

P1 - Econometria III

Pedro Mendes *

Dados

```
flights <- list(  
  flightsbr::read_flights(date = 2016),  
  flightsbr::read_flights(date = 2018)  
) |>  
  data.table::rbindlist()  
  
flights <- flights[  
  nm_pais_destino == "BRASIL" & ds_servico_tipo_linha == "PASSAGEIRO",  
  .(  
    data = dt_referencia,  
    passag_pagos = nr_passag_pagos,  
    kg_bagagem_excesso, kg_payload, km_distancia  
  )  
][  
  ,  
  data := lubridate::ymd(data)  
][  
  ,  
  purrr::map(.SD, collapse::fsum),  
  .SDcols = passag_pagos:km_distancia,  
  by = "data"  
] |>  
  tibble::as_tibble()  
  
head(flights)
```

```
## # A tibble: 6 x 5  
##   data          passag_pagos kg_bagagem_excesso kg_payload km_distancia  
##   <date>          <int>          <int>          <int>          <int>  
## 1 2016-01-02      289201          60647      39577389      2522831  
## 2 2016-01-01      239562          56759      37164088      2165028  
## 3 2016-01-03      328902          66891      43596294      2634334  
## 4 2016-01-04      360089          94393      48326140      2796921  
## 5 2016-01-08      331791          88175      47990916      2790186  
## 6 2016-01-05      343607         101003      45909254      2651594
```

*Fonte: <https://www.bbc.com/portuguese/brasil-39226346>

Modelo

Pretende testar o efeito da cobrança por bagagem no avião no número de passageiros pagantes em vôos comerciais domésticos no Brasil¹. Para isso, o estudo utiliza dados do pacote `flightsbr`².

Variáveis utilizadas³:

- *passag_pagos*: Número de passageiros que ocupam assentos comercializados ao público e que geram receita, com a compra de assentos, para a empresa de transporte aéreo. A variável será usada como proxy para o volume de vendas.
- *kg_bagagem_excesso*: Excesso de bagagem. É a quantidade total de bagagem que excede o limite de peso acordado entre a empresa aérea e o passageiro (adquirido antes ou depois da compra do bilhete), verificada no momento do despacho, expressa em quilogramas. É esperado que a variável tenha um efeito negativo sobre o número de passageiros pagos pois, segundo a hipótese do setor, o pagamento adicional por bagagens em excesso elevava o custo da passagem. Espera-se também que o efeito do volume de bagagens em excesso sobre o número de passageiros pagantes diminua após a implementação da política.
- *kg_payload*: Capacidade Payload, expressa em quilogramas, usada como variável de controle.
- *km_distancia*: Distância, expressa em quilômetros, entre os aeródromos de origem e destino da etapa, considerando a curvatura do planeta terra. É usada como variável de controle.

Dummies de período

```
set.seed(55871)

data_model <- flights |>
  dplyr::mutate(
    d_anos = dplyr::case_when(
      lubridate::year(data) == "2016" ~ 0,
      TRUE ~ 1
    )
  ) |>
  dplyr::group_by(d_anos) |>
  dplyr::slice_sample(n = 80) |> # sem cálculo de tamanho de amostra
  dplyr::ungroup()

head(data_model)
```

```
## # A tibble: 6 x 6
##   data      passag_pagos kg_bagagem_excesso kg_payload km_distancia d_anos
##   <date>         <int>         <int>         <int>         <int> <dbl>
## 1 2016-04-17      240814          66580      34699133      1990860     0
## 2 2016-02-13      223920          59465      35486773      2149916     0
## 3 2016-12-02      279427          57268      39633164      2326555     0
## 4 2016-04-19      250925          50423      38388057      2088379     0
## 5 2016-11-08      249447          83333      38162401      2144001     0
## 6 2016-09-30      266776          54500      38459485      2283126     0
```

¹Fonte: <https://www.bbc.com/portuguese/brasil-39226346>

²Fonte: <https://github.com/ipeaGIT/flightsbr>

³Fonte: <https://www.anac.gov.br/assuntos/setor-regulado/empresas/envio-de-informacoes/descricao-de-variaveis>

