

# Resampling With Permutation

eCornell

7/26/2021

Scenario: Use this R Markdown file as a guide to resampling with permutation. This example compares student exam scores across two different course sections.

##Step 1: Create a sample data set.

```
# Create the example data frame:
dat = data.frame(Score = c(10, 20, 30, 40, 50, 60, 70), Name = c("Alice", "Bob", "Catie", "Dave", "Eve", "Frank", "George"))
dat # Examine the data set
```

```
##   Score   Name Section
## 1    10  Alice      A
## 2    20   Bob      A
## 3    30  Catie      A
## 4    40   Dave      A
## 5    50   Eve      B
## 6    60  Frank      B
## 7    70 George      B
```

```
set.seed(1) #set the seed for reproducibility
```

##Step 2: Resample with permutation. To do permutation resampling in R, use the `sample()` function with `replace = FALSE`.

```
1:7 #the vector you'll use for permutation
```

```
## [1] 1 2 3 4 5 6 7
```

```
perm.id = sample(1:7, size = 7, replace = FALSE) # the permuted vector
perm.id # note that this vector is no longer in numerical order
```

```
## [1] 1 4 7 2 5 3 6
```

```
dat.perm = dat # make the permuted data set equal to the original data set
dat.perm$Score = dat$Score[perm.id] #shuffle *only* the scores
dat.perm #examine the permuted data frame
```

```
##   Score   Name Section
## 1    10  Alice      A
## 2    40   Bob      A
## 3    70  Catie      A
## 4    20   Dave      A
## 5    50   Eve      B
## 6    30  Frank      B
## 7    60 George      B
```

Note that after permutation, there is a mix of low and high scores in each section. The average scores of Sections A and B are no longer very different.