Setting Seed in R

Learning Objectives

1. Why Set the Seed
2. How to Set the Seed
3. Rejection and Rounding Samplers

1. Why Set the Seed

Oftentimes, we want to randomly split our data into training and test sets, or run simulation studies. These involve generating random numbers. When we want to be able to reproduce the same results, we can use `set.seed()`, where we specify a number with the function.

2. How to Set the Seed

For example, I want to sample 5 integers from 1 to 100 randomly. Running the following two lines of code will always produce the same 5 integers:

```
set.seed(1)
sample.int(100,5)
```

```
## [1] 68 39  1 34 87
```

Everyone should be getting the following 5 integers: 68, 39, 1, 34, 87.

If you got 27, 37, 57, 89, 20, please read on to see how you can get the "correct" set of integers.

3. Rejection and Rounding Samplers

I won't go into the technical details of how random numbers are generated, but there are various techniques to do so, among which there is rejection sampling and rounding sampling.

In R versions 3.6.0 and later, the default technique is rejection sampling. Older versions of R used rounding sampling. Occasionally, I have had students who are using a newer version of R but are still getting a different set of numbers; for some reason, their default is rounding sampling. To select the right sampling, you will need to add an extra line before using 'set.seed():

```
RNGkind(sample.kind = "Rejection")
set.seed(1)
sample.int(100,5)
```

```
## [1] 68 39  1 34 87
```

You can actually use the old rounding sampling, if you want to reproduce results that were based on older versions of R:

```
RNGkind(sample.kind = "Rounding")
```

```
## Warning in RNGkind(sample.kind = "Rounding"): non-uniform 'Rounding' sampler
## used
```

```
set.seed(1)
sample.int(100,5)
```

```
## [1] 27 37 57 89 20
```

Notice the warning message produced. It was discovered that the rounding sampler didn't produce the integers uniformly, which is what we need. So the rounding sampler shouldn't be used, except to reproduce results based on older versions of R.

I believe the 1st edition of our textbook was based on the old rounding sampler.