6.3 Data summary for Exercise 6.3

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>sum</th>
<th>st dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_i$</td>
<td>14</td>
<td>425300</td>
<td>3127.3</td>
</tr>
<tr>
<td>$y_i$</td>
<td>14</td>
<td>62400</td>
<td>792.96</td>
</tr>
<tr>
<td>$y_i - rx_i$</td>
<td>14</td>
<td>0</td>
<td>611.52</td>
</tr>
</tbody>
</table>

$$r = \frac{\sum y_i}{\sum x_i} = \frac{62400}{425300} = .147$$

$$2\sqrt{V(r)} = 2\sqrt{\left(\frac{N-n}{nN}\right)\left(\frac{1}{\bar{x}^2}\right)s_r^2}$$
$$= 2\sqrt{\frac{150-14}{14(150)}} \frac{611.52}{(425300/14)} = .0102$$

6.9 Data summary for exercise 6.9

$x_i =$ photo count, $y_i =$ ground count

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>mean</th>
<th>st dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_i$</td>
<td>10</td>
<td>23.4</td>
<td>11.4717</td>
</tr>
<tr>
<td>$y_i$</td>
<td>10</td>
<td>30.6</td>
<td>14.8489</td>
</tr>
<tr>
<td>$y_i - rx_i$</td>
<td>10</td>
<td>0.00</td>
<td>3.4751</td>
</tr>
</tbody>
</table>

$$r = \frac{\sum y_i}{\sum x_i} = \frac{\bar{y}}{\bar{x}} = \frac{30.6}{23.4} = 1.31$$

$$\hat{\tau}_y = r\tau_x = 1.31(4200) = 5492.31$$

$$\hat{V}(\hat{\tau}_y) = \tau_x^2\left(\frac{N-n}{nN}\right)\frac{1}{\mu_x^2}s_r^2 = N^2\left(\frac{N-n}{nN}\right)s_r^2$$

$$B = 2\sqrt{\hat{V}(\hat{\tau}_y)} = 2(200)\sqrt{\frac{200-10}{10(200)}} \cdot 3.4751 = 428.44$$
The regression equation is: \( \text{Ground} = 1.13 + 1.26 \ \text{Photo} \)

Then,

\[
\hat{y}_{yL} = N\hat{\mu}_{yL} = N\left[\bar{y} + b(\mu_x - \bar{x})\right]
\]

\[
\hat{y}_{yL} = 200\left[\frac{306}{10} + 1.26\left(\frac{4200}{200} - \frac{234}{10}\right)\right] = 5515.50
\]

\[
\hat{V}(\hat{\mu}_{yL}) = \frac{N-n}{nN} \text{MSE}
\]

\[
= \frac{200 - 10}{10(200)} = 13.2
\]

\[
B = 2\sqrt{\hat{V}(\hat{y}_{yL})} = 2\sqrt{\hat{V}(N\hat{\mu}_{yL})} = 2\sqrt{\hat{V}(\hat{\mu}_{yL})} = 2(200)(1.1216) = 448.6
\]
6.23 (a) Data summary for Exercise 6.23(a)

USE DATA AT RIGHT FOR THIS PROBLEM

\[ y = 1989 \text{ incomes}, \quad x = 1980 \text{ incomes} \]

<table>
<thead>
<tr>
<th></th>
<th>( n )</th>
<th>mean</th>
<th>st dev</th>
<th>SE mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_i )</td>
<td>6</td>
<td>40.83</td>
<td>19.93</td>
<td>8.14</td>
</tr>
<tr>
<td>( y_i )</td>
<td>6</td>
<td>54.5</td>
<td>29.4</td>
<td>12.0</td>
</tr>
<tr>
<td>( y_i - rx_i )</td>
<td>6</td>
<td>-0.00</td>
<td>7.19</td>
<td>2.94</td>
</tr>
</tbody>
</table>

\[ r = \frac{\bar{y}}{\bar{x}} = \frac{54.50}{40.83} = 1.335 \]

\[ \hat{\gamma}_y = r \hat{\gamma}_x = (1.335)(674) = 895.75 \]

\[ \hat{V}(\hat{\gamma}_y) = \frac{N^2}{nN} \left( \frac{N-n}{nN} \right) \]  
\[ = (19)^2 \left( \frac{19-6}{19} \right)(2.94)^2 \]

\[ B = 2\sqrt{\hat{V}(\hat{\gamma}_y)} = 2(46.20) = 92.40 \]

