**R Class : Set Theory and Sampling**  
  
1. Sequence of Integers

A sequence of integers can be created using the “:” operator, specifying the upper and lower bounds on either side.

2. Set Theory  
R has the following commands that it would use for set theory - (Also quite useful for some simulations)

* Union (all values in the data sets combined)
* Intersection (values in both data sets)
* Set Difference (values in one data set but not the other)

|  |
| --- |
| X = 5:10 Y = 8:12  #X #[1] 5 6 7 8 9 10  #Y #[1] 8 9 10 11 12  union(X,Y) #[1] 5 6 7 8 9 10 11 12  intersect(X,Y) #[1] 8 9 10  setdiff(X,Y)  #[1] 5 6 7  setdiff(Y,X)  #[1] 11 12 |

### 3. Sampling

The **R** command we use to perform sampling is **sample().**  
We specify the dataset that we wish to sample from, and how many sample values to take.

|  |
| --- |
| > sample(X,2) [1] 6 5 > > sample(X,1);sample(X,1);sample(X,1); [1] 6 [1] 8 [1] 5 |

When x is a single value, the function **sample()** behaves differently.

|  |
| --- |
| > Y=c(4) > > sample(Y,1) [1] 2 >  > sample(Y,2) [1] 3 1 |

### 4. Sampling With Replacement and Without Replacement

Can we allow a value to be chosen more than once? Consider the following examples:

* Generate a quick pick : pick 6 numbers from 1 to 42. ( Same number can’t be selected more than once)
* Generate five values from a die (Same number can be selected more than once).

We call this sampling without replacement , and sampling with replacement respectively.

By default, the command uses sampling without replacement. To specify sampling with replacement – add the argument “**replace=true**”.

|  |
| --- |
| > Lotto = 1:42 > Dice = 1:6 >  > sample(Lotto,6) [1] 38 25 34 30 22 29 >  > sample(Dice,5,replace = TRUE) [1] 4 3 2 3 3 > |

### 5. Controlling precision of a value

Consider the value pi. When called, R presents a value to 6 decimal places.

Often this level of precision is not necessary.

|  |
| --- |
| > pi  [1] 3.141593  > floor(pi)  [1] 3  > ceiling(pi)  [1] 4  > round(pi,2)  [1] 3.14  > round(pi,3)  [1] 3.142 |

As you can see, we have many different options when it comes to controlling the precision. If we want to round to the next highest, or next lowest, integers, we can use either the ceiling function or floor function.

We can use the round function to specify the appropriate precision to a suitable number of decimal places.