P8130_hw3_wl2829

Wentong

10/15/2021

Problem 1

Read the dataset and save the sample.

```
population = read.csv("./ce8130entire.csv")
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
```

```
set.seed(1000)
A = population %>%
  group_by(sex) %>%
  sample_n(100)
```

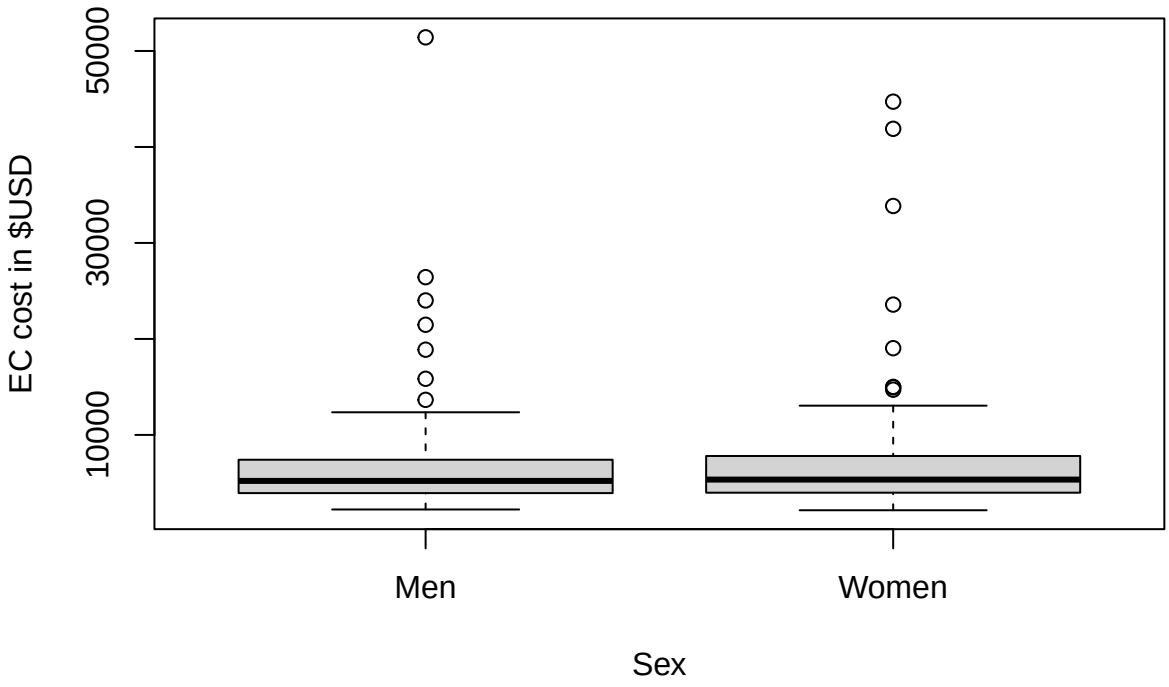
Problem 2

Save a separate sample.

