You will not receive anywhere near full credit on these problems unless you write down any computations you perform (even if you perform them by calculator). Many calculators have functions to compute various statistical values; answers obtained this way will receive little or no credit. Rather, you must show how to obtain the relevant values without a calculator from the definition. You will not be graded down if you do not simplify your answers algebraically, as long as they are correct.

1. At a pier on the beach, the water level (relative to a given height) was measured at four different times during the day. The levels obtained were: 2 cm, -1 cm, 3 cm, 5 cm. Find the median water level (of these measurements). Also, find the variance in these measurements.

A solution: In ascending order, the data are -1 cm, 2 cm, 3 cm, 5 cm. Since there is an even number of data points, the median is the mean of the middle two, so the median equals \((2+3)/2 = 5/2 = 2.5\). To find the variance, we need first to find the mean, which is by definition \((-1 + 2 + 3 + 5)/4 = 9/4 = 2.25\). Also by definition then, the variance equals 
\[
\frac{(-1 - 2.25)^2 + (2 - 2.25)^2 + (3 - 2.25)^2 + (5 - 2.25)^2}{3}
\]
[If you really want to compute this on your calculator, you will find that it equals 6.25.]

2. In 1798, Henry Cavendish measured the density of the earth several times using an instrument called a torsion balance. The distribution of his measurements is shown in the histogram below.

![Density of the Earth Measurements](image)

Comment on the overall pattern of the distribution (including, as usual, information about shape, center, and spread) and on the presence or lack of outliers.
A solution: As for the shape of the distribution, it is hard to say whether it should be considered more or less symmetric around 5.5 with an outlier around 4.0, or whether it should be considered left skewed. [Either answer would be acceptable.] In any case, it appears to be unimodal, with the single mode being a little above 5 g/cm$^3$, maybe about 5.2 g/cm$^3$. As for center, its median appears to be at about 5.4 g/cm$^3$ or 5.5 g/cm$^3$ (since about half the values seem to fall to the left of there). Its mean would be a bit lower, maybe around 4.8 g/cm$^3$ or so, because of the measurement near 4.0 g/cm$^3$ that shifts the “center of balance” to the left. As for spread, it is mostly clustered between 5.2 g/cm$^3$ and 5.9 g/cm$^3$, but the overall spread is from about 4.0 g/cm$^3$ to about 5.9 g/cm$^3$. While most of the rest of the data seem to fit the overall pattern pretty well, there is a clear outlier at about 4.0 g/cm$^3$.

3. The price of a certain stock was recorded at the beginning of the week for six weeks to be $1.28, $1.32, $1.50, $1.41, $1.43, and $1.62. Find the mean and median of these measurements, as well as their standard deviation.

A solution: The mean of these measurements is, by definition:

\[
(1.28 + 1.32 + 1.50 + 1.41 + 1.43 + 1.62) / 6 \text{ dollars} = 1.43 \text{ dollars.}
\]

To find the median, we first put the data in ascending order: $1.28, \$1.32, \$1.41, \$1.43, \$1.50, \$1.62. Since there are six data points, the median is the mean of the middle two, which is $(\$1.41 + \$1.43)/2 = \$1.42$. The standard deviation is the square root of the variance, so we first compute the variance. By definition, it is

\[
((1.28-1.43)^2 + (1.32 – 1.43)^2 + (1.50 – 1.43)^2 + (1.41 – 1.43)^2 + (1.43 – 1.43)^2 + (1.62 – 1.42)^2) / 5.
\]

The standard deviation is the square root of this number. [If you actually compute them out on your calculator, you should find that the variance is $.0152$, and that the standard deviation is $.123$.]
4. For a collection of crickets, the average number of chirps they produced per second was measured (over a certain time period). The data obtained are shown in the following histogram.

![Cricket Chirping](image)

Comment on the pattern of this distribution (including, as usual, shape, center, and spread) as well as the existence or lack of outliers.

