VARIANCE AND STANDARD DEVIATION

The **variance** or **standard deviation** of the data values give a measure of the spread of the data.

If a discrete random variable X has k possible values $x_1, x_2, x_3, ..., x_k$
with probabilities $p_1, p_2, p_3, ..., p_k$

then:
- the **expected value** or **mean** is $\mathbf{E(X) = \mu = \sum x_i p_i}$
- the **variance** is $\mathbf{Var(X) = \sigma^2 = \sum (x_i - \mu)^2 p_i = E[(X - \mu)^2]}$
- the **standard deviation** is $\boldsymbol{\sigma = \sqrt{\sum (x_i - \mu)^2 p_i}}$.

Find the standard deviation for this distribution.

x_i	1	2	3	4	5
p_i	0.23	0.38	0.21	0.13	0.05

The mean $\mu = \sum x_i p_i = 2.39$

The standard deviation

$$\begin{aligned}\sigma &= \sqrt{\sum (x_i - \mu)^2 p_i} \\ &= \sqrt{(1 - 2.39)^2 \times 0.23 + (2 - 2.39)^2 \times 0.38 + ... + (5 - 2.39)^2 \times 0.05} \\ &\approx 1.12\end{aligned}$$

An alternative formula for the standard deviation is $\boxed{\boldsymbol{\sigma = \sqrt{\sum x_i{}^2 p_i - \mu^2}.}}$

For example, for the roll of a die:

$$\mu = \sum x_i p_i = 1(\tfrac{1}{6}) + 2(\tfrac{1}{6}) + 3(\tfrac{1}{6}) + 4(\tfrac{1}{6}) + 5(\tfrac{1}{6}) + 6(\tfrac{1}{6}) = 3.5$$

and $$\sigma = \sqrt{\sum x_i{}^2 p_i - \mu^2} = \sqrt{1^2(\tfrac{1}{6}) + 2^2(\tfrac{1}{6}) + 3^2(\tfrac{1}{6}) + 4^2(\tfrac{1}{6}) + 5^2(\tfrac{1}{6}) + 6^2(\tfrac{1}{6}) - (3.5)^2}$$

$$\approx 1.708$$

Use $\sigma^2 = \sum(x_i - \mu)^2 p_i$ to show that $\sigma^2 = \sum x_i^2 p_i - \mu^2$.

$$\begin{aligned}
\sigma^2 &= \sum(x_i - \mu)^2 p_i \\
&= (x_1 - \mu)^2 p_1 + (x_2 - \mu)^2 p_2 + \ldots\ldots + (x_n - \mu)^2 p_n \\
&= (x_1^2 - 2x_1\mu + \mu^2)p_1 + (x_2^2 - 2x_2\mu + \mu^2)p_2 + \ldots\ldots + (x_n^2 - 2x_n\mu + \mu^2)p_n \\
&= (x_1^2 p_1 + x_2^2 p_2 + x_3^2 p_3 + \ldots\ldots + x_n^2 p_n) - 2\mu(x_1 p_1 + x_2 p_2 + \ldots\ldots + x_n p_n) \\
&\quad + \mu^2(p_1 + p_2 + p_3 + \ldots\ldots + p_n)
\end{aligned}$$

Now $p_1 + p_2 + \ldots\ldots + p_n = 1$

$$\begin{aligned}
\therefore \quad \sigma^2 &= \sum x_i^2 p_i - 2\mu\left(\sum x_i p_i\right) + \mu^2(1) \\
&= \sum x_i^2 p_i - 2\mu(\mu) + \mu^2 \qquad \left\{\text{since } \sum x_i p_i = \mu\right\} \\
&= \sum x_i^2 p_i - \mu^2 \quad \text{as required}
\end{aligned}$$

Hence $E\left(\left[X - \mu\right]^2\right) = E\left(X^2\right) - \mu^2$.

The previous slide showed that $Var(X) = E[(X-\mu)^2] = E(X^2) - \mu^2$

This is the form the Nelson text book uses. Either form can be used.

