Measure of Spread

• Several different measures to describe the variability in the data

• The range of the data is simply the difference between the largest and the smallest observation.
Sample Variance and Standard Deviation

- The range defined above can be very sensitive to outliers and does not capture the “average” deviation of the data from the mean.
- For every data point one can define the deviation of that data point from the mean as:

\[ d_i = (x_i - \bar{x}) \]
Sample Variance and Standard Deviation

• The average deviation for any data is always 0.
• Based on the deviations the variance is defined as:

\[ s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \]
Sample Variance and Standard Deviation

- The sample Standard Deviation is defined as the square root of the Variance:

\[ s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}} \]
Properties of Variance and Standard Deviation

• Variance is translation invariant. If:

\[ y_i = x_i + c, \quad i = 1, \ldots, n \]

\[ s^2_x = s^2_y \]
Properties of Variance and Standard Deviation

• If the data is scaled by a constant that is:

\[ y_i = cx_i, \quad i = 1, \ldots, n \]

Then

\[ s^2_y = c^2 s^2_x \quad \quad S_y = c S_x \]
Coefficient of Variation

• Coefficient of variation is defined as % ratio of standard deviation to the mean:

$$cv = 100 \times \frac{S_x}{\bar{x}}$$
Grouped Data

• Rather than using the raw data sometimes it convenient to group the data in to separate intervals.
• Each group intervals as an associated midpoint $z_i$ and frequency $f_i$.
• Using this grouped data we can define the Grouped means and variance:
Grouped Mean

- The Grouped mean is defined as:

\[ \bar{x}_g = \frac{\sum_{i=1}^{k} f_i z_i}{\sum_{i=1}^{k} f_i} \]
Grouped Variance

- The Grouped Variance is defined as:

\[ s_g^2 = \frac{\sum_{i=1}^{k} f_i (z_i - \bar{x}_g)^2}{\sum_{i=1}^{k} f_i - 1} \]

\[ s_g^2 = \frac{\sum_{i=1}^{k} f_i (z_i - \bar{x}_g)^2}{n - 1} \]