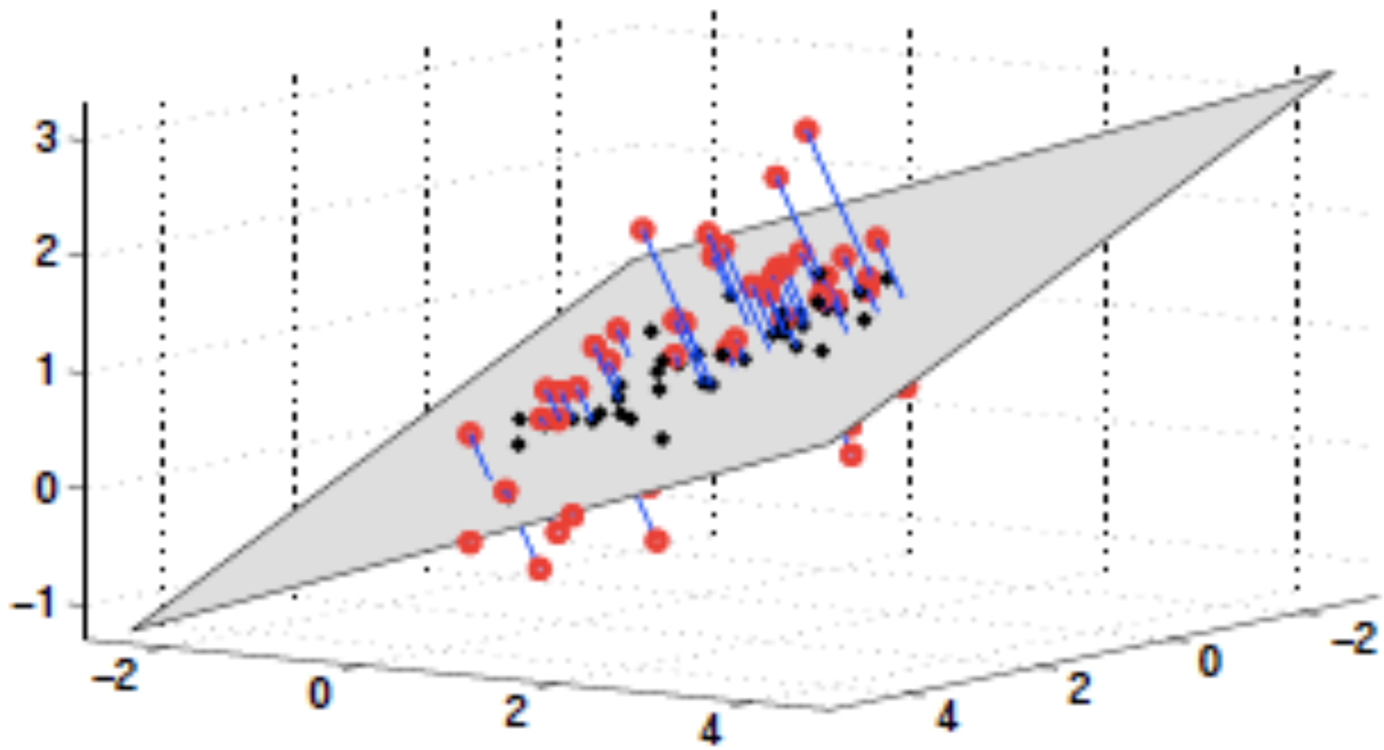


Aprendizado supervisionado 1

Especialização em Ciência de dados, IM/UFRJ
Thaís C O Fonseca
DME/UFRJ

Modelos de regressão linear múltipla

Exemplo

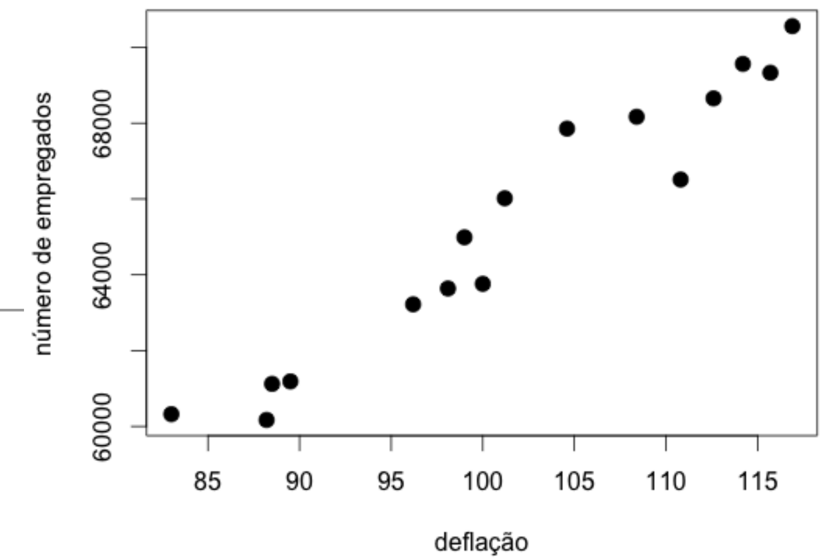
Considere novamente o exemplo sobre emprego. Além do índice econômico temos disponível também

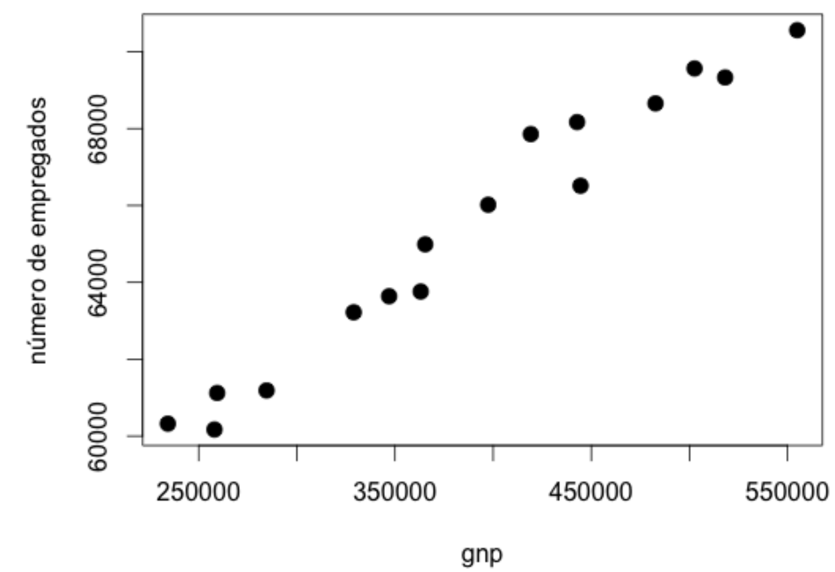
- o GNP, o produto nacional bruto.

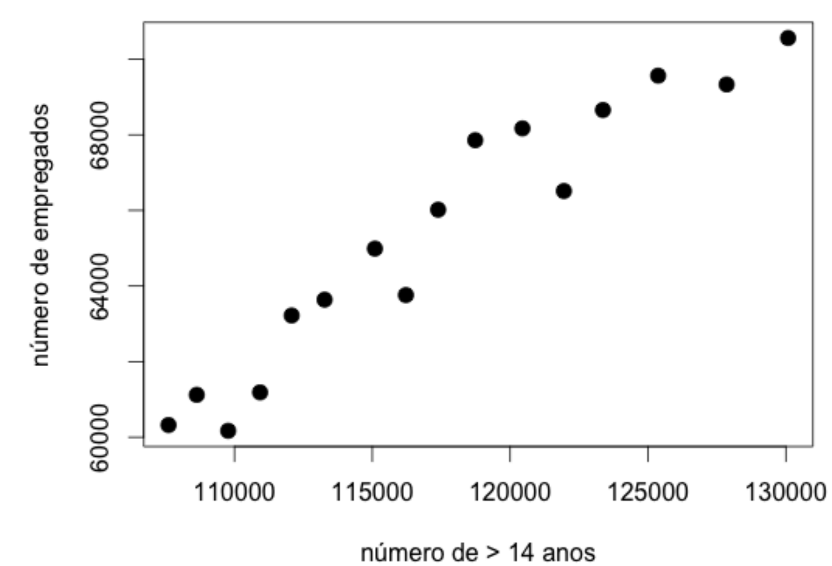
- o número de pessoas com mais de 14 anos de idade.

Seja X1 o índice econômico de deflação, Y o número de pessoas empregadas, X2 o GNP e X3 o número de pessoas com mais de 14 anos

Existe relação entre Y e as covariáveis X1, X2 e X3?







Modelo de regressão linear múltipla

O modelo geral é dado por

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \epsilon.$$

Nesse caso,

$$E(y \mid x) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k.$$

$\rightarrow$ Dizemos que a regressão é linear pois $E(y \mid x)$ depende de forma linear de $\beta_0, \beta_1, \ldots, \beta_k$.

O modelo abaixo é linear?

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x 2 + \beta_3 x_1 x_2 + \epsilon$$

Sim!!!! Basta reescrever

$x_3 = x_1 x_2$.

