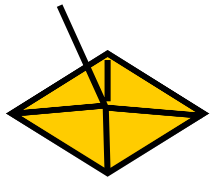
THE SAMPLING DISTRIBUTION



Consider a simple 4-sided fair spinner with numbers 1, 2, 3, 4:

The probability density function is written

x	1	2	3	4
P(X = x)	¼	¼	¼	¼

E(X) = 2.5 and Var(X) = 1.25 for this discrete uniform distribution.

If we spin the spinner twice we could get results like (1, 4), (2, 2), (3, 1) …… The sampling distribution of the **sum** would give all possible values of the **sum** with their associated probabilities. A two-way table would help to set out the probability density function:

+	1	2	3	4
1	2	3	4	5
2	3	4	5	6
3	4	5	6	7
4	5	6	7	8

x + x	2	3	4	5	6	7	8
P	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

The **mean** of each of the possible values would be obtained by dividing the sums by 2. The probabilities of these **means** would be the same as the probabilities of the sums.

$\frac{x+x}{2}$	1	1.5	2	2.5	3	3.5	4
P	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

This is the sampling distribution of the means. I.e the distribution of $\bar{X}$

Its mean, E($\bar{X}$) can be shown to be 2.5 and the variance is 0.625

If we consider all possible samples of **size 3** and their respective means we get a distribution with mean 2.5 and variance $\frac{5}{12}$ (0.417).

If we consider all possible samples of **size n** and their respective means we get a distribution with mean 2.5 and variance $\frac{1}{n}$ Var(X).

This is not just confined to a spinner: If X is a random variable with mean **μ** and variance σ^2 then the distribution of the sample mean for a sample of size n, has mean **μ** and variance $\frac{\sigma^2}{n}$.

A sample of size n would give a **total** of $X_1 + X_2 + X_3 + \ldots\ldots X_n$ and the associated mean ($\bar{X}$) would be $(X_1 + X_2 + X_3 + \ldots\ldots X_n)/n$

$$E(\bar{X}) = E\left[(X_1 + X_2 + X_3 + \ldots\ldots X_n)/n\right] = \frac{1}{n}\left[E(X_1) + E(X_2) + E(X_3) + \ldots\ldots E(X_n)\right] = \frac{1}{n}\left[\mu + \mu + \mu + \mu + \ldots\mu\right] = \frac{1}{n}(n\mu) = \mu$$

$$Var(\bar{X}) = Var\left[(X_1 + X_2 + X_3 + \ldots\ldots X_n)/n\right] = \frac{1}{n^2}\left[Var(X_1) + Var(X_2) + Var(X_3) + \ldots Var(X_n)\right]$$

$$= \frac{1}{n^2}\left[\sigma^2 + \sigma^2 + \sigma^2 + \ldots \sigma^2\right] = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}$$

The **central limit theorem** states that if X is a random variable from **any** distribution with mean μ **and variance** σ^2 **then providing the sample is large enough, the distribution of the sample mean** is approximately normal with the same mean as the original distribution but with a variance of $\frac{\sigma^2}{n}$. where n is the sample size. The larger the sample size the smaller the variance $\frac{\sigma^2}{n}$ and the better the approximation.

$\sqrt{\frac{\sigma^2}{n}}$ is the **standard deviation of the sampling distribution**, written as $\frac{\sigma}{\sqrt{n}}$ and with its own name: **The standard error of the mean**.

HYPOTHESIS TESTING USING THE NORMAL DISTRIBUTION

The age of head teachers in the united kingdom follows the normal distribution with mean 50 and standard deviation 8.

It is suspected that head teachers are getting younger so a sample of size 16 is taken and the sample mean $\bar{x}$ has been computed as 46.

Ho: $\mu = 50$

H1: $\mu < 50$

$\bar{X} \sim N(50, \frac{8^2}{16})$. **The critical value at the 5% level for the normal distribution and a one-tailed test is –1.65**

The test statistic $z = \frac{46-50}{2} = -2$ **and since –2 < -1.65, we reject the null hypothesis.**

The evidence suggests heads are getting younger.