

# Sampling with and without replacement

## Student Activity

7 8 9 10 11 12



TI-Nspire



Investigation



Student



60 min

## Introduction

Suppose that a random sample of 3 marbles is to be drawn from an urn containing 5 blue marbles, 3 green marbles and 2 red marbles. In this activity, you will investigate the probability that the sample contains 0, 1, 2 or 3 blue marbles, when the sample is selected *with replacement* and when the sample is selected *without replacement*. A single trial in these experiments consists of selecting a sample of 3 marbles from the urn.

Selection **with replacement** means that a single marble is drawn from the urn, its colour is noted, and is returned to the urn before the next marble is drawn.

Selection **without replacement** means that each marble drawn from the urn is set aside and is **not** returned to the urn before the next marble is drawn.



### Teacher Tip:

Before distributing the 'Repl\_noRepl' tns file, it might be instructive for students to create a simple simulation in a Notes page, such as the one on the right, to explore the 'randSamp' and 'countIf' commands. Each time the 'replace' or 'noreplace' MathsBox is evaluated, a new sample is drawn.

```

Simulation
Explore 'randSamp' and 'countIf'
{b,b,b,b,b,g,g,g,r,r} → urn
▶ {b,b,b,b,b,g,g,g,r,r}
replace:=randSamp(urn,3,0) ▶ {r,b,b}
noreplace:=randSamp(urn,3,1) ▶ {b,g,r}
countIf(replace,b) ▶ 2
countIf(noreplace,b) ▶ 1
    
```

## Simulating random sampling with replacement and without replacement

Open the TI-Nspire file named 'Repl\_noRepl' and read the instructions on Pages 1.2 and 1.3.

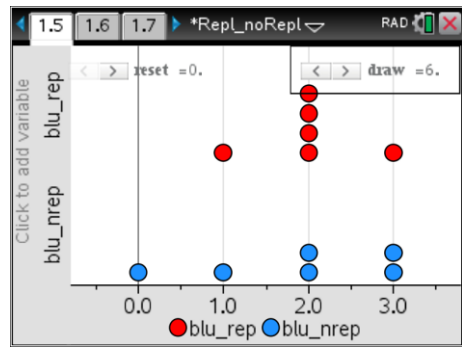
On Page 1.4, each increment on the slider labelled 'draw' carries out a single trial in the simulations, for selection with replacement and without replacement. The outcome of each trial appears as the green output, below the sliders. For example, the outcome {r,r,g} represents a sample containing 2 red and 1 green marble, so the quantity of blue marbles in the sample is shown as {0}.

```

1.2 1.3 1.4 ▶ Repl_noRepl
draw = 18. reset = 0.
Select 3 marbles with replacement
repl:=randSamp(urn,3,0) · sign(draw+1) ▶ {b,b,g}
Select 3 marbles without replacement
norepl:=randSamp(urn,3,1) · sign(draw+1)
▶ {r,r,g}
{countIf(repl,b)} → b_rep ▶ {2}
{countIf(norepl,b)} → b_rep ▶ {0}
    
```

A simple program captures the results of individual trials, so that the quantity of blue marbles can be displayed as sampling is repeated many times.

These results are displayed as dot-plots and histograms on Pages 1.5 and 1.6, respectively. Each time a new sample is drawn, using the 'draw' slider, the displays of data change interactively to incorporate the new result. 'blu\_rep' denotes the quantity of blue marbles in a sample selected *with* replacement, and 'blu\_nrep' denotes the quantity of blue marbles in a sample selected *without* replacement.



## Estimated probabilities from simulating random sampling

On Page 1.5, use the 'draw' slider to simulate drawing 100 samples, with and without replacement. Use the results displayed in the dot-plot and in the histogram to complete the following tables in Questions 1 and 2.

### Question: 1

Complete Table 1 for the estimated probability that a sample, drawn **with** replacement, will contain 0, 1, 2 or 3 blue marbles.

**Table 1: With replacement**

Quantity of blue marbles	0	1	2	3
Estimated probability				

### ANSWER

Results will vary. Typical results are shown in the screenshot on the right.

	blu_nrep	blues	prob_rep	prob_nrep
1	0.	0	0.12	0.09
2	1.	1	0.37	0.35
3	2.	2	0.41	0.38
4	1.	3	0.1	0.18
5	2.			

### Question: 2

Complete Table 2 for the estimated probability that a sample, drawn **without** replacement, will contain 0, 1, 2 or 3 blue marbles.

**Table 2: Without replacement**

Quantity of blue marbles	0	1	2	3
Estimated probability				

### ANSWER

Results will vary. Typical results are shown in the screenshot on the right.

