Chapter1

Haoran Xu

2025-01-06

```
if(!require("remotes"))
  install.packages("remotes",
                   repos = "https://cran.rstudio.org")
```

```
## Loading required package: remotes
```

```
if(!require("discrtr"))
  remotes::install_github("paezha/discrtr")
```

```
## Loading required package: discrtr
```

```
# renv::install("tidyverse")
# renv::install("dplyr")
# renv::install("mlogit")
# renv::install("stargazer")
# renv::install("dfidx")

library(discrtr)
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)
library(mlogit)
```

```
## Loading required package: dfidx
```

```
##
## Attaching package: 'dfidx'
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
library(readr)
library(stargazer)
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.
##  R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

```
# renv::install("tinytex")
library(tinytex)
```

```
# Read a csv file data and name the object
mc_mode_choice <- read_csv(system.file("extdata",
                                       "mc_commute.csv",
                                       package = "discrtr"),
                           show_col_types = FALSE)

summary(mc_mode_choice)
```

```
##   RespondentID           choice         avcycle           avwalk
##  Min.   :566872636   Min.   :1.000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:567814188   1st Qu.:2.000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :568682048   Median :2.000   Median :0.0000   Median :1.0000
##  Mean   :570566454   Mean   :2.618   Mean   :0.2747   Mean   :0.6613
##  3rd Qu.:574925212   3rd Qu.:3.000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :587675235   Max.   :4.000   Max.   :1.0000   Max.   :1.0000
##      avhsr             avcar          timecycle            timewalk
##  Min.   :0.0000   Min.   :0.0000   Min.   :     0.29   Min.   :     1.00
##  1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:     3.79   1st Qu.:    13.66
##  Median :1.0000   Median :1.0000   Median :     5.83   Median :    20.00
##  Mean   :0.9608   Mean   :0.5472   Mean   : 34014.86   Mean   : 37364.73
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:100000.00   3rd Qu.:100000.00
##  Max.   :1.0000   Max.   :1.0000   Max.   :100000.00   Max.   :100000.00
##    accesshsr     waitingtimehsr     transfer         timehsr
##  Min.   : 0.00   Min.   : 0.00   Min.   :     0   Min.   :     1.00
##  1st Qu.: 2.48   1st Qu.:10.23   1st Qu.:     0   1st Qu.:     4.75
##  Median : 6.21   Median :10.23   Median :     0   Median :    10.00
##  Mean   :11.06   Mean   :10.25   Mean   :  3925   Mean   :  3940.57
##  3rd Qu.:12.42   3rd Qu.:10.23   3rd Qu.:     1   3rd Qu.:    25.00
##  Max.   :62.11   Max.   :50.00   Max.   :100000   Max.   :100000.00
##     timecar          parking            vehind          owncycle
##  Min.   :     1   Min.   :0.00000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:     8   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :    30   Median :0.00000   Median :0.0000   Median :0.0000
##  Mean   : 45283   Mean   :0.08358   Mean   :0.2565   Mean   :0.4513
##  3rd Qu.:100000   3rd Qu.:0.00000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :100000   Max.   :1.00000   Max.   :1.0000   Max.   :1.0000
##      gender            work            visa             age
##  Min.   :0.0000   Min.   :0.000   Min.   :0.0000   Min.   :17.00
##  1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:1.0000   1st Qu.:20.00
##  Median :0.0000   Median :0.000   Median :1.0000   Median :21.00
##  Mean   :0.4012   Mean   :0.492   Mean   :0.9622   Mean   :22.08
##  3rd Qu.:1.0000   3rd Qu.:1.000   3rd Qu.:1.0000   3rd Qu.:23.00
##  Max.   :1.0000   Max.   :1.000   Max.   :1.0000   Max.   :60.00
##       solo            shared          family           child
##  Min.   :0.0000   Min.   :0.000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :1.000   Median :0.0000   Median :0.0000
##  Mean   :0.1272   Mean   :0.625   Mean   :0.2478   Mean   :0.2115
```

