Problem 3:

**Part a:** We compute the number of dredge rows for each stratum by dividing the area of the stratum by 0.039 acres. We treat those numbers as the population sizes $N_1, \ldots, N_4$ and apply stratified sampling theory. Our estimate of the total in the $h$th stratum is given by

$$
\hat{t}_h = N_h \bar{y}_h,
$$

where $N_h$ is the population size for the $h$th stratum and $\bar{y}_h$ is the stratum sample mean. Our estimate of the variance of the total $\hat{t}_h$ is given by

$$
\hat{\text{Var}}(\hat{t}_h) = N_h^2 \left( 1 - \frac{n_h}{N_h} \right) \frac{s_h^2}{n_h},
$$

where $s_h^2$ is the sample variance for the $h$th stratum. We table the relevant information below.

<table>
<thead>
<tr>
<th>$h$</th>
<th>$N_h$</th>
<th>$n_h$</th>
<th>$\bar{y}_h$</th>
<th>$s_h^2$</th>
<th>$\hat{t}_h$</th>
<th>$\hat{\text{Var}}(\hat{t}_h)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5713</td>
<td>4</td>
<td>0.44</td>
<td>0.068</td>
<td>2513.7</td>
<td>554,463.8</td>
</tr>
<tr>
<td>2</td>
<td>1272</td>
<td>6</td>
<td>1.17</td>
<td>0.042</td>
<td>1488.2</td>
<td>11,272.5</td>
</tr>
<tr>
<td>3</td>
<td>1288</td>
<td>3</td>
<td>3.92</td>
<td>2.146</td>
<td>5049.0</td>
<td>1,183,933.9</td>
</tr>
<tr>
<td>4</td>
<td>5072</td>
<td>5</td>
<td>1.80</td>
<td>0.794</td>
<td>9129.6</td>
<td>4,081,132.1</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18,180.5</td>
<td>5,830,802.3</td>
</tr>
</tbody>
</table>

We estimate the total as $\hat{t} = \hat{t}_1 + \hat{t}_2 + \hat{t}_3 + \hat{t}_4$. Since the estimates for the separate strata are independent, we can add the corresponding variance estimates. Thus, our estimate of the total is 18,180 bushels, and the corresponding standard error is

$$
SE(\hat{t}) = \sqrt{5,830,802.3} \approx 2,415 \text{ bushels}.
$$

**Part b:** Here we follow the same procedure that we used in Part a.

<table>
<thead>
<tr>
<th>$h$</th>
<th>$N_h$</th>
<th>$n_h$</th>
<th>$\bar{y}_h$</th>
<th>$s_h^2$</th>
<th>$\hat{t}_h$</th>
<th>$\hat{\text{Var}}(\hat{t}_h)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8274</td>
<td>8</td>
<td>0.63</td>
<td>0.083</td>
<td>5212.6</td>
<td>709,576.2</td>
</tr>
<tr>
<td>2</td>
<td>5072</td>
<td>5</td>
<td>0.40</td>
<td>0.046</td>
<td>2028.8</td>
<td>236,438.4</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7,241.4</td>
<td>946,014.6</td>
</tr>
</tbody>
</table>

Our estimate of the total is 7,241 bushels, and the corresponding standard error is

$$
SE(\hat{t}) = \sqrt{946,014.6} \approx 973 \text{ bushels}.
$$
Problem 6:

Part a: The key result that we use here is the fact that (see p.107 in the text) the optimal sample size \( n_h \) for the \( h \)th stratum satisfies

\[
n_h \propto \frac{N_h S_h}{\sqrt{c_h}},
\]

where \( s_h \) is the standard deviation in the \( h \)th stratum and \( c_h \) is the cost of sampling one unit from the \( h \)th stratum. Let stratum 1 consist of the households with telephones, and let stratum 2 consist of the households without telephones. We then have that if \( N \) is the total population, then \( N_1 = 0.9N \) and \( N_2 = 0.1N \). Also, since the variances in the two strata are similar, we have that \( S_1 = S_2 \). Finally, since all households are being interviewed in person, we have that \( c_1 = c_2 = $30 \).

Thus, we have that

\[
n_1 \propto \frac{0.9 N S_1}{\sqrt{30}},
\]

while

\[
n_2 \propto \frac{0.1 N S_1}{\sqrt{30}},
\]

meaning that

\[
\frac{n_1}{n_2} = \frac{0.9 N S_1}{\sqrt{30}} \cdot \frac{\sqrt{30}}{0.1 N S_1} = 9.
\]

We also know that since the amount available for sampling is $15,000, we must have that

\[
15,000 = n_1(30) + n_2(30).
\]

Since \( n_1 = 9n_2 \), this means that \( 15,000 = 9n_2(30) + n_2(30) = 300n_2 \). Thus,

\[
n_2 = \frac{15,000}{300} = 50,
\]

and \( n_1 = 9n_2 = 450 \).

