Problem Set 4

Topics in Advanced Econometrics (ResEcon 703)
University of Massachusetts Amherst

Solutions

Rules

Email a single .pdf file of your problem set writeup, code, and output to mwoerman@umass.edu by the date and time above. You may work in groups of up to three and submit one writeup for the group, and I strongly encourage you to do so. You can use any "canned" routine (e.g., lm(), glm(), and mlogit()) for this problem set.

Data

Download the file camping_dataset.zip from the course website. This zipped file contains the dataset camping.csv, which you will use for this problem set. This dataset contains simulated data on the state park choice of 1000 visitors who camped at one of five Massachusetts State Parks. See the file camping_description.txt for a description of the variables in the dataset.

```
### Load packages for problem set
library(tidyverse)
library(mlogit)
```

```
## Load dataset
data_camping <- read_csv('camping.csv')

## Rows:  5000 Columns:  8
## - Column specification --------------------------
## Delimiter:  ","
## chr  (1):  park
## dbl  (7):  camper_id, park_id, visit, mountain, beach, cost, time
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

Problem 1: Generalized Extreme Value Models

We are studying how campers at Massachusetts State Parks choose the park where they camp, which will assist the Massachusetts Department of Conservation and Recreation (DCR) in their planning. In particular, they want to understand how campers value their time to travel to the park and the setting—mountain or beach—of the park. Additionally, DCR is considering an increase to the camping fee at Mount Greylock due to the cost of maintaining those camp sites, and they want to know how this change would affect park visitation patterns.

a. Model the camping park choice as a multinomial logit model. Express the representative utility of each alternative as a linear function of its cost, time, and setting—mountain or beach—with common parameters on each variable. That is, the representative utility to camper n from park j is

$$V_{nj} = \beta_1 C_{nj} + \beta_2 T_{nj} + \beta_3 M_j$$

where C_{nj} is the cost to camper n of traveling to and camping at park j, T_{nj} is the time for camper n to travel to park j, M_j is a binary indicator if park j is in the mountains, and the β parameters are to be estimated. Importantly, do not include alternative-specific intercepts because β_3 would not be identified. (Reminder: the `mlogit()` function from the `mlogit` package estimates a multinomial logit model, but the data must first be converted to an indexed data frame using the `dfidx()` function from the `dfidx` package. See the Week 4 slides or the `mlogit` vignettes at `cran.r-project.org/web/packages/mlogit/index.html` for information on specifying a `formula` for the `mlogit()` function.)

```r
## Convert dataset to dfidx format
data_dfidx <- dfidx(data = data_camping, shape = 'long',
                    choice = 'visit', idx = c('camper_id', 'park_id'))
# Model camping park visit as a multinomial logit
model_1a <- mlogit(formula = visit ~ cost + time + mountain | 0,
                   data = data_dfidx)
```

i. Report the estimated parameters and standard errors from this model. Briefly interpret these results. For example, what does each parameter mean?

```r
## Summarize model results
summary(model_1a)
```

```
##
## Call:
## mlogit(formula = visit ~ cost + time + mountain | 0, data = data_dfidx,
##     method = "nr")
##
## Frequencies of alternatives:choice
##     1     2     3     4     5
## 0.274 0.202 0.209 0.167 0.148
##
## nr method
## 5 iterations, 0h:0m:0s
## g'(-H)^-1g = 3.77E-05
## successive function values within tolerance limits
##
## Coefficients :
##             Estimate  Std. Error z-value  Pr(>|z|)
## cost     -0.01478389  0.00368970 -4.0068 6.155e-05 ***
## time     -0.00163201  0.00040778 -4.0022 6.275e-05 ***
## mountain -0.33105186  0.17595101 -1.8815    0.0599 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log-Likelihood: -1524.2
```

The cost and time parameters—β_1 and β_2, respectively—are statistically significant and can be interpreted as marginal utilities. The cost of camping at a park reduces the utility of camping there, and the time spent traveling to the park reduces the utility of camping there. The mountain parameter, β_3, is statistically significant at the 10% level and is interpreted as the relative utility of camping in the mountains. The campers in our dataset obtain less utility, *ceteris paribus*, from camping in the mountains than from camping at the beach.

ii. Calculate the dollar value that a camper places on each hour spent traveling and the dollar value that a camper places on camping in the mountains (relative to camping at the beach).