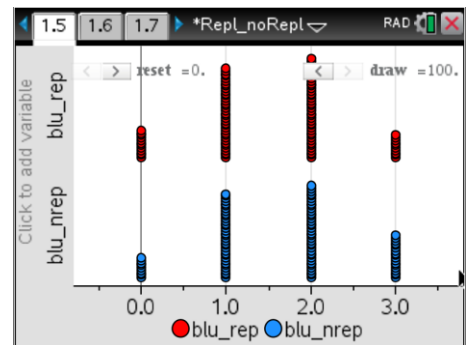
	_nrep	blues	prob_rep	prob_nrep
1	0.	0	0.12	0.09
2	1.	1	0.37	0.35
3	2.	2	0.41	0.38
4	1.	3	0.1	0.18
5	2.			

### Question: 3

What are some key similarities and differences for the data obtained 'with replacement' and 'without replacement'? Consider features such as the shapes and heights of the data displays, and the estimated probabilities displayed in Tables 1 and 2.

### ANSWER

Typical results are shown in the screenshot on the right. Students might note that the shapes are similar, and fairly symmetrical, in both cases. Students might, for example, observe that 'with replacement', there is greater chance of zero blue / all blue than 'without replacement'.



### Question: 4

Discuss why selecting the samples with replacement and without replacement lead to these differences, including in the estimated probabilities displayed in Tables 1 and 2.

### ANSWER

Typical answer might refer to the fact that in the case of 'with replacement', it is possible for the same marble to be selected more than once in a single trial.

## Variability of results for repeated simulations

On Page 1.5, click the 'reset' slider to clear the data. Simulate drawing a new set of 100 trials. Actual results are stored in the spreadsheet on page 1.7

	_nrep	blu_nrep	blues	prob...	prob...
1	2.	3.	0	0.	0.166667
2	2.	2.	1	0.1666...	0.166667
3	2.	2.	2	0.6666...	0.333333
4	1.	1.	3	0.1666...	0.333333
5	2.	3.			1.

### Question: 5

Compare the data obtained from the new simulations with the results of the previous simulations. What similarities and differences are there?

#### ANSWER

While actual numerical values will vary between simulations, the general pattern will remain much the same for repeated simulations.

The simulation results are summarised as lists on Page 1.7. The orange column shows the estimated probabilities of 0, 1, 2 or 3 blue marbles for sampling with replacement; the yellow column shows this information for sampling without replacement.

### Question: 6

For the new set of 100 trials, compare the estimated probabilities on your Page 1.7 with that obtained by other students. Suggest ways in which the variability of results between successive simulations could be reduced.

#### ANSWER

In general, increasing the number of trials per simulation will reduce the variability between successive simulations.

## Increasing the number of trials in the simulation

On Page 2.1, the list 'urn' is defined as {b,b,b,b,b,g,g,g,r,r}, and the commands 'randSamp(urn,3,0)' and 'randSamp(urn,3,1)' are contained within the nested instruction of the form 'countIf(randSamp(urn,3,x),b)'.

### Question: 7

Explain the purpose of the commands:

- 'randSamp(urn,3,0)' and 'randSamp(urn,3,1)'



#### ANSWER

'randSamp(urn,3,0)' selects, **with** replacement, a random sample of size 3 from the list 'urn'.  
'randSamp(urn,3,1)' selects, **without** replacement, a random sample of size 3 from the list 'urn'.

- 'countIf(randSamp(urn,3,0),b)' and 'countIf(randSamp(urn,3,1),b)'.

#### ANSWER

'countIf(randSamp(urn,3,0),b)' and 'countIf(randSamp(urn,3,1),b)' count the quantity of the item 'b' in the respective random samples

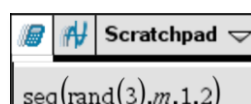
Open the Scratchpad of your TI-Nspire by pressing the  key. Explore the 'seq' command by inputting and evaluating the following. Press the  key to return to the main document.



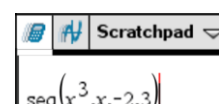
```
seq(hello,k,1,5)
```



```
rand(3)
```



```
seq(rand(3),m,1,2)
```



```
seq(x3,x,-2,3)
```

**Question: 8**

Page 2.1 contains the commands 'seq(countlf(randSamp(urn,3,0),b),k,1,1000)' and 'seq(countlf(randSamp(urn,3,1),b),k,1,1000)'. Explain the purpose of these commands.

