Loci example

load libraries

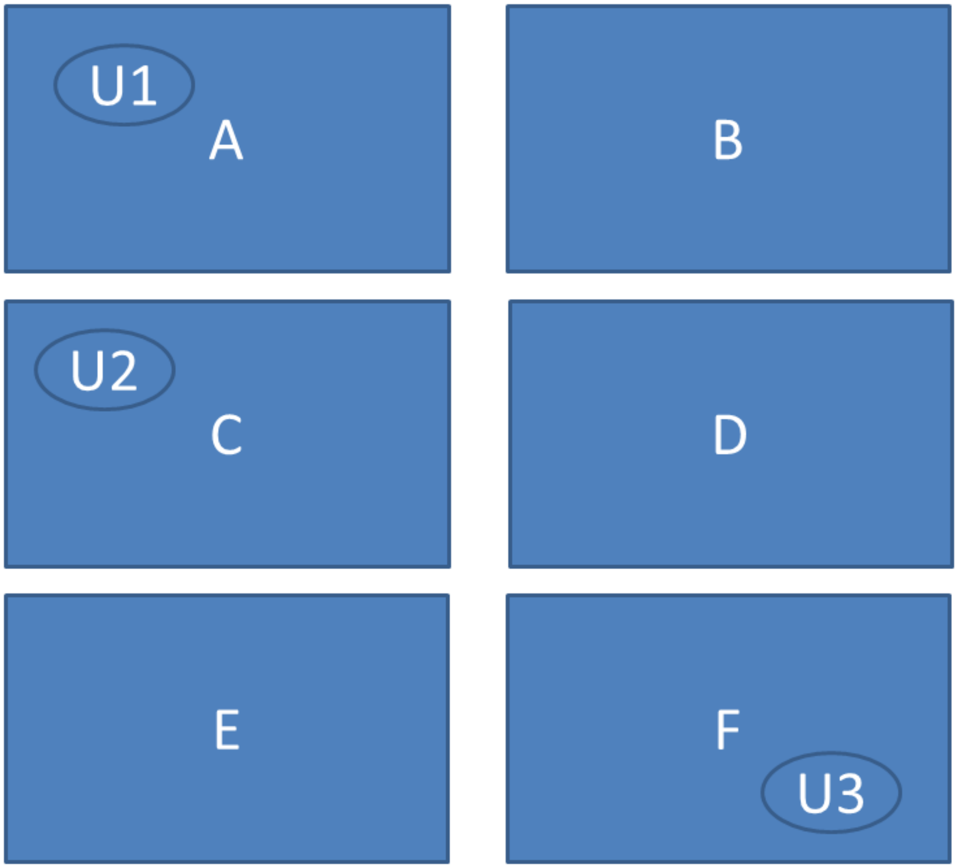
```
library(tidyverse)
library(mlogit)
library(ggrepel)
```

Introduction

The purpose of this notebook is to give an example how you can calculate the LOCI index.

Consider a simple University market consisting of three universities. Each university offers a Economics Msc programme. In this market we have students choosing their master program based on their preferences. Their preferences are determined by their own characteristics and on the university's characteristics. Let us say that an important factor for their preferences are the distance between where they are currenty living and the university and their gender.

Schematic discription of the market:



Currently, the prices are fixed by the government. Next year the government want to free the prices. We have been asked to do an analyses on the competitve pressure that the universities currently have on each other.

data

Let say we have acces to a dataset that contains the university choices that students have made in the past year, including their relevant characteristics.

Let's read them.

```
dt <- read.csv2("logitdata.csv")
```

Let's check it out

```
head(dt,9)
```

```
##   studentID zipcode gender  alternative time choice
## 1         1       A      1 University.1    0      1
## 2         1       A      1 University.2   20      0
## 3         1       A      1 University.3   70      0
## 4         2       B      1 University.1   20      1
## 5         2       B      1 University.2    0      0
## 6         2       B      1 University.3   55      0
## 7         3       C      1 University.1   30      0
```

```
## 8          3       C       1 University.2   20      0
## 9          3       C       1 University.3   40      1
```

For each student, has her unique ID: “studentID”. We can observe that for each student we have her “zipcode”, “gender” (1 is male), and the “time” (traveltime) from her zipcode to each University. We also have a column “choice”. When choice==1, it means that the student has chosen that corresponding university.

Let’s dig deeper into the data.

