Data Science Math Skills
Paul Bendich and Daniel Egger
Duke University

Sigma Notation: Mean and Variance

Video companion

1 Introduction

Important equations for this video:

$$
\begin{aligned}
X &= \{x_1, ..., x_n\} \\
\mu_x &= \frac{1}{n}\sum_{i=1}^{n} x_i \\
{\sigma_x}^2 &= \frac{1}{n}\left[\sum_{i=1}^{n} (x_i - \mu_x)^2\right]
\end{aligned}
$$

The symbol μ_x is the "mean of x," and ${\sigma_x}^2$ is the "variance of x." The standard deviation is denoted σ_x.

2 Mean

Example:

$$
\begin{aligned}
Z &= \{1, 5, 12\} \\
|Z| &= 3 \\
\mu_z &= \frac{1+5+12}{3} = \frac{18}{3} = 6
\end{aligned}
$$

The mean μ_z is also denoted $\mu(z)$ or simply μ.

Symbolic example:

$$
\begin{aligned}
Y &= \{y_1, y_2, y_3, y_4\} \\
\mu_y &= \frac{1}{4}(y_1 + y_2 + y_3 + y_4) \\
&= \frac{1}{4}\left(\sum_{i=1}^{4} y_i\right)
\end{aligned}
$$

In general, suppose you have a set

$$X = \{x_1, x_2, ..., x_n\},$$

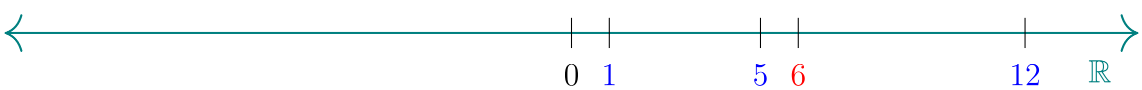
then the mean of X is

$$\mu_x = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right).$$

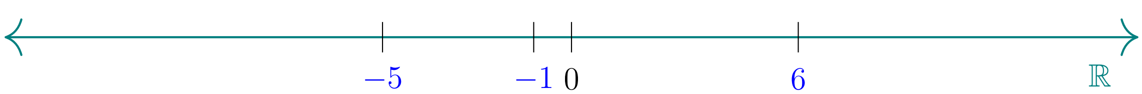
The variable i is a counter. The variable n is a number, which tells you when to stop counting.

3 Mean centering

$$Z = \{1, 5, 12\}$$
$$\mu_z = 6$$



$$\begin{aligned} Z' &= \{1-6, 5-6, 12-6\} \\ &= \{-5, -1, 6\} \\ \mu_{z'} &= 0 \end{aligned}$$

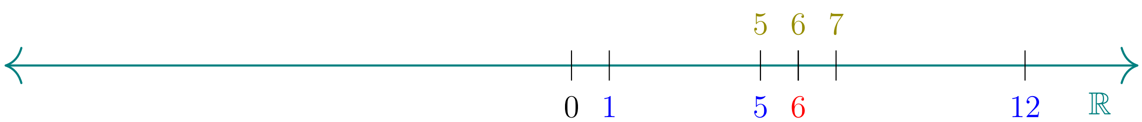


Mean centering data produces a new data set, which has the same relationships, but the mean is zero.

4 Variance

$$Z = \{1, 5, 12\}$$
$$\mu_z = 6$$

$$W = \{5, 6, 7\}$$
$$\mu_w = 6$$



Set Z (blue) is more "spread out" than set W (olive).

If $X = \{x_1, ..., x_n\}$, the variance of X is

$$\sigma_x{}^2 = \frac{1}{n}\left[\sum_{i=1}^{n}(x_i - \mu_x)^2\right].$$

The standard deviation is given by

$$\sigma_x = \sqrt{\sigma_x{}^2}.$$

Z and W have the same mean, but Z is more spread out, so σ_z should be greater than σ_w.

$$\begin{aligned}
\sigma_w^2 &= \frac{1}{3}\left[\sum_{i=1}^{3}(w_i - \mu_w)^2\right] \\
&= \frac{1}{3}\left[(5-6)^2 + (6-6)^2 + (7-6)^2\right] \\
&= \frac{1}{3}\left[(-1)^2 + 0^2 + 1^2\right] \\
&= \frac{2}{3} \\
\sigma_w &= \sqrt{\frac{2}{3}}
\end{aligned}$$

$$
\begin{aligned}
\sigma_z^2 &= \frac{1}{3}\left[\sum_{i=1}^{3}(z_i-\mu_z)^2\right] \\
&= \frac{1}{3}\left[(1-6)^2+(5-6)^2+(12-6)^2\right] \\
&= \frac{1}{3}\left[(-5)^2+(-1)^2+6^2\right] \\
&= \frac{62}{3} \\
\sigma_w &= \sqrt{\frac{62}{3}}
\end{aligned}
$$

$\sigma_z^2 \gg \sigma_w^2$, so Z is much more spread out than W.