Laboratory 2 - Simulated Sampling

Andrew Judell-Halfpenny

September 14, 2018

```
> library(knitr);library(magrittr); library(markdown); library(dplyr)
> library(ggplot2)
> # cache chunks and do not tidy ggplot2 examples code
> #opts_chunk$set(tidy = FALSE, cache = TRUE)

> # r samp
> #There are a number of ways to get random numbers in R from a variety of distributions.
> #For our simulations, let's start with a sample of 40 individuals that's from a population
>
> set.seed(323)
> samp <- rnorm(n = 40, mean = 10, sd = 3)
> par(mar = c(4, 4, 1, 1), mgp = c(2, 1, 0), cex = 0.8)
> plot(cars, pch = 20, col = 'darkgray')
> fit <- lm(dist ~ speed, data = cars)
> abline(fit, lwd = 2)
>

> #This is *incredibly* useful in a wide variety of context. For example, let's say we wante
>
> # r sample_mtcars
> mtcars[sample(1:nrow(mtcars), 5), ]

                mpg cyl  disp  hp drat    wt  qsec vs am gear carb
Mazda RX4 Wag 21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
Porsche 914-2 26.0   4 120.3  91 4.43 2.140 16.70  0  1    5    2
Toyota Corona 21.5   4 120.1  97 3.70 2.465 20.01  1  0    3    1
Volvo 142E    21.4   4 121.0 109 4.11 2.780 18.60  1  1    4    2
Fiat 128      32.4   4  78.7  66 4.08 2.200 19.47  1  1    4    1

> # Of course, for that kind of thing, `dplyr` has you covered with
>
> #r sample_dplyr
> # sample_n(mtcars, 5)
>
> #  use a porportions - so, here's a random sampling of 10 percent of mtcars
>
```

```
> #r sample_dplyr_frac
> #sample_frac(mtcars, 0.1)
> sample(1:3, 15, replace = TRUE)

 [1] 3 1 2 2 3 2 1 2 3 3 3 1 3 3 3

> ### 3.2 Using `group_by()` for simulation
>
> sampSim <- data.frame(samp_size = rep(3:50, times = 10))
> # Take the data frame
> samp_size <- sampSim %>%
+    rowwise() %>%
+      mutate(samp.mean = mean(rnorm(samp_size,
+                                    mean = 10, sd = 3)),
+      samp_mean_from_sample = mean(sample(samp, size = samp_size, replace=T)))
> plot(samp.mean ~ samp_size, data=samp_size)
> # Also draw a mean from the samp vector with n = samp_size, with replacement
>
> # For each individual simulation
>
>
> #      group_by(sim_number) %>%
> # Use samp_size to get the mean of a random normal population with mean 10 and sd 3
> #  mutate(mean_pop = mean(rnorm(samp_size, mean = 10, sd = 3)),
>
> # Also draw a mean from the samp vector with n = samp_size, with replacement
> #          mean_sample = mean(sample(samp, size = samp_size, replace=T))) %>%
> # Cleanup
> #    ungroup()
>
> #### 3.3.1 Faded Examples.
> # Lets try this out, and have you fill in whats missing in these faded examples.
> # r faded_sim, eval=FALSE
> #Some preperatory material
> set.seed(42)
> samp <- rnorm(100, 10, 3)
> sampSim <- data.frame(samp_size = rep(3:50, times = 10))
> sampSim$sim_number = 1:nrow(sampSim)
> #Mean simulations
> sampSim %>%
+    group_by(sim_number) %>%
+    mutate(mean_pop = mean(rnorm(samp_size, mean = 10, sd = 3)),
+          mean_sample = mean(sample(samp, size = samp_size, replace=T))) %>%
+    ungroup()

# A tibble: 480 x 4
   samp_size sim_number mean_pop mean_sample
```

```
           <int>        <int>      <dbl>        <dbl>
1             3            1       11.2         6.04
2             4            2       10.6         9.23
3             5            3       10.2         9.22
4             6            4        9.58        6.93
5             7            5        7.83       11.8
6             8            6        8.49        9.26
7             9            7       10.4         9.35
8            10            8        9.21       11.8
9            11            9        9.47        9.37
10           12           10       10.9         9.61
# ... with 470 more rows

>
```