Notação matricial

Para uma amostra de tamanho n, $y_1, \ldots, y_n$, defina:

$$\mathbf{y} = (y_1, \ldots, y_n)' \rightarrow \text{ vetor n por 1 de observáveis}$$

$$\beta = (\beta_0, \beta_1, \ldots, \beta_k)' \rightarrow \text{ vetor (p=k+1) por 1 de coeficientes}$$

$$\epsilon = (\epsilon_1, \ldots, \epsilon_n)' \rightarrow \text{ vetor n por 1 de erros}$$

$$\mathbf{X} = (\mathbf{1}, x_1, \ldots, x_k) \rightarrow \text{ matriz n por p de covariáveis}$$

O modelo é dado por

$$\boxed{\mathbf{y} = \mathbf{X}\beta + \epsilon}$$

Interpretação dos coeficientes

- Vimos que para cada $X_{i1}, \ldots X_{ik}$ temos

$$E(y \mid X) = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_k X_{ik}$$

Isso significa que para $X_{i1} = 0, \ldots X_{ik} = 0$ temos

$$E(y \mid X) = \beta_0.$$

Nível comum!

Então β_0 é interpretado como o valor esperado de y para o valor 0 das covariáveis.

Interpretação dos coeficientes

- Considere $X = X^*$ então

$$E(y \mid X^*) = \beta_0 + \beta_1 X^*_{i1} + \beta_2 X^*_{i2} + \ldots + \beta_k X^*_{ik}.$$

Considere também $X_{i1} = X^*_{i1} + 1$ e $X_{i(-1)} = X^*_{i(-1)}$

$$E(y \mid X^*_{i1} + 1, X^*_{i(-1)}) = \beta_0 + \beta_1 (X^*_{i1} + 1) + \beta_2 X^*_{i2} \ldots + \beta_k X^*_{ik}.$$

Fazendo a diferença temos

$$E(y \mid X^*_{i1} + 1, X^*_{i(-1)}) - E(y \mid X = X^*) = \beta_1.$$

Então β_1 é interpretado como a mudança no valor esperado de y quando acrescentamos 1 unidade em X_1 para as outras variáveis fixas em uma constante qualquer.

O mesmo para x_j!

Método de mínimos quadrados

Considere uma amostra de tamanho n.

Queremos encontrar o modelo ajustado

$$\hat{\mathbf{y}} = X\hat{\beta} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k$$

que minimiza a soma de quadrados dos erros. Então queremos encontrar β que minimize a função

$$S(\beta) = (\mathbf{y} - X\beta)'(\mathbf{y} - X\beta) = \sum_{i=1}^{n}(y_i - (\beta_0 + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_k x_{ik}))^2.$$

Será satisfeito quando as colunas de X não forem linearmente dependentes!

Isto é, queremos minimizar

$$S(\beta) = \mathbf{y}'\mathbf{y}' - \beta'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{X}\beta$$

Resultando em

$$\boxed{\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}}$$

Quando a matriz $(X'X)^{-1}$ existe!!!

Algumas definições

- Modelo ajustado

$$\hat{\mathbf{y}} = X\hat{\beta}$$

Podemos reescrever

$$\hat{\mathbf{y}} = X(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = H\mathbf{y}$$

Matriz H: importante na análise de resíduos.

- Resíduos

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - X\hat{\beta} = \mathbf{y} - H\mathbf{y} = (I_n - H)\mathbf{y}$$

Suposições do modelo de regressão múltipla

As suposições são as mesmas do modelo simples:

1. Erros tem média zero. $E(\epsilon) = 0$
2. Variância constante. $V(\epsilon) = \sigma^2$
3. Erros não correlacionados. $Cor(\epsilon_i, \epsilon_j) = 0$

Vimos que o modelo para k covariáveis é dado por:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \epsilon$$

Método de máxima verossimilhança

$$
\begin{aligned}
L(\beta, \sigma^2; y, x) &= f(y_1, \ldots, y_n; x) \\
&= (2\pi\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta) \right\}
\end{aligned}
$$

Para encontrar o EMV devemos maximizar a função acima.

Isso equivale a encontrar o estimador de mínimos quadrados para β.

Os valores de β, σ^2 que maximizam a função acima são

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{\mathbf{-1}}\mathbf{X}'\mathbf{y}$$

$$\hat{\sigma}^2 = \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{n}$$

Exemplo de modelagem (regressão)

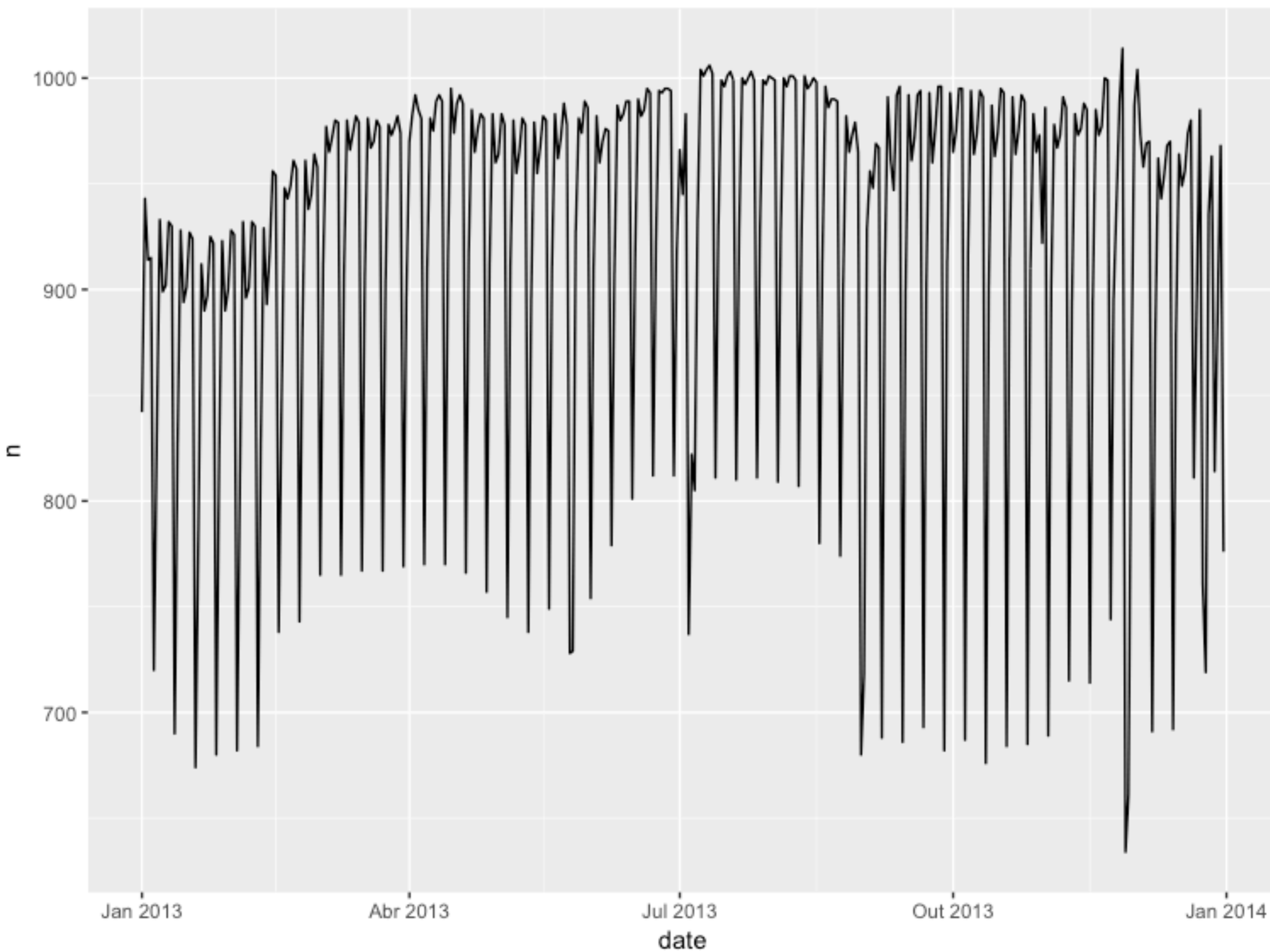
O que afeta o número diário de voos em NY

- Vamos considerar os dados do número de voos por dia que saem da cidade de Nova York.