**ANSWER**

'seq(countlf(randSamp(urn,3,0),b),k,1,1000)' selects, **with** replacement, a random sample of size 3 from the list 'urn', and counts the quantity of the item 'b' in the sample. It repeats this for  $k = 1$  to  $k = 1000$ , thereby generating 1000 trials. 'seq(countlf(randSamp(urn,3,1),b),k,1,1000)' is analogous, except that the sampling is **without** replacement..

Each increment of the sliders labelled 'simulate' on Pages 2.1, 2.2 and 2.4 result in a new simulation of 1000 trials, for sampling with replacement and without replacement.

**Question: 9**

Observe the histograms on Page 2.2 as the 'simulate' slider is used for repeated simulations.

**ANSWER**

The variability should be less than in Problem 1 (100 trials), due to the greater number of trials.

Comment on the variability between repeated simulations, as indicated by changes in the relative heights of the columns in the histograms.

**Question: 10**

From the heights of the columns of the histograms, estimate the probability of 0, 1, 2 or 3 blue marbles in samples selected with replacement and without replacement. From the table on Page 2.3, check your answer.

**ANSWER**

Results will vary. Typical results are shown in the screenshot on the right.

	_nrep	blues	prob_rep	prob_nrep
1	0.	0	0.12	0.09
2	1.	1	0.37	0.35
3	2.	2	0.41	0.38
4	1.	3	0.1	0.18
5	2.			

## Comparing theoretical probabilities with the simulation results

### Question: 11

Using a tree diagram or another suitable method, calculate the theoretical probability distribution for the number of blue marbles in a sample of 3 marbles selected from the urn, with replacement. Show these probabilities in the second row of Table 3.

Table 3: With replacement

Quantity of blue marbles	0	1	2	3
Theoretical probability	$\frac{1}{8} = 0.125$	$\frac{3}{8} = 0.375$	$\frac{3}{8} = 0.375$	$\frac{1}{8} = 0.125$
Estimated probability*	Typical result 0.1237	Typical result 0.3786	Typical result 0.3704	Typical result 0.1234
Percentage error	Typical result 1.04% $\frac{ 0.125 - 0.1237 }{0.125 / 100}$	Typical result 0.96% $\frac{ 0.375 - 0.3786 }{0.375 / 100}$	Typical result 1.23% $\frac{ 0.375 - 0.3704 }{0.375 / 100}$	Typical result 1.28% $\frac{ 0.125 - 0.1234 }{0.125 / 100}$

\* Estimated from 10 000 trials

### Question: 12

Using a tree diagram or another suitable method, calculate the theoretical probability distribution for the number of blue marbles in a sample of 3 marbles selected from the urn, without replacement. Show these probabilities in the second row of Table 4.

Table 4: Without replacement

Quantity of blue marbles	0	1	2	3
Theoretical probability	$\frac{1}{12} \approx 0.0833$	$\frac{5}{12} \approx 0.4167$	$\frac{5}{12} \approx 0.4167$	$\frac{1}{12} \approx 0.0833$
Estimated probability*	Typical result 0.0794	Typical result 0.4180	Typical result 0.4182	Typical result 0.0845
Percentage error	Typical result 4.72% $\frac{ 1/12 - 0.0794 }{1/1200}$	Typical result 0.32% $\frac{ 5/12 - 0.4180 }{5/1200}$	Typical result 0.37% $\frac{ 5/12 - 0.4167 }{5/1200}$	Typical result 1.40% $\frac{ 1/12 - 0.0845 }{1/1200}$

\* Estimated from 10 000 trials

Use the 'simulate' slider to obtain the results of repeated simulations, and observe these results on Page 2.3. These results are captured by a program on Page 2.5, and the mean of the repeated simulation results is displayed in the table on Page 2.4.

On Page 2.4, click the 'refresh' slider twice, then use the 'simulate' slider to accumulate the results of 10 repeated simulations, of 1000 trials each, for sampling with replacement and without replacement.

### Question: 13

In the third rows of Tables 3 and 4, record the estimated probabilities from Page 2.4, correct to four decimal places, based on 10,000 trials. Calculate the magnitude of the percentage errors in the estimated probabilities, correct to two decimal places, and record these percentage errors in the fourth rows of Tables 3 and 4.

## ANSWER

Results will vary. Typical results are shown in the screenshot on the right. Typical calculations are shown in Tables 3 and 4.

	blues	replace	no_replace	D
1	0	0.1237	0.07935	Trials
2	1	0.3786	0.41795	10000.
3	2	0.37035	0.4182	
4	3	0.12735	0.0845	

C2 =mean(nrep\_1b)

simulate =10. Refresh