# *Variance*

Let $X$ be a **discrete random variable**, then the variance of $X$ is;

$$\text{Var}(X) = \sum (x-\mu)^2\, p(x)$$

where $p(x) = \text{P}(X = x) \geq 0$

$$\mu = \text{E}(X)$$

***Note:*** $x-\mu$ is the **deviation** of $x$ from $\mu$

Var($X$) is a measure of spread

e.g. Using the data of the milk marketing survey, find the variance of the number of litres of milk consumed in a week by a family.

| $x$ | **0** | **1** | **2** | **3** | **4** | **5** | **Σ** |
|---|---|---|---|---|---|---|---|
| $p(x)$ | $\frac{2}{25}$ | $\frac{5}{25}$ | $\frac{9}{25}$ | $\frac{5}{25}$ | $\frac{3}{25}$ | $\frac{1}{25}$ | 1 |
| $xp(x)$ | 0 | $\frac{5}{25}$ | $\frac{18}{25}$ | $\frac{15}{25}$ | $\frac{12}{25}$ | $\frac{5}{25}$ | 2.2 |
| $(x-\mu)^2$ | $\frac{121}{25}$ | $\frac{36}{25}$ | $\frac{1}{25}$ | $\frac{16}{25}$ | $\frac{81}{25}$ | $\frac{196}{25}$ | — |
| $(x-\mu)^2 p(x)$ | $\frac{242}{625}$ | $\frac{180}{625}$ | $\frac{9}{625}$ | $\frac{80}{625}$ | $\frac{243}{625}$ | $\frac{196}{625}$ | 1.52 |

$\mu$

Var($X$)
$= \sum (x-\mu)^2 p(x)$
$= 1.52$

***alternatively***

$$\mathrm{Var}(X) = \sum (x-\mu)^2\, p(x)$$
$$= \mathrm{E}\left[(X-\mu)^2\right]$$
$$= \mathrm{E}\left(X^2 - 2\mu X + \mu^2\right)$$
$$= \mathrm{E}\left(X^2\right) - 2\mu \mathrm{E}(X) + \mu^2$$
$$= \mathrm{E}\left(X^2\right) - 2\mu \times \mu + \mu^2$$

$$\mathrm{Var}(X) = \mathrm{E}\left(X^2\right) - \mu^2$$
$$\text{where } \mathrm{E}\left(X^2\right) = \sum x^2 p(x)$$

| $x$ | **0** | **1** | **2** | **3** | **4** | **5** | **Σ** |
|---|---|---|---|---|---|---|---|
| $p(x)$ | $\frac{2}{25}$ | $\frac{5}{25}$ | $\frac{9}{25}$ | $\frac{5}{25}$ | $\frac{3}{25}$ | $\frac{1}{25}$ | 1 |
| $xp(x)$ | 0 | $\frac{5}{25}$ | $\frac{18}{25}$ | $\frac{15}{25}$ | $\frac{12}{25}$ | $\frac{5}{25}$ | 2.2 ← $\mu$ |
| $x^2p(x)$ | 0 | $\frac{5}{25}$ | $\frac{36}{25}$ | $\frac{45}{25}$ | $\frac{48}{25}$ | $\frac{25}{25}$ | 6.36 ← $\mathrm{E}(X^2)$ |

$$\mathrm{Var}(X) = \mathrm{E}\left(X^2\right) - \mu^2$$
$$= 6.36 - (2.2)^2 = \underline{1.52}$$

# *Standard Deviation (σ)*

If the random variable, $X$, is measured in units
then variance $(\mathrm{E}(X^2) - \mu^2)$ is measured in units$^2$.

