Notes and jottings column

Sample Means

Sampling distribution of the mean: In most practical situations it is impossible to measure values for a complete population, and statisticians must therefore deal with samples. How are the mean and standard deviation of a sample related to those for the parent population? Clearly, if you take a number of samples then the means of these samples will themselves form a distribution: this is called the *sampling distribution of the* mean. Using μ and σ^2 for the population mean and variance, and n for the sample size, it can be shown that the sampling distribution also has mean μ but has variance $\frac{\sigma^2}{n}$. The sampling distribution of the means is given the symbol $\bar{X}$, so we can write:

$$E(\bar{X}) = \mu, \quad Var(\bar{X}) = \frac{\sigma^2}{n}$$

Note that the larger n is, the lower the variance. It follows that larger samples give better estimates of the population mean. Fairly obvious, perhaps, but this is the proof that larger samples give more accurate results.

Standard error of the mean: The standard deviation of the sample means is $\frac{\sigma}{\sqrt{n}}$ and this is knows as *the standard error of the mean*.

Central limit theorem: If the parent population is Normally distributed then, not surprisingly, so is the sampling distribution of the mean. In other words, if you take a large number of samples, you would find the means of these samples would be themselves Normally distributed, centred on the population mean. However, if the parent population is *not* Normal, the sampling distribution of the mean is *still* approximately Normal, and becomes more so the larger the sample size n. This result is called the *Central Limit Theorem.*

TI-83
Using the DISTR menu, the calculation is:
normalcdf(11,1E99,10,√0.5)
The 1E99 effectively gives the top end of the range as ∞

A sample of size 8 is taken from a population with distribution N(10,4) Find the probability that the sample mean is more than 11.

$\bar{X}$ will be normally distributed (because the parent population is), so

$Var(\bar{X}) = \frac{\sigma^2}{n} = \frac{4}{8} = 0.5$ Thus $\bar{X} \sim N(10,0.5)$.

Then, using either Normal probability tables or your calculator, we find that P($\bar{X}$ >11) = 0.0786.

Pooled estimators: In the Statistics section of the core syllabus, you learnt that the sample mean is a good estimator of the population mean, and that $\sqrt{\frac{n}{n-1}}\, s_n$ serves as an estimate of the population SD (where s_n is the standard deviation of the sample).

If you have two samples of size n and m , then unbiased estimators of the population parameters can be calculated as:

$$\mu = \frac{n\bar{X}_n + m\bar{X}_m}{n+m}, \quad \sigma^2 = \frac{nS^2{}_n + mS^2{}_m}{n+m-2}$$