```
##  3rd Qu.:0.0000   3rd Qu.:1.000   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.000   Max.   :1.0000   Max.   :1.0000
##  primary_caregiver      LAT             LONG            DAUID
##  Min.   :     0    Min.   :43.08   Min.   :-80.09   Min.   :35250031
##  1st Qu.:100000    1st Qu.:43.25   1st Qu.:-79.92   1st Qu.:35250540
##  Median :100000    Median :43.26   Median :-79.91   Median :35250670
##  Mean   : 75218    Mean   :43.25   Mean   :-79.90   Mean   :35250612
##  3rd Qu.:100000    3rd Qu.:43.26   3rd Qu.:-79.90   3rd Qu.:35250677
##  Max.   :100000    Max.   :43.28   Max.   :-79.64   Max.   :35250970
##       mhi            dwell_den            lum             st_den
##  Min.   : 0.000   Min.   :    0.0   Min.   :0.0000   Min.   : 0.00
##  1st Qu.: 4.577   1st Qu.:  488.7   1st Qu.:0.2808   1st Qu.:10.36
##  Median : 5.491   Median :  950.0   Median :0.4501   Median :14.29
##  Mean   : 6.168   Mean   : 1373.0   Mean   :0.4183   Mean   :13.27
##  3rd Qu.: 7.556   3rd Qu.: 1688.6   3rd Qu.:0.5038   3rd Qu.:16.18
##  Max.   :14.595   Max.   :45209.9   Max.   :0.9081   Max.   :25.22
##    inter_den        SF_P_ratio        side_den      Shelters_SD
##  Min.   :  0.00   Min.   :0.0000   Min.   : 0.00   Min.   :0.00000
##  1st Qu.: 25.82   1st Qu.:0.2309   1st Qu.:18.19   1st Qu.:0.00000
##  Median : 41.04   Median :0.2709   Median :22.63   Median :0.00000
##  Mean   : 52.09   Mean   :0.2625   Mean   :24.18   Mean   :0.04433
##  3rd Qu.: 73.08   3rd Qu.:0.3134   3rd Qu.:35.70   3rd Qu.:0.00000
##  Max.   :645.86   Max.   :0.8808   Max.   :59.41   Max.   :1.00000
##    Shelters_D       Shelters_A      Shelters_SA
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.0000   Median :0.0000   Median :0.00000
##  Mean   :0.2289   Mean   :0.3576   Mean   :0.02762
##  3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.00000
```

```
# the weird thing is: only by adding "results = `asis`" will the PDF doc be rendered out the correct ta

# `stargazer()` takes as an input a data frame
stargazer(as.data.frame(mc_mode_choice[,1:5]),
          type = "latex",   # change the type to text, html, or latex depending on the desired output
          header = FALSE,
          title = "Example of a table with summary statistics",   # do not print package version info i
          omit.summary.stat = c("N",
                                "median"),   # summary statistics to omit from output
          font.size = "small")
```

Table 1: Example of a table with summary statistics

Statistic	Mean	St. Dev.	Min	Max
RespondentID	570,566,454.000	3,786,118.000	566,872,636	587,675,235
choice	2.618	0.845	1	4
avcycle	0.275	0.447	0	1
avwalk	0.661	0.473	0	1
avhsr	0.961	0.194	0	1

```
# Function `factor()` is used to convert a variable (which could be character or numeric)
# into a factor, that is, a label or category; when we want a factor to be ordered (treated
# as an ordinal variable) we specify argument ordered = TRUE. Non-ordinal variables by default
```

```r
# are displayed alphabetically, but changing their order when specifying the labels changes
# the order they are displayed _without necessarily making them ordinal_
mc_mode_choice$choice <- factor(mc_mode_choice$choice,
                                labels = c("Cycle",
                                           "Walk",
                                           "HSR",
                                           "Car"))

class(mc_mode_choice$choice)
```

```
## [1] "factor"
```

```r
class(mc_mode_choice$timecycle)
```

```
## [1] "numeric"
```

```r
mc_mode_choice[2, 2]
```

```
## # A tibble: 1 x 1
##   choice
##   <fct>
## 1 HSR
```

```r
mc_mode_choice$choice[2]
```

```
## [1] HSR
## Levels: Cycle Walk HSR Car
```

```r
mc_mode_choice[["choice"]][2]
```

```
## [1] HSR
## Levels: Cycle Walk HSR Car
```

```r
mc_mode_choice[2:5, 7:8]
```

```
## # A tibble: 4 x 2
##   timecycle timewalk
##       <dbl>    <dbl>
## 1      3.73     12.8
## 2 100000    100000
## 3      5.83     20
## 4      5.83     20
```

