Learning R Live!

Curtis Petersen

6/7/2017

Create a data set

```
# Set seed to ensure reproducability
        set.seed(123)

# Create data
        n <- 1000 # Number of individuals, 1,000
        Age = 20 + 60 * runif(n) # Create each individuals age; generate uniformly distributed number b
        Gender = runif(n) < 0.5 # Generate indicator for gender so about 50% of the population is femal
        X1 = rnorm(n, mean = 20, sd = 5) # Generate random variable from normal distribution, with mean
        X2 = runif(n) < 0.3 # Generate an indicator for values from the unicorm distribution under 0.3
        P = 20 + 5*rnorm(n) # Generate Predictor that is 5 * a number from the normal distribution + 20

# Connect all these vectors into single object
        Data <- cbind(P, Age, Gender, X1, X2) # Bind these objects together as a matrix
        Data = as.data.frame(Data) # Convert the above matrix to a data frame
```

Wrangle some data

Method 1

```
cbind("Mean" = mean(Data$P), "Min" =  min(Data$P), "Max" = max(Data$P))
```

```
##          Mean      Min      Max
## [1,] 20.00302 4.354559 37.10547
```

Summarize Method 2

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
# Explore data
        Data %>%
                group_by(Gender) %>%
```

```
                summarise("n" = n(),
                          "Mean X1" = mean(X1),
                          "SD X1" = sd(X1)
                          )
```

```
## # A tibble: 2 x 4
##   Gender     n `Mean X1`  `SD X1`
##    <dbl> <int>     <dbl>    <dbl>
## 1      0   497  20.01193 4.859892
## 2      1   503  20.41033 5.225223
```

```
# Means of all variables in data frame

        # Get names of all columns
        dat_cov <- names(Data)

        # Create summary table
        Data %>%
                group_by(Gender) %>%
                select(one_of(dat_cov)) %>%
                summarise_all(funs(mean(., na.rm = T), sd(., na.rm = T)))
```

```
## # A tibble: 2 x 9
##   Gender   P_mean Age_mean  X1_mean   X2_mean     P_sd   Age_sd    X1_sd
##    <dbl>    <dbl>    <dbl>    <dbl>     <dbl>    <dbl>    <dbl>    <dbl>
## 1      0 19.95874 49.22137 20.01193 0.3179074 4.966017 17.18427 4.859892
## 2      1 20.04677 50.44463 20.41033 0.2803181 5.059133 17.30838 5.225223
## # ... with 1 more variables: X2_sd <dbl>
```

This is some text! Hey looking at the table above, I can see that the Gender distribution is about equal, and the mean and SD of X1 are approximatly equal as well!