Report Value of Time

Ludmilla Mattos e Rafael Morais

# Introdução

Exploratória dos dados de escolha entre viagens de Avião e Trem de Alta Velocidade - TAV.

# Dados

Os dados foram coletados….

[Contextualizar com parte do que está no relatório .-of-time-final-aeroportos.pdf]

## Importa e tratamento

library(tidyverse)  
library(readr)  
  
# import and consolidade data ---------------------------------------------  
files <- list.files(path = '../data', pattern = '\*.csv', full.names = T)  
  
dados1 <- read\_csv2(files[1], locale=locale(encoding="latin1")) %>% mutate(fonte = 'Amostra\_1')  
dados2 <- read\_csv2(files[2], locale=locale(encoding="latin1")) %>% mutate(fonte = 'Amostra\_2')  
dados3 <- read\_csv2(files[3], locale=locale(encoding="latin1")) %>% mutate(fonte = 'Amostra\_3')  
  
dados2 <- dados2 %>%   
 mutate('Tempo\_Av-TAV' = Tempo\_avião - Tempo\_TAV)  
  
dados3 <- dados3 %>%   
 mutate('Tempo\_Av-TAV' = Tempo\_avião - Tempo\_TAV)  
  
colunas <- c('fonte', 'SbjNum', 'Opção\_escolhida',   
 'Tempo\_Av-TAV', 'Tempo\_avião', 'Tempo\_TAV',  
 'Valor\_avião', 'Valor\_TAV',  
 'Q\_2', 'Q\_57', 'Rota TAV' )  
  
convert\_money <- function(x){  
 y <- as.numeric(gsub(',00','', gsub("([R\\$.])","", x)))  
 return(y)  
}  
  
dados\_all <- dados1 %>% select(colunas) %>%   
 bind\_rows(dados2 %>% select(colunas)) %>%   
 bind\_rows(dados3 %>% select(colunas)) %>%   
 mutate(Opção\_escolhida = ifelse(Opção\_escolhida == 'Não sabe/ Não respondeu'  
 ,'NSR', Opção\_escolhida)  
 ,dif\_av\_tav = convert\_money(Valor\_avião) - convert\_money(Valor\_TAV)  
 ,valor\_aviao = convert\_money(Valor\_avião)  
 ,usou\_milha = case\_when(Q\_57 == 'Comprada' ~ 'Comprada'  
 ,TRUE ~ 'Milhas') ) %>%   
 group\_by(`Rota TAV`, fonte) %>%   
 mutate(z\_valor = (valor\_aviao - mean(valor\_aviao))/sd(valor\_aviao) ) %>%   
 ungroup()  
  
knitr::kable(head(dados\_all[,2:6]))

| SbjNum | Opção\_escolhida | Tempo\_Av-TAV | Tempo\_avião | Tempo\_TAV |
| --- | --- | --- | --- | --- |
| 8078429 | Avião | -12 | 118 | 130 |
| 8078429 | NSR | 18 | 118 | 100 |
| 8078708 | Trem de alta velocidade | -98 | 149 | 247 |
| 8078708 | Trem de alta velocidade | -40 | 149 | 189 |
| 8079114 | Avião | -5 | 131 | 136 |
| 8079114 | Avião | 26 | 131 | 105 |

## Escolhas por aeroportos

Por aeroporto:

dados\_all %>%   
 distinct(SbjNum, Q\_2, Opção\_escolhida) %>%   
 group\_by(Q\_2) %>%   
 summarise(TOT = n()  
 ,Av = sum(ifelse(Opção\_escolhida == 'Avião', 1, 0))  
 ,p\_Av = round(Av/TOT, 3) ) %>%   
 arrange(desc(p\_Av)) %>%   
 knitr::kable()

| Q\_2 | TOT | Av | p\_Av |
| --- | --- | --- | --- |
| SBUL | 953 | 537 | 0.563 |
| SBBH | 704 | 386 | 0.548 |
| SBCF | 1193 | 535 | 0.448 |
| SBGR | 943 | 368 | 0.390 |
| SBKP | 570 | 220 | 0.386 |
| SBSP | 1697 | 648 | 0.382 |
| SBCT | 989 | 358 | 0.362 |
| SBGL | 506 | 180 | 0.356 |
| SBUR | 345 | 122 | 0.354 |
| SBRJ | 1512 | 428 | 0.283 |
| SBRP | 704 | 173 | 0.246 |