A summary of the data

```r
summary(dt)
```

```
##    studentID         zipcode              gender    alternative
##  Min.   :  1.00   Length:504         Min.   :0.0   Length:504
##  1st Qu.: 42.75   Class :character   1st Qu.:0.0   Class :character
##  Median : 84.50   Mode  :character   Median :0.5   Mode  :character
##  Mean   : 84.50                      Mean   :0.5
##  3rd Qu.:126.25                      3rd Qu.:1.0
##  Max.   :168.00                      Max.   :1.0
##       time           choice
##  Min.   : 0.00   Min.   :0.0000
##  1st Qu.:20.00   1st Qu.:0.0000
##  Median :20.00   Median :0.0000
##  Mean   :26.11   Mean   :0.3333
##  3rd Qu.:35.00   3rd Qu.:1.0000
##  Max.   :70.00   Max.   :1.0000
```

Number of students per University

```r
dt %>%
  filter(choice==1) %>%
  group_by(alternative) %>%
  summarise(n(), .groups = 'drop')
```

```
## # A tibble: 3 x 2
##   alternative  `n()`
##   <chr>        <int>
## 1 University.1    50
## 2 University.2    64
## 3 University.3    54
```

Number of students in each zipcode

```r
dt %>%
  filter(choice == 1) %>%
  group_by(zipcode) %>%
  summarise(n(), .groups = 'drop')
```

```
## # A tibble: 6 x 2
##   zipcode `n()`
##   <chr>   <int>
## 1 A          28
## 2 B          28
## 3 C          28
## 4 D          28
## 5 E          28
## 6 F          28
```

Number of students per gender

```
dt %>%
  filter(choice == 1) %>%
  group_by(gender) %>%
  summarise(n(), .groups = 'drop')
```

```
## # A tibble: 2 x 2
##   gender `n()`
##    <int> <int>
## 1      0    84
## 2      1    84
```

Underlying utility and calculate Loci

In this case we could have the following specification for the utility that a student of type t would receive if she would go to university j

$$V_{tj} = \beta_1 distance_{tj} + \beta_2 Gender_t distance_{tj} + \varepsilon_{tj}$$

Remember, the LOCI formula in terms of observed market shares is

$$\Lambda_j = \sum_t w_{tj}(1 - s_{tj})$$

```
weighted by the relative importance of each consumer type
```

$$w_{tj} = \frac{N_t s_{tj}}{\sum_t N_t s_{tj}}$$

Let's calculate the loci step by step.

first we want to create the micromarkets. Given that we have information that gender and traveltime play a role the in choice process of the students, we could make the micromarket based on student zipcode and gender.

```
micromarkets  <- dt %>%
    group_by(zipcode, gender,alternative) %>%
  summarise(n = sum(choice), .groups = 'drop') %>%
  ungroup()
```

```
head(micromarkets, 12)
```

```
## # A tibble: 12 x 4
##    zipcode gender alternative      n
##    <chr>    <int> <chr>        <int>
##  1 A            0 University.1    10
##  2 A            0 University.2     4
##  3 A            0 University.3     0
##  4 A            1 University.1    10
##  5 A            1 University.2     4
##  6 A            1 University.3     0
##  7 B            0 University.1     3
##  8 B            0 University.2    11
```

```
##   9 B            0 University.3     0
## 10 B            1 University.1     4
## 11 B            1 University.2    10
## 12 B            1 University.3     0
```

For each micromarket (based on zipcode and gender), we have now the number of students (N) that have chosen each alternative.

For the loci we have to calculate the shares s and weights w. For example, the share of university 1 in micromarket "A0" is defined by $10/(10 + 4 + 0)$. Let's calculate it for all students.

First, calculate the total N per micromarkt.

```
micromarkets <- micromarkets %>%
  group_by(zipcode, gender) %>%
  mutate(N = sum(n)) %>%
 ungroup()

head(micromarkets)
```

```
## # A tibble: 6 x 5
##   zipcode gender alternative      n     N
##   <chr>    <int> <chr>        <int> <int>
## 1 A            0 University.1    10    14
## 2 A            0 University.2     4    14
## 3 A            0 University.3     0    14
## 4 A            1 University.1    10    14
## 5 A            1 University.2     4    14
## 6 A            1 University.3     0    14
```

Secondly, calculate the shares.

```
micromarkets <- micromarkets %>%
  mutate(s = round(n / N, 2))

head(micromarkets)
```

```
## # A tibble: 6 x 6
##   zipcode gender alternative      n     N     s
##   <chr>    <int> <chr>        <int> <int> <dbl>
## 1 A            0 University.1    10    14 0.71
## 2 A            0 University.2     4    14 0.290
## 3 A            0 University.3     0    14 0
## 4 A            1 University.1    10    14 0.71
## 5 A            1 University.2     4    14 0.290
## 6 A            1 University.3     0    14 0
```