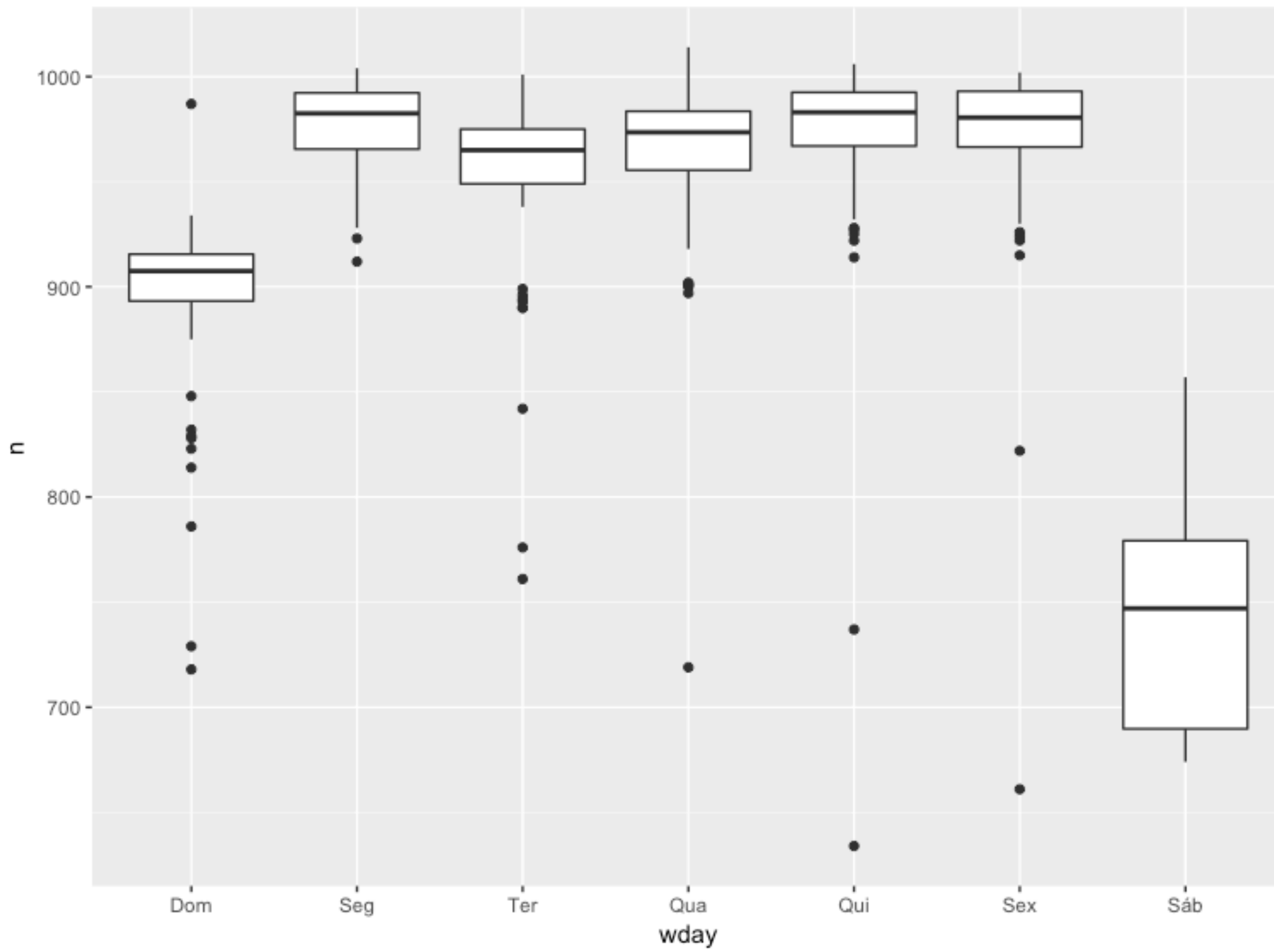
- O conjunto de dados é de pequeno porte: 365 linhas e 2 colunas.

```
# A tibble: 365 x 2
   date           n
   <date>     <int>
 1 2013-01-01   842
 2 2013-01-02   943
 3 2013-01-03   914
 4 2013-01-04   915
 5 2013-01-05   720
 6 2013-01-06   832
 7 2013-01-07   933
 8 2013-01-08   899
 9 2013-01-09   902
10 2013-01-10   932
# ... with 355 more rows
```

Gráfico da série temporal do número de voos



Dados por dia da semana



- Claro efeito de fim de semana, em particular no sábado.
- Maior parte das viagens devem ser a trabalho.
- Note que o efeito de sábado é bem menor (até mesmo quando comparado a domingo).

- Vamos ajustar um modelo com efeito de dia da semana. Qual é esse modelo?

Modelo com efeito de dia da semana

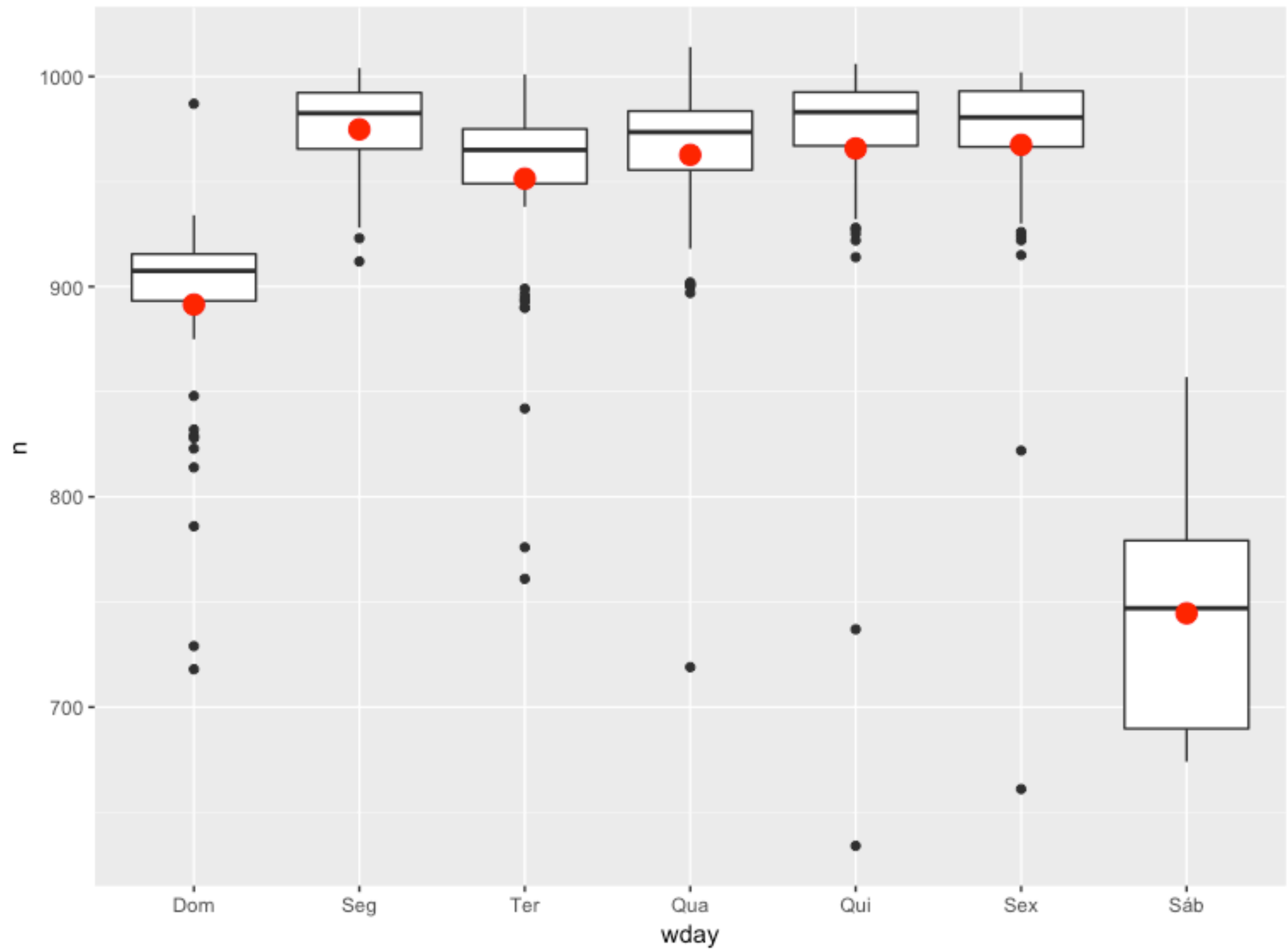
Neste exemplo temos variáveis dummies indicadoras do dia da semana, com $I_{mon} = 1$ se segunda-feira e 0 caso contrário.

$$y = X\beta + \epsilon,$$

onde $X = (I_{sun}, I_{mon}, I_{tue}, I_{wed}, I_{thu}, I_{fri}, I_{sat})$.

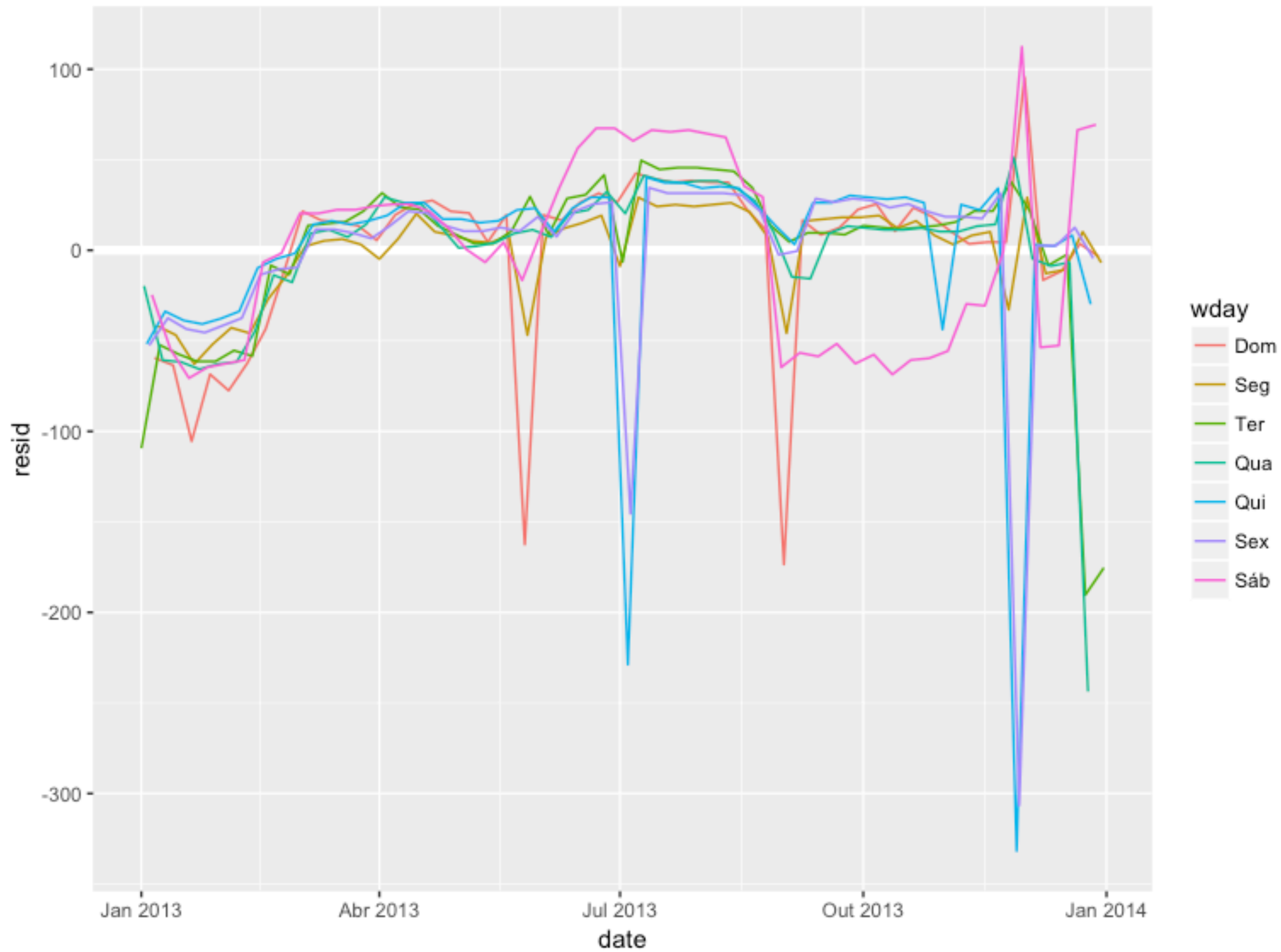
```
Coefficients:
daily$wdaySun  daily$wdayMon  daily$wdayTue  daily$wdayWed  daily$wdayThu
        891.5          974.8          951.4          962.7          965.8
daily$wdayFri  daily$wdaySat
        967.5          744.6
```

