1. For the frequency distribution shown to the right, provide these statistics: \( M, s^2, \) and \( s. \)

2. For the following data, which are rank ordered for your convenience, plot a frequency polygon based on equal sized intervals (use between 8 and 12 intervals in all). How would you describe the general shape of this distribution?

3. Dr. Diablo obtains a sample of 10 scores and computes \( s^2, \) which turns out to be equal to zero. Because of this fact, what must be true about the scores in this sample?

4. Below are two frequency histograms, each representing a population of scores. Each bar represents the frequency of a single score value, not an interval of score values. The lowest score is the same in both distributions and the highest score is the same in both distributions. Which one of the statements below is true?
   (a) \( \sigma \) is larger in distribution A
   (b) \( \sigma \) is larger in distribution B, but we cannot tell how much larger
   (c) \( \sigma \) is twice as large in distribution B as in distribution A
   (d) \( \sigma \) is equal for distributions A and B
   (e) one cannot draw a conclusion about the relative size of \( \sigma \) for these distributions

Formulas:
\[
M = \frac{\sum X}{N} \quad s^2 = \frac{\sum (X-M)^2}{N-1} \quad \sigma^2 = \frac{\sum (X-\mu)^2}{N}
\]
6. Suppose you have two distributions of data each consisting of 50 scores. One is a positively skewed distribution, with a minimum score of 10, a mode of 15, and a maximum score of 35. The other distribution is negatively skewed, with a minimum score of 25, a mode of 45, and a maximum score of 50. If the scores in these distributions are combined to form a single distribution, what would be its general shape? Select one option.
(a) uniform  (b) normal  (c) bimodal  (d) positively skewed  (e) negatively skewed

7. Suppose that the distributions in question 6 each had $s^2 = 8.6$. What would be the value of $s^2$ for the distribution that results from combining these two distributions? Select one option.
(a) 8.6  (b) more than 8.6  (c) less than 8.6  (d) cannot tell without more information

8. For the data in the frequency distribution shown to the right, compute the mode, median, and mean.

<table>
<thead>
<tr>
<th>Score</th>
<th>Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
</tr>
</tbody>
</table>

9. Compute $s^2$ and $s$ for the following set of scores.

6 8 8 9 11 12

10. Suppose a sample of 20 scores has $M = 14.0$ and $s = 4.2$. Now suppose that an additional score, equal to 14, is added to the sample. What would be the value of $s$ for this sample of 21 scores? Select one option.
(a) 4.2  (b) less than 4.2  (c) more than 4.2  (d) cannot tell without more information
11. When computing sample variance, the sum of squared deviations from the mean is divided by $N - 1$ instead of $N$. This procedure is followed because
(a) $\mu$ is not known
(b) it ensures that $s^2$ will equal $\sigma^2$
(c) it ensures that $s^2$ will never be less than $\sigma^2$
(d) across many samples, the average value of $s^2$ will be closer to $\sigma^2$ than if $N$ were used
(e) two of the above

12. Use R to generate a random sample of 50 scores (rounded to the nearest whole number) from a normal distribution with $\mu = 30$ and $\sigma = 8$. Generate a frequency histogram with 10 intervals for this sample. Use the information you learned in Assignment 1 to carry out these steps. Compute the mean, variance, and standard deviation for this sample. The R commands for computing these statistics are `mean`, `median`, `var`, and `sd`. Each of these commands will take the name of the data variable holding the sample of scores as an argument:

> mean(X)

where X is the arbitrary name of the data variable (you can use a different name if you like). What would be the median of the original population? Compare how well the sample values for the mean and median approximate the corresponding population values.

Now draw a random sample from the same population, but this time make the sample size 500 instead of 50. Construct the histogram and compute the descriptive statistics as before. Look at how closely the sample results for the mean and median fit the population values in these two cases ($N = 50$ vs. $N = 500$).

13. Use the R commands shown in class and in the lecture notes to draw 100,000 random samples from a normal distribution with $\mu = 100$ and $\sigma = 15$. In the first case, make the samples of size 10 (small sample size). Once the samples have been drawn, compute the mean of all the sample variances to see how well this average conforms to the known population variance of $15^2 = 225$. With so many samples taken from the population, the average value of all those sample variances should come very close to $\sigma^2$. Also produce a histogram with 100 intervals for the set of sample variances. Note how variable are the values in this histogram. Repeat this procedure, but now with sample size set at 200 instead of just 10. Everything else stays the same. Don't forget to create a variable to hold the sample variances and set it equal to null before you start the simulation. What do you notice about the variability of the sample variances in this case (with large sample size), relative to the first case (small sample size)?