Plan for today: Given data we will calculate the variance and standard deviation, the lower and upper quartiles, range, interquartile range, five-number summary, introduce the notion of outliers, state the Chebyshev rule and check it for the data on Research and Development.
Let us recall the notions of the sample mean and sample variance
\[
\bar{x} = \frac{1}{n} (x_1 + \ldots + x_n);
\]

\[
s^2 = \frac{1}{n-1} \left( (x_1 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2 \right).
\]

When our data represents a population, the population mean and variance are denoted by \( \mu \) and \( \sigma^2 \) respectively and are defined as follows
\[
\mu = \frac{1}{n} (x_1 + \ldots + x_n);
\]

\[
\sigma^2 = \frac{1}{n} \left( (x_1 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2 \right).
\]
Problem. Use the idea of shortcut calculation of the mean and variance to find the mean and variance of the following data: 20006, 20009, 20012. Hint: represent our sequence $Y$ in the form $Y=X+20000$.

PROBLEM 2.63, p.77. Consider the following sample of 5 measurements: 2, 1, 1, 0, 3. Find $\bar{x}$, median $m$, range, variance and standard deviation.

Answers:
$\bar{x} = 1.4$, $\text{Range} = \max - \min = 3$, $s^2 = 1.3$, $s = 1.14$

While calculating $s$ we can use the following shortcut formula

$$s^2 = \frac{\sum_{k=1}^{n} x_k^2 - n\bar{x}^2}{n-1}.$$
Problem. The following data are the numbers of ounces of silver per ton of ore for a mine

<table>
<thead>
<tr>
<th></th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>

Find: the median, quartiles ($Q_U$ and $Q_L$), interquartile range $Q_U - Q_L$, and the five number summary.

Answers: $n=19$, $m=40$, $Q_L = 32$, $Q_U = 49$, $IQR=Q_U - Q_L = 17$.

Homework: Text: pp.73-80.
Exercises: 2.57-2.63, 2.74 (no Empirical rule part).
Chebyshev’s Rule

At least $\frac{3}{4}$ of the observations will fall within 2 standard deviations of the mean (i.e. within the interval $(\bar{x} - 2s, \bar{x} + 2s)$);

At least $\frac{8}{9}$ of the observation will fall within 3 standard deviations of the mean (i.e. within the interval $(\bar{x} - 3s, \bar{x} + 3s)$);

In general, at least $1 - \frac{1}{k^2}$ of the observations will fall within $k$ standard deviation of the mean (i.e. within the interval $(\bar{x} - ks, \bar{x} + ks)$) ($k$ is supposed to be greater than 1).
Table 2.8 (page 80) R&D Percentages for 50 companies

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>13.5</td>
<td>9.5</td>
<td>8.2</td>
<td>6.5</td>
<td>8.4</td>
<td>8.1</td>
<td>6.9</td>
<td>7.5</td>
<td>10.5</td>
<td>13.5</td>
</tr>
<tr>
<td>7.2</td>
<td>7.1</td>
<td>9.0</td>
<td>9.9</td>
<td>8.2</td>
<td>13.2</td>
<td>9.2</td>
<td>6.9</td>
<td>9.6</td>
<td>7.7</td>
</tr>
<tr>
<td>9.7</td>
<td>7.5</td>
<td>7.2</td>
<td>5.9</td>
<td>6.6</td>
<td>11.1</td>
<td>8.8</td>
<td>5.2</td>
<td>10.6</td>
<td>8.2</td>
</tr>
<tr>
<td>11.3</td>
<td>5.6</td>
<td>10.1</td>
<td>8.0</td>
<td>8.5</td>
<td>11.7</td>
<td>7.1</td>
<td>7.7</td>
<td>9.4</td>
<td>6.0</td>
</tr>
<tr>
<td>8.0</td>
<td>7.4</td>
<td>10.5</td>
<td>7.8</td>
<td>7.9</td>
<td>6.5</td>
<td>6.9</td>
<td>6.5</td>
<td>6.8</td>
<td>9.5</td>
</tr>
</tbody>
</table>

13.5
9.5
8.2
6.5
8.4
8.1
6.9
7.5
10.5
13.5
7.2
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 11.3 | 5.6 | 10.1 | 8.0 | 8.0 | 8.5 | 11.7 | 7.1 | 7.7 | 9.4 | 6.0 | 6.0 | 8.0 | 8.0 | 7.4 | 10.5 | 7.8 | 7.9 | 6.5 | 6.9 | 6.9 | 6.5 | 6.5 | 6.8 | 6.8 | 9.5 |
\( \bar{x} = 8.49, \ s = 1.98, \ (\bar{x} - 2s, \bar{x} + 2s) \) is the interval within 2s of the mean

The interval is (4.53, 12.45). It contains 47 of the 50 observations, or 94% which is in accordance with the Chebyshev’s rule.

Analogously, the 3-standard-deviation interval around the mean is (2.55, 14.43), and it contains all the observations.
Quiz 1 will be held on Monday, July 13. The following topics will be covered:

1. Histogram and Stem-and-Leaf Display (pp. 51-53), Lower and Upper Quartiles (p.91). Outliers (p.97). Constructions of the histogram and stem-and-leaf (p.53) are typical possible problems.
