

Report1

James Yang, Eunho Colin Pi

4/18/2019

```
library(survey)
```

```
## Loading required package: grid
## Loading required package: Matrix
## Loading required package: survival
##
## Attaching package: 'survey'
## The following object is masked from 'package:graphics':
##
##      dotchart
```

```
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
```

```
library(ggplot2)
```

```
set.seed(23)
```

```
# sample 10 grids from each stratum (4 * 9 grids to choose from for each stratum)
```

```
s1 <- 1:36
```

```
s2 <- 1:36
```

```
unit1 <- sample(s1, 10) # 21 8 12 24 27 14 29 30 33 32
```

```
unit2 <- sample(s2, 10) # 32 25 14 11 28 5 16 18 27 17
```

```
# use sampled data from each grid to get estimate of total population
```

```
sample_data1 <- c(122,125,105,130,111,119,114,99,113,107)
```

```
sample_data2 <- c(61, 53, 57, 65, 58, 76, 56, 51, 67, 56)
```

```
crowd <- data.frame(strata=c(rep(1, 10), rep(2, 10)))
```

```
crowd$size <- c(sample_data1, sample_data2)
```

```
crowd$n <- 10
```

```
crowd$N <- 36
```

```
crowd$wts <- 36/10
```

```
# create survey object and compute estimates
```

```

crowd.strat <- svydesign(id = ~1, weights = ~wts, fpc = ~N, strata = ~strata, data = crowd)
svytotal(~size, crowd.strat)

##      total      SE
## size  6282 117.81
# compare variance for both samples
var(sample_data1)

## [1] 92.05556
var(sample_data2)

## [1] 56.22222
crowd$strata = as.factor(crowd$strata)

ggplot(data = crowd) +
  geom_boxplot(aes(x = strata, y = size))

```

