Usar distribuição normal dos efeitos aleatórios pq eh padrão

Usar numero de Halton = 300 (jessica) e 500 (luiza)

Na base completa fizemos o ajuste para renda (dos que não informaram)

Testes de base completa:

Geral

Por renda

Por distância

Por motivo

Por renda e distância

Por motivo e renda

Teste por renda declarada:

Geral

Teste dos que mudaram de opinião:

Geral

Por renda

Por motivo

## Análise dos Dados

Assim como na pesquisa online de passageiros de rodovias, a análise dos dados da pesquisa de aeroportos foi realizada a partir da aplicação do modelo de escolha discreta mixed logit, metodologia que está descrita no item do presente relatório.

### Estimativa do modelo mixed logit

Para determinar o valor do tempo e o valor do tempo de atraso, as variáveis aleatórias consideradas no modelo são: tarifa, tempo e tempo de atraso, doravante denominados custo, tempo e atraso. Para que o cálculo do WTP seja realizado de forma direta, dividindo os coeficientes dos parâmetros, a distribuição estatística deles deverá ser a mesma.

O modelo foi rodado na base completa para testar as distribuições normal e normal truncada, mantendo o número de sorteios fixo em 45, e os indicadores avaliados foram *o log-likelihood* (quanto maior, melhor o ajuste)e o *McFadden R²* (quanto mais próximo de zero, melhor o ajuste). O valor dos indicadores foi muito próximo para ambas as distribuições, com a distribuição normal truncada possuindo o melhor ajuste, como pode ser observado na tabela x abaixo. Além da análise estatística, considerando que é ilógico pagar um valor maior por um tempo maior tanto de viagem como de atraso, faz mais sentido utilizarmos a distribuição normal truncada, que foi definida como a opção a ser utilizada.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Distribuição** | **Log-likelihood** | **McFadden** | **tempo rodagem** | **β Tempo** | **β Atraso** | **β Valor** |
| normal | -8699,7 | 0,29155 | 7:22:54 | 0,0587804 | 0,3538219 | 0,0305437 |
| n truncada | -8699,8 | 0,29154 | 8:45:46 | 0,0567261 | 0,2955336 | 0,0276418 |

Da mesma forma, considerando os mesmo indicadores, foi testado o número (R) de sorteios de Halton a serem utilizados. O modelo foi rodado na base completa com R=300, R= 500, R=700 e R=1000. Os resultados se encontram na Tabela X abaixo. Definiu-se que o número de sorteios a ser utilizado seria 700, já que possui um valor de *log-likelihood* um pouco maior, mesmo levando um pouco mais de tempo para ser executado.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sorteios** | **Log-likelihood** | **McFadden** | **Tempo modelo** | **β Tempo** | **β Atraso** | **β Valor** |
| 300 | -8630,5 | 0,29718 | 21:49:38 | 0,0881277 | 0,5302869 | 0,0310146 |
| 500 | -8625 | 0,29763 | 10:54:02 | 0,0900109 | 0,5865719 | 0,0324134 |
| 700 | -8622,4 | 0,29785 | 15:5:41 | 0,0899307 | 0,5833774 | 0,0354024 |
| 1000 | -8626,6 | 0,2975 | 12:37:38 | 0,0909818 | 0,5838489 | 0,0362049 |

### Modelo com Base completa

Tabela 1 - Modelo Completo

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variáveis** | **Coeficiente** | **Std. Error** | **z-value** | **Pr(>|z|)** |
| **Média** |  |  |  |  |
| Intercepto | 1,687 | 0,165 | 10,201 | 0 |
| Tempo | 0,090 | 0,016 | 5,621 | 1,90E-08 |
| Atraso | 0,583 | 0,092 | 6,371 | 1,88E-10 |
| Valor | 0,035 | 0,005 | 6,576 | 4,82E-11 |
| **SD** |  |  |  |  |
| Tempo | 0,413 | 0,063 | 6,565 | 5,21E-11 |
| Atraso | 4,051 | 0,613 | 6,605 | 3,98E-11 |
| Valor | 0,462 | 0,072 | 6,454 | 1,09E-10 |
| **Estatísticas** |  | | | |
| *Log-likelihood* | -8622,4 | | | |
| McFadden R² | 0,29785 | | | |
| **Resultados** |  | | | |
| WTP Tempo | 152,41 | | | |
| WTP Atraso | 988,71 | | | |

## Análise dos Dados

Sob o mesmo processo de análise da pesquisa online de passageiros de rodovias, na avaliação do estudo em aeroportos tem-se a aplicação do Modelo de Escolha Discreta *Mixed Logit*, metodologia descrita no item do presente relatório.

