are too small; these must be 30 and 40. Thus, 40 < \( \mu \). One lower bound is too large; it must be 44. Thus \( \mu < 44 \). The remaining CI’s are correct; thus, 22 \( \leq \mu \leq 48 \) and 35 \( \leq \mu \leq 73 \). Using the logic of ‘and,’ 40 < \( \mu < 44 \).

(b) Numbering the CI’s, left-to-right, 1–5, we see that the pairs 14, 15 and 25 do not overlap. Thus, the possibilities for the three correct CI’s are: 123 or 234 or 345. Apply the logic of ‘and’ to each triple and we get:
\[ 22 \leq \mu \leq 30 \text{ or } 35 \leq \mu \leq 40 \text{ or } 44 \leq \mu \leq 48. \]
19. (a) One $U$ is smaller than 800, so one CI is too small. Two $L$’s are larger than 800, so two CI’s are too large. The remaining 17 CI’s must be correct.

(b) This means that 793, 796, 833 and 855 are too large, but 786 is not too large. Thus,

$$786 \leq \mu < 793.$$  

(c) This means that 771 and 803 are too small. Thus,

$$\mu > 803.$$  

(d) This means that 774 is not too large. Thus,

$$\mu \geq 774.$$  

20. (a) One $U$ is smaller than 495, so one CI is too small. Two $L$’s are larger than 495, so two CI’s are too large. The remaining 22 CI’s must be correct.

(b) This means that 493, 498, 553 and 568 are too small, but 574 is not too small. Thus,

$$568 < \mu \leq 574.$$  

(c) This means that 447 is not too large. Thus,

$$447 \leq \mu.$$  

(d) This means that 553 is too small. Thus,

$$553 < \mu.$$  

21. Let $t_B$ denote Bob’s observed value of the test statistic. Define $t_C$ and $t_D$ in a similar manner. We see that

$$t_B = \frac{120 - 90}{s/\sqrt{n}} = 30/d.$$  

We don’t know the value of $d$ except that $d > 0$. Thus, $t_B > 0$. Similarly,

$$t_C = 45/d > 0 \text{ and } t_D = -45/d = -t_C < 0.$$  

(a) Both A and C are possible. (B is impossible because the area to the left of $t_B$ is larger than 0.5 because the t-curves are symmetric around 0.)

(b) Even though Carl’s $t$ is larger than Bob’s, Carl has a larger P-value. This can occur only if C is Carl’s alternative and A is Bob’s alternative.

(c) C. See the answer to (b).

(d) David’s P-value is one-half of Carl’s; hence, it is 0.1221.

22. Let $t_C$ denote Carol’s observed value of the test statistic. Define $t_D$ and $t_E$ in a similar manner. We see that

$$t_C = \frac{150 - 180}{s/\sqrt{n}} = -30/d.$$  

We don’t know the value of $d$ except that $d > 0$. Thus, $t_C < 0$. Similarly,

$$t_D = -20/d > 0 \text{ and } t_E = 30/d = -t_C > 0.$$  

(a) Both B and C are possible. (A is impossible because the area to the right of $t_C$ is larger than 0.5 because the t-curves are symmetric around 0.)

(b) Even though Diane’s $t$ is larger than Carol’s, Diane has a smaller P-value. This can occur only if C is Carol’s alternative and B is Diane’s alternative.

(c) B. See the answer to (b).

(d) Edith’s P-value is one-half of Carol’s; hence, it is 0.1180.

23. We have

$$s_p^2 = \frac{4(144) + 9(324)}{4 + 9} = \frac{3492}{13} = 268.6154.$$  

Thus, $s_p = \sqrt{268.6154} = 16.39.$
24. We have 
\[ s_p^2 = \frac{15(225) + 8(100)}{15 + 8} = \frac{4175}{23} = 181.5217. \]
Thus, \( s_p = \sqrt{181.5217} = 13.47. \)

25. It is 
\[ t = \frac{22 - 14}{\sqrt{15\sqrt{(1/20) + (1/5)}}} = \frac{8}{1.9365} = 4.131. \]

26. It is 
\[ t = \frac{35 - 43}{\sqrt{25\sqrt{(1/15) + (1/12)}}} = \frac{-8}{1.9365} = -4.131. \]

27. (a) For independent samples,
\[ s_p^2 = [(13.28)^2 + (21.47)^2] / 2 = 318.66. \]
Thus, \( s_p = 17.85. \) The 95% CI is \( (89.30 - 93.30) \pm \)
\[ 2.101(17.85)\sqrt{1/10 + 1/10} = \]
\[ -4.00 \pm 16.77 = [-20.77, 12.77]. \]
(b) For paired data the 95% CI is 
\[ -4.00 \pm 2.262(31.70/\sqrt{10}) = \]
\[-4.00 \pm 22.68 = [-26.68, 18.68]. \]
(c) The CI in (b) is much wider; pairing was not effective.
(d) The interval \([d(3), d(8)]\) is \([-23, 19]\). It is the CI for the median of the population of differences. (Not necessarily the same thing as the difference of the population medians.)