Regressão

$$\text{passag_pagos} = \beta_0 + \delta_0(\text{d_ano}) + \beta_1(\text{kg_bagagem_excesso}) + \delta_1(\text{kg_bagagem_excesso} \cdot \text{d_ano}) + \beta_2(\text{kg_payload}) + \beta_3(\text{kg_distancia})$$

```
model_1 <- data_model |>
  fixest::feols(
    passag_pagos ~ d_ano + (d_ano * kg_bagagem_excesso) + kg_payload + km_distancia
  )

summary(model_1)
```

```
## OLS estimation, Dep. Var.: passag_pagos
## Observations: 160
## Standard-errors: IID
##
##              Estimate   Std. Error   t value   Pr(>|t|)
## (Intercept)   -64479.082755 11377.816202 -5.667088 6.9553e-08 ***
## d_ano          12599.541037   9179.629152  1.372554 1.7189e-01
## kg_bagagem_excesso -0.064419    0.111764 -0.576385 5.6520e-01
## kg_payload      0.002015     0.000662  3.044000 2.7462e-03 **
## km_distancia    0.115394     0.013692  8.428067 2.3349e-14 ***
## d_ano:kg_bagagem_excesso -0.639574    0.320170 -1.997608 4.7520e-02 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 11,799.7   Adj. R2: 0.874258
```

Teste para heterocedasticidade:

```
olsrr::ols_test_breusch_pagan(
  lm(
    passag_pagos ~ d_ano + (d_ano * kg_bagagem_excesso) + kg_payload + km_distancia,
    data_model
  )
)
```

```
##
## Breusch Pagan Test for Heteroskedasticity
## -----
## Ho: the variance is constant
## Ha: the variance is not constant
##
##              Data
## -----
## Response : passag_pagos
## Variables: fitted values of passag_pagos
##
##      Test Summary
## -----
## DF          =      1
## Chi2         =    0.1718356
## Prob > Chi2  =    0.6784856
```

Baseado no teste de Breusch Pagan, não se pode afirmar que as variâncias são diferentes.

Teste de Chow para quebras estruturais no tempo:

```
car::linearHypothesis(  
  lm(  
    passag_pagos ~ d_ano + (d_ano * kg_bagagem_excesso) + kg_payload + km_distancia,  
    data_model  
  ),  
  hypothesis.matrix = c("d_ano = 0", "d_ano:kg_bagagem_excesso = 0"),  
  test = "F"  
)
```

```
## Linear hypothesis test  
##  
## Hypothesis:  
## d_ano = 0  
## d_ano:kg_bagagem_excesso = 0  
##  
## Model 1: restricted model  
## Model 2: passag_pagos ~ d_ano + (d_ano * kg_bagagem_excesso) + kg_payload +  
## km_distancia  
##  
## Res.Df      RSS Df Sum of Sq      F Pr(>F)  
## 1      156 2.2919e+10  
## 2      154 2.2277e+10  2 641600503 2.2176 0.1123
```

```
car::linearHypothesis(  
  lm(  
    passag_pagos ~ d_ano + (d_ano * kg_bagagem_excesso) + kg_payload + km_distancia,  
    data_model  
  ),  
  hypothesis.matrix = c("d_ano:kg_bagagem_excesso = 0"),  
  test = "F"  
)
```

```
## Linear hypothesis test  
##  
## Hypothesis:  
## d_ano:kg_bagagem_excesso = 0  
##  
## Model 1: restricted model  
## Model 2: passag_pagos ~ d_ano + (d_ano * kg_bagagem_excesso) + kg_payload +  
## km_distancia  
##  
## Res.Df      RSS Df Sum of Sq      F Pr(>F)  
## 1      155 2.2855e+10  
## 2      154 2.2277e+10  1 577250992 3.9904 0.04752 *  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Baseado nos testes acima, não existe uma diferença de nível significativo. Porém, existe uma diferença de inclinação significativa.

Conclusões a partir do modelo

- Não existiu uma alteração relevante para o número de passageiros pagantes ao se comparar os períodos de 2016 e 2018.
- Neste período, o efeito do volume de bagagens em excesso em vôos comerciais domésticos se tornou significativo e aumentou. Isso pode ser interpretado da seguinte forma: após a implementação do pagamento a parte por bagagem excedente em 2017, um aumento no volume de tais itens afeta mais o número de pagantes em vôos, se tornando um parâmetro mais sensível na escolha do consumidor.

Limitações do modelo

- O modelo se baseia em variáveis-proxy para tentar captar um efeito do aumento da cobrança por bagagens excedentes por parte de passageiros. Apesar do bom ajuste do modelo, o ideal seria utilizar variáveis que tratassem explicitamente de questões monetárias.