### Estimativa do modelo mixed logit

A fim de determinar a predisposição a pagar pela economia de tempo (WTP Tempo) e pelo “não-atraso” (WTP Atraso), consideram-se as variáveis aleatórias: tarifa, tempo de viagem e tempo de atraso, doravante denominados custo, tempo e atraso. Para que o cálculo do WTP seja realizado de forma direta, por meio da divisão dos coeficientes dos parâmetros, é necessário que tenham a mesma distribuição estatística.

O Modelo de Escolha Discreta *Mixed Logit* foi empregado primeiramente na base geral sem nenhuma restrição, e duas distribuições diferentes para os parâmetros aleatórios foram testadas: distribuição normal e normal-truncada. O número de sorteios de Halton foi fixado em 45. Os indicadores de ajuste considerados foram o *log-likelihood* e o McFadden R².

O R-Quadrado de McFadden é um parâmetro que determina o quanto num modelo de dados binários individuais as covariáveis viabilizam a respeito da previsão acurada dos resultados. De acordo com McFadden, valores de 0,2 a 0,4 indicam excelente ajuste de modelo. Para um modelo de regressão logística, o valor de log verossimilhança é sempre negativo devido à contribuição de probabilidade de cada observação que varia 0 e 1. Nesse indicador, quanto maior seu valor, melhor o ajuste.

Ao observar os valores encontrados na Tabela 1, têm-se resultados similares para ambas as distribuições. É propício o uso da distribuição normal-truncada, mesmo com pouca diferença na qualidade do ajuste, visto que é ilógico pagar um valor maior por um tempo maior tanto de viagem como de atraso.

Tabela 1 – Resultados dos testes de distribuição

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Distribuição** | **Log-likelihood** | **McFadden** | **tempo rodagem** | **β Tempo** | **β Atraso** | **β Valor** |
| normal | -8699,7 | 0,29155 | 7:22:54 | 0,0587804 | 0,3538219 | 0,0305437 |
| n truncada | -8699,8 | 0,29154 | 8:45:46 | 0,0567261 | 0,2955336 | 0,0276418 |

Da mesma forma que o estudo anterior, tem-se a verificação do número ideal de sorteios de Halton (R) utilizando a base geral. Os valores tesados foram: R=300, R= 500, R=700 e R=1000. Os resultados se encontram na Tabela 2. O número de sorteios definiu-se em R=700, em consequência ao valor máximo encontrado dos indicadores *log-likelihood* e McFadden R², mesmo que o tempo de processamento seja maior. Nota-se que após esse número de ensaio a qualidade de ajuste é inferior.

Tabela 2 - Teste de sorteios de Halton

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sorteios** | **Log-likelihood** | **McFadden** | **Tempo modelo** | **β Tempo** | **β Atraso** | **β Valor** |
| 300 | -8630,5 | 0,29718 | 21:49:38 | 0,0881277 | 0,5302869 | 0,0310146 |
| 500 | -8625 | 0,29763 | 10:54:02 | 0,0900109 | 0,5865719 | 0,0324134 |
| 700 | -8622,4 | 0,29785 | 15:5:41 | 0,0899307 | 0,5833774 | 0,0354024 |
| 1000 | -8626,6 | 0,2975 | 12:37:38 | 0,0909818 | 0,5838489 | 0,0362049 |

Visto que a distribuição estatística e o número de sorteios foram estabelecidos, o próximo passo é a execução do Modelo *Mixed Logit* na base geral e, posteriormente, nas bases filtradas por características individuais. Para garantir uma destacada qualidade dos resultados, a significância estatística definiu-se em 95%. Os resultados serão apresentados e analisados nas seguintes subseções.

### Modelo com Base completa

Nesta primeira etapa tem-se 18274 observações de 9166 respondentes. A apresenta os coeficientes e valores de propensão a pagar pela economia de tempo (WTP Tempo) e pela confiabilidade de não haver atraso (WTP Atraso). Todas as variáveis são significantes com o auxílio de um grande número de observações. Nota-se que a disposição a pagar para não ter atraso é consideravelmente maior que a disposição a pagar pelo tempo, o que indica que usuários do transporte aeroviário valorizam mais a confiabilidade do serviço do que o tempo de deslocamento (velocidade do transporte). O desvio padrão (SD) do Modelo *Mixed Logit* indica a significância da heterogeneidade individual de preferências e sensibilidades. No modelo aplicado, tem-se relevância dos parâmetros de escala sobre seus respectivos coeficientes aleatórios, ou seja, existe uma diversidade de favoritismo de escolha de um respondente para o outro de acordo com essas variáveis.