```
## Calculate value of time and mountain park
coef(model_1a)[2:3] / coef(model_1a)[1] * c(60, -1)

##       time   mountain
##   6.623449 -22.392742
```

These campers value their time spent traveling to the park at $6.62 per hour, and they value camping in the mountains at $22.39 less than camping at the beach.

iii. Calculate the elasticity of choosing each park with respect to the cost of camping at Mount Greylock (`park_id == 1`) for each camper; that is, 5 alternatives × 1000 campers = 5000 elasticities. For each park, report the mean of its elasticity with respect to the cost of camping at Mount Greylock. Describe how these elasticities and substitution patterns relate to an important property of the logit model. (Reminder: the `fitted()` function with argument `type = 'probabilities'` calculates the choice probabilities of each alternative for each decision maker.)

```
## Calculate mean elasticities with respect to the cost of alternative 1
data_camping %>%
  filter(park_id == 1) %>%
  mutate(prob = fitted(model_1a, type = 'probabilities')[, 1],
         own_elas = coef(model_1a)[1] * cost * (1 - prob),
         cross_elas = -coef(model_1a)[1] * cost * prob) %>%
  summarize(own_elas = mean(own_elas),
            cross_elas = mean(cross_elas))

## # A tibble: 1 x 2
##   own_elas cross_elas
##      <dbl>      <dbl>
## 1   -0.676      0.223
```

The mean elasticity of camping at Mount Greylock with respect to its cost is -0.676, and the mean elasticity of camping at any of the other four parks with respect to the cost of camping at Mount Greylock is 0.223. This model implies that campers will substitute to the other park in proportion to their observed visits. In other words, campers will substitute to other parks with no consideration for whether those parks share attributes with Mount Greylock. This proportional substitution is an example of the rigid substitution patterns imposed by the logit model.

b. The multinomial logit model of part (a) is not the best model for this setting if a camper's unobserved (and random) utility includes an individual preference for the mountains or the beach, which would create correlations among parks with the same setting. Model the camping park choice as a nested logit model with two nests, one for each park setting: mountains and beach. As in part (a), model the representative utility for park j as

$$V_{nj} = \beta_1 C_{nj} + \beta_2 T_{nj} + \beta_3 M_j$$

where C_{nj} is the cost to camper n of traveling to and camping at park j, T_{nj} is the time for camper n to travel to park j, M_j is a binary indicator if park j is in the mountains, and the β parameters are to be estimated. (Reminder: the `mlogit()` function from the `mlogit` package estimates a nested logit model if you use the `nests` argument to specify nests as a named list.)

```r
## Model camping park visit as a nested logit
model_1b <- mlogit(formula = visit ~ cost + time + mountain | 0,
                   data = data_dfidx,
                   nests = list(mountain = 1:2, beach = 3:5))
```

i. Report the estimated parameters and standard errors from this model. Briefly interpret these results. For example, what does each parameter mean?

```r
## Summarize model results
summary(model_1b)
```

```
##
## Call:
## mlogit(formula = visit ~ cost + time + mountain | 0, data = data_dfidx,
##     nests = list(mountain = 1:2, beach = 3:5))
##
## Frequencies of alternatives:choice
##     1     2     3     4     5
## 0.274 0.202 0.209 0.167 0.148
##
## bfgs method
## 12 iterations, 0h:0m:0s
## g'(-H)^-1g = 1.65E-07
## gradient close to zero
```