Por trecho:

dados\_all %>%   
 distinct(SbjNum, `Rota TAV`, Opção\_escolhida) %>%   
 group\_by(`Rota TAV`) %>%   
 summarise(TOT = n()  
 ,Av = sum(ifelse(Opção\_escolhida == 'Avião', 1, 0))  
 ,p\_Av = round(Av/TOT, 3) ) %>%   
 arrange(desc(p\_Av)) %>%   
 knitr::kable()

| Rota TAV | TOT | Av | p\_Av |
| --- | --- | --- | --- |
| SBCF - SBUR | 7 | 7 | 1.000 |
| SBCF - SBRP | 6 | 5 | 0.833 |
| SBUL - SBCF | 223 | 181 | 0.812 |
| SBUL - SBBH | 141 | 110 | 0.780 |
| SBCF - SBUL | 64 | 48 | 0.750 |
| SBBH - SBUL | 205 | 151 | 0.737 |
| SBUR - SBCF | 60 | 44 | 0.733 |
| SBCF - SBCT | 65 | 41 | 0.631 |
| SBRP - SBBH | 72 | 45 | 0.625 |
| SBCT - SBCF | 86 | 53 | 0.616 |
| SBBH - SBRP | 121 | 74 | 0.612 |
| SBBH - SBUR | 7 | 4 | 0.571 |
| SBRP - SBCF | 21 | 12 | 0.571 |
| SBGR - SBBH | 72 | 40 | 0.556 |
| SBUR - SBBH | 20 | 11 | 0.550 |
| SBGL - SBCT | 130 | 70 | 0.538 |
| SBCT - SBRJ | 90 | 47 | 0.522 |
| SBUL - SBGR | 112 | 55 | 0.491 |
| SBGR - SBCF | 239 | 117 | 0.490 |
| SBCF - SBGR | 300 | 144 | 0.480 |
| SBSP - SBCF | 287 | 136 | 0.474 |
| SBCF - SBSP | 259 | 117 | 0.452 |
| SBCT - SBGL | 146 | 64 | 0.438 |
| SBCF - SBKP | 136 | 59 | 0.434 |
| SBBH - SBGR | 190 | 82 | 0.432 |
| SBRJ - SBCT | 78 | 33 | 0.423 |
| SBGR - SBGL | 154 | 65 | 0.422 |
| SBKP - SBCT | 109 | 46 | 0.422 |
| SBKP - SBGL | 121 | 51 | 0.421 |
| SBKP - SBRP | 19 | 8 | 0.421 |
| SBUL - SBKP | 132 | 55 | 0.417 |
| SBBH - SBKP | 181 | 75 | 0.414 |
| SBKP - SBUR | 10 | 4 | 0.400 |
| SBUL - SBSP | 326 | 130 | 0.399 |
| SBSP - SBCT | 277 | 109 | 0.394 |
| SBSP - SBGL | 156 | 61 | 0.391 |
| SBCT - SBKP | 201 | 76 | 0.378 |
| SBKP - SBUL | 16 | 6 | 0.375 |
| SBSP - SBUL | 102 | 38 | 0.373 |
| SBKP - SBBH | 27 | 10 | 0.370 |
| SBSP - SBRJ | 799 | 292 | 0.365 |
| SBKP - SBRJ | 141 | 51 | 0.362 |
| SBGR - SBCT | 206 | 74 | 0.359 |
| SBKP - SBCF | 115 | 41 | 0.357 |
| SBGL - SBCF | 77 | 27 | 0.351 |
| SBRJ - SBKP | 160 | 56 | 0.350 |
| SBRJ - SBCF | 295 | 100 | 0.339 |
| SBCF - SBRJ | 266 | 90 | 0.338 |
| SBGL - SBRP | 18 | 6 | 0.333 |
| SBUL - SBRP | 18 | 6 | 0.333 |
| SBUR - SBGR | 92 | 30 | 0.326 |
| SBRP - SBGL | 156 | 48 | 0.308 |
| SBCT - SBRP | 10 | 3 | 0.300 |
| SBGL - SBKP | 113 | 33 | 0.292 |
| SBRP - SBRJ | 29 | 8 | 0.276 |
| SBKP - SBJF | 11 | 3 | 0.273 |
| SBCF - SBGL | 85 | 23 | 0.271 |
| SBGL - SBSP | 74 | 20 | 0.270 |
| SBGR - SBRJ | 178 | 48 | 0.270 |
| SBGR - SBUL | 15 | 4 | 0.267 |
| SBCT - SBGR | 231 | 61 | 0.264 |
| SBGL - SBGR | 93 | 24 | 0.258 |
| SBGR - SBRP | 78 | 20 | 0.256 |
| SBRJ - SBRP | 4 | 1 | 0.250 |
| SBRJ - SBSP | 782 | 193 | 0.247 |
| SBCT - SBSP | 225 | 54 | 0.240 |
| SBRJ - SBGR | 171 | 40 | 0.234 |
| SBRJ - SBSJ | 22 | 5 | 0.227 |
| SBUR - SBKP | 173 | 37 | 0.214 |
| SBCF - SBJF | 5 | 1 | 0.200 |
| SBSP - SBRP | 76 | 12 | 0.158 |
| SBRP - SBSP | 153 | 24 | 0.157 |
| SBRP - SBUL | 28 | 4 | 0.143 |
| SBRP - SBGR | 197 | 28 | 0.142 |
| SBRP - SBKP | 48 | 4 | 0.083 |
| SBGL - SBSJ | 1 | 0 | 0.000 |
| SBGR - SBUR | 1 | 0 | 0.000 |
| SBKP - SBSJ | 1 | 0 | 0.000 |
| SBUL - SBUR | 1 | 0 | 0.000 |

