RMIT UNIVERSITY

# SAMPLING DISTRIBUTIONS

A ***sampling distribution*** is the probability distribution for the means of all samples of size n from a given distribution. The sampling distribution will be normal distributed with parameters $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$, if either

- the population from which the samples are drawn is normally distributed, or
- the samples are large ($n \geqq 30$)

where $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ [for large samples]

NB:
- the sampling distribution has the same centre as the population
- the measure of variability of a sampling distribution, $\sigma_{\bar{x}}$, is called the ***standard error***. The distribution of means is not as spread out as the values in the population from which the sample was drawn.
- if we do not know the population standard deviation we approximate with the sample standard deviation: $s_{\bar{x}} \cong \sigma_{\bar{x}}$ and $\frac{s}{\sqrt{n}} \cong \frac{\sigma}{\sqrt{n}}$)

Consider the little 'population' of values P = {1 2 3 4 5}
This population has $\mu = 3$ and $\sigma = 1.41$
If a sample of size n = 3 was drawn from this population it could be any one of…
(1 2 3) (1 2 4) (1 2 5) (1 3 4) (1 3 5) (1 4 5) (2 3 4) (2 3 5) (2 4 5) (3 4 5)
The means of each of the samples, and a histogram of the distribution of means, are shown in the table and graph below:

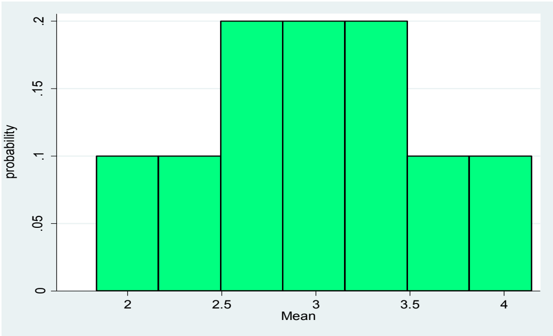
| Sample | Mean |
|---|---|
| 1 2 3 | $\bar{x} = 2$ |
| 1 2 4 | $\bar{x} = 2.33$ |
| 1 2 5 | $\bar{x} = 2.67$ |
| 1 3 4 | $\bar{x} = 2.67$ |
| 1 3 5 | $\bar{x} = 3$ |
| 1 4 5 | $\bar{x} = 3.33$ |
| 2 3 4 | $\bar{x} = 3$ |
| 2 3 5 | $\bar{x} = 3.33$ |
| 2 4 5 | $\bar{x} = 3.67$ |
| 3 4 5 | $\bar{x} = 4$ |

$\bar{\bar{x}} = 3$
$\sigma_{\bar{x}} = 0.61$

The sampling distribution of the means for samples of size 3 is:

| $\bar{X}$ | 2 | 2.33 | 2.67 | 3 | 3.33 | 3.67 | 4 |
|---|---|---|---|---|---|---|---|
| $P(\bar{X} = \bar{x})$ | 0.1 | 0.1 | 0.2 | 0.2 | 0.2 | 0.1 | 0.1 |

Even though this sample is small, and the population is not normally distributed (though it is symmetric) the sampling distribution is reasonably normally distributed:




We can see that the mean of the sampling distribution (the mean of all the means) is the same as the population mean, $\bar{\bar{x}} = \mu = 3$. But the variability in the sampling distribution is less than that of the population: $\sigma_{\bar{x}} = 0.61$ and $\sigma = 1.41$. Because larger samples, or those drawn from normally distributed populations, will follow a normal distribution we can use the properties of normal distributions to find probabilities relating to samples: $Z_{\bar{x}} = \dfrac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

## Example

*The shire of Bondara has 1200 preschoolers. The mean weight of pre-schoolers is known to be 18kg with a standard deviation of 3kg. What is the probability that a random sample of 50 preschoolers will have a mean weight more than 19kg?*

$n = 50$, $\mu = 18$ and $\sigma = 3$

The sampling distribution of the means for samples of size 50 will have $\mu_{\bar{x}} = \mu = 18$, and standard error, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{50}} = 0.42$.

$$Z_{\bar{x}} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{19 - 18}{\frac{3}{\sqrt{50}}} = 2.38$$

$\Pr(\bar{x} > 19) = \Pr(Z_{\bar{x}} > 2.38)$
$= 1 - 0.9913$ [from tables]
$= 0.0087$

## Exercise

1. List all samples of size 2 for the population {1, 2, 3, 4, 5, 6}. What is the probability of obtaining a sample mean of less than 3?

2. Samples of size 40 are drawn from a population with $\mu = 50$ and $\sigma = 5$.
   (a) What are the mean and standard error of the sampling distribution?
   (b) What is the probability that a particular sample has a mean less than 48.5?

3. If IQ in the general population of secondary students is known to follow a normal distribution with $\mu = 100$ and $\sigma = 10$,
   (a) find the mean and standard error for a random samples of size 100.
   (b) To test whether a secondary school is representative of the general population a sample of 100 students from that school is chosen. What is the probability of the mean IQ being more than 105?
   (c) What would be your conclusion?

### Answers

1. 4/15
2. (a) $\mu_{\bar{x}} = 50$ and $\sigma_{\bar{x}} = 0.79$ (b) 0.0288
3. (a) $\mu_{\bar{x}} = 100$ and $\sigma_{\bar{x}} = 1$ (b) 0.00003 (c) either the sample was not random (perhaps all the smartest students were in the sample) or this school has a higher IQ than the general population.