```
##
## Coefficients :
##                Estimate  Std. Error z-value  Pr(>|z|)
## cost        -0.00607442  0.00151849 -4.0003 6.326e-05 ***
## time        -0.00146685  0.00024578 -5.9681 2.400e-09 ***
## mountain    -0.19944206  0.09847759 -2.0253   0.04284 *
## iv:mountain  0.27206879  0.05790001  4.6989 2.615e-06 ***
## iv:beach     0.31647461  0.06018706  5.2582 1.455e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log-Likelihood: -1500.9
```

The cost, time, and mountain parameters—β_1, β_2, and β_3, respectively—are interpreted as they were in part (a), although the mountain parameter is now statistically significant at the 5% level. This model contains two additional parameters, the coefficients on the inclusive value for each nest, which are statistically significant. These parameters represent the independence of random utility within each nest, with a value of 1 indicating full independence and a value close to 0 indicating high dependence. Thus, we conclude there is some dependence within each of the nests. That is, the random utility that a camper obtains from camping at each of the parks in the mountains is correlated, as is the random utility from camping at each of the parks at the beach. This result is consistent with each camper having an individual preference for the mountains or the beach beyond the average preference represented by β_3, the common mountain parameter.

ii. The model in part (a) is effectively imposing a restriction on the model in part (b). Write the null hypothesis that is imposed by the model in part (a) and describe this hypothesis in words. Conduct a likelihood ratio test to test this null hypothesis. Do you reject this null hypothesis? What is the p-value of the test? Briefly interpret the result of this test. (Reminder: the `lrtest()` function performs a likelihood ratio test.)

```
## Conduct likelihood ratio test of the models in parts b and d
lrtest(model_1a, model_1b)
```

```
## Likelihood ratio test
##
## Model 1: visit ~ cost + time + mountain | 0
## Model 2: visit ~ cost + time + mountain | 0
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1   3 -1524.2
## 2   5 -1500.9  2 46.564  7.739e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The null hypothesis is that the inclusive value coefficients equal 1 for both nests:

$$H_0\colon\ \lambda_{mountain} = \lambda_{beach} = 1$$

In words, this null hypothesis imposes that there is no correlation among the random utility terms, or that the random utility of every alternative is i.i.d. But if a camper has an individual preference for camping in the mountains or at the beach beyond the average preference represented by β_3, this hypothesis would not hold. We reject this hypothesis with a p-value of approximately 0, indicating that we have strong statistical evidence to conclude there are correlations among the random utility terms within each nest.

iii. Calculate the dollar value that a camper places on each hour spent traveling and the dollar value that a camper places on camping in the mountains (relative to camping at the beach).

```
## Calculate value of time and mountain park
coef(model_1b)[2:3] / coef(model_1b)[1] * c(60, -1)
```

```
##     time mountain
##  14.4888 -32.8331
```

These campers value their time spent traveling to the park at $14.49 per hour, and they value camping in the mountains at $32.83 less than camping at the beach.

iv. Calculate the elasticity of choosing each park with respect to the cost of camping at Mount Greylock (`park_id == 1`) for each camper; that is, 5 alternatives × 1000 campers = 5000 elasticities. For each park, report the mean of its elasticity with respect to the cost of camping at Mount Greylock. Compare these elasticities to those you found in part (a) and describe any important differences.

```
## Calculate choice probabilities for every alternative
probs_1b <- fitted(model_1b, type = 'probabilities')
## Calculate mean elasticities with respect to the cost of alternative 1
data_camping %>%
  filter(park_id == 1) %>%
  mutate(prob = probs_1b[, 1],
         prob_nest = rowSums(probs_1b[, 1:2]),
         prob_cond = prob / prob_nest,
         own_elas = coef(model_1b)[1] * cost *
           ((1 / coef(model_1b)[4]) -
              ((1 - coef(model_1b)[4]) /
                 coef(model_1b)[4] * prob_cond) -
              prob),
         cross_elas_mountain = -coef(model_1b)[1] * cost * prob *
           (1 + ((1 - coef(model_1b)[4]) /
                   (coef(model_1b)[4] * prob_nest))),
         cross_elas_beach = -coef(model_1b)[1] * cost * prob) %>%
  summarize(own_elas = mean(own_elas),
            cross_elas_mountain = mean(cross_elas_mountain),
            cross_elas_beach = mean(cross_elas_beach))
```

```
## # A tibble: 1 x 3
##   own_elas cross_elas_mountain cross_elas_beach
##      <dbl>               <dbl>            <dbl>
## 1   -0.709               0.649           0.0952
```

The mean elasticity of camping at Mount Greylock with respect to its cost is -0.709; the mean elasticity of camping at October Mountain with respect to the cost of camping at Mount Greylock is 0.649; and the mean elasticity of camping at any of beach parks with respect to the cost of camping at Mount Greylock is 0.095. This model implies that campers will substitute to October Mountain in much greater proportion than to the beach parks. These elasticities are more intuitive than the elasticities in part (a)—which imposed proportional substitution—if campers have individual preferences for camping in the mountains.