(b) The regression equation is: \( y = -3.94 + 1.43x \)

\[ \hat{\gamma}_{yL} = N\hat{\gamma}_{yL} = N[\bar{y} + b(\bar{X} - \bar{X})] \]

\[ \hat{\gamma}_{yL} = 19 \left[ \frac{327}{6} + 1.43 \left( \frac{674}{19} - \frac{245}{6} \right) \right] = 889.88 \]

\[ \hat{V}(\hat{\gamma}_{yL}) = \frac{N-n}{nN} \text{MSE} \]
\[ = \]
\[ = \frac{19-6}{6(19)}(60) \]

\[ B = 2\sqrt{\hat{V}(\hat{\gamma}_{yL})} = 2\sqrt{\hat{V}(N\hat{\gamma}_{yL})} = 2N\sqrt{\hat{V}(\hat{\gamma}_{yL})} = 2(19)(2.6154) = 99.38 \]
(c) The plot of 1989 values versus 1980 values shows a linear pattern that could be modeled by a straight line nearly through the origin, so both ratio and regression methods work well (and give nearly the same answers). The difference method is a little off because the slope of the regression line is quite far from unity.

<table>
<thead>
<tr>
<th>1980 x</th>
<th>1989 y</th>
<th>$d_i = y_i - x_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>26</td>
<td>5</td>
</tr>
<tr>
<td>63</td>
<td>91</td>
<td>28</td>
</tr>
<tr>
<td>35</td>
<td>47</td>
<td>12</td>
</tr>
<tr>
<td>60</td>
<td>70</td>
<td>10</td>
</tr>
<tr>
<td>16</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>50</td>
<td>76</td>
<td>26</td>
</tr>
</tbody>
</table>

$$\sum_{i=1}^{6} d_i = 82, \quad \sum_{i=1}^{6} d_i^2 = 1730 \quad s_d^2 = 121.87$$

$$\hat{\tau}_{y_D} = N\hat{\mu}_{y_D} = N(\mu_x + d) = 19\left(\frac{674}{19} + \frac{82}{6}\right) = 933.67$$

$$\widehat{V}(\tau_{y_D}) = \widehat{V}(N\hat{\mu}_{y_D}) = N^2\widehat{V}(\hat{\mu}_{y_D}) = N^2(1 - \frac{n}{N})\frac{s_d^2}{n}$$

$$= (19)^2\left(1 - \frac{6}{19}\right)\frac{121.87}{6} = 5017$$

$$B = 2\sqrt{5017} = 141.66$$

(d) The plot of 1989 values versus 1980 values shows a linear pattern that could be modeled by a straight line nearly through the origin, so both ratio and regression methods work well (and give nearly the same answers). The difference method is a little off because the slope of the regression line is quite far from unity.
7.3  
(a) \( N = 40, \ k = 10, \ n = 4 \)

<table>
<thead>
<tr>
<th>sample</th>
<th>sample elements</th>
<th>( \hat{p} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1   11  21  31</td>
<td>0.75</td>
</tr>
<tr>
<td>2</td>
<td>2   12  22  32</td>
<td>1.00</td>
</tr>
<tr>
<td>3</td>
<td>3   13  23  33</td>
<td>0.75</td>
</tr>
<tr>
<td>4</td>
<td>4   14  24  34</td>
<td>0.75</td>
</tr>
<tr>
<td>5</td>
<td>5   15  25  35</td>
<td>0.25</td>
</tr>
<tr>
<td>6</td>
<td>6   16  26  36</td>
<td>0.00</td>
</tr>
<tr>
<td>7</td>
<td>7   17  27  37</td>
<td>0.00</td>
</tr>
<tr>
<td>8</td>
<td>8   18  28  38</td>
<td>0.00</td>
</tr>
<tr>
<td>9</td>
<td>9   19  29  39</td>
<td>0.25</td>
</tr>
<tr>
<td>10</td>
<td>10  20  30  40</td>
<td>0.25</td>
</tr>
</tbody>
</table>

