Distribution of the sample means of random variable *X*

Examples of random variable *X*:
The IQ of a person in a population (e.g. the Perth population)
The length of a fish in a lake

Population:
Random variable *X* with *any type* of distribution
Population mean μ of *X* is known
Population standard deviation σ of *X* is known

μ and σ are constants for a particular population

Take random samples *A*, *B*, *C*, … of the same size n from the population.

A random sample [x_1 x_2 x_3 …………... x_n] of size *n*.

In random sample *A*, the mean (average) value of the sample contents is called the sample mean of *A* and it is denoted by $\bar{x}_A$. It is calculated by $\bar{x}_A = \frac{x_{A1} + x_{A2} + \cdots + x_{An}}{n}$.

In random sample *B*, $\bar{x}_B = \frac{x_{B1} + x_{B2} + \cdots + x_{Bn}}{n}$, etc.

The sample mean depends on the content of a sample, .: it varies from sample to sample, .: we can consider $\bar{x}_A$, $\bar{x}_B$, … as values of a random variable denoted by $\bar{X}$.

Notations: $\bar{X}$ to represents the sample mean random variable and $\bar{x}$ represents the value of $\bar{X}$ for a random sample.

Irrespective of the distribution of *X* in the population, if the sample size is large enough,

(1) the distribution of the sample mean $\bar{X}$ is approximately normal

(2) the mean of $\bar{X}$ is given by $\mathrm{E}(\bar{X})$ which is exactly equal to μ, the population mean

(3) the standard deviation of $\bar{X}$ is given by $\mathrm{sd}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$ where σ is the standard deviation of *X* in the population

In the discussion above, the population constants μ and σ of *X* are known or given. We use them to predict the distribution of $\bar{X}$ in the random samples.

Inference about the population from a sample

Usually we don't know what the population μ and σ are.
To learn about the population we take a random sample of size *n* from it. $\bar{x}$ of the random sample can be used as an estimate of the population mean μ. We infer about the population from a random sample.

A random sample [x_1 x_2 x_3 …………... x_n] of size *n*.

Calculate the sample mean, one of the statistics of the sample, $\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$, if it is not known or given.

The value of $\bar{x}$ is a reasonable estimate of the population mean μ of random variable *X*. The larger the sample size *n*, the better is the estimation.

A better alternative to $\bar{x}$ as an estimator of μ is to give an interval of *X* values that we are 95% sure contains the population mean μ, meaning about 95 out of 100 confidence intervals of *X* values calculated from the random samples contain μ.
This interval is called a 95% confidence interval for μ.

Its calculation is $\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}},\ \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$ approximately.

For a 99% confidence interval, $\left(\bar{x} - 2.85\frac{\sigma}{\sqrt{n}},\ \bar{x} + 2.85\frac{\sigma}{\sqrt{n}}\right)$

Note 1: The larger the sample size *n*, the distribution is closer to normal and .: the confidence interval is more precise.

Note 2:
$\Pr(-1.96 < Z < 1.96) \approx \frac{95}{100}$, $\Pr(-2.85 < Z < 2.85) \approx \frac{99}{100}$
where random variable Z has a standard normal distribution.

Note 3:
In general, an *A*% confidence interval is approximately
$\left(\bar{x} - z\frac{\sigma}{\sqrt{n}},\ \bar{x} + z\frac{\sigma}{\sqrt{n}}\right)$ where $\Pr(-z < Z < z) \approx \frac{A}{100}$, $z > 0$

Note 4:
For constant sample size *n*, the higher the required confidence level, the wider is the required interval.

For a required confidence level, the larger the sample size *n*, the narrower is the required interval.

However, the population standard deviation σ is generally unknown.
In the absence of any other information, the sample standard deviation s can be used instead of σ in the calculation e.g.
$\left(\bar{x} - 1.96\frac{s}{\sqrt{n}},\ \bar{x} + 1.96\frac{s}{\sqrt{n}}\right)$ for the 95% confidence interval.

Example 1 (2016 VCAA SM Sample Exam 2 SECTION A Q 19)

The mean study score for a large VCE study is 30 with a standard deviation of 7. A class of 20 students may
be considered as a random sample drawn from this cohort.
The probability that the class mean for the group of 20 exceeds 32 is

A. 0.1007
B. 0.3875
C. 0.3993
D. 0.6125
E. 0.8993

$\text{sd}\left(\bar{X}\right)=\frac{\sigma}{\sqrt{n}}=\frac{7}{\sqrt{20}}\approx 1.56525$, $\mu=30$

Normal: $\Pr\left(\bar{X}>32\right)\approx 0.1007$ A