Problem 2: Mixed Logit Model

The models in problem 1 have common parameters for all campers in the dataset. In reality, however, some or all of these parameters are likely to vary by camper for unobserved reasons. DCR is interested in understanding this heterogeneity and how it could affect park visitation patterns.

a. Model the camping park choice as a mixed logit model. Express the representative utility of each alternative as a linear function of its cost, time, and setting—mountain or beach—with random coefficients on each variable. That is, the representative utility to camper n from park j is

$$V_{nj} = \beta_{1n} C_{nj} + \beta_{2n} T_{nj} + \beta_{3n} M_j$$

where C_{nj} is the cost to camper n of traveling to and camping at park j, T_{nj} is the time for camper n to travel to park j, M_j is a binary indicator if park j is in the mountains. Model all three β coefficients as random with a normal distribution:

$$\beta_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$$
$$\beta_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$$
$$\beta_3 \sim \mathcal{N}(\mu_3, \sigma_3^2)$$

Estimate this model using 100 draws for simulation (`R = 100`) and set a seed of 703 for replication (`seed = 703`). (Reminder: the `mlogit()` function from the `mlogit` package estimates a mixed logit model if you use the `rpar` argument to specify the random coefficients as a named vector.)

```
## Model camping park visit as a mixed logit
model_2a <- mlogit(formula = visit ~ cost + time + mountain | 0,
                   data = data_dfidx,
                   rpar = c(cost = 'n', time = 'n', mountain = 'n'),
                   R = 100, seed = 703)
```

i. Report the estimated parameters and standard errors from this model. Briefly interpret these results. For example, what does each parameter mean?

```
## Summarize model results
summary(model_2a)

##
## Call:
## mlogit(formula = visit ~ cost + time + mountain | 0, data = data_dfidx,
##     rpar = c(cost = "n", time = "n", mountain = "n"), R = 100,
##     seed = 703)
##
## Frequencies of alternatives:choice
##     1     2     3     4     5
## 0.274 0.202 0.209 0.167 0.148
##
## bfgs method
## 17 iterations, 0h:0m:11s
## g'(-H)^-1g = 3.19E-07
## gradient close to zero
##
## Coefficients :
##              Estimate Std. Error z-value  Pr(>|z|)
## cost        -0.0214846  0.0042303 -5.0788 3.799e-07 ***
## time        -0.0061720  0.0010827 -5.7007 1.193e-08 ***
## mountain    -0.8444033  0.3083382 -2.7386  0.006171 **
## sd.cost      0.0126612  0.0166059  0.7624  0.445791
## sd.time      0.0036772  0.0015353  2.3951  0.016616 *
## sd.mountain -5.6000997  1.1536954 -4.8541 1.210e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log-Likelihood: -1504.5
##
## random coefficients
##          Min.      1st Qu.       Median         Mean      3rd Qu. Max.
## cost     -Inf -0.030024408 -0.021484581 -0.021484581 -0.012944754  Inf
## time     -Inf -0.008652227 -0.006171992 -0.006171992 -0.003691758  Inf
## mountain -Inf -4.621613130 -0.844403300 -0.844403300  2.932806531  Inf
```

The cost coefficient, β_1, is interpreted as the marginal utility of the cost of the camping trip, or the negative of the marginal utility of money. We model this coefficient as being random with a normal distribution, meaning that different campers can have a different marginal utility of money. We estimate that this coefficient has a mean of -0.021 and a standard deviation of 0.013. The standard deviation parameter is not statistically significant, however, so we cannot conclude that this coefficient has any variance, suggesting it may be better to model it as a fixed coefficient. The other parameters have similar interpretations, and all other parameters are statistically significant. The time coefficient parameters, μ_2 and σ_2^2, indicate that the marginal utility of time traveling to camp is normally distributed with a mean of -0.0062 and a standard deviation of 0.0037. The mountain coefficient parameters, μ_3 and σ_3^2, indicate that the utility obtained by camping in the mountains, relative to camping at the beach and *ceteris paribus*, is normally distributed with a mean of -0.84 and a standard deviation of 5.6.