Tabela 3 - Modelo Completo

|  |  |  |
| --- | --- | --- |
| **Variáveis** | **Coeficiente** | **Std. Error** |
| **Média** |  |  |
| Intercepto | 1,687 | 0,165 |
| Tempo | 0,090 | 0,016 |
| Atraso | 0,583 | 0,092 |
| Valor | 0,035 | 0,005 |
| **SD** |  |  |
| Tempo | 0,413 | 0,063 |
| Atraso | 4,051 | 0,613 |
| Valor | 0,462 | 0,072 |
| **Estatísticas** |  | |
| *Log-likelihood* | -8622,4 | |
| McFadden R² | 0,29785 | |
| **Resultados** |  | |
| WTP Tempo | R$152,41 | |
| WTP Atraso | R$988,71 | |

Para fins de comparação, foram excluídos da base de dados os respondentes que optaram por não declarar renda e os que declararam não possuir renda familiar. As 3658 observações retiradas representam 20% da base. A apresenta os resultados, com valores substancialmente superiores aos encontrados na .

Tabela 4 - Respondentes que declararam renda

|  |  |  |
| --- | --- | --- |
| **Variáveis** | **Coeficiente** | **Std. Error** |
| **Média** |  |  |
| Intercepto | 1,834 | 0,180 |
| Tempo | 0,111 | 0,020 |
| Atraso | 0,262 | 0,047 |
| Valor | 0,012 | 0,004 |
| **SD** |  |  |
| Tempo | 0,472 | 0,073 |
| Atraso | 2,703 | 0,425 |
| Valor | 0,399 | 0,064 |
| **Estatísticas** |  | |
| *Log-likelihood* | -6899,4 | |
| McFadden R² | 0,29321 | |
| **Resultados** |  | |
| WTP Tempo | R$567,05 | |
| WTP Atraso | R$1337,31 | |

### Modelo considerando características específicas

Na segunda etapa, tem-se a aplicação do modelo por distintas características de viagem e particularidades socioeconômicas. O objetivo é verificar diferenças de disposição a pagar e significâncias estatísticas por faixa de renda familiar, distância e motivo de viagem.

As faixas de renda familiar do questionário foram agrupadas em quatro níveis: Renda 1, até 5 salários mínimos; Renda 2, mais de 5 a 10 salários mínimos; Renda 3, mais de 10 a 20 salários mínimos; e Renda 4, acima de 20 salários mínimos.

Realizou-se um ajuste nos respondentes que declararam não possuir renda ou que preferiram não declarar sua renda, a fim de que essas entrevistas não fossem descartadas. O processo consiste numa reversão à média, ou seja, considerou-se que estes têm renda familiar média no nível 3 (de R$7.241 até R$14.480). Os resultados obtidos são apresentados na Tabela 4.

Os coeficientes de tempo e atraso no nível Renda 1 não apresentaram significância estatística, provavelmente por não haver observações suficientes (cerca de 14% da amostra), resultado esperado pelo custo elevado do transporte aeroviário. O valor do tempo de atraso é superior ao valor do tempo de viagem em todas as faixas de renda seguintes.

Tabela 4 - Por Renda

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Renda 1** | | **Renda 2** | | **Renda 3** | | **Renda 4** | |
| **Variáveis** | **Coef** | **Std. Error** | **Coef** | **Std. Error** | **Coef** | **Std. Error** | **Coef** | **Std. Error** |
| **Média** |  |  |  |  |  |  |  |  |
| Intercepto | 1,262 | 0,365 | 1,849 | 0,381 | 1,429 | 0,232 | 2,694 | 0,457 |
| Tempo | 0,077\* | 0,039 | 0,111 | 0,041 | 0,092 | 0,026 | 0,063 | 0,024 |
| Atraso | 0,047\* | 0,050 | 0,449 | 0,153 | 0,629 | 0,161 | 0,691 | 0,199 |
| Valor | 0,053 | 0,021 | 0,059 | 0,017 | 0,020 | 0,007 | 0,022 | 0,008 |
| **SD** |  |  |  |  |  |  |  |  |
| Tempo | 0,476 | 0,204 | -0,478 | 0,151 | 0,406 | 0,101 | 0,343 | 0,094 |
| Atraso | 4,500 | 1,950 | 4,441 | 1,414 | 4,624 | 1,139 | 2,436 | 0,674 |
| Valor | 0,496 | 0,216 | 0,449 | 0,149 | 0,475 | 0,120 | 0,350 | 0,098 |
| **Estatísticas** |  | |  | |  | |  | |
| *Log-likelihood* | -1222,6 | | -1838,3 | | -4076,8 | | -1382,8 | |
| McFadden R² | 0,30248 | | 0,32343 | | 0,28464 | | 0,28915 | |
| **Resultados** |  | |  | |  | |  | |
| WTP Tempo | R$87,39\* | | R$113,48 | | R$274,07 | | R$169,21 | |
| WTP Atraso | R$54,09\* | | R$458,73 | | R$1873,92 | | R$1846,96 | |