Let's do the weights w now. The weight of micromarket t for university j is the share of micromarket t in the total volume of university j. First we calculate per university the total number of students that have chosen that university.

```
micromarkets <- micromarkets %>%
    group_by(alternative) %>%
    mutate(N_University = sum(n)) %>%
    ungroup()

head(micromarkets)
```

```
## # A tibble: 6 x 7
```

```
##   zipcode gender alternative      n     N     s N_University
##   <chr>    <int> <chr>        <int> <int> <dbl>        <int>
## 1 A            0 University.1    10    14 0.71            50
## 2 A            0 University.2     4    14 0.290           64
## 3 A            0 University.3     0    14 0               54
## 4 A            1 University.1    10    14 0.71            50
## 5 A            1 University.2     4    14 0.290           64
## 6 A            1 University.3     0    14 0               54
```

Now calculate the weights

```
micromarkets <- micromarkets %>%
  mutate(w = round(n / N_University, 2))

head(micromarkets)
```

```
## # A tibble: 6 x 8
##   zipcode gender alternative      n     N     s N_University     w
##   <chr>    <int> <chr>        <int> <int> <dbl>        <int> <dbl>
## 1 A            0 University.1    10    14 0.71            50  0.2
## 2 A            0 University.2     4    14 0.290           64  0.06
## 3 A            0 University.3     0    14 0               54  0
## 4 A            1 University.1    10    14 0.71            50  0.2
## 5 A            1 University.2     4    14 0.290           64  0.06
## 6 A            1 University.3     0    14 0               54  0
```

We can now calculate

$$\Lambda_j = \sum_t w_{tj}(1 - s_{tj})$$

```
result_observed <- micromarkets %>%
    group_by(alternative) %>%
    summarise(Loci = sum(w*(1-s)), .groups = 'drop')

result_observed
```

```
## # A tibble: 3 x 2
##   alternative   Loci
##   <chr>        <dbl>
## 1 University.1 0.504
## 2 University.2 0.420
## 3 University.3 0.318
```

Estimate demand model and calculate Loci

A alternative way to calculate the LOCI is th first estimate a logit model and predict per student the probability that she would choose University.1, University.2 and University.3. There could be different reasons to do this. Maybe we want to consider a counterfactual situation (where for example individuals do not react on prices or quality). Or maybe we want to smooth out the probabilities to create smaller micromarkets. In general, estimating a choice model gives you more insight into which factors are important for the choice process.

We use the package "mlogit". We need to set the data into the "mlogit" format.

```
dt_mlogit <- mlogit.data(dt,
                         alt.levels = c(" University.1", "University.2", "University.3"),
```

```
                              id.var = "studentID")

head(dt_mlogit)
```

```
## ~~~~~~~
##  first 10 observations out of 504
## ~~~~~~~
##    studentID zipcode gender  alternative time choice          alt    idx
## 1          1       A      1 University.1    0      1 University.1 1:ty.1
## 2          1       A      1 University.2   20      0 University.2 1:ty.2
## 3          1       A      1 University.3   70      0 University.3 1:ty.3
## 4          2       B      1 University.1   20      1 University.1 2:ty.1
## 5          2       B      1 University.2    0      0 University.2 2:ty.2
## 6          2       B      1 University.3   55      0 University.3 2:ty.3
## 7          3       C      1 University.1   30      0 University.1 3:ty.1
## 8          3       C      1 University.2   20      0 University.2 3:ty.2
## 9          3       C      1 University.3   40      1 University.3 3:ty.3
## 10         4       D      1 University.1   20      1 University.1 4:ty.1
##
## ~~~ indexes ~~~~
##    chid          alt
## 1     1  University.1
## 2     1  University.2
## 3     1  University.3
## 4     2  University.1
## 5     2  University.2
## 6     2  University.3
## 7     3  University.1
## 8     3  University.2
## 9     3  University.3
## 10    4  University.1
## indexes:  1, 2
```

Estimate a simple conditional logit model, where we interact traveltime with gender. We thus have to following specefication in the logit model for each student i

$$V_{ij} = \beta_1 distance_{ij} + \beta_2 Gender_i distance_{ij} + \varepsilon_{ij}$$

```
m <- mlogit(choice ~ time + time:gender-1, data = dt_mlogit)
summary(m)
```

```
##
## Call:
## mlogit(formula = choice ~ time + time:gender - 1, data = dt_mlogit,
##     method = "nr")
##
## Frequencies of alternatives:choice
##  University.1  University.2  University.3
##       0.29762       0.38095       0.32143
##
## nr method
## 6 iterations, 0h:0m:0s
## g'(-H)^-1g = 3.33E-05
## successive function values within tolerance limits
```

```
##
## Coefficients :
##              Estimate Std. Error z-value  Pr(>|z|)
## time        -0.089289   0.015895 -5.6175 1.937e-08 ***
## time:gender  0.014043   0.020936  0.6708    0.5024
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log-Likelihood: -120.14
```

Let us predict the choices in steps.