where \( \hat{p} \) is the proportion of delinquent accounts in the sample

The probability distribution of \( \hat{p} \) is

<table>
<thead>
<tr>
<th>( \hat{p} )</th>
<th>0.00</th>
<th>0.25</th>
<th>0.75</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p(\hat{p}) )</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.1</td>
</tr>
</tbody>
</table>

\[
E(\hat{p}) = \sum \hat{p} p(\hat{p}) = 0(.3) + (.25)(.3) + (.75)(.3) + 1(.1) = .4
\]
\[
E(\hat{p}^2) = \sum \hat{p}^2 p(\hat{p}) = (0)^2(.3) + (.25)^2(.3) + (.75)^2(.3) + (1)^2(.1) = .2875
\]
\[
V(\hat{p}) = E(\hat{p}^2) - (E(\hat{p}))^2 = .2875 - .4^2 = .1275
\]
(b) \( N = 40, \ k = 5, \ n = 8 \)

<table>
<thead>
<tr>
<th>Sample</th>
<th>Sample Elements</th>
<th>( \hat{p} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1 6 11 16 21 26 31 36</td>
<td>0.375</td>
</tr>
<tr>
<td>2</td>
<td>2 7 12 17 22 27 32 37</td>
<td>0.500</td>
</tr>
<tr>
<td>3</td>
<td>3 8 13 18 23 28 33 38</td>
<td>0.375</td>
</tr>
<tr>
<td>4</td>
<td>4 9 14 19 24 29 34 39</td>
<td>0.500</td>
</tr>
<tr>
<td>5</td>
<td>5 10 15 20 25 30 35 40</td>
<td>0.250</td>
</tr>
</tbody>
</table>

where \( \hat{p} \) is the proportion of delinquent accounts in the sample

The probability distribution of \( \hat{p} \) is

<table>
<thead>
<tr>
<th>( \hat{p} )</th>
<th>( p(\hat{p}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0.2</td>
</tr>
<tr>
<td>0.375</td>
<td>0.4</td>
</tr>
<tr>
<td>0.50</td>
<td>0.4</td>
</tr>
</tbody>
</table>

\[
E(\hat{p}) = \sum \hat{p}p(\hat{p}) = .25(.2) + .375(.4) + .5(.4) = .4
\]

\[
E(\hat{p}^2) = \sum \hat{p}^2p(\hat{p}) = (.25)^2(.2) + (.375)^2(.4) + (.5)^2(.4) = .16875
\]

\[
V(\hat{p}) = E(\hat{p}^2) - (E(\hat{p}))^2 = .16875 - (.4)^2 = .00875
\]

See top of next page for 7-3(c)

7.6 \( N = 1800 \quad n = 36 \quad \sum y_i = 430.01 \)

\[
s^2 = .0062
\]

\[
\hat{\mu} = \bar{y} = \frac{430.01}{36} = 11.94
\]

\[
\hat{V}(\hat{\mu}) = \frac{s^2}{n} \left( \frac{N-n}{N} \right) = \frac{.0062}{36} \left( \frac{1800-36}{1800} \right)
\]

\[
B = 2\sqrt{\hat{V}(\hat{\mu})} = .026
\]

7.18 \( N = 15200 \quad n = 304 \quad \sum y_i = 88 \)

\[
\hat{p}_{xy} = \frac{\sum y_i}{n} = \frac{88}{304} = .2895
\]

\[
\hat{\tau}_{xy} = N\hat{p}_{xy} = N \frac{\sum y_i}{n} = 15200 \left( \frac{88}{304} \right) = 4400
\]

\[
B = 2\sqrt{\hat{V}(\hat{\tau}_{xy})} = 2N\sqrt{\hat{V}(\hat{p}_{xy})} = 2N \sqrt{\frac{\hat{p}_{xy} \hat{q}_{xy}}{n-1} \left( \frac{N-n}{N} \right)}
\]

\[
= 2(15200) \sqrt{\left( \frac{.2895}{303} \right) \left( \frac{15200-304}{15200} \right)} = 784.08
\]
Problem 7-3(c)