Modelo com efeito de dia da semana



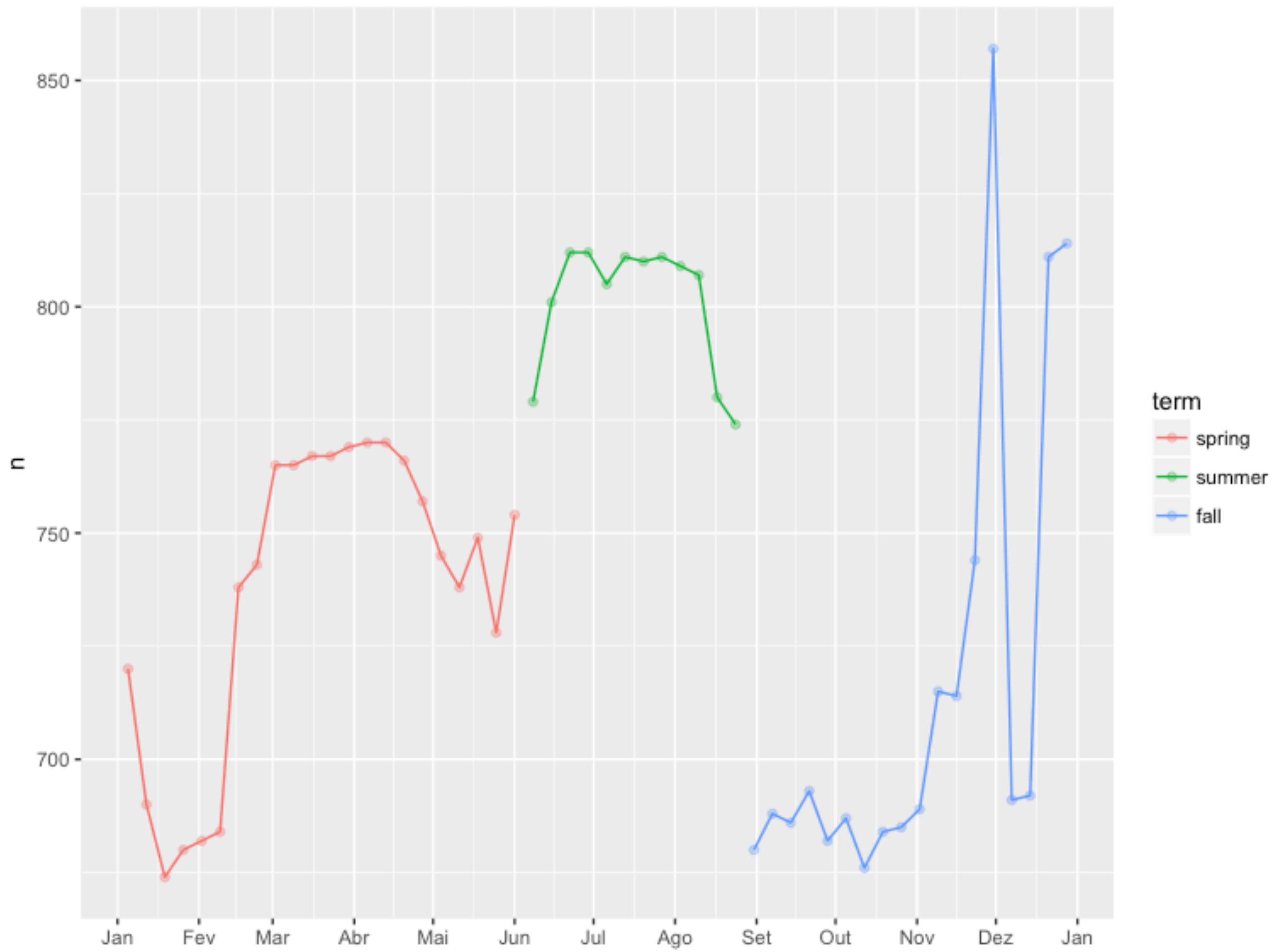
- Dados e previsão usando modelo com efeito de dia da semana.

Resíduos



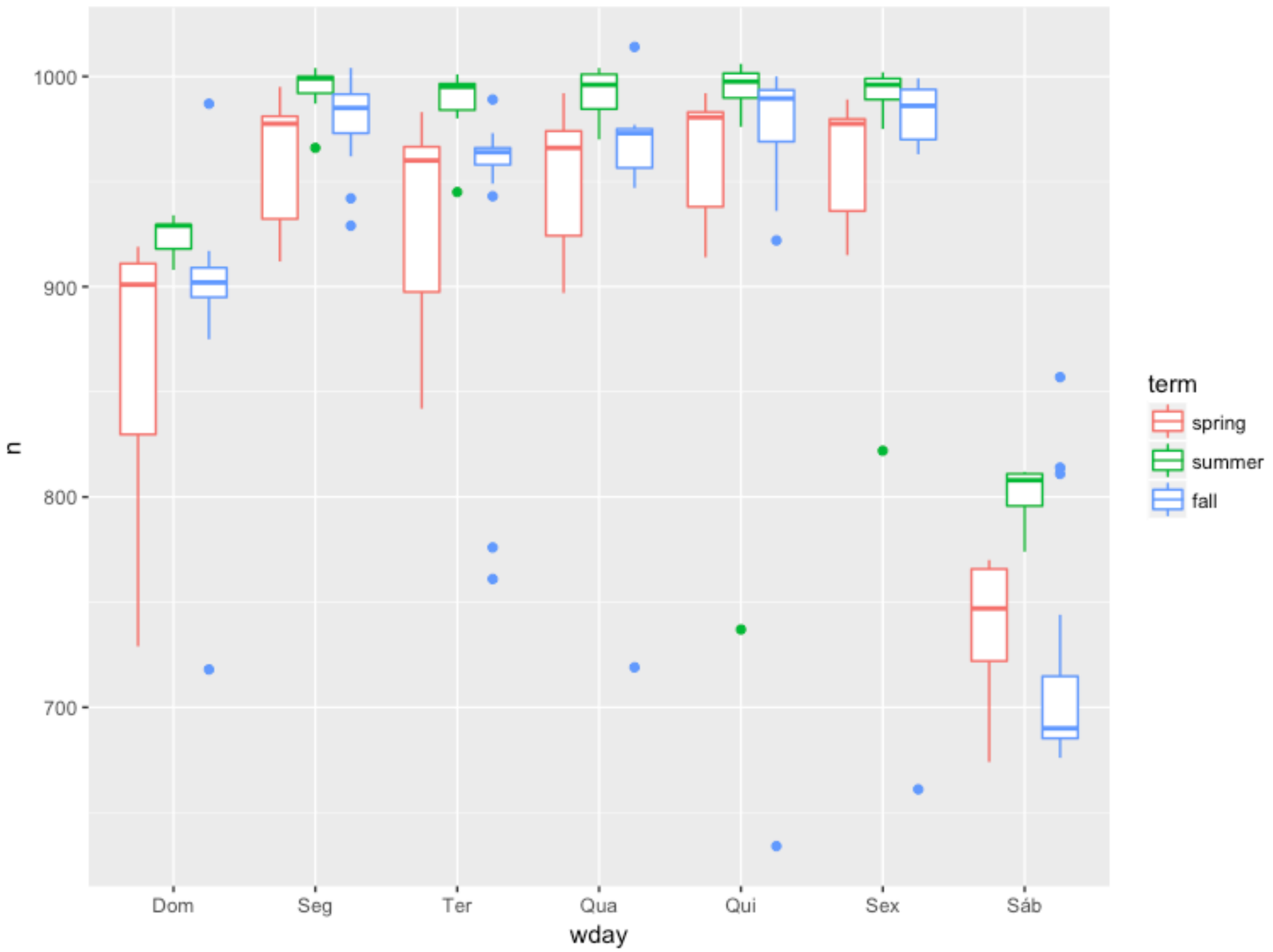
- Resíduos por dia da semana.
- Note que o comportamento no sábado difere dos demais.

