Lab 10

Lab Overview

This lab will look at sampling with unequal weights using natural gas deliveries. The dataset contains natural gas deliveries (in million cubic feet) for 56 utilities in Montana, Wyoming, Idaho, Utah, Colorado, North Dakota, and South Dakota.

```
NG <- read_csv('http://www.math.montana.edu/ahoegh/teaching/stat446/EIA_NG.csv')
NG
```

```
## # A tibble: 56 x 4
##    Company                         total2018 total2017 weight
##    <chr>                               <dbl>     <dbl>  <dbl>
##  1 PUB SERVICE CO OF COLORADO         253050.   225147. 0.231
##  2 QUESTAR GAS COMPANY               185972.   175728. 0.180
##  3 COLORADO INTERSTATE GAS COMPANY LLC 105885.    93805. 0.0962
##  4 NORTHWESTERN ENERGY                80145.    76020. 0.0780
##  5 INTERMOUNTAIN GAS COMPANY          68606.    70353. 0.0722
##  6 BLACK HILLS ENERGY                 62054.    51519. 0.0528
##  7 MONTANA DAKOTA UTILITIES CO        49411.    44459. 0.0456
##  8 COLORADO SPRINGS UTILITIES         40102.    32914. 0.0338
##  9 WBI ENERGY TRANSMISSION INC        29335.    25123. 0.0258
## 10 ATMOS ENERGY CORPORATION           18016.    17137. 0.0176
## # ... with 46 more rows
```

Rather than explicitly calculating the variance for each scenario, we will use the repeated sampling technique from earlier in the class.

1. (4 points)

Take a simple random sample of size 20 and estimate the total natural gas delivered in this region for the year 2018.

2. (4 points)

Now repeat the process from part 1 and take 1000 different samples of size 20 and plot the resultant point estimator along with true value.

3. (4 points)

Take a sample of size 20 with unequal probability using the `weight` column in the dataset. Then estimate the total natural gas delivered in this region for the year 2018 using the Horvitz-Thompson framework.

4. (4 points)

Now repeat the process from part 3 and take 1000 different samples of size 20 and plot the resultant point estimator along with true value.

5. (4 points)

Comment on the strengths and weaknesses of the approaches in part 1 and part 3.