Sample Distribution

Some Definitions

Population is the whole set of measurements or quantities we want to sample

Sample is a subset of the population

Example

10,000 = UoM Student population
If we are interested in heights then
height population = 10,000
If we are interested in male heights then male height population = 4500
A sample is a subset of the above we can have a random sample or a biased sample
Random sample

Each quantity or measurement in the population has the same probability of being selected.

→ uniform distribution

Biased sample

Probabilities are not the same for each measurement.

Note: When using random sampling we need to devise a method that ensures equal probability of being selected.

E.g. give an identity number to each entity and use a uniform random generator.
Sample Size

This is another important parameter.

0 'I am using a sample size of 20.'
+ 'That too small, I'm using n=30.'
0 'Why?'
+ 'It sounds right, doesn't it!'

Intuitively n=30 is more accurate than n=20.

But how much accurate?

The answer to this question will be given later...
Sample Distribution

Finally...
Suppose our population has
mean = μ
std dev = σ

for our measurement, say of height,
this implies that the distribution is continuous, but not necessarily normal, as below

\[
\text{population distribution}
\]

Suppose we take a random sample of size n.
We then calculate
sample mean = \( \bar{x} \)
std dev = s
We repeat the above experiment. Then we have two more values for \( \bar{x} \) and \( s \)

\[ \bar{x}_1 \text{ and } s_1, \]
\[ \bar{x}_2 \text{ and } s_2 \]

If we repeat the above for say 40 times then we have a large number of

Sample means = \( \{ \bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots, \bar{x}_{40} \} \)

We can therefore plot a distribution of the sample means

This distribution will have a mean and standard deviation

mean = \( \mu_{\bar{x}} \)

std dev = \( \sqrt{\bar{x}} \)
What is the connection between $(\mu, \sigma, \text{pdf})$ for population and $(\bar{x}, \frac{1}{\sqrt{n}} \text{pdf}\bar{x})$ for sampling results?

From statistical theory we have some results.

1) $\mu_{\bar{x}} = \mu$

2) $\frac{1}{\sqrt{n}}$  

3) If $n$ is large, pdf$\bar{x}$ is approximately normal, irrespective of shape of pdf population.

4) If population is normal, then pdf$\bar{x}$ is normal for all $n$. 
Graphical view

\[ \bar{X} = \mu \]

Result (3) is related to the central limit theorem.

It also emphasises the importance of the normal distribution.