Número de voos aos sábados no “term"



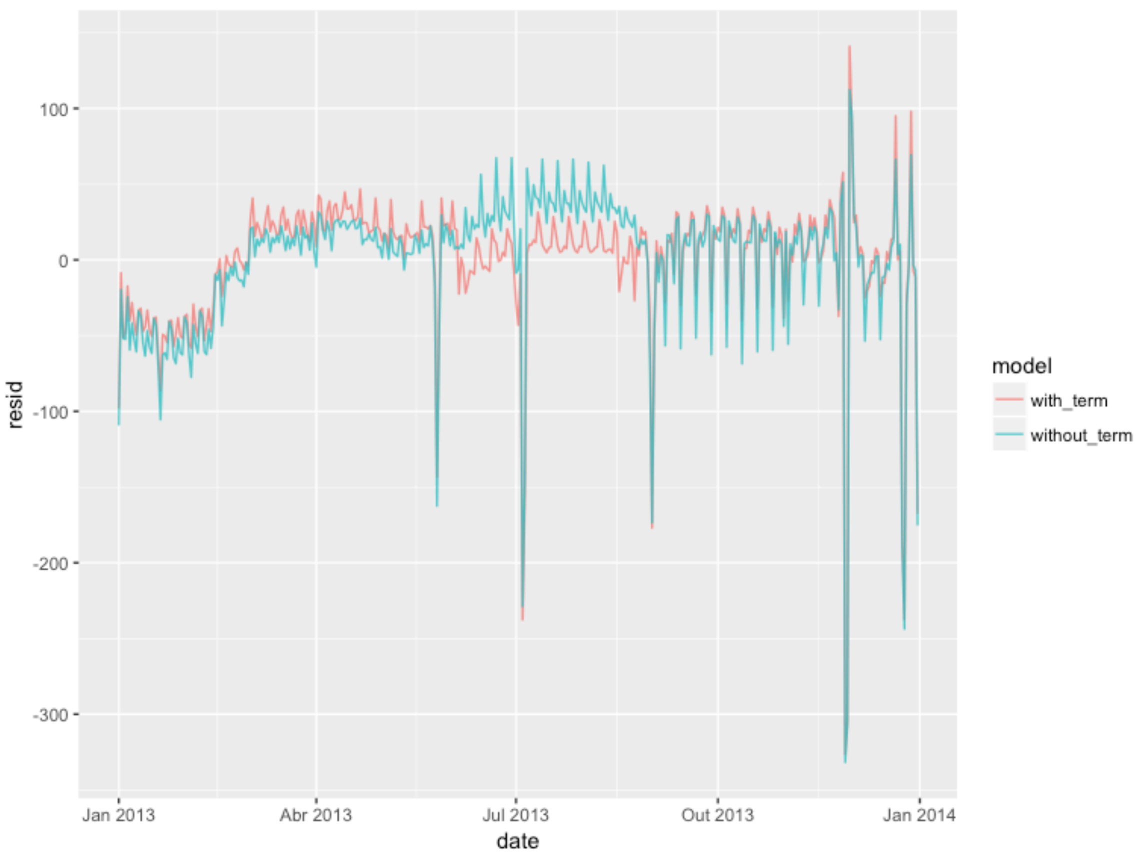
- Note os comportamentos bem distintos ao longo do ano.

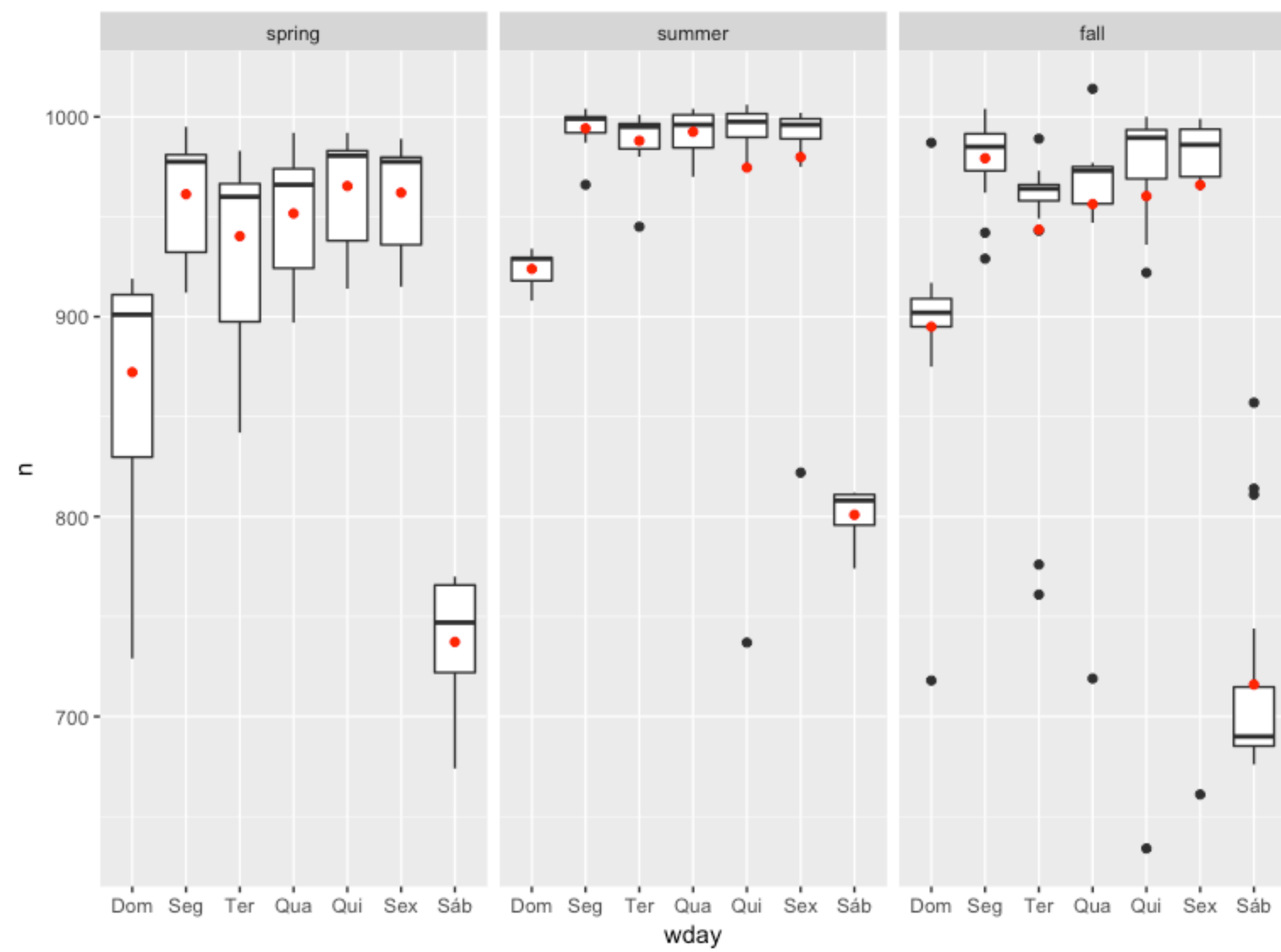
Para os outros dias da semana



- O summer term tem número de voo mais altos do que os outros terms.
- Esse efeito deve ser incluído no modelo.