b. It is easier to calculate how campers value their time to travel to the park and the setting—mountain or beach—of the park when cost has a fixed (not-random) coefficient. Model the camping park choice as a mixed logit model with a fixed coefficient on cost. Express the representative utility of each alternative as a linear function of its cost, time, and setting—mountain or beach—with random coefficients on time and mountain. That is, the representative utility to camper n from park j is

$$V_{nj} = \beta_1 C_{nj} + \beta_{2n} T_{nj} + \beta_{3n} M_j$$

where C_{nj} is the cost to camper n of traveling to and camping at park j, T_{nj} is the time for camper

n to travel to park j, M_j is a binary indicator if park j is in the mountains. Model β_1 as a fixed coefficient and β_2 and β_3 as random with a normal distribution:

$$\beta_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$$
$$\beta_3 \sim \mathcal{N}(\mu_3, \sigma_3^2)$$

Estimate this model using 100 draws for simulation (`R = 100`) and set a seed of 703 for replication (`seed = 703`).

```r
## Model camping park visit as a mixed logit with fixed cost coefficient
model_2b <- mlogit(formula = visit ~ cost + time + mountain | 0,
                   data = data_dfidx,
                   rpar = c(time = 'n', mountain = 'n'),
                   R = 100, seed = 703)
```

i. Report the estimated parameters and standard errors from this model. Briefly interpret these results. For example, what does each parameter mean?

   ```r
   ## Summarize model results
   summary(model_2b)
   ```

   ```
   ## 
   ## Call:
   ## mlogit(formula = visit ~ cost + time + mountain | 0, data = data_dfidx, 
   ##     rpar = c(time = "n", mountain = "n"), R = 100, seed = 703)
   ## 
   ## Frequencies of alternatives:choice
   ##     1     2     3     4     5 
   ## 0.274 0.202 0.209 0.167 0.148 
   ## 
   ## bfgs method
   ## 16 iterations, 0h:0m:21s 
   ## g'(-H)^-1g = 0.00651 
   ## successive function values within tolerance limits 
   ## 
   ## Coefficients :
   ##                Estimate  Std. Error z-value  Pr(>|z|)    
   ## cost        -0.01872287  0.00410751 -4.5582 5.159e-06 ***
   ## time        -0.00487803  0.00089451 -5.4533 4.945e-08 ***
   ## mountain    -0.82439901  0.28712024 -2.8713  0.004088 ** 
   ## sd.time      0.00194078  0.00100759  1.9262  0.054084 .  
   ## sd.mountain  4.67359368  1.00229284  4.6629 3.118e-06 ***
   ## ---
   ## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
   ## 
   ## Log-Likelihood: -1503.2
   ## 
   ## random coefficients
   ```

```
##          Min.      1st Qu.       Median         Mean      3rd Qu. Max.
## time     -Inf -0.006187071 -0.004878034 -0.004878034 -0.003568997  Inf
## mountain -Inf -3.976690042 -0.824399010 -0.824399010  2.327892022  Inf
```

The cost coefficient, β_1, is now fixed at a value of -0.019, indicating this value is the marginal utility of cost for all campers. The interpretation of the other parameters is similar to that in part (a). The time coefficient parameters, μ_2 and σ_2^2, indicate that the marginal utility of time traveling to camp is normally distributed with a mean of -0.0049 and a standard deviation of 0.0019. The mountain coefficient parameters, μ_3 and σ_3^2, indicate that the utility obtained by camping in the mountains, relative to camping at the beach and *ceteris paribus*, is normally distributed with a mean of -0.82 and a standard deviation of 4.7.

ii. We can test if β_1 is a fixed or random coefficient. Write the null hypothesis of your test and describe this hypothesis in words. Conduct a likelihood ratio test to test this null hypothesis. Do you reject this null hypothesis? What is the p-value of the test? Briefly interpret the result of this test.

```
## Conduct likelihood ratio test of the models in parts a and b
lrtest(model_2a, model_2b)
```

```
## Likelihood ratio test
##
## Model 1: visit ~ cost + time + mountain | 0
## Model 2: visit ~ cost + time + mountain | 0
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1   6 -1504.5
## 2   5 -1503.2 -1 2.5395      0.111
```