\*não possuem significância estatística (Pr >0,05)

A distância de viagem apresentada no questionário era a distância real a ser viajada pelo passageiro. Então, foram criadas três faixas de distância: até 500km, mais de 500km a 800 km e mais de 800 km. Os resultados encontrados (veja ) não foram relevantes em geral, apenas o valor do tempo para viagens de longa distância (nível Distância 3) apresentou significância estatística.

Tabela 5 - Por Distância

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Distância 1** | | **Distância 2** | | **Distância 3** | |
| **Variáveis** | **Coeficiente** | **Std. Error** | **Coeficiente** | **Std. Error** | **Coeficiente** | **Std. Error** |
| **Média** |  |  |  |  |  |  |
| Intercepto | 1,298 | 0,138 | 1,069 | 0,276 | 4,363 | 1,618 |
| Tempo | 0,122 | 0,021 | 0,050 | 0,020 | 0,026 | 0,012 |
| Atraso | 0,837 | 0,143 | 0,303 | 0,105 | -0,018\* | 0,028 |
| Valor | 0,002\* | 0,006 | -0,015\* | 0,010 | 0,019 | 0,009 |
| **SD** |  |  |  |  |  |  |
| Tempo | 0,551 | 0,078 | 0,291 | 0,107 | 0,178 | 0,067 |
| Atraso | 4,582 | 0,745 | 1,303 | 0,455 | 0,526 | 0,228 |
| Valor | 0,218 | 0,055 | 1,249 | 0,414 | 0,211 | 0,076 |
| **Estatísticas** |  | |  | |  | |
| *Log-likelihood* | -3978,8 | | -3157,3 | | -1279,6 | |
| McFadden R² | 0,25796 | | 0,29301 | | 0,28863 | |
| **Resultados** |  | |  | |  | |
| WTP Tempo | R$3843,33\* | | - R$206,32\* | | R$83,64 | |
| WTP Atraso | R$26312,19\* | | - R$1243,42\* | | -R$57,29\* | |

\*não possuem significância estatística (Pr >0,05)

Considerou-se pertinente separar os motivos de viagem entre “trabalho”, incluindo negócios e evento profissional, e “não-trabalho”, que é composto por todos os motivos restantes. Todos os valores encontrados possuem significância estatística. Como é possível perceber na Tabela 6, o Valor do Tempo de Trabalho (R$253,88) é quase o triplo do Valor do Tempo de Não-trabalho (R$87,83). Quando comparados os Valores de Atraso em ambos, essa proporção é ainda maior, ultrapassando os 600%. Assim como nos casos anteriores, o Valor do Atraso é superior ao Valor do Tempo. Porém, quando a viagem é por motivo de trabalho, essa diferença é ainda maior (723%).

Tabela 6 - Por motivo

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Trabalho** | | **Não-trabalho** | |
| **Variáveis** | **Coeficiente** | **Std. Error** | **Coeficiente** | **Std. Error** |
| **Média** |  |  |  |  |
| Intercepto | 2,102 | 0,211 | 0,933 | 0,152 |
| Tempo | 0,095 | 0,019 | 0,035 | 0,009 |
| Atraso | 0,684 | 0,117 | 0,117 | 0,028 |
| Valor | 0,022 | 0,005 | 0,024 | 0,005 |
| **SD** |  |  |  |  |
| Tempo | 0,391 | 0,064 | 0,169 | 0,039 |
| Atraso | 3,683 | 0,603 | 1,817 | 0,358 |
| Valor | 0,394 | 0,066 | 0,340 | 0,067 |
| **Estatísticas** |  | |  | |
| *Log-likelihood* | -4803,5 | | -3789,8 | |
| McFadden R² | 0,31210 | | 0,27201 | |
| **Resultados** |  | |  | |
| WTP Tempo | R$253,88 | | R$87,83 | |
| WTP Atraso | R$1836,08 | | R$289,97 | |

\*não possuem significância estatística (Pr >0,05)

Foram realizados outros testes, como utilizar a base somente com os respondentes que escolheram trocar de modo de transporte, classificação por motivo e renda, por distância e renda, por renda declarada e distância, dentre outros. Nesses casos, os resultados obtidos não apresentaram relevância estatística e, portanto, não serão discutidos no presente documento (resultados em anexo).