28. (a) For independent samples,
\[ s_p^2 = [(27.20)^2 + (27.05)^2] / 2 = 735.77. \]
Thus, \( s_p = 27.13. \) The 98% CI is \( (193.25 - 189.25) \pm \)
\[ 2.624(27.13)\sqrt{1/8 + 1/8} = \]
\[ 4.00 \pm 35.59 = [-31.59, 39.59]. \]
(b) For paired data the 98% CI is 
\[ 4.00 \pm 2.998(4.31/\sqrt{8}) = \]
\[ 4.00 \pm 4.57 = [-0.57, 8.57]. \]
(c) The CI in (a) is much wider; pairing was very effective.
(d) The interval \([d(2), d(7)]\) is \([0, 8]\). It is the CI for the median of the population of differences.

29. (a) Here is the table, with reasoning after.

<table>
<thead>
<tr>
<th></th>
<th>Drive</th>
<th>Don’t Drive</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>660</td>
<td>540</td>
<td>1200</td>
</tr>
<tr>
<td>Male</td>
<td>340</td>
<td>460</td>
<td>800</td>
</tr>
<tr>
<td>Total</td>
<td>1000</td>
<td>1000</td>
<td>2000</td>
</tr>
</tbody>
</table>

The total of 2000 is given. Sixty percent of 2000 gives us 1200 females. Fifty-five percent of 1200 gives us 660 females who drive. Let \( w \) denote the number who drive. We are given that \( 660 = 0.66w \), which gives \( w = 1000 \) drivers. The other numbers are obtained by subtracting and adding.

(b) Just divide each entry by 2000:

<table>
<thead>
<tr>
<th></th>
<th>Drive</th>
<th>Don’t Drive</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>0.33</td>
<td>0.27</td>
<td>0.60</td>
</tr>
<tr>
<td>Male</td>
<td>0.17</td>
<td>0.23</td>
<td>0.40</td>
</tr>
<tr>
<td>Total</td>
<td>0.50</td>
<td>0.50</td>
<td>1.00</td>
</tr>
</tbody>
</table>
(c) We want

\[ P(D|M) = 0.17/0.40 = 0.425. \]

(d) We want

\[ P(F|D^c) = 0.27/0.50 = 0.54. \]

30. (a) Here is the table, with reasoning after.

<table>
<thead>
<tr>
<th>Disease</th>
<th>Positive</th>
<th>Negative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present</td>
<td>0.045</td>
<td>0.005</td>
<td>0.050</td>
</tr>
<tr>
<td>Absent</td>
<td>0.038</td>
<td>0.912</td>
<td>0.950</td>
</tr>
<tr>
<td>Total</td>
<td>0.083</td>
<td>0.917</td>
<td>1.000</td>
</tr>
</tbody>
</table>

The row total of 0.050 is given. We obtain 0.950 by subtracting. Ninety percent of 0.050 gives 0.045. Four percent of 0.950 is 0.038. The remaining numbers are found by adding and subtracting.

(b) \( P(B) = 0.083 \), from the table.

(c) \( P(A|B) = P(AB)/P(B) = 0.045/0.083 = 0.542. \)

31. (a) \( 2415/3000 = 0.805. \)

(b) \( 450/3000 = 0.150. \)

(c) \( 225/3000 = 0.075. \)

(d) \( 90/3000 = 0.030. \)

(e) \( 360/585 = 0.615. \)

(f) \( 2325/2550 = 0.912. \)

(g) \( (90 + 225)/3000 = 0.105. \)

(h) \( 360/(360 + 2325) = 0.134. \)

(i) \( (360 + 2550)/3000 = 0.970. \)

32. (a) \( 900/5000 = 0.180. \)

(b) \( 4250/5000 = 0.850. \)

(c) \( 4025/5000 = 0.805. \)

(d) \( 675/5000 = 0.135. \)

33. (a) Below is the completed table.

<table>
<thead>
<tr>
<th></th>
<th>( B )</th>
<th>( B^c )</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A )</td>
<td>40</td>
<td>40</td>
<td>80</td>
</tr>
<tr>
<td>( A^c )</td>
<td>30</td>
<td>90</td>
<td>120</td>
</tr>
<tr>
<td>Total</td>
<td>70</td>
<td>130</td>
<td>200</td>
</tr>
</tbody>
</table>

(b) \( 0.9059. \)

(c) \( 2(0.1410) = 0.2820. \)

34. (a) Below is the completed table.