Part b: We follow the same strategy as in Part a. The one difference is that the costs are now different. We have that \( c_1 = 10 \), while \( c_2 = 40 \). Thus, we know that

\[
n_1 \propto \frac{0.9 N S_1}{\sqrt{10}} \quad \text{and} \quad n_2 \propto \frac{0.1 N S_1}{\sqrt{40}},
\]

meaning that

\[
\frac{n_1}{n_2} = \frac{0.9 N S_1}{\sqrt{10}} \cdot \frac{\sqrt{40}}{0.1 N S_1} = 18.
\]

We also know that \( 15,000 = n_1(10) + n_2(40) \). Since \( n_1 = 18n_2 \), this means that

\[
15,000 = 18n_2(10) + n_2(40) = 220n_2.
\]

Hence \( n_2 = \frac{15,000}{220} = 68 \), while \( n_1 = 18n_2 = 1224 \).
Problem 10:

Part a: Our estimation here follows the same pattern that was used in Problem 3. We table the relevant information on the four strata below.

<table>
<thead>
<tr>
<th>h</th>
<th>$N_h$</th>
<th>$n_h$</th>
<th>$\bar{y}_h$</th>
<th>$s^2_h$</th>
<th>$\hat{t}_h$</th>
<th>$\hat{\text{Var}}(\hat{t}_h)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Bio)</td>
<td>102</td>
<td>7</td>
<td>3.14</td>
<td>6.81</td>
<td>320.3</td>
<td>9,427.0</td>
</tr>
<tr>
<td>2 (Phy)</td>
<td>310</td>
<td>19</td>
<td>2.11</td>
<td>8.21</td>
<td>654.1</td>
<td>38,980.2</td>
</tr>
<tr>
<td>3 (Soc)</td>
<td>217</td>
<td>13</td>
<td>1.23</td>
<td>4.36</td>
<td>266.9</td>
<td>14,846.8</td>
</tr>
<tr>
<td>4 (Hum)</td>
<td>178</td>
<td>11</td>
<td>0.45</td>
<td>0.87</td>
<td>80.1</td>
<td>2,351.1</td>
</tr>
<tr>
<td>Total</td>
<td>807</td>
<td>50</td>
<td>1321.4</td>
<td>65,605.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We thus estimate that the total number of refereed publications by faculty members in the college is about 1,321 papers. The corresponding standard error is

$$SE(\hat{t}) = \sqrt{65,605.1} = 256$$

Part b: In Problem 8 of Chapter 2, our estimate of the total was 1436 papers, and the corresponding standard error was 296. Thus, both the point estimate and the standard error obtained in Part a are smaller than the values previously obtained.

Part c: The parameter of interest is a proportion. We proceed by first estimating the total number of faculty with no refereed publications, then dividing our estimate by the total number of faculty. In computing $\hat{\text{Var}}(\hat{t}_h)$, we use the fact that for a variable that can only take the values 0 and 1,

$$\hat{\text{Var}}(\hat{t}_h) = N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{\hat{p}_h(1 - \hat{p}_h)}{n_h - 1},$$

where $\hat{p}_h$ is the proportion in the $h$th stratum. We table the relevant information on the four strata below.

<table>
<thead>
<tr>
<th>h</th>
<th>$N_h$</th>
<th>$n_h$</th>
<th>$\bar{y}_h = \hat{p}_h$</th>
<th>$\hat{t}_h$</th>
<th>$\hat{\text{Var}}(\hat{t}_h)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Bio)</td>
<td>102</td>
<td>7</td>
<td>1/7</td>
<td>14.6</td>
<td>197.8</td>
</tr>
<tr>
<td>2 (Phy)</td>
<td>310</td>
<td>19</td>
<td>10/19</td>
<td>163.1</td>
<td>1249.4</td>
</tr>
<tr>
<td>3 (Soc)</td>
<td>217</td>
<td>13</td>
<td>9/13</td>
<td>150.2</td>
<td>785.8</td>
</tr>
<tr>
<td>4 (Hum)</td>
<td>178</td>
<td>11</td>
<td>8/11</td>
<td>129.5</td>
<td>589.6</td>
</tr>
<tr>
<td>Total</td>
<td>807</td>
<td>50</td>
<td>457.4</td>
<td>2822.6</td>
<td></td>
</tr>
</tbody>
</table>

Our estimate for the population proportion is then $\hat{p} = \frac{457.4}{807} = 0.567$. The corresponding standard error estimate is

$$SE(\hat{p}) = \frac{SE(\hat{t})}{\sqrt{N}} = \frac{\sqrt{2822.6}}{807} = 0.066.$$

Part d: In both of the computations we made in this example, stratification increased the precision, though the increase was not particularly large. The standard error for our estimate of the
total number of refereed publications dropped to 256 from 296, while the standard error for our estimate of the proportion of faculty with no refereed publications dropped to 0.066 from 0.069. We expect stratified sampling to outperform simple random sampling when the stratum means are very different. In this case, there is strong evidence that faculty in the natural sciences, particularly the biological sciences, tend to have more refereed publications than faculty in the social sciences and humanities. Thus, the stratum means are in fact different. If we ran this survey again, we could probably increase the precision even further by sampling at different rates in the different strata.

**Problem 12:**

To solve this problem, again use the result that the optimal sample size $n_h$ for the $h$th stratum satisfies

$$n_h \propto \frac{N_h S_h}{\sqrt{c_h}}.$$  

Since the costs are the same from one stratum to the next, we have that

$$n_h \propto N_h S_h.$$  

Now apply the same strategy used in Problem 6.