Simple random sample of size 4
28, 40, 26, 13 (past due).
\[ \hat{p} = \frac{1}{4}; \hat{V}(\hat{p}) = \frac{\hat{p}(1-\hat{p})}{n} = \frac{.25 \times (.75)}{4} = .04688 \]
Note that since \( p = .4 \), for simple random samples of size 4,
\[ V(\hat{p}) = \frac{(4)(.6)}{4} = .06. \]

Simple random sample of size 8
4 (past due), 40, 8, 16, 5, 10, 35, 15
\[ \hat{p} = \frac{1}{8}; \hat{V}(\hat{p}) = \frac{125 \times (.875)}{8} = .01367. \]
Note that since \( p = .4 \), for simple random samples of size 8,
\[ V(\hat{p}) = \frac{(8)(.6)}{8} = .03. \]

The pattern of past due accounts is cyclic since the past due accounts are the lower numbered accounts in each department. When the population of responses is cyclic, systematic sampling in general is worse than simple random sampling.

1-in-10 systematic samples which result in samples of size 4 may result in a sample selected from the higher account numbers in each department and therefore have no past due accounts in the sample. Note in part (a) that \( \hat{p} = 0 \) with probability 0.3. A 1-in-10 sample may also result in all sampled accounts being past due as seen in part (a). As seen above, simple random samples of size \( n = 4 \) have smaller variance.

1-in-5 systematic samples result in samples of size 8 and so are more likely to have both past due and up-to-date accounts. The systematic sample size in this case is large to enable it to do better than simple random samples of size 8.
Another Chapter 7 Problem.

A newspaper reporter is writing a story about a recent significant two-day rain event in central North Carolina. She would like to obtain an accurate estimate of the mean rainfall across the Piedmont region. Since the reporter does not have time to obtain the rainfall amount recorded at all of the N = 60 weather stations in the Piedmont, she decides to select a 1-in-5 systematic sample from a list of the weather stations. She then will call the weather stations in her sample to obtain their rainfall amounts $y_i$. There are 5 possible samples, called cluster samples, that can be selected; each possible sample will have $n = 12$ weather stations.

a. Suppose that the rainfall amounts at the weather stations selected by the reporter are $y_1, y_2, \ldots, y_{12}$ and that $s^2 = 1.343$. Calculate an estimate of the variance $V(\bar{y}_{sy})$ assuming a randomly ordered population (include the finite population correction factor).

**Answer**

When the population is randomly ordered we can estimate the variance of $\bar{y}_{sy}$ using the estimated variance of the sample mean $\bar{y}$ obtained from a simple random sample: $\hat{V}(\bar{y}_{sy}) = \left(1 - \frac{n}{N}\right)\frac{s^2}{n} = (.8)\frac{1.343}{12} = .08953$.

The reporter is concerned that the population of weather stations on her list is not randomly ordered since the bound on the error of estimation resulting from the above estimated variance is 0.6 inches, rather large for an estimated rainfall amount.

The actual variance $V(\bar{y}_{sy})$ depends on $\rho$, a measure of the correlation among elements in the same cluster. An analysis of the rainfall amounts at all N = 60 weather stations results in the ANOVA table below.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F-Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>4</td>
<td>0.69546</td>
<td>0.173865</td>
<td>0.13236161</td>
<td>0.9699</td>
</tr>
<tr>
<td>Error</td>
<td>55</td>
<td>62.245833</td>
<td>1.3135606</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>59</td>
<td>72.941293</td>
<td>1.3135606</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

b. Use the ANOVA table to calculate the value of $\rho$

**Answer**

$$\rho \approx \frac{MSE}{MST} = \frac{0.174 - 1.236}{11(1.236)} = -.0781$$

where $\text{MST} = \frac{SST}{N-1}$

c. Calculate $V(\bar{y}_{sy})$.

**Answer**

$$V(\bar{y}_{sy}) = \frac{s^2}{n} \left[1 + \left(n - 1\right)\rho\right] = \frac{72.941}{12(60)} \left[1 + 11(-.0781)\right] = .0143$$