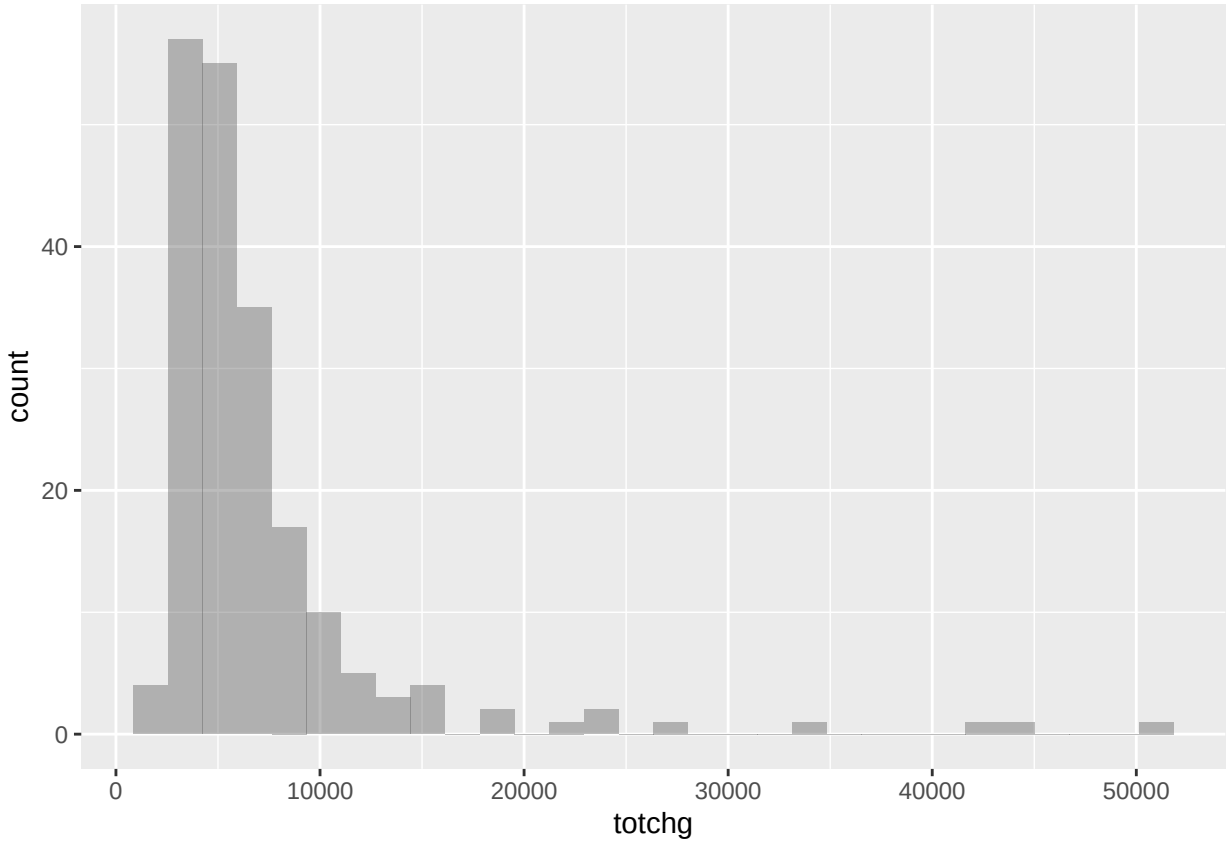
```
set.seed(1000)
B = population %>%
  group_by(sex) %>%
  sample_n(30)
```

Problem 3

```
boxplot(A$totchg ~ A$sex, names=c("Men","Women"), ylab ="EC cost in $USD", xlab ="Sex")
library(ggplot2)
```




```
ggplot(A, aes(x = totchg)) +
geom_histogram(aes(color = sex, fill = sex),
position = "identity", bins = 30, alpha = 0.4)
```



Problem 4

```
A$sample = "Sample A"
B$sample = "Sample B"
A_B = rbind(A, B)

library(Rmisc)
```

```
## Loading required package: lattice

## Loading required package: plyr

## ------------------------------------------------------------------------------

## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)

## ------------------------------------------------------------------------------

##
## Attaching package: 'plyr'
```

```
## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

```
A_B_summary <- summarySE(A_B, measurevar="totchg", groupvars=c("sex","sample"))
A_B_summary$sex <- as.factor(A_B_summary$sex)
p_dodge = position_dodge(0.1) # move them .05 to the left and right
plot = ggplot(A_B_summary, aes(x=sample, y=totchg, colour=sex)) +
geom_errorbar(aes(ymin=totchg-ci, ymax=totchg+ci), width=.1, position=p_dodge) +
geom_point(position=p_dodge)
```

How do they differ, which confidence intervals are wider? Explain why.

Problem 5

```
var.test(totchg ~ sex, data = A)
```

```
##
##  F test to compare two variances
##
## data:  totchg by sex
## F = 0.80342, num df = 99, denom df = 99, p-value = 0.2779
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.540577 1.194076
## sample estimates:
## ratio of variances
##          0.8034238
```

Problem 6

```
A_men =
  A %>%
  filter(sex == 1)

A_women =
  A %>%
  filter(sex == 2)

t.test(A_men$totchg, A_women$totchg, paired = TRUE, alternative = "l")
```

```
##
##  Paired t-test
##
## data:  A_men$totchg and A_women$totchg
## t = -0.57949, df = 99, p-value = 0.2818
## alternative hypothesis: true difference in means is less than 0
```

```
## 95 percent confidence interval:
##      -Inf 946.6114
## sample estimates:
## mean of the differences
##                 -507.49
```

Problem 7

```
res <- t.test(totchg ~ sex, data = A, var.equal = FALSE)
```

Problem 8

```
population_man =
  population %>%
  filter(sex == 1)

population_woman =
  population %>%
  filter(sex == 2)
man_mean = mean(population_man$totchg)
woman_mean = mean(population_woman$totchg)
mean_dif = man_mean - woman_mean
```