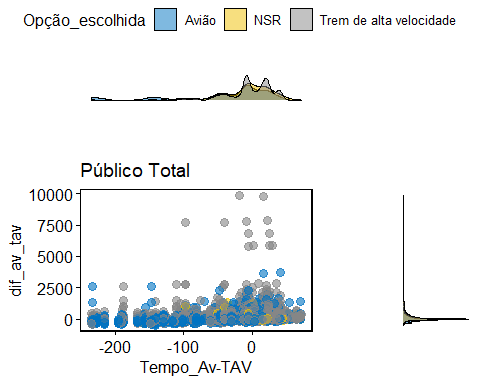
## Distribuição de escolhas

Abaixo o gráfico de dispersão das escolhas dado as variações de preço e tempo para cada tipo de viagem.

library(ggpubr)

## Warning: package 'ggpubr' was built under R version 4.1.3

sp <- ggscatter(dados\_all, x = "Tempo\_Av-TAV", y = "dif\_av\_tav",  
 color = "Opção\_escolhida", palette = "jco",  
 size = 3, alpha = 0.6, title = "Público Total")+  
 border()   
# Marginal density plot of x (top panel) and y (right panel)  
xplot <- ggdensity(dados\_all, "Tempo\_Av-TAV", fill = "Opção\_escolhida",  
 palette = "jco")  
yplot <- ggdensity(dados\_all, "dif\_av\_tav", fill = "Opção\_escolhida",   
 palette = "jco")+  
 rotate()  
# Cleaning the plots  
yplot <- yplot + clean\_theme()   
xplot <- xplot + clean\_theme()  
# Arranging the plot  
ggarrange(xplot, NULL, sp, yplot,   
 ncol = 2, nrow = 2, align = "hv",   
 widths = c(2, 1), heights = c(1, 2),  
 common.legend = TRUE)



### Identificando método de pagamento

Abaixo os volumes de uso de milhas aéreas na compra de passagens:

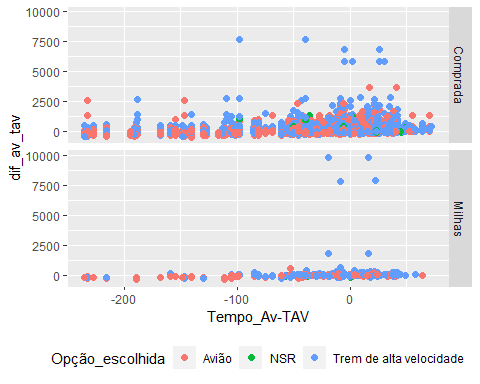
knitr::kable(table(dados\_all$fonte,dados\_all$usou\_milha)/2)

|  | Comprada | Milhas |
| --- | --- | --- |
| Amostra\_1 | 2724 | 177 |
| Amostra\_2 | 2853 | 216 |
| Amostra\_3 | 3298 | 163 |

# metade pois cada passageiro responde duas vezes

Cerca de 5% dos passageiros declaram terem comprado a passagem por milhas.

dados\_all %>%   
 ggplot( aes(x=`Tempo\_Av-TAV`, y=dif\_av\_tav, color=Opção\_escolhida)) +  
 geom\_point(size=2) +  
 facet\_grid(usou\_milha ~ .) +   
 theme(legend.position="bottom")



## Passageiros com mudança de escolha

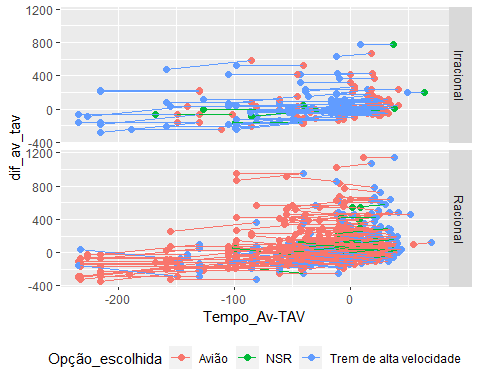
# Identificando mudança de escolha ----------------------------------------  
d = dados\_all %>%   
 select(Opção\_escolhida, SbjNum) %>%   
 mutate(ind = 1) %>%   
 group\_by(SbjNum, Opção\_escolhida) %>%   
 summarise(FREQ = n())