```r
time.Cycle.clean <- mc_mode_choice$timecycle[mc_mode_choice$timecycle != 100000]
class(time.Cycle.clean)
```

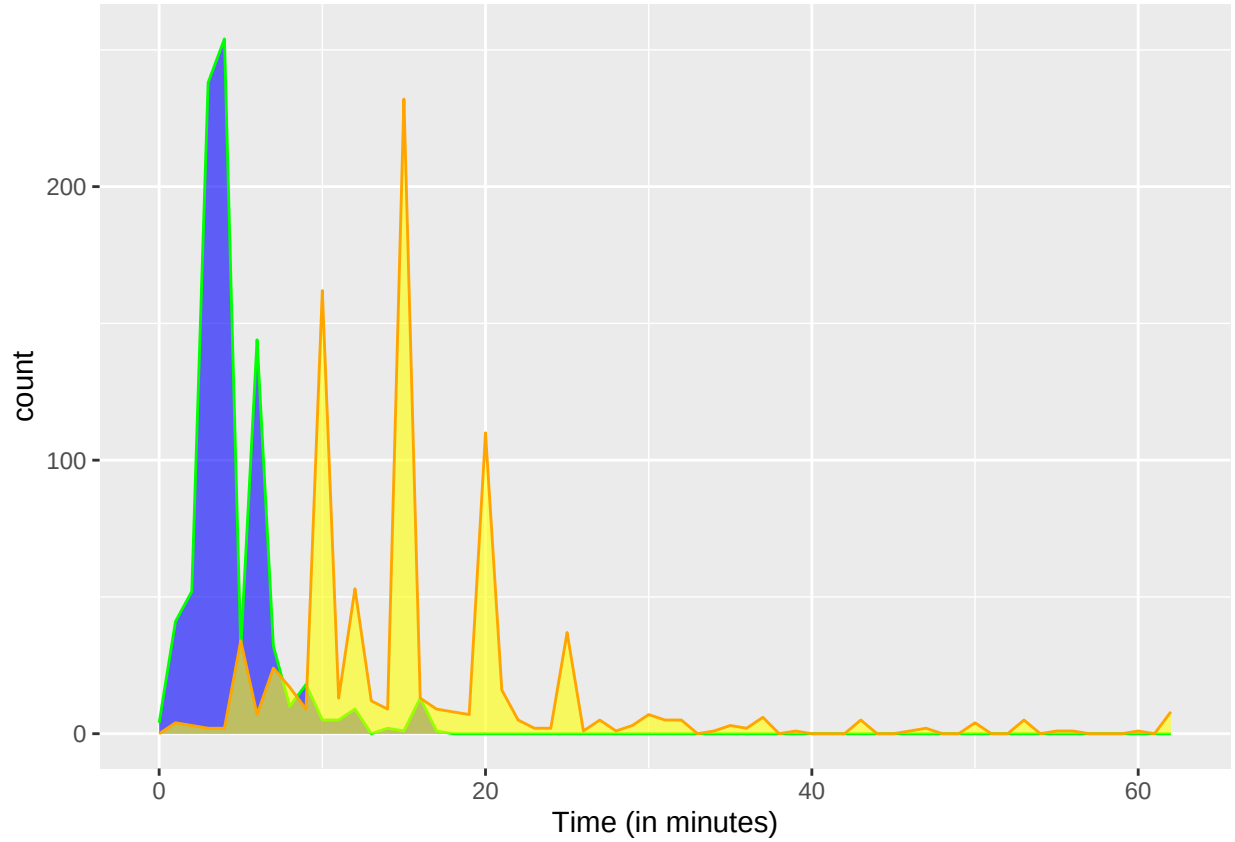
```
## [1] "numeric"
```

```r
# a better and more effient way using {dplyr}

time.Active.clean <- mc_mode_choice %>%  # Pipe data frame `mc_mode_choice`
  select(c("timecycle",   # Select columns from the data frame that was piped
           "timewalk")) %>%
  filter(timecycle != 100000 & timewalk != 100000)   # Filter observations that are _not_ 100000

summary(time.Active.clean)
```

```
##    timecycle          timewalk
##  Min.   : 0.2914   Min.   : 1.00
```

```
##  1st Qu.: 2.9141   1st Qu.:10.00
##  Median : 4.3711   Median :15.00
##  Mean   : 4.5852   Mean   :16.10
##  3rd Qu.: 5.8282   3rd Qu.:20.00
##  Max.   :17.4845   Max.   :62.11
```

```
ggplot(data = time.Active.clean) +
  geom_area(aes(x = timecycle),
            stat = "bin",    #  binned for smoothing
            binwidth = 1,  # smaller means less smoothing
            fill = "blue",
            color = "green",
            # alpha controls opacity (0 means transparent, 1 means solid)
            alpha = 0.6) +
  geom_area(aes(x = timewalk),
            stat = "bin",
            binwidth = 1,
            fill = "yellow",
            color = "orange",
            alpha = 0.6) +
  xlab("Time (in minutes)")
```




```
# pipe operator: %>%

mc_mode_choice %>%
  select(c("choice", "side_den")) %>%
  summary()
```

```
##    choice       side_den
##  Cycle: 48   Min.   : 0.00
##  Walk :711   1st Qu.:18.19
##  HSR  :336   Median :22.63
##  Car  :281   Mean   :24.18
##              3rd Qu.:35.70
##              Max.   :59.41
```

```
mc_mode_choice %>%
  ggplot(aes(x = choice,
             y = side_den)) +
  geom_boxplot()
```