# Statistics 2B (Practical)

*Compiled by Mr. V van Appel, Department of Statistics - University of Johannesburg*
*Practical video done by Mr. T Hansragh, Department of Statistics - University of Johannesburg*

## Practical 3: Sampling Distributions

https://youtu.be/G1D42EW83xo

## 3.1 Sampling distribution

To assess the accuracy of an estimate of a parameter, the probability distribution of the statistic of interest is used to place probabilistic bounds on the sampling error. The probability distribution associated with all the possible values that a statistic can assume is called the sampling distribution of the statistic.

As an example, consider the sample mean (a statistic) $\bar{X}$: suppose 10 students are randomly selected from the population of students in South Africa and their mean age is computed. If this process were repeated three times, it is unlikely that any of the computed sample means would be identical. Similarly, it is also not likely that any of the three computed sample means would be exactly equal to the population mean. However, these sample means are typically used to estimate the unknown population mean. To assess the accuracy of the estimate (sampled value) the sampling distribution of the sample mean is used.

Recall: If $X$ is a random variable with mean, $\mu$, and variance, $s^2$, and if a random sample $X_1, X_2, \ldots, X_n$ is taken, then the expected value and variance of the sample mean, $\bar{X}$ are respectively

$$E(\bar{X}) = \mu_{\bar{X}} = \mu \tag{3.1}$$

$$V(\bar{X}) = \sigma^2_{\bar{X}} = \frac{\sigma^2}{n} \tag{3.2}$$

## 3.2 Sampling from a finite distribution

**Example 3.1** *Consider an experiment where two balls are randomly selected from an urn containing six numbered balls.*

- *Case 1: Suppose the sampling is done with replacement*
- *Case 2: Suppose the sampling is done without replacement*

*Use R to list the exact sampling distributions of the sample mean, $\bar{X}$, and the sample variance, $s^2$, for both cases.*

*Then create graphs that compare these four distributions.*

### 3.2.1 Case 1: Random sampling

When sampling is done with replacement, the outcomes can be seen as a random sample of size 2 drawn from a discrete uniform distribution.

Population parameters: $\mu = 3.5$ and $\sigma^2 = 2.9166$

Random sampling - Possible samples of size 2 with $\bar{X}$ and $s^2$ for each sample.

| $(x_1, x_2)$ | $\bar{X}$ | $s^2$ | $(x_1, x_2)$ | $\bar{X}$ | $s^2$ |
|---|---|---|---|---|---|
| (1,1) | 1.0 | 0.0 | (4,1) | 2.5 | 4.5 |
| (1,2) | 1.5 | 0.5 | (4,2) | 3.0 | 2.0 |
| (1,3) | 2.0 | 2.0 | (4,3) | 3.5 | 0.5 |
| (1,4) | 2.5 | 4.5 | (4,4) | 4.0 | 0.0 |
| (1,5) | 3.0 | 8.0 | (4,5) | 4.5 | 0.5 |
| (1,6) | 3.5 | 12.5 | (4,6) | 5.0 | 2.0 |
| (2,1) | 1.5 | 0.5 | (5,1) | 3.0 | 8.0 |
| (2,2) | 2.0 | 0.0 | (5,2) | 3.5 | 4.5 |
| (2,3) | 2.5 | 0.5 | (5,3) | 4.0 | 2.0 |
| (2,4) | 3.0 | 2.0 | (5,4) | 4.5 | 0.5 |
| (2,5) | 3.5 | 4.5 | (5,5) | 5.0 | 0.0 |
| (2,6) | 4.0 | 8.0 | (5,6) | 5.5 | 0.5 |
| (3,1) | 2.0 | 2.0 | (6,1) | 3.5 | 12.5 |
| (3,2) | 2.5 | 0.5 | (6,2) | 4.0 | 8.0 |
| (3,3) | 3.0 | 0.0 | (6,3) | 4.5 | 4.5 |
| (3,4) | 3.5 | 0.5 | (6,4) | 5.0 | 2.0 |
| (3,5) | 4.0 | 2.0 | (6,5) | 5.5 | 0.5 |
| (3,6) | 4.5 | 4.5 | (6,6) | 6.0 | 0.0 |

Random sampling - Sampling distribution of $\bar{X}$ (sample mean)

| $\bar{X}$ | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 | 4.5 | 5 | 5.5 | 6 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $f(\bar{X})$ | 1/36 | 2/36 | 3/36 | 4/36 | 5/36 | 6/36 | 5/36 | 4/36 | 3/36 | 2/36 | 1/36 |

