<u>CENTRAL LIMIT THEOREM</u>

A consultancy undertakes systems development projects. From previous experience it has been ascertained that the mean profit per project is£24,000 with a standard deviation of £8,000. Assuming the Normal distribution can be applied:

(a) What is the probability that the **mean** profit for 16 projects is greater than £26,000?

Here a question is being asked about the **mean from a sample of size 16.**

We need to look at the distribution for the mean of all such samples. This is the **sampling distribution of the means.**

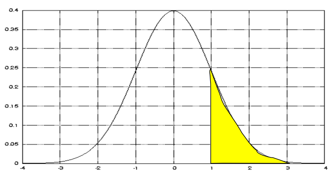
The **central limit theorem** states that this distribution is approximately normal with the same mean as the population distribution but with a variance of $\frac{\sigma^2}{n}$, where σ^2 is the population variance and n is the sample size. The larger the sample size the smaller the variance $\frac{\sigma^2}{n}$ and the better the approximation. $\sqrt{\frac{\sigma^2}{n}}$ is the **standard deviation of the sampling distribution**.

It can be written as $\frac{\sigma}{\sqrt{n}}$ and has its own name: **The standard error of the mean**.

To answer the question: The sample mean has the distribution $\overline{X} \sim N(24000, \sqrt{\frac{8000^2}{16}})$ ie $N(24000,2000)$

$$\mathbf{Pr(\bar{x} \geq 26{,}000)}$$

Standardize: $\mathbf{Pr(} z \geq \frac{26{,}000 - 24{,}000}{2{,}000}\mathbf{)} = \mathbf{Pr(} z \geq 1\mathbf{)} =$



$= \mathbf{1 - \Phi(1) = 1 - 0.8413}$

<u>**= 0.1587**</u>

What if the population variance is unknown? **We have to estimate it using the sample values**

The **most efficient estimator** of the population mean is $\bar{x} = \frac{\sum x}{n}$ but the most efficient estimator of the population variance is

$\frac{\sum (x-\bar{x})^2}{n-1}$ **notice the divisor (n – 1) for estimation.** The variance of a set of sample data is $\frac{\sum (x-\bar{x})^2}{n} = \frac{\sum (x^2) - n\bar{x}^2}{n}$

The observations 32, 44, 45, 21, 18, 56, 27, 31, 57, 28, 39, 49, 22, 15, 33 have a **mean** of $\frac{517}{15}$ = **34.5** and a **variance** of $\frac{20329-17853.75}{15}$.

The **estimate** of the population variance has the same numerator but must be divided by (n-1) 14. $\frac{20329-17853.75}{14}$ = **176.8**

For a set of data given as a grouped frequency table: Using class mid-points to find unbiased estimates for the mean and variance of the population: **We may have to work in units of 10000.**

Expenses claims (£)	Frequency (f)	mid-point	mid-point^2	mid-point x frequency	mid-point^2 x frequency
0-	5	500	250000	2500	1250000
1000-	18	1500	2250000	27000	40500000
2000-	38	2500	6250000	95000	237500000
3000-	45	3500	12250000	157500	551250000
4000-	10	4500	20250000	45000	202500000
5000-	6	5500	30250000	33000	181500000
6000-	0	6500	42250000	0	0

Totals: **122** **360000** **1214500000**

Unbiased estimates: **mean** $= \frac{360000}{122} =$ **<u>2951</u>** **variance** $= \frac{1214500000-(122\times 2951^2)}{121} = \frac{152080000}{121} =$ **1256860**

And the unbiased (most efficient) estimate for the population standard deviation is $\sqrt{1256860}$ = **<u>1121</u>**
Two standard deviations either side of the mean give us the interval 709 ________ 5193 (Almost all the distribution).