Note: The problems are taken from the Exercises in Scheaffer, Mendenhall, Ott, Gerow (2011), 7th edition, unless otherwise noted. For each problem, please explain your reasoning clearly. It is not acceptable to only provide your final result.

**Homework 1** (Due Mon, Feb 3):
2.1, 2.2, 2.3, 2.5, 2.7, 3.5, 3.6

Additional problems: See the Probability Problems posted on the course webpage.

**Homework 2** (Due Mon, Feb 17, extended to Thu, Feb 20):
4.1, 4.2, 4.14, 4.15, 4.17, 4.19, 4.30, 4.44

The following questions from Sampling Activity, Random Rectangles [found at the end of Exercises in Chapter 4, page 112-113 of the textbook]:

- Questions 1, 2;
- Questions 5: use R to draw your random sample, and also compute the sample mean, bound on error of estimation, and confidence interval for the population mean area of the rectangles;
- Additional question: repeat Question 5 for 5 times, and report your 5 sample means and 5 confidence intervals. What proportions of your confidence intervals cover the truth (computed from the table in Questions 8)?

[Hint: In R, need to create a vector manually to store the population data shown in the table on page 113 of the textbook.]

Additional problem I: For simple random sampling, prove the following results.

1) Show that each pair of units, $i$ and $j (i \neq j)$, has probability \( \frac{n(n-1)}{N(N-1)} \) of being selected.

2) Show that
\[
E(s^2) = \frac{N}{N-1} \sigma^2,
\]
where $s^2$ is the sample variance and $\sigma^2$ the population variance.

Additional problem II: The following table gives the distribution of the number of Apple products used by a class of 60 students:

<table>
<thead>
<tr>
<th># of Apple products</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td># of students</td>
<td>15</td>
<td>24</td>
<td>10</td>
<td>6</td>
<td>5</td>
</tr>
</tbody>
</table>
Use R to complete the following exercises. Please submit your R codes and R output showing your answers.

1) Draw a simple random sample of size 15, and compute the sample mean and confidence interval for the class mean number of Apple products used by the students.

2) Repeat Exercise 1) for 1000 times, and plot a histogram of your 1000 sample means. Find the mean and variance of the 1000 sample means, and compare them to the theoretical mean and variance of the sample mean.

3) From Exercise 2), also compute the confidence interval for each of your 1000 samples. What proportions of your 1000 confidence intervals cover the truth?

[Hint: Need to modify the R codes posted on the course webpage accordingly, to handle the non-binary data here.]

Homework 3 (Due Thu, March 5, extended to Mon, March 9):
5.2, 5.5, 5.13, 5.14, 5.15, 5.21, 5.30, 5.43, 5.44, 5.45

Sampling from Real Populations [page 165 in the textbook]:
5.3, 5.5

Note: Please show your codes and output in R. The dataset CARS93 is posted on the course website, and can be read into R as follows:

```r
data <- read.table(file="cars93.txt", sep="\t", header=TRUE)
```

Additional problem I: For stratified random sampling, let \( \mu \) be the population mean, \( \sigma^2 \) be the population variance, and \( S^2 = N/(N - 1)\sigma^2 \). Show that

\[
S^2 = \frac{1}{N-1} \sum_{i=1}^{L} \{(N_i - 1)S_i^2 + N_i(\mu_i - \mu)^2\},
\]

where \( \mu_i \) is the stratum population mean, \( \sigma_i^2 \) is the stratum population variance, and \( S_i^2 = N_i/(N_i - 1)\sigma_i^2 \) for \( i = 1, \ldots, L \).

Homework 4 (Due Mon, April 6):
6.1, 6.2, 6.4, 6.5, 6.9 (use a ratio estimator), 6.10 (ignore the question “What pattern ...”), 6.16, 6.25, 6.26

Additional problem I: Let \( \bar{x} \) and \( \bar{y} \) be the sample means, \( s_x^2, s_y^2 \), and \( s_{xy} \) be the sample
variances and covariance, $\hat{b} = s_{xy} / s^2_x$ and $\hat{\rho} = s_{xy} / (s_x s_y)$. Show that

$$\frac{1}{n-1} \sum_{i \in \text{sample}} \{y_i - \bar{y} - \hat{b}(x_i - \bar{x})\}^2 = s^2_y + \hat{b}^2 s^2_x - 2\hat{b} x s_y = s^2_y - \hat{b}^2 s^2_x = (1 - \hat{\rho}^2) s^2_y.$$

Sampling from Real Populations (the datasets are posted on the course website):

6.1 (see the note below), 6.5

Note: Please show your codes and output in R. The datasets can be read into R by the following codes:

```r
data1 <- read.table(file="temps.txt", sep="\t", header=TRUE)
data2 <- read.table(file="schools.txt", sep="\t", header=TRUE)
```

For 6.1, use the sample size 20. For 6.1, (b), the problem should be stated as follows.

Estimate (with an error bound on) the average March precipitation for all stations, using the sample data on January and March precipitation and the additional knowledge that the average January precipitation for all stations is 2.526136 inches. Compare your answers using ratio, regression, and difference estimation.

**Homework 5** (Due Thu, Apr 16):

8.2, 8.3, 8.4, 8.5, 8.8, 8.9, 8.20, 8.21, 8.22

Note: You may choose to use R to perform the computation, instead of by hand. In either case, please show enough details of your solutions. The datasets for some selected exercises are posted on the course website.

Sampling from Real Populations:

8.4 (a), (c), (d)

Note: Please show your codes and output in R. Please ignore Question (b) for systematic sampling. For Questions (c) and (d), please draw random samples in the two ways stated there, i.e., with either rows or columns (as you consider appropriate) as strata for (c) and as clusters for (d), and compare your answers. You should use R to draw the samples and then compute the estimates (by either hands or R) using the tree grid in the textbook. For your convenience, the R code to read the tree grid into R is given in the course website.
**Homework 6** (Due Thu, Apr 24):

Note: You may choose to use R to do the computation, instead of by hand. In either case, please show enough details of your solutions. The datasets for some selected exercises are posted on the course website.

Sampling from Real Populations:
Return to the Problem 8.4. You are asked to sample a total of 30 trees to estimate the proportion of diseased trees. Use two-stage cluster sampling, with either rows or columns as clusters, as follows.

1) Draw 10 rows, and then draw 3 cells from each selected row.

2) Use the sample data from 1) to estimate the variance that would be obtained if drawing 6 rows and then 5 cells from each selected row.

3) Draw 3 columns, and then draw 10 cells from each selected column.

4) Use the sample data from 3) to estimate the variance that would be obtained if using stratified sampling: drawing 6 cells from each of the 5 columns.