## `summarise()` has grouped output by 'SbjNum'. You can override using the  
## `.groups` argument.

prop.table(table(d$FREQ))

##   
## 1 2   
## 0.135429 0.864571

mudou\_ideia <- unlist(unique(d[d$FREQ==1,'SbjNum']))  
  
choices <- dados\_all %>%   
 filter(SbjNum %in% mudou\_ideia) %>%   
 arrange(SbjNum, Opção\_escolhida) %>%   
 group\_by(SbjNum) %>%   
 mutate(choice = row\_number()  
 ,opcao\_num = case\_when(Opção\_escolhida == "Avião" ~ 3  
 ,Opção\_escolhida == "NSR" ~ 2  
 ,Opção\_escolhida == "Trem de alta velocidade" ~ 1)  
 ) %>%   
 ungroup() %>%   
 select(fonte, SbjNum, `Tempo\_Av-TAV`, dif\_av\_tav, Opção\_escolhida, opcao\_num, choice)   
  
choices <- choices %>% filter(choice == 1) %>%   
 left\_join(choices %>% filter(choice == 2), suffix = c(".1", ".2"), by = c("SbjNum" = "SbjNum", 'fonte' = 'fonte'))  
  
choices <- choices %>%   
 mutate(irracional = case\_when((dif\_av\_tav.1 - dif\_av\_tav.2 >= 0 &  
 `Tempo\_Av-TAV.1` - `Tempo\_Av-TAV.2` > 0 &  
 opcao\_num.1 - opcao\_num.2 > 0 ) ~ 1   
 , TRUE ~ 0)  
 )  
  
  
dados\_all <- dados\_all %>%   
 mutate(escolhas = case\_when(SbjNum %in% mudou\_ideia ~ 'mudou'  
 ,TRUE ~ 'igual' )  
 ,tp\_escolha = case\_when(SbjNum %in% choices[choices$irracional == 1,]$SbjNum ~ 'Irracional',  
 TRUE ~ 'Racional')  
 )  
  
dados\_all %>%   
 filter(escolhas == 'mudou') %>%   
 ggplot( aes(x=`Tempo\_Av-TAV`, y=dif\_av\_tav, color=Opção\_escolhida)) +  
 geom\_point(size=2) +   
 geom\_line(aes(group = SbjNum)) +   
 facet\_grid(tp\_escolha ~ .) +   
 theme(legend.position="bottom")



Proporção de tipo de escolha por aeroporto de coleta:

dados\_all %>%   
 distinct(SbjNum, Q\_2, escolhas, tp\_escolha) %>%   
 group\_by(Q\_2) %>%   
 summarise(TOT = n()  
 ,mudou = sum(ifelse(escolhas == 'mudou', 1, 0))  
 ,p\_mudou = round(mudou/TOT, 3)  
 ,racional = sum(ifelse(tp\_escolha == 'Racional' & escolhas == 'mudou', 1, 0))  
 ,p\_racional = round(racional/TOT, 3)) %>%   
 arrange(desc(p\_racional)) %>%   
 knitr::kable()

| Q\_2 | TOT | mudou | p\_mudou | racional | p\_racional |
| --- | --- | --- | --- | --- | --- |
| SBCT | 877 | 112 | 0.128 | 82 | 0.094 |
| SBBH | 641 | 63 | 0.098 | 49 | 0.076 |
| SBCF | 1091 | 102 | 0.093 | 78 | 0.071 |
| SBGR | 875 | 68 | 0.078 | 59 | 0.067 |
| SBRJ | 1408 | 104 | 0.074 | 95 | 0.067 |
| SBSP | 1576 | 121 | 0.077 | 104 | 0.066 |
| SBKP | 537 | 33 | 0.061 | 32 | 0.060 |
| SBGL | 473 | 33 | 0.070 | 26 | 0.055 |
| SBUL | 926 | 27 | 0.029 | 24 | 0.026 |
| SBUR | 336 | 9 | 0.027 | 5 | 0.015 |
| SBRP | 691 | 13 | 0.019 | 10 | 0.014 |

# Método

Contextualizar modelo de escolha discreta (logistico) e cáclulo de WTP

Multinomial logit?

Formula do WTP:

# Resultados

Inicialmente descritivas seguidos pelo modelo em si.

## Descritivas

2/0  
teste

## Modelo