The null hypothesis is that the variance of the random β_1 coefficient equals 0:

$$H_0\colon\ \sigma_1^2 = 0$$

In words, this null hypothesis imposes that there is no heterogeneity in this cost coefficient, so all campers have the same marginal utility of cost or income. We fail to reject this hypothesis with a p-value of 0.11, so we conclude that the cost coefficient, β_1, is fixed for all campers.

iii. Calculate the dollar value that a camper places on each hour spent traveling and the dollar value that a camper places on camping in the mountains (relative to camping at the beach). Because we have distributions for β_2 and β_3, these dollar values will also be distributions. Report the mean and standard deviation of each of these dollar value distributions. Briefly interpret these results.

```
## Calculate distribution of the value of time
c(coef(model_2b)[2] / coef(model_2b)[1] * 60,
  abs(coef(model_2b)[4]) / -coef(model_2b)[1] * 60) %>%
  setNames(c('time', 'sd.time'))
```

```
##      time   sd.time
## 15.632328  6.219499
```

```
## Calculate distribution of the value of a mountain park
c(coef(model_2b)[3] / -coef(model_2b)[1],
  abs(coef(model_2b)[5]) / -coef(model_2b)[1]) %>%
  setNames(c('mountain', 'sd.mountain'))
```

```
##    mountain sd.mountain
##   -44.03166   249.61950
```

These campers have a heterogeneous valuation of their time spent traveling to the park. These time values are normally distributed with a mean of $15.63 and a standard deviation of $6.22. Similarly, the dollar value of camping in the mountains, relative to camping at the beach and *ceteris paribus*, is normally distributed with a mean of -$44.03 and a standard deviation of $249.62.

iv. Calculate the proportion of campers who have a positive value of camping in the mountains (relative to camping at the beach).

```
## Calculate proportion of visitors with a positive value of mountain parks
1 - pnorm(q = 0,
          mean = coef(model_2b)[3] / -coef(model_2b)[1],
          sd = abs(coef(model_2b)[5]) / -coef(model_2b)[1])
```

```
## [1] 0.4299918
```

Of the 1000 campers in this dataset, 43% have a positive valuation of camping in the mountains relative to camping at the beach. That is, *ceteris paribus*, 43% of these campers would prefer to camp in the mountains than at the beach.

c. DCR is considering an increase to the camping fee at Mount Greylock, which would increase the cost by $20 for each camper in our dataset. Use your parameter estimates from part (b) to simulate this counterfactual.

```
## Create counterfactual camping dataset
data_camping_counter <- data_camping %>%
  mutate(cost = if_else(park_id == 1, cost + 20, cost))
## Convert dataset to dfidx format
data_counter_dfidx <- dfidx(data = data_camping_counter, shape = 'long',
                            choice = 'visit', idx = c('camper_id', 'park_id'))
```

i. How many fewer campers—of the 1000 campers in this dataset—do you expect will camp at Mount Greylock because of this fee increase? How many more campers do you expect will camp at each of the other four parks?

```
## Calculate aggregate choices using observed data
agg_choices_obs_2b <- predict(model_2b, newdata = data_dfidx)
## Calculate aggregate choices using counterfactual data
agg_choices_counter_2b <- predict(model_2b, newdata = data_counter_dfidx)
## Calculate difference between aggregate choices
colSums(agg_choices_counter_2b - agg_choices_obs_2b)
```

```
##          1          2          3          4          5
## -47.108365  32.020511   5.906957   4.544644   4.636252
```

Due to this camping fee increase at Mount Greylock, we would expect 47.1 fewer campers at that park. We would also expect an additional 32.0 campers at the other mountain park, October Mountain, and approximately only 5 or 6 additional campers at each of the beach parks, as reported above.

ii. How do you expect this increased camping fee at Mount Greylock will affect the economic surplus of the 1000 campers in this dataset?

```r
## Calculate log-sum values using observed data
logsum_obs_2b <- logsum(model_2b, data = data_dfidx)
## Calculate log-sum values using counterfactual data
logsum_counter_2b <- logsum(model_2b, data = data_counter_dfidx)
## Calculate change in consumer surplus from subsidy
sum((logsum_counter_2b - logsum_obs_2b)) / -coef(model_2b)[1]

##      cost
## -4302.969
```

This increased camping fee is expected to reduce the economic surplus of these 1000 campers by a total of roughly $4303.