AS Statistics – Chapter 2 – Measures of Location and Spread

Measures of Location

1. The **mode** or **modal class** is the value or class that occurs most often.

2. The **median** is the middle value when all the data values are put in order.

3. The **mean** is $\frac{sum\ of\ data\ values}{number\ of\ data\ values}$ can be calculated using the formula $\bar{x} = \frac{\Sigma x}{n}$.

4. When choosing the best average to use, bear the following in mind:
   Mode: Used for qualitative data, or quantitative data with a single mode or two modes (bimodal).
   Median: Used for quantitative data. Usually used when there are extreme values, as they do not affect it.
   Mean: Used for quantitative data. Uses all the data so gives a true measure. Affected by extreme values.

5. For data given in a frequency table, the mean can be calculated using the formula $\bar{x} = \frac{\Sigma fx}{\Sigma f}$.

6. If you're given a list of data, remember these definitions for the **quartiles**:
   The **lower quartile** is the median of the list of all values *below the median*
   The **upper quartile** is the median of the list of all values *above the median*

   These shortcuts can also help, but you'll need to remember them very carefully:

   To find the **lower quartile** for discrete data, divide n by 4.
   If this is a whole number, the lower quartile is halfway between this data point and the one above.
   If it is not a whole number, round *up* and pick this data point.

   To find the **upper quartile** for discrete data, find $\frac{3}{4}$ of n.
   If this is a whole number, the lower quartile is halfway between this data point and the one above.
   If it is not a whole number, round *up* and pick this data point.

7. For grouped data, use **interpolation** to estimate the median, quartiles and percentiles.
   When using interpolation, you are assuming the data is **evenly distributed** within each class.

   For the quartiles:

   $$Q_1 = \frac{n}{4}th \text{ data value} \qquad Q_2 \text{ (median)} = \frac{n}{2}th \text{ data value} \qquad Q_3 = \frac{3n}{4}th \text{ data value}$$

   For the percentiles:

   $$\text{1st percentile } = \frac{n}{100}th \text{ data value}, \qquad \text{2nd percentile } = \frac{2n}{100}th \text{ data value}, \qquad etc \ldots$$

Measures of Spread

8. The **interquartile range (IQR)** is the difference between the upper and lower quartiles, $Q_3 - Q_1$.

9. The **interpercentile range** is the difference between the values for two given percentiles.

10. **Variance** $\sigma^2 = \frac{\Sigma(x-\bar{x})^2}{n} = \frac{\Sigma x^2 - n\bar{x}^2}{n} = \frac{S_{xx}}{n}$ where $S_{xx} = \Sigma(x-\bar{x})^2 = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$

11. **Standard deviation** $\sigma = \sqrt{\text{Variance}} = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \sqrt{\frac{\Sigma x^2 - n\bar{x}^2}{n}} = \sqrt{\frac{S_{xx}}{n}}$

12. You can use these versions of the above formulae for grouped data presented in a frequency table:

$$\sigma^2 = \frac{\Sigma f(x-\bar{x})^2}{\Sigma f} = \frac{\Sigma f x^2 - n\bar{x}^2}{n} \qquad \sigma = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f}} = \sqrt{\frac{\Sigma f x^2 - n\bar{x}^2}{n}}$$

Coding

13. If data is coded using the formula $y = \frac{x-a}{b}$,

- The mean of the coded data is $\bar{y} = \frac{\bar{x}-a}{b}$
- The standard deviation of the coded data is given by $\sigma_y = \frac{\sigma_x}{b}$

Notice that subtraction (or addition) only affects the mean, but division (or multiplication) affects both.

This is also true for data coded using $y = a + bx \rightarrow \bar{y} = a + b\bar{x}$ and $\sigma_y = b\sigma_x$