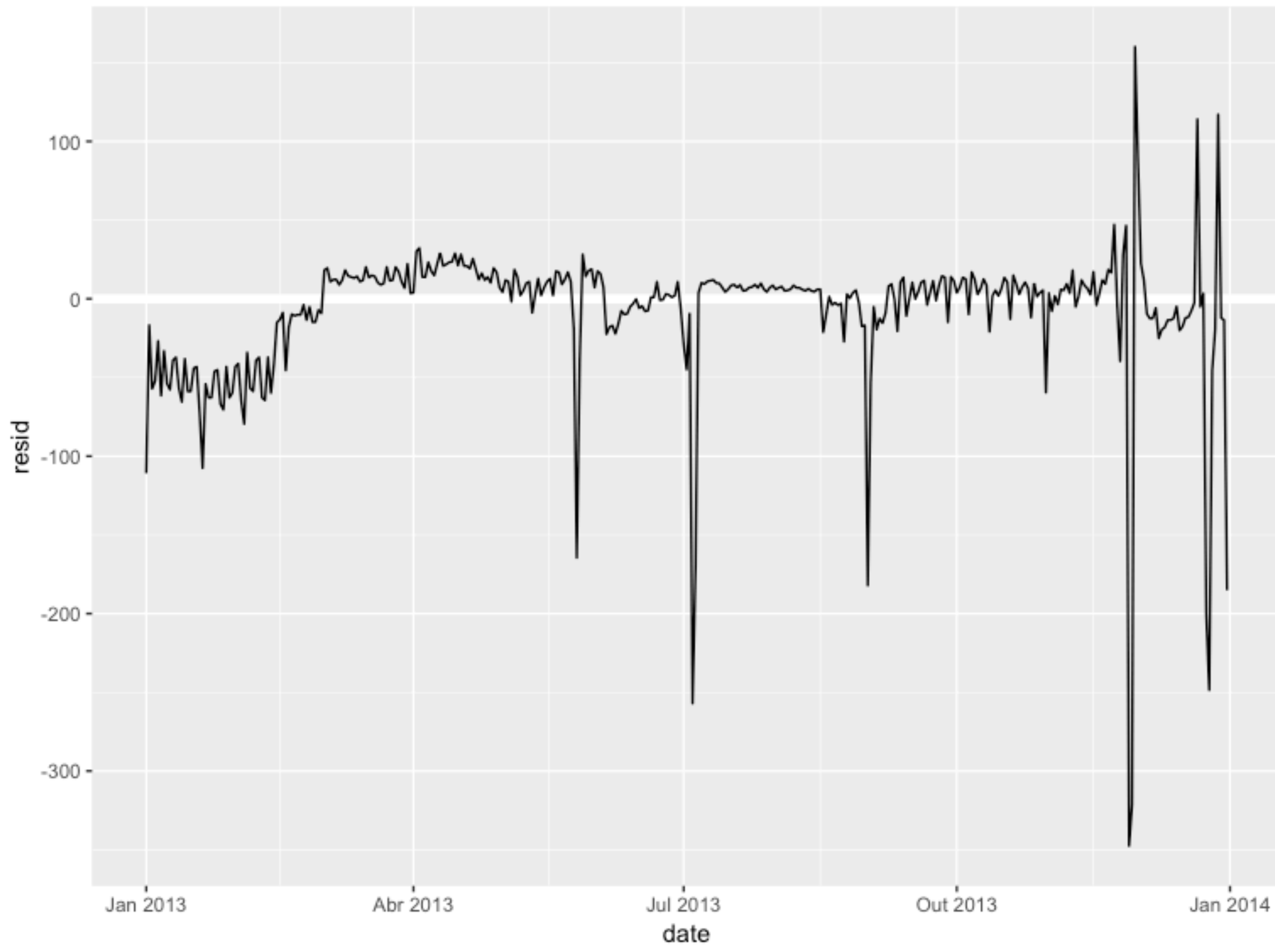
Modelo com term





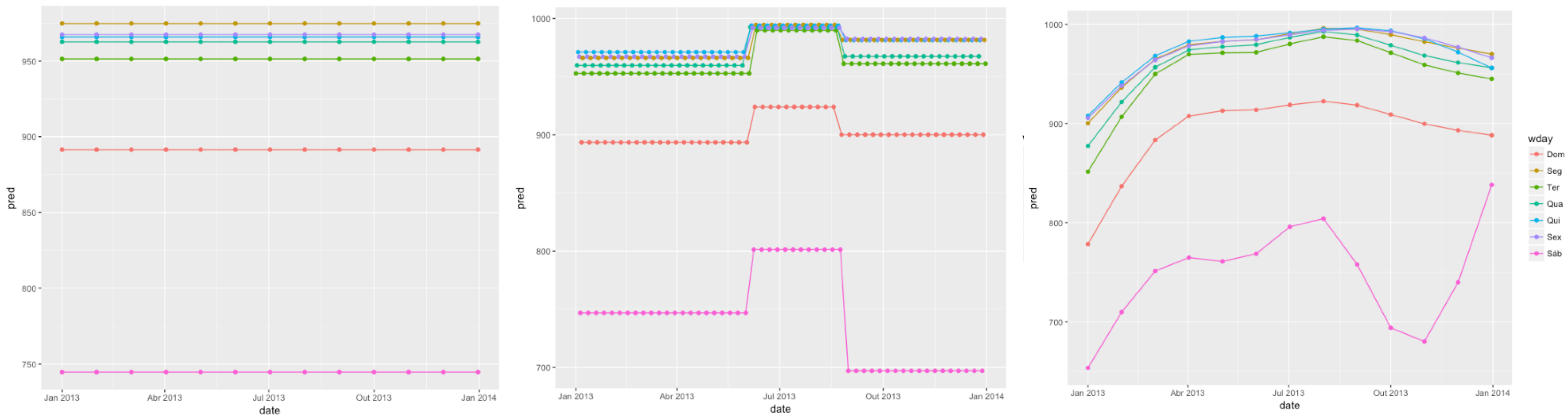
- As médias estimadas são afetadas pelos outliers nos dados.

Resíduos



- Agora não vemos mais os efeitos de dia da semana e dos terms.
- Conseguimos ver grandes outliers devido a feriados nacionais. E também uma tendência ao longo do ano.

Melhorando o modelo



- Gráfico: Modelo ajustado com efeito do dia da semana, com efeito do dia da semana e term e modelo que usa splines (veremos mais a frente esse ajuste).

- Note que o modelo de regressão mantém a interpretabilidade.

Alguns comentários

- Note que algum pensamento crítico foi necessário nesse exemplo para que variáveis explicativas razoáveis fossem sugeridas.

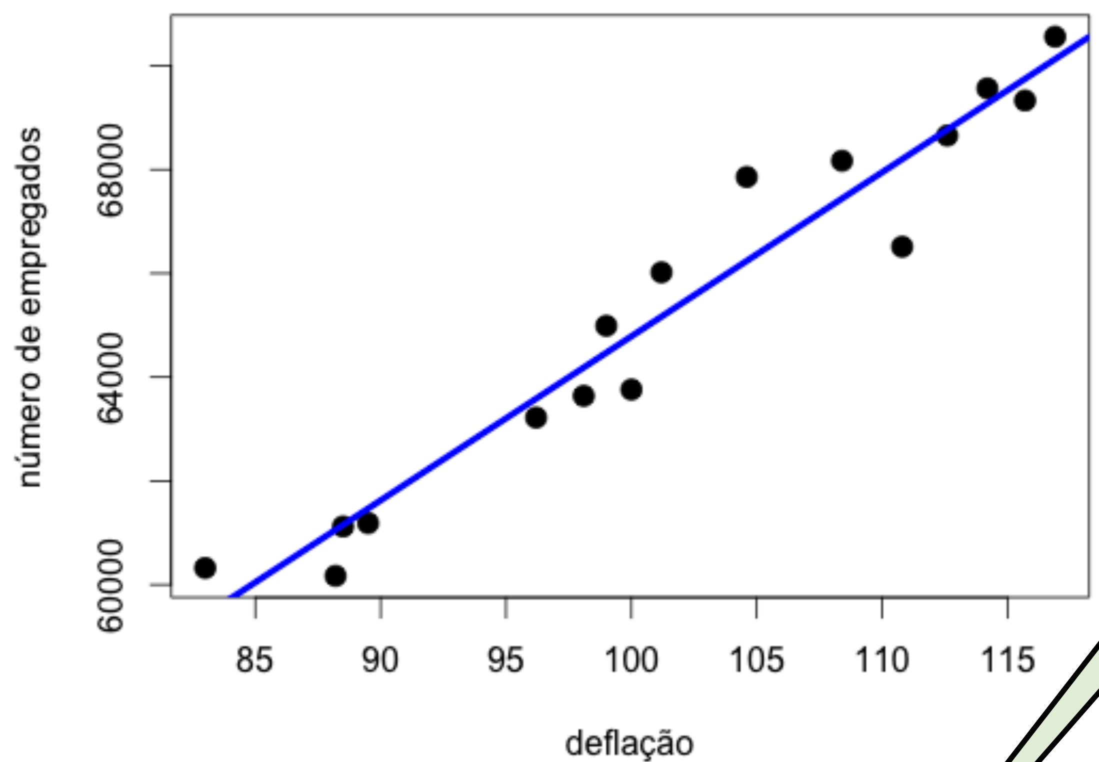
- Um método totalmente automático pode não considerar esses nuances que o cientista de dados fazendo a modelagem deve considerar.

- Isso é verdade mesmo em contexto puramente preditivos.

- O modelo considera observações não correlacionadas, porém temos uma série temporal. O modelo poderia ser melhorado para incluir correlação temporal.

Regressão: contexto de Estatística x contexto de Machine learning

Contexto estatístico



Foco na interpretação dos coeficientes, verificação das hipóteses.

- O ajuste da regressão é bom?
- O modelo parece útil para previsão?
- As suposições básicas são válidas?

Os erros têm média 0, variância constante e são não correlacionados?

Contexto de machine learning

- Iremos focar no objetivo preditivo, isto é, no caso de regressão linear onde $\mathbf{x} = (x_1, \ldots, x_p)$ temos

$$g(\mathbf{x}) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_p x_p$$

- Após ajustar o modelo utilizando $\mathbf{x_1}, \ldots, \mathbf{x}_N$, para novos dados temos

$$g(\mathbf{x}_{N+1}) \approx y_{N+1}, \ldots, g(\mathbf{x}_{N+k}) \approx y_{n+k}$$

- Escolheremos o modelo com o melhor poder preditivo.

> - O modelo é utilizado apenas para criar bons algoritmos para prever bem novas observações.
> - Muitas vezes, não há modelo probabilístico explícito por trás dos algoritmos utilizados.

Função de risco

- Seja $g(\mathbf{x})$ a função utilizada para previsão.
- Para uma nova observação $(\mathbf{x}, y)$, podemos definir o risco de um modelo preditivo por
$$R_g = E[y - g(\mathbf{x})]^2$$
- Esta função de risco é a perda quadrática esperada e será dita risco quadrático.
- Quanto menor o risco melhor o modelo, segundo a perda escolhida.

Função de risco

- Podemos mostrar que

$$R_g = E[y - g(\mathbf{x})]^2$$
$$= E\{[y - E(y \mid x)]^2\} + E\{[E(y \mid x) - g(\mathbf{x})]^2\}$$

- Condicional a $\mathbf{x}$, temos que $g(x) = E(y \mid x)$ minimiza o risco.

- Se X visto como aleatório temos o mesmo resultado com a esperança também tomada na distribuição de $\mathbf{x}$.