Using the estimated coefficients we can calculate the "value" that student i receive when she would choose university j.

$$\hat{U}_{ij} = \hat{\beta}_1 distance_{ij} + \hat{\beta}_2 Gender_i distance_{ij}$$

Using these utilities we can calculate the probability that student t choose university j, denoted by $prob_{ij}$.

$$prob_{ij} = \frac{\exp(U_{ij})}{\sum_{g \in G} \exp(U_{ig})}$$

In our case we have that $G = University.1, University.2, University.3$

First get the estimated coeficients

```
coeficients<-as.numeric(summary(m)$coefficients)

coeficients
```

```
## [1] -0.08928894  0.01404346
```

Second, using our main data, calculate for each student and university $\hat{U}_{ij}$

```
dt <- dt %>%
  mutate(utility = round(exp(coeficients[1] * time + coeficients[2] * time * gender), 4))

head(dt)
```

```
##   studentID zipcode gender  alternative time choice utility
## 1         1       A      1 University.1    0      1  1.0000
## 2         1       A      1 University.2   20      0  0.2220
## 3         1       A      1 University.3   70      0  0.0052
## 4         2       B      1 University.1   20      1  0.2220
## 5         2       B      1 University.2    0      0  1.0000
## 6         2       B      1 University.3   55      0  0.0159
```

Next, we calculate $prob_{ij}$

```
dt <- dt %>%
  group_by(studentID) %>%
  mutate(utilitySum = sum(utility)) %>%
  ungroup() %>%
  mutate(prob = round(utility / utilitySum, 4))

head(dt)
```

```
## # A tibble: 6 x 9
##   studentID zipcode gender alternative   time choice utility utilitySum   prob
##       <int> <chr>    <int> <chr>        <int>  <int>   <dbl>      <dbl>  <dbl>
## 1         1 A            1 University.1     0      1  1            1.23 0.815
## 2         1 A            1 University.2    20      0  0.222        1.23 0.181
## 3         1 A            1 University.3    70      0  0.0052       1.23 0.0042
## 4         2 B            1 University.1    20      1  0.222        1.24 0.179
## 5         2 B            1 University.2     0      0  1            1.24 0.808
## 6         2 B            1 University.3    55      0  0.0159       1.24 0.0128
```

Now we can calculate the loci, following the steps that we have taken above.

Make the micromarket based on student zipcode and gender

```
micromarkets_pred  <- dt %>%
    group_by(zipcode, gender, alternative) %>%
  summarise(n = sum(prob), .groups = 'drop') %>%
  ungroup()
```

Calculate the total per micromarkt

```
micromarkets_pred <- micromarkets_pred %>%
  group_by(zipcode, gender) %>%
  mutate(N = sum(n)) %>%
 ungroup()
```

Calculate the shares

```
micromarkets_pred <- micromarkets_pred %>%
  mutate(s = round(n / N, 2))
```

Calculate total per university

```
micromarkets_pred <- micromarkets_pred %>%
    group_by(alternative) %>%
    mutate(N_University = sum(n)) %>%
    ungroup()
```

Now calculate the weights

```
micromarkets_pred <- micromarkets_pred %>%
  mutate(w = round(n / N_University, 2))
```

We can now calculate

$$\Lambda_j = \sum_t w_{tj}(1 - s_{tj})$$

based on estimated probabilities

```
result_predicted <- micromarkets_pred %>%
    group_by(alternative) %>%
    summarise(Loci = sum(w * (1-s)), .groups = 'drop')

result_predicted
```

```
## # A tibble: 3 x 2
##   alternative   Loci
##   <chr>        <dbl>
## 1 University.1 0.474
## 2 University.2 0.417
```

```
## 3 University.3 0.365
```

We will compare the results of the observed market shares versus the predicted market shares.

```r
graph_data <- left_join(result_observed, result_predicted, by = "alternative")
```

Plot:

```r
ggplot(data = graph_data, aes(x = Loci.x, y = Loci.y, label = alternative)) +
  geom_point(color = "firebrick", size = 2) +
  geom_abline(intercept = 0, slope = 1, color = "blue") +
  xlim(0, 1) +
  ylim(0,1) +
  geom_text_repel() +
  xlab("observed loci") +
  ylab("predicted loci") +
  theme_classic()
```