Standard deviation is used to measure spread using the same units as the random variable.

$$\sigma = \sqrt{\mathrm{Var}(X)}$$

***or***

$$\sigma^2 = \mathrm{Var}(X)$$

e.g. $\sigma = \sqrt{1.52}$

$= 1.23288...$

$= \underline{1.23}$ (to 2 dp)

# *Standardising Data*

Data standardisation is the process of making sure your data set can be compared to other data sets

**Original data**

$$\mathrm{E}(X) = \mu \quad \text{and} \quad \mathrm{Var}(X) = \sigma^2$$

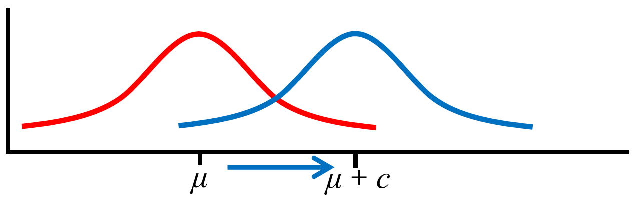
**adding a constant amount to each piece of data**

$$Y = X + c$$

$$\begin{aligned}\mathrm{E}(Y) &= \mathrm{E}(X + c)\\ &= \mathrm{E}(X) + \mathrm{E}(c)\\ &= \mu + c\end{aligned}$$

$$\begin{aligned}\mathrm{Var}(Y) &= \mathrm{Var}(X + c)\\ &= \mathrm{E}\left[(X + c)^2\right] - (\mu + c)^2\\ &= \mathrm{E}\left(X^2 + 2cX + c^2\right) - \left(\mu^2 - 2c\mu + c^2\right)\\ &= \mathrm{E}\left(X^2\right) + 2c\mathrm{E}(X) + c^2 - \mu^2 + 2c\mu - c^2\\ &= \mathrm{E}\left(X^2\right) + 2c\mu + c^2 - \mu^2 + 2c\mu - c^2\\ &= \mathrm{E}\left(X^2\right) - \mu^2\\ &= \mathrm{Var}(X) = \sigma^2\end{aligned}$$

simply adding a constant amount does not change the shape of the data, it only translates it *c* units

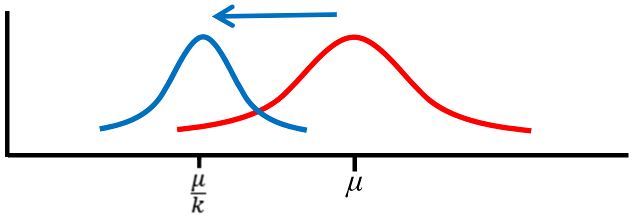


**dividing each piece of data by a constant amount**

$$Z = \frac{X}{k}$$

$$\mathrm{E}(Z) = \mathrm{E}\left(\frac{X}{k}\right) = \frac{\mathrm{E}(X)}{k} = \frac{\mu}{k}$$

dividing by a constant amount changes the shape of the data, as well as altering the mean




$$\mathrm{Var}(Z) = \mathrm{Var}\left(\frac{X}{k}\right) = \mathrm{E}\left[\left(\frac{X}{k}\right)^2\right] - \left(\frac{\mu}{k}\right)^2 = \frac{\mathrm{E}(X^2)}{k^2} - \frac{\mu^2}{k^2} = \frac{\mathrm{E}(X^2) - \mu^2}{k^2} = \frac{\mathrm{Var}(X)}{k^2} = \frac{\sigma^2}{k^2} = \left(\frac{\sigma}{k}\right)^2$$

**z-scores**

A common standardisation is to "normalise" the data by creating a z-score

$$z = \frac{x-\mu}{\sigma}$$

$$\mathrm{E}(Z) = \mathrm{E}\left(\frac{X-\mu}{\sigma}\right)$$

$$= \frac{\mathrm{E}(X-\mu)}{\sigma}$$

$$= \frac{\mu-\mu}{\sigma}$$

$$= 0$$

$$\mathrm{Var}(Z) = \mathrm{Var}\left(\frac{X-\mu}{\sigma}\right)$$

$$= \frac{\mathrm{Var}(X-\mu)}{\sigma^2}$$

$$= \frac{\sigma^2}{\sigma^2} = 1$$

$$\therefore \mu_Z = 0 \quad \text{and} \quad \sigma_Z = 1$$

**Exercise 13C; 1, 2, 3bd, 4, 6, 8, 9, 10, 11, 12abc, 13**