Modelagem e previsão

Modelo global e modelo local (perda quadrática)

Outras relações mais flexíveis que a regressão linear poderiam ser usadas aqui.

- Vimos que $E(y \mid x)$ minimiza a perda quadrática.
- Podemos pensar em duas soluções:

  (i) Regressão: $\hat{g}(x) = \hat{\beta}_0 + \hat{\beta}_1, \ldots + \hat{\beta}_p x_p$;

  (ii) Vizinhos próximos: $\hat{g}(x) = media(y_i \mid x_i \in V_k(x))$.

- Aproxima-se a media teórica usando a média empírica de dados de treinamento.
- Considere-se apenas os k vizinhos mais próximos.

Para N grande a média empírica irá convergir para a teórica. Porém a **taxa de convergência é mais lenta quando p cresce**.

Aprendizado supervisionado

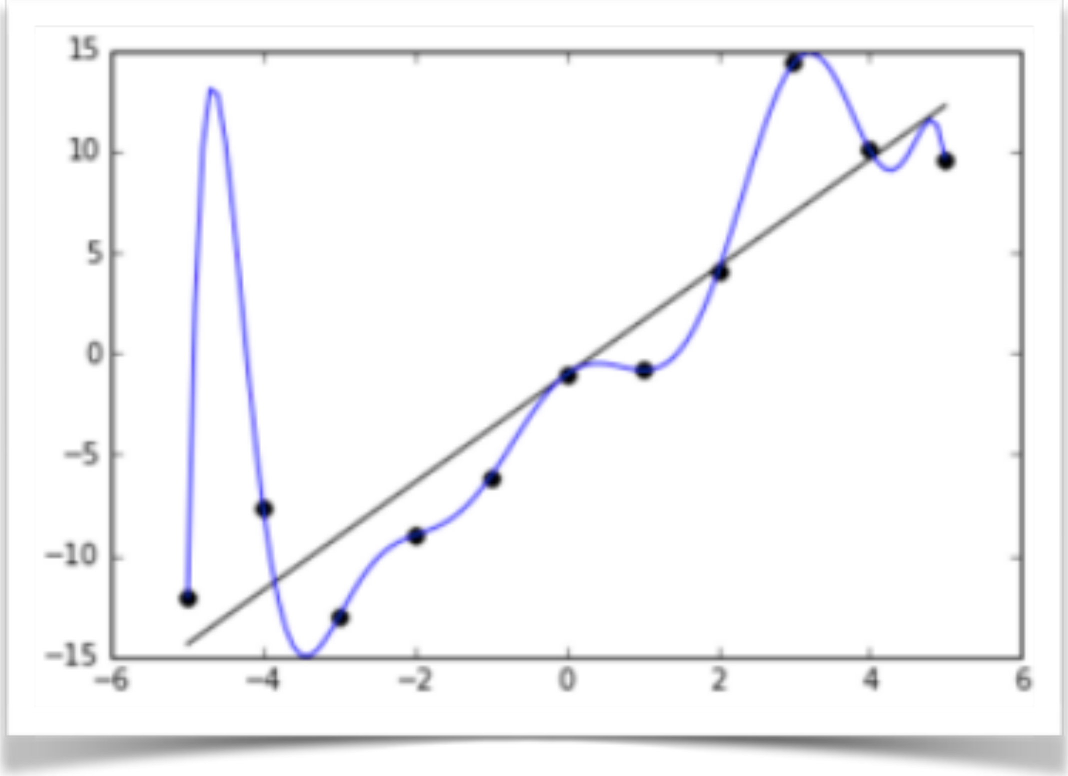
- Considere interesse em estimar $g(\mathbf{x})$ usando dados $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)$ e a relação
$$y = g(\mathbf{x}) + \epsilon.$$
- Dados de treinamento $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)$ alimentam um algoritmo que produz outputs $\hat{g}(x_i)$;
- O erro $y_i - \hat{g}(x_i)$ realimenta o algoritmo de modo a aperfeiçoá-lo.

Dados de treinamento e teste

- Note que se usamos apenas os dados de treinamento para escolher o melhor modelo de regressão podemos ter sobreajuste.

- Por exemplo um polinômio com n-1 graus ajusta os dados perfeitamente.

- Mas esse modelo não seria útil para prever fora da amostra de treinamento.



Dados de treinamento e teste

- Definimos então um conjunto de teste.
- Este não é usado para estimar parâmetros, apenas para calcular o risco do modelo.
- Por exemplo,

$$R_g = \sum_{j=1}^{M} (y_{N+j} - \hat{g}(\mathbf{x_{N+j}}))^2$$

Dados de teste

- Pode ser escolhido de forma aleatória dentre todos os dados observados.

- Pode-se definir por exemplo 70% dos dados para treinamento e 30% para teste.

- Validação cruzada (leave-one-out):

$$R_g = \sum_{i=1}^{N+M} (y_i - \hat{g}_{-i}(\mathbf{x_i}))^2,$$

onde g_{-i} foi estimada usando todos os dados menos a i-ésima observação.

Previsão fora da amostra de treinamento

Previsão no caso de regressão múltipla

Vimos que no caso de regressão simples um valor predito é dado por

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0.$$

Vimos também que extrapolação pode ser perigoso.

Por exemplo, o comportamento fora do intervalo de x's observados pode ser muito diferente de uma reta!

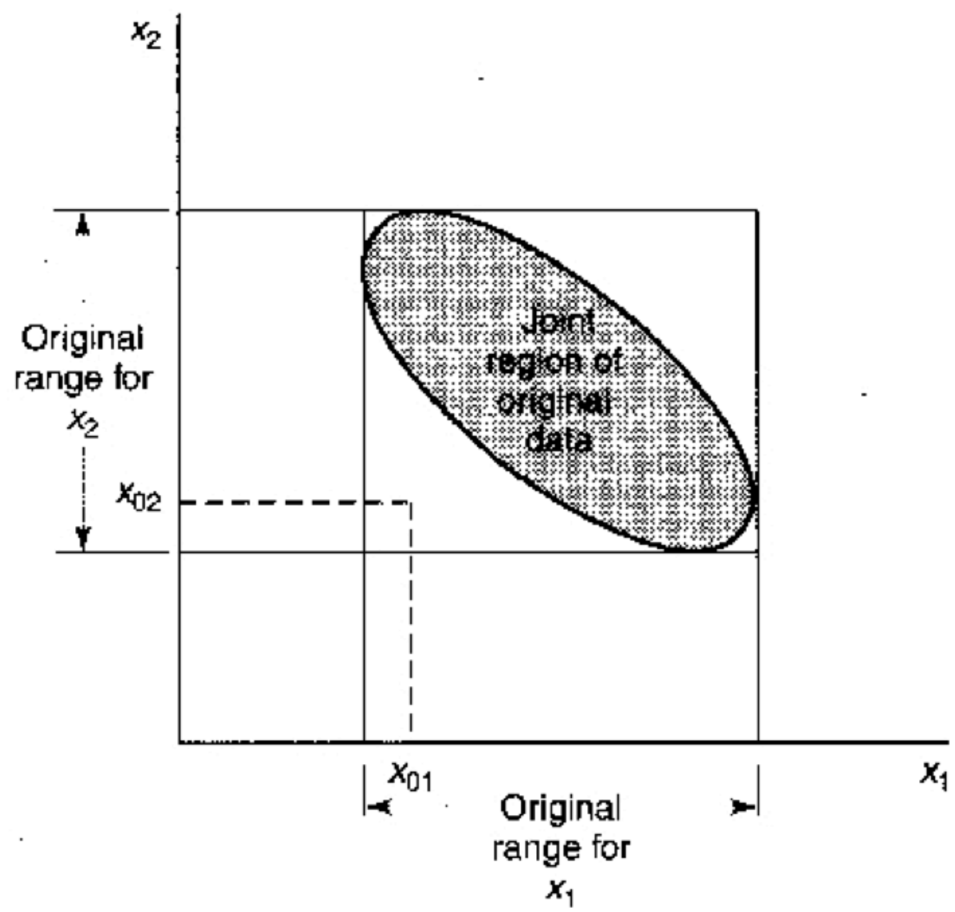
Se a previsão é feita fora do intervalo $(min(x), max(x))$ chamamos extrapolação.

No caso de regressão múltipla, temos $\mathbf{x_1, x_2, \ldots, x_k}$.

Como saber se estamos extrapolando?

Região de previsão

- Em regressão múltipla é difícil saber se x_0 está ou não dentro do domínio dos x's observados ou se trata-se de uma extrapolação.
- Testar cada componente x_{0j} não é suficiente!



Região de previsão

Considere a matriz $H = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'$.

$h_{max} = max\{h_{ii}\}$ está no limite da região do domínio observado de X.

Portanto, para verificar se x_0 está dentro da região de previsão basta verificar se

$$h_{00} = x_0'(\mathbf{XX})^{-\mathbf{1}}x_0 < h_{max}.$$

Caso contrário temos uma extrapolação!

Prevendo para um novo valor de x = x_0

```
################## previsão ##################
beta.hat = ajuste2$coef

range(X[,2]) ##### 234289 554894
range(X[,3]) ##### 107608 130081
X0 = c(1,300000,119000)

y0.hat = t(X0)%*%beta.hat #### 59137.06
```

Mas temos uma extrapolação???

```
############## calculando h00 #################

XX = cbind(rep(1,16),X[,-1])
max(hatvalues(ajuste2)) #### 0.3384269
t(X0)%*%solve(t(XX)%*%XX)%*%X0 ####  4.66992
```

SIM!!

