Measures of Location and Spread Cheat Sheet

Stats/Mech Year 1

Measures of central tendency

A measure of central tendency describes the centre of the data. You need to decide of the best measure to use in particular situations.

The mode or modal class is the value of class which occurs most often. This is used when data is qualitative or quantitative with one mode or two modes (bimodal). It is not informative if each value only occurs once.

The median is the middle value when the data values are put in order. This is used for quantitative data and usually used when there are extreme values as they are unaffected.

The mean can be calculated using:

$$\bar{x} = \frac{\Sigma x}{n}$$

Where $\bar{x}$ (x bar) is the mean,
Σx is the sum of the data values,
n is the number of data values

For data given in a cumulative frequency table, the mean can be calculated using:

$$\bar{x} = \frac{\Sigma xf}{\Sigma f}$$

Where Σfx is the sum of the products of the data values and their frequencies,
Σf is the sum of frequencies

The mean is used for quantitative data. It uses all values in the data therefore it gives a true measure of data. However, it is affected by extreme values.

You can calculate the mean, class containing median and modal class for continuous data presented in a grouped frequency table by finding the midpoint of each class interval.

Other measures of location

The median(Q_2) splits the data into two equal halves (50%).
The lower quartile (Q_1) is one quarter of the way through the dataset.
The upper quartile (Q_3) is three quarters of the way through the dataset.

Percentiles split the data set into 100 parts. The 10th percentile is one-tenth of the way through the data, for example. 10% of data values are less than the 10th percentile and 90% are greater.

To find lower and upper quartiles for discrete data:

1. Divide n by 4. (lower quartile) OR
   Find $\frac{3}{4}$ of n. (upper quartile)
2. If this is a whole number, the lower or upper quartile is the midpoint between this data point and the number above. If it is not, round up and pick this number.

When data is presented in a grouped frequency table, you can use interpolation to estimate the medians, quartiles, and percentiles. This method assumes that the data values are evenly distributed within each class.

$Q_1 = \frac{n}{4}$th data value
$Q_2 = \frac{n}{2}$th data value
$Q_3 = \frac{3n}{4}$th data value

Measures of spread

Measures of spread shows how spread out the data is. They are also known as measures of dispersion or measures of variation.

- Range
  The difference between largest and smallest values in the dataset.
- Interquartile range (IQR)
  The difference between upper and lower quartile.
- Interpercentile range
  Difference between the values of two given percentiles.

Variance (σ^2) and standard deviation (σ)

The variance also shows how spread out the data is. There are 3 versions of the formulae used to find variance:

1. $\sigma^2 = \frac{\Sigma(x-\bar{x})^2}{n}$
2. $\sigma^2 = \frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2$ — Easier to use when given raw data
   You can remember this as "mean of squares minus square of means"
3. $\sigma^2 = \frac{S_{xx}}{n}$
   Where $S_{xx} = \Sigma(x-\bar{x})^2 = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$ — S_{xx} is a summary statistic used to simplify formula
   You can use calculator to calculate S_{xx}

Standard deviation is the square root of variance.

$$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2} = \sqrt{\frac{S_{xx}}{n}}$$

For grouped data presented in frequency table:

$$\sigma^2 = \frac{\Sigma f(x-\bar{x})^2}{\Sigma f} = \frac{\Sigma fx^2}{\Sigma f} - \left(\frac{\Sigma fx}{\Sigma f}\right)^2$$

$$\sigma = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f}} = \sqrt{\frac{\Sigma fx^2}{\Sigma f} - \left(\frac{\Sigma fx}{\Sigma f}\right)^2}$$

Where f is the frequency of each group and Σf is the total frequency

Example 1: Shamsa records the time spent out of school during lunch hour to the nearest minute, x, of the students in her year in the table below. Calculate the standard deviation.

Time spent out of school (min)	35	36	37	38
Frequency	3	17	29	34

1. Find Σfx^2, Σfx and Σf
   $\Sigma fx^2 = 3 \times 35^2 + 17 \times 36^2 + 29 \times 37^2 + 34 \times 38^2$
   $= 114504$
   $\Sigma fx = 3 \times 35 + 17 \times 36 + 29 \times 37 + 34 \times 38$
   $= 3082$
   $\Sigma f = 3 + 17 + 29 + 34 = 83$
2. Use formula for grouped data in frequency table to find variance:
   $\sigma^2 = \frac{114504}{83} - \left(\frac{3082}{83}\right)^2 = 0.74147\ldots$
3. Square root variance to find standard deviation:
   $\sigma = \sqrt{0.74147\ldots} = 0.861$ (3s.f.)

Coding

Each value in the data can be coded to give a new set of values, which is easier to work with. Coding also changes different statistics in different ways.

If data is coded using the formula $y = \frac{x-a}{b}$, where a and b are constants that you have to choose or given in the question:

- Mean of coded data: $\bar{y} = \frac{\bar{x}-a}{b}$
  Rearrange the formula to find original mean: $\bar{x} = b\bar{y} + a$
- Standard deviation of coded data: $\sigma_y = \frac{\sigma_x}{b}$
  Rearrange the formula to find original standard deviation: $\sigma_x = b\sigma_y$

Example 2: A scientist measures the temperature, x°C, at five different points of a nuclear reactor. Her results are given below:
332°C, 355°C, 306°C, 317°C, 340°C

a. Use the coding $y = \frac{x-300}{10}$ to code this data.

Substitute each value into x to get coded data, y.

Original data, x	332	355	306	317	340
Coded data, y	3.2	5.5	0.6	1.7	4.0

b. Calculate the mean and standard deviation of the coded data.
$\Sigma y = 15$, $\Sigma y^2 = 59.74$
$\bar{y} = \frac{15}{5} = 3$
$\sigma_y^2 = \frac{59.74}{5} - \left(\frac{15}{5}\right)^2 = 2.948$
$\sigma_y = \sqrt{2.948} = 1.72$ (3s.f.)

c. Calculate the mean and variance of the original data using your answers from part b.
$3 = \frac{\bar{x}-300}{10}$ so $\bar{x} = 330$°C
$1.72 = \frac{\sigma_x}{10}$ so $\sigma_x = 17.2$°C (3s.f.)