A die is rolled and X = the square of the score. Calculate the mean, variance and standard deviation of X.

x	1	4	9	16	25	36
$P(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$E(X) = 1 \times \frac{1}{6} + 4 \times \frac{1}{6} + 9 \times \frac{1}{6} + 16 \times \frac{1}{6} + \cdots$$

$$\mu = 15\frac{1}{6}$$

$$E(X^2) = 1 \times \frac{1}{6} + 16 \times \frac{1}{6} + 81 \times \frac{1}{6} + \cdots$$

$$= 379\,\frac{1}{6}$$

$$Var(X) = E(X^2) - \mu^2$$

$$= \frac{2275}{6} - \left(\frac{91}{6}\right)^2$$

$$= \frac{5369}{36}$$

$$\approx 149.139$$

$$\sigma = \sqrt{\frac{5369}{36}}$$

$$\approx 12.212$$

The mean is $15\,\frac{1}{6}$, the variance is about 149.1 and the standard deviation is about 12.2.

The number of coins given in change to customers at a fast food restaurant are recorded below.

Coin	5c	10c	20c	50c	$1	$2
Number	143	151	215	52	141	98

a Estimate the probability distribution of X, the value of coins given in change.

b Estimate the expected value, variance and standard deviation of X.

x	0.05	0.1	0.2	0.5	1	2
$P(x)$	0.1788	0.1888	0.2688	0.0650	0.1763	0.1225

$$E(X) \approx 0.05 \times 0.1778 + 0.1 \times 0.1888 + \cdots$$

$$\mu \approx 0.5353$$

$$SD(X) = \sqrt{Var(X)}$$

$$\approx 0.6395$$

$$E(X^2) \approx (0.05)^2 \times 0.1788 + (0.1)^2 \times 0.1888 + \cdots$$

$$\approx 0.6956$$

$$Var(X) = E(X^2) - \mu^2$$

$$= 0.6956 - (0.5353)^2$$

$$\approx 0.4090$$

The expected value is about 53 cents, the variance is about 0.41 and the standard deviation is about 64 cents.

How to use a calculator to determine the mean and variance of a discrete probability functions.
https://www.youtube.com/watch?v=lpGxDQZuSa4

1 For each probability distribution, find:

i the mean μ **ii** the variance σ^2 **iii** the standard deviation σ.

a

x	1	2	3
$P(X = x)$	0.3	0.4	0.3

b

x	0	1	2	3
$P(X = x)$	0.2	0.4	0.1	0.3

Check your answers using technology.

1 **a** **i** $\mu = \sum x_i p_i = 1(0.3) + 2(0.4) + 3(0.3) = 2$

ii
$$\sigma^2 = \sum (x_i - \mu)^2 p_i$$
$$= (1-2)^2 \times 0.3 + (2-2)^2 \times 0.4 + (3-2)^2 \times 0.3$$
$$= 0.6$$

iii
$$\sigma = \sqrt{\sum (x_i - \mu)^2 p_i}$$
$$= \sqrt{0.6}$$
$$\approx 0.775$$

b **i** $\mu = \sum x_i p_i = 0(0.2) + 1(0.4) + 2(0.1) + 3(0.3) = 1.5$

ii
$$\sigma^2 = \sum (x_i - \mu)^2 p_i$$
$$= (0-1.5)^2 \times 0.2 + (1-1.5)^2 \times 0.4 + (2-1.5)^2 \times 0.1 + (3-1.5)^2 \times 0.3$$
$$= 1.25$$

iii
$$\sigma = \sqrt{\sum (x_i - \mu)^2 p_i}$$
$$= \sqrt{1.25}$$
$$\approx 1.12$$

2 Consider the probability distribution alongside.

a Find the value of k.

b Find the mode of the distribution.

c Find the mean μ.

d Find the standard deviation σ.

x_i	2	4	10	20
p_i	k	0.05	0.35	$3k$

2 **a** Since this is a probability distribution, $\sum_{i=1}^{4} p_i = 1$

$$\therefore \quad k + 0.05 + 0.35 + 3k = 1$$
$$\therefore \quad 4k + 0.4 = 1$$
$$\therefore \quad 4k = 0.6$$
$$\therefore \quad k = 0.15$$

b $3k = 0.45$

$\therefore$ the value 20 has the highest probability of occurring, so this is the mode of the distribution.