*A solution*: The shape of this distribution is pretty hard to describe. Depending on what you make of the downturn around 16.5 chirps/s, you might describe it either as being unimodal with a mode at about 17.0 or 17.5 chirps/s (thinking that the little dip is not particularly significant) or maybe bimodal with modes at about 15.0 and 17.5 chirps/s. It is certainly not symmetric, and it is not skewed left. It might be viewed as being skewed right if we take the little upturn at about 17.5 chirps/s as not being significant. As for the center, it looks like there are about 13 data points, with the middle point (hence the median) being between about 15.5 and 16 chirps/s. The mean (or “balance point”) is probably a little further to the right, maybe around 16.5 chirps/s, since the data are spread more widely in that direction. As for the spread, the data are mostly concentrated between about 14 and 17.5 chirps/s, but the overall spread of the data is from about 14.0 chirps/s to about 20.0 chirps/s.

5. A student measures the heights of 9 of her friends (in inches) to be: 61, 72, 68, 70, 73, 64, 64, and 74. Give the five-number summary of her data.

*A solution*: For the five-number summary, we need the minimum, maximum, median, and first and third quartiles. Putting the data in ascending order helps: 61, 64, 64, 68, 70, 72, 73, 74. The minimum then is 61, and the maximum 74. Since there is an even number of data points, the median is the mean of the middle two, which is $(68+70)/2 = 69$. By our definition, the first quartile is the median of the data to the left of the median, which is the median of the data 61, 64, 64, 68. This again has an even number of values, so the median is the mean of the middle two, which is $(64 + 64)/2 = 64$. For the third quartile, it is by definition the median of the data to the right of the median, so the median of the data 70, 72, 73, 74. This has an even number of values, so the median is the mean of the middle two, namely $(72+73)/2 = 72.5$. The five-number summary then is:
6. During November 1998, the lengths and widths of 88 Puget Sound Butter Clams were collected from the beach at Alki Point in west Seattle. The distribution of the ratio of length to width of these clams is shown in the histogram below.

Comment on the pattern of the distribution (shape, center, and spread, as usual) and the existence or lack of outliers.

A solution: I would describe the shape as being symmetric with a single outlier at around 1.65 to 1.7. However, it could be argued that the shape is a bit right skewed. [Either answer would be acceptable.] In any case though, the distribution is unimodal with its single peak around 1.3 to 1.35. As for center, the median appears to be somewhere between 1.25 and 1.3, with about half the values lying to the left of there. The mean is a little higher, perhaps around 1.35, because the value around 1.7 shifts the “center of balance” to the right. As for spread, most of the data are clustered between about 1.15 and 1.45, but the overall range is from about 1.15 to about 1.7. The value around 1.7 is the sole outlier.

7. Two people conduct phone surveys asking how many oranges people have in their refrigerators at present. The first person obtains data of 0, 0, 1, 4, and 5, while the second person obtains data of 0, 0, 0, 2, and 8. Which data set has the larger variance?

A solution: To find the two variances, we first must find the two means. By definition, they are \((0 + 0 + 1 + 4 + 5) / 5 = 2\) and \((0 + 0 + 0 + 2 + 8)/5 = 2\). By definition then, the two variances are:

\[
\frac{((0 - 2)^2 + (0 - 2)^2 + (1 - 2)^2 + (4 - 2)^2 + (5 - 2)^2)}{4} \quad \text{and} \quad \frac{((0 - 2)^2 + (0 - 2)^2 + (0 - 2)^2 + (2 - 2)^2 + (8 - 2)^2)}{4}.
\]
Computing these explicitly (by hand or by calculator) gives variances of $22/4 = 5.5$ and $48/4 = 12$, so the second data set has the larger variance.

8. In 1999, a group from the University of Washington collected data on the temperatures of Washington streams. The distribution of their data is shown in the histogram below.

Comment on the pattern of the distribution (meaning, as usual, shape, center, and spread) and the presence or lack of outliers.

A solution: This distribution appears to be somewhat right skewed, since the “tail” swings wide in that direction but not the other. The distribution is unimodal, with its single mode being at about 13.5 or 14.0 degrees C. The median appears to be around 16.0 to 16.5 degrees C, since about half the values appear to be to the left of there. The mean is a little further to the right, because all the data points far to the right shift the “center of balance” that direction. Most of the data are between about 11.5 and 21.5 degrees C, but the overall range is from about 9.5 to 27.5 degrees C. There are no clear outliers.