Task 1

Ariella Fuzaylov

4/10/2021

```
Employee_A_data=read.csv("Employee_A_data.csv", header=TRUE)
```

Given

- $N = 40,041$
- $n = 6,000$
- SRS of $6,000$ reviews

Subtask 1:

Estimate average rating

```
N = 40041
n= 6000
mu<-sum(Employee_A_data$Rating)/n
```

$\hat{\mu} = 4.2226667$

Confidence interval

```
srs_design = svydesign(id=~1,data=Employee_A_data, fpc=rep(N,n))

svymean(x=~Rating,design = srs_design)
```

```
##          mean     SE
## Rating 4.2227 0.0125
```

```
conf= confint(svymean(x=~Rating,design = srs_design))
conf
```

```
##            2.5 %   97.5 %
## Rating 4.198217 4.247116
```

```r
Employee_A_data%>% group_by(Branch)%>%
  summarise(n= n(), Mean= mean(Rating),StD=sd(Rating))
```

```
## # A tibble: 3 x 4
##   Branch                    n  Mean   StD
##   <chr>                 <int> <dbl> <dbl>
## 1 Disneyland_California  2769  4.40 0.952
## 2 Disneyland_HongKong    1321  4.21 0.937
## 3 Disneyland_Paris       1910  3.98 1.19
```

```r
Employee_A_data%>% group_by(continent)%>%summarise(n= n(), Mean= mean(Rating),StD=sd(Rating))
```

```
## # A tibble: 5 x 4
##   continent     n  Mean   StD
##   <chr>     <int> <dbl> <dbl>
## 1 Africa       66  4.15 1.15
## 2 Americas   2413  4.32 1.01
## 3 Asia        987  4.27 0.932
## 4 Europe     1772  4.04 1.15
## 5 Oceania     762  4.28 0.997
```