<table>
<thead>
<tr>
<th></th>
<th>( B )</th>
<th>( B^c )</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A )</td>
<td>45</td>
<td>55</td>
<td>100</td>
</tr>
<tr>
<td>( A^c )</td>
<td>40</td>
<td>160</td>
<td>200</td>
</tr>
<tr>
<td>Total</td>
<td>85</td>
<td>215</td>
<td>300</td>
</tr>
</tbody>
</table>

(b) \( 0.0753. \)

(c) \( 2(0.0753) = 0.1506. \)

35. The probabilities are:

\[ p_1 = 14/100 = 0.14 \text{ and } p_2 = 16/400 = 0.04. \]

Thus, the relative risk is \( 0.14/0.04 = 3.50 \)

and the odds ratio is

\[ \frac{0.14(0.96)}{0.86(0.04)} = 3.907. \]

36. The probabilities are:

\[ p_1 = 16/100 = 0.16 \text{ and } p_2 = 20/500 = 0.04. \]

Thus, the relative risk is \( 0.16/0.04 = 4 \)

and the odds ratio is

\[ \frac{0.16(0.96)}{0.84(0.04)} = 4.571. \]
37. The point estimate is
\[
\frac{100(460)}{40(400)} = 2.875.
\]
38. The point estimate is
\[
\frac{100(660)}{40(600)} = 2.75.
\]
39. The slope is
\[
b_1 = r(s_y/s_x) = 0.30(25/10) = 0.75.
\]
The intercept is
\[
b_0 = \bar{y} - b_1 \bar{x} = 200 - 0.75(60) = 155.
\]
Thus, the regression line is
\[
\hat{y} = 155 + 0.75x.
\]
40. The slope is
\[
b_1 = r(s_y/s_x) = -0.60(30/15) = -1.2.
\]
The intercept is
\[
b_0 = \bar{y} - b_1 \bar{x} = 250 + 1.2(80) = 346.
\]
Thus, the regression line is
\[
\hat{y} = 346 - 1.2x.
\]
41. For the given case,
\[
-10 = 250 - \hat{y} \text{ or } \hat{y} = 260.
\]
Thus, we have two points on the regression line: (120,260) and (100,300).
Thus, the slope is
\[
(300-260)/(100-120) = 40/(-20) = -2.
\]
The intercept satisfies
\[
300 = b_0 - 2(100) \text{ or } b_0 = 500.
\]
The regression line is
\[
\hat{y} = 500 - 2x.
\]
42. For the given case,
\[
20 = 295 - \hat{y} \text{ or } \hat{y} = 275.
\]
Thus, we have two points on the regression line: (75,275) and (60,245).
Thus, the slope is
\[
(275 - 245)/(75 - 60) = 30/15 = 2.
\]
The intercept satisfies
\[
245 = b_0 + 2(60) \text{ or } b_0 = 125.
\]
The regression line is
\[
\hat{y} = 125 + 2x.
\]
43. (a) \(A = 13\); select ‘Area right of;’ \(B = 1.234\). The P-value is \(C\).
(b) \(A = 16\); select ‘Area right of;’ \(C = 0.05\). The \(t\) for the CI is \(B\).
(c) \(A = 20\); select ‘Area left of;’ \(B = -2.587\). The P-value is \(C\).
(d) \(A = 13\); select ‘Area right of;’ \(C = 0.005\). The \(t\) for the CI is \(B\).
(e) \(A = 25\); select ‘Area left of;’ \(B = -2.03\). The P-value is \(2C\).
44. (a) \(A = 21\); select ‘Area right of;’ \(B = 3.234\). The P-value is \(2C\).
(b) \(A = 26\); select ‘Area right of;’ \(C = 0.10\). The \(t\) for the CI is \(B\).
(c) \(A = 19\); select ‘Area left of;’ \(B = -4.59\). The P-value is \(C\).
(d) \(A = 19\); select ‘Area right of;’ \(C = 0.01\). The \(t\) for the CI is \(B\).
(e) \(A = 32\); select ‘Area right of;’ \(B = 2.837\). The P-value is \(C\).
45. First, \(df = n - 2 = 22 - 2 = 20\). Thus, \(t^* = 2.086\). Thus, the 95% CI is
\[
4.05 \pm 2.086(0.8774) = 4.05 \pm 1.83 = [2.22, 5.88].
\]
46. Again, $t^* = 2.086$. Thus, the 95% CI is

$$192.33 \pm 2.086(4.32) = 192.33 \pm 9.01 = [183.32, 201.34].$$

47. First, we need the variance of the predicted value:

$$ (2.41)^2 + (11.30)^2 = 133.4981. $$

Thus, the SE of the predicted value is 11.55. The 95% PI is

$$208.54 \pm 2.086(11.55) = 208.54 \pm 24.09 = [184.45, 232.63].$$

This is a very wide interval!

48. (a) $\hat{y} = -91.44 + 4.0537(79) = 228.80$.

Or, you can take the $\hat{y}$ for $x = 78$, which is 224.76 from the output, and add one slope, 4.05, to obtain 228.81.

(b) From the table, we see that $\hat{y} = 220.70$. Thus, the residual is $230 - 220.70 = 9.30$.

(c) We are given

$$ -12.59 = 200 - \hat{y}. $$

Thus, $\hat{y} = 212.59$. Now you have a choice. You can solve this last equation for $x$ or you can look at the computer output. From the latter, clearly $x = 75$.

49. The correlation coefficient is the positive square root of $R^2$ because the slope of the regression line is positive. Thus, $r = \sqrt{0.516} = 0.718$.

50. (a) This means that the residual is positive; by counting there are 12 positive residuals.