Random sampling - Sampling distribution of $s^2$ (sample variance) Note the following:

| $s^2$ | 0 | 0.5 | 2 | 4.5 | 8 | 12.5 |
|---|---|---|---|---|---|---|
| $f(s^2)$ | 6/36 | 10/36 | 8/36 | 6/36 | 4/36 | 2/36 |

- The formulae and the sampling distribution of $\bar{X}$ is: $\mu_{\bar{X}} = E(\bar{X}) = 3.5$ and $\sigma^2_{\bar{X}} = 1.4583$
- biased estimate: $E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2 = 1.4583$
- unbiased estimate: $E(s^2) = \sigma^2 = 2.9166$

#### How to do with R

**Try doing this in R. The solution will be made available after the Practical class.**

### 3.2.2 Case 2: Simple random sampling

When sampling is performed without replacement, the outcomes can be viewed as a simple random sample of size 2 drawn from a discrete uniform distribution. Now fewer samples $\binom{6}{2} = 15$ are possible, but each sample is equally likely to be drawn.

Simple random sampling - Possible samples of size 2 with $\bar{X}$ and $s^2$

| $(x_1, x_2)$ | $\bar{X}$ | $s^2$ |
|---|---|---|
| (1,2) | 1.5 | 0.5 |
| (1,3) | 2.0 | 2.0 |
| (1,4) | 2.5 | 4.5 |
| (1,5) | 3.0 | 8.0 |
| (1,6) | 3.5 | 12.5 |
| (2,3) | 2.5 | 0.5 |
| (2,4) | 3.0 | 2.0 |
| (2,5) | 3.5 | 4.5 |
| (2,6) | 4.0 | 8.0 |
| (3,4) | 3.5 | 0.5 |
| (3,5) | 4.0 | 2.0 |
| (3,6) | 4.5 | 4.5 |
| (4,5) | 4.5 | 0.5 |
| (4,6) | 5.0 | 2.0 |
| (5,6) | 5.5 | 0.5 |

Random sampling - Sampling distribution of $\bar{X}$ (sample mean)

| $\bar{X}$ | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 | 4.5 | 5 | 5.5 |
|---|---|---|---|---|---|---|---|---|---|
| $f(\bar{X})$ | 1/15 | 1/15 | 2/15 | 2/15 | 3/15 | 2/15 | 2/15 | 1/15 | 1/15 |

Random sampling - Sampling distribution of $s^2$ (sample variance)

| $s^2$ | 0.5 | 2 | 4.5 | 8 | 12.5 |
|---|---|---|---|---|---|
| $f(s^2)$ | 5/15 | 4/15 | 3/15 | 2/15 | 1/15 |

Note the following:

- Population parameters: $\mu = 3.5$ and $\sigma^2 = 2.9166$
- Mean of the sampling distribution of $\bar{X} : \mu_{\bar{X}} = E(\bar{X}) = 3.5$
- Variance of the sampling distribution of $\bar{X} : \sigma^2_{\bar{X}} = E[(\bar{X} - \mu_{\bar{X}})^2] = 1.16666$
- Expected value of sample variance: $E(s^2) = 3.5$
- Summary results for sampling without replacement (finite population):

$$\mu_{\bar{X}} = \mu \tag{3.3}$$

$$\sigma^2_{\bar{X}} = \frac{\sigma^2}{n} \times \frac{N-n}{N-1} \tag{3.4}$$

$$E(\hat{\sigma}^2) = \sigma^2 \left(\frac{n-1}{n}\right)\left(\frac{N}{N-1}\right) \tag{3.5}$$

where an unbiased estimate of the population variance is:

$$E\left[s^2 \times \frac{N-1}{N}\right] = \sigma^2 \tag{3.6}$$

Note: The expected value of the sample mean is $\mu$, for both sampling with or without replacement. However, the other parameters change. (Another simple example is discussed in Sleuth (p.10)).

**How to do with R**

**Try doing this in R. The solution will be made available after the Practical class.**

### 3.2.3 Sampling from an infinite distribution

Consider a population (with an infinite distribution) from which $k$ random samples, each of size $n$, are taken. In general these samples will result in $k$ different values, $\bar{X}_1, \ldots, \bar{X}_k$, for the sample mean $\bar{X}$. If $k$ is very large (theoretically infinite), the relative frequency distribution (histogram) of $\bar{X}_i$, $i = 1, \ldots, k$, will represent the sampling distribution of $\bar{X}$.

Note that the sampling distribution of $\bar{X}$ will typically not the same as the distribution of the $X$, from which the original observations come.

See Wackerly for the theoretical calculation of the sampling distribution of several statistics. Note that simulation can be used to empirically verify the results obtained (see for example Verzani, section 7).