c
$$\mu = \sum x_i p_i$$
$$= 2(0.15) + 4(0.05) + 10(0.35) + 20(0.45)$$
$$= 13$$

d
$$\sigma = \sqrt{\sum (x_i - \mu)^2 p_i}$$
$$= \sqrt{(2-13)^2 \times 0.15 + (4-13)^2 \times 0.05 + (10-13)^2 \times 0.35 + (20-13)^2 \times 0.45}$$
$$= \sqrt{47.4}$$
$$\approx 6.88$$

4 A country exports crayfish to overseas markets. The buyers are prepared to pay high prices when the crayfish arrive still alive.
Let X be the number of deaths per dozen crayfish. The probability distribution for X is given by:

x	0	1	2	3	4	5	> 5
$\mathrm{P}(X = x)$	0.54	0.26	0.15	k	0.01	0.01	0.00

a) Find k
b) Over a long period, what is the mean number of deaths per dozen crayfish?
c) Find the standard deviation for the distribution.

a Since this is a probability distribution, $\sum p_i = 1$

$$\therefore \quad 0.54 + 0.26 + 0.15 + k + 0.01 + 0.01 = 1$$
$$\therefore \quad k + 0.97 = 1$$
$$\therefore \quad k = 0.03$$

b
$$\begin{aligned}\mu &= \sum x_i p_i \\ &= 0(0.54) + 1(0.26) + 2(0.15) + 3(0.03) + 4(0.01) + 5(0.01) \\ &= 0.26 + 0.30 + 0.09 + 0.04 + 0.05 \\ &= 0.74\end{aligned}$$

So, over a long period the mean number of deaths per dozen crayfish is 0.74.

c

$$\begin{aligned}\sigma &= \sqrt{\sum (x_i - \mu)^2 p_i} \\ &= \sqrt{(0-0.74)^2 \times 0.54 + (1-0.74)^2 \times 0.26 + (2-0.74)^2 \times 0.15 + \ldots . + (5-0.74)^2 \times 0.01} \\ &\approx 0.996\end{aligned}$$

5 A die is numbered 1, 1, 2, 3, 3, 3. Let X be the result when the die is rolled once.

a Construct the probability distribution for X.

b Find the mean μ.

c Find the standard deviation for the distribution.

5 **a**

x	1	2	3
$\mathrm{P}(X = x)$	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{3}{6}$

b
$$\begin{aligned}\mu &= \sum x_i p_i \\ &= 1\left(\tfrac{2}{6}\right) + 2\left(\tfrac{1}{6}\right) + 3\left(\tfrac{3}{6}\right) \\ &= \tfrac{2}{6} + \tfrac{2}{6} + \tfrac{9}{6} \\ &= \tfrac{13}{6} \\ &\approx 2.17\end{aligned}$$

c $\sigma = \sqrt{\sum (x_i - \mu)^2 p_i}$

6 A random variable X has the probability distribution function $P(x) = \frac{x^2 + x}{20}$, $x = 1, 2, 3$.
For this distribution, calculate the:

a) The μ b) The σ

6 $P(x) = \frac{x^2 + x}{20}$ for $x = 1, 2, 3$

x	1	2	3
$P(x)$	$\frac{2}{20} = 0.1$	$\frac{6}{20} = 0.3$	$\frac{12}{20} = 0.6$

$$
\begin{aligned}
\mu &= \sum x_i p_i \\
&= 1(0.1) + 2(0.3) + 3(0.6) \\
&= 2.5
\end{aligned}
$$

$$
\begin{aligned}
\sigma &= \sqrt{\sum (x_i - \mu)^2 p_i} \\
&= \sqrt{(1 - 2.5)^2 \times 0.1 + (2 - 2.5)^2 \times 0.3 + (3 - 2.5)^2 \times 0.6} \\
&\approx 0.671
\end{aligned}
$$

EXERCISE 2.07 (PAGE 81)

- Question 2, 3, 6, 10, 12, 14, 17