Exemplo: nem sempre o modelo correto tem o melhor poder preditivo

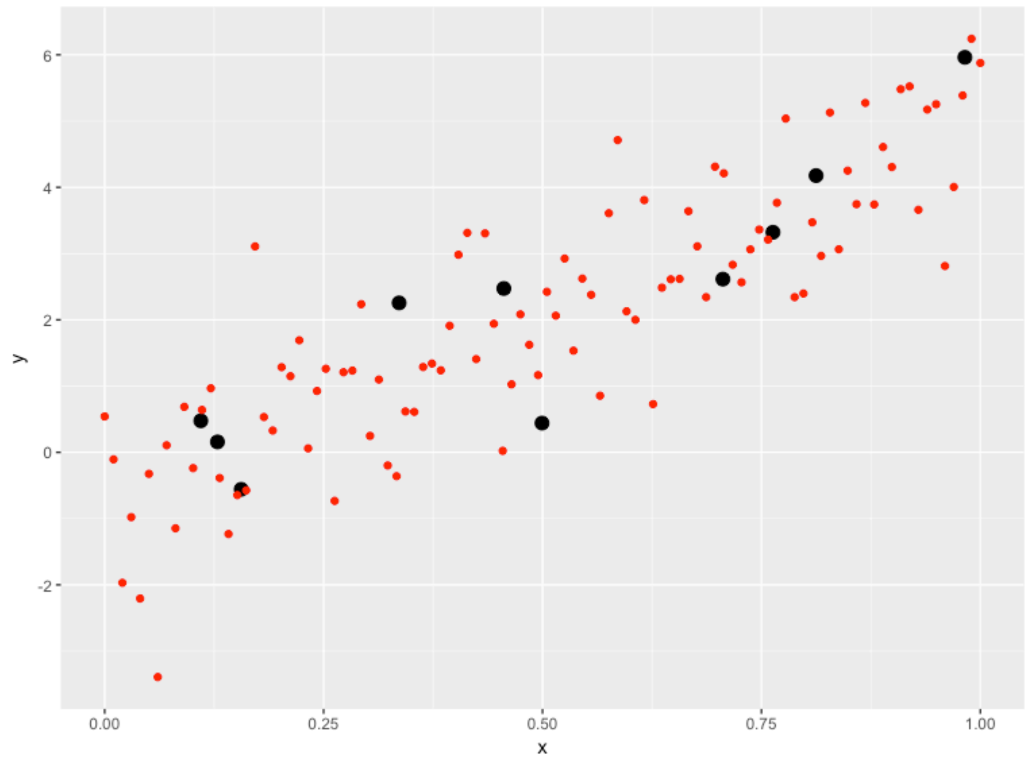
- Considere o modelo de regressão polinomial

  (i) $y = x^t\beta + \epsilon$;

  (ii) $\beta = (0, 3, 2, 0.2, 0.1, 0.1)$ e $x = (1, x, x^2, \ldots, x^5)$;

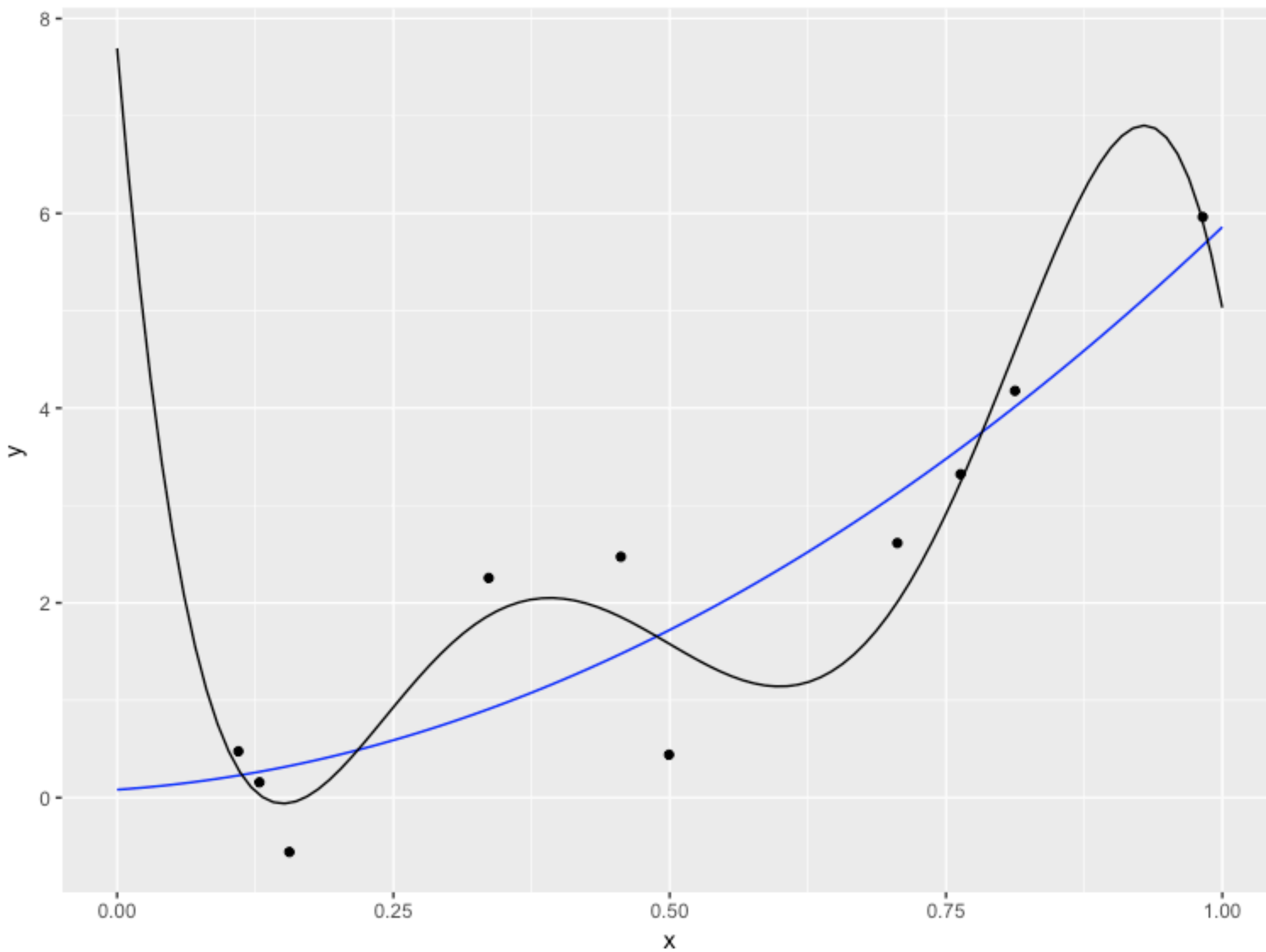
  (iii) $\epsilon \sim N(0, 1)$.

- Vamos ajustar um modelo polinomial de ordem 2 e o modelo correto de ordem 5.



Neste exemplo, temos 10 pontos de treinamento e 100 pontos de teste.

Exemplo: nem sempre o modelo correto tem o melhor poder preditivo



- Para 100 pontos de teste o erro quadrático médio do modelo de ordem 2 é **111** enquanto para o modelo de ordem 5 é **437**.

Seleção de variáveis

Métodos forward, backward e stepwise

- O número de modelos possíveis no modelo de regressão quando temos p covariáveis é 2^p.
- Quando p cresce é inviável selecionar variáveis apenas olhando para análises descritivas ou ajustando todos os modelos. E.g. $2^{10} = 1024$.
- Método forward, backward e stepwise: algoritmos sequenciais nos quais uma variável por vez é introduzida no modelo.

  A variável é escolhida dependendo de algum critério de ajuste.

  O número de modelos testados neste algoritmo é $1 + p(p+1)/2$.

Ideia dos algoritmos

- **Seleção Forward:** começa com nenhum preditor; sequencialmente adiciona o preditor mais importante no modelo; parada quando não há melhora na medida de interesse.

- **Seleção Backward:** começa com todos os preditores no modelo e sequencialmente remove o preditor que menos contribui para o modelo; parada quando todos os preditores do modelo são significativos.

- **Seleção stepwise:** combina seleção forward com backward; começa com nenhum preditor; sequencialmente adicional o preditor que mais contribui para o modelo; após incluir uma variável ele remove variáveis que possivelmente não são mais necessárias no modelo.

Exemplo: marketing

- Vamos analisar o banco de dados marketing do pacote datarium.
- A base contém o impacto nas vendas do gasto com 3 tipos de propagandas: youtube, facebook e jornal. Dimensão: 200 por 3.
- **Objetivo do ajuste**: medir o impacto do gasto nas vendas.
- Vamos dividir os dados em treinamento e teste.

	youtube	facebook	newspaper	sales
1	276.12	45.36	83.04	26.52
2	53.40	47.16	54.12	12.48
3	20.64	55.08	83.16	11.16
4	181.80	49.56	70.20	22.20

```
set.seed(123)
training.samples <- marketing$sales %>%
  createDataPartition(p = 0.8, list = FALSE)
train.data  <- marketing[training.samples, ]
test.data <- marketing[-training.samples, ]
```

Exemplo: marketing

- Modelo para todas as variáveis.

```
Model1 <- lm(sales ~ youtube + facebook + newspaper, data = train.data)
summary(model)
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.391883   0.440622   7.698 1.41e-12 ***
youtube     0.045574   0.001592  28.630  < 2e-16 ***
facebook    0.186941   0.009888  18.905  < 2e-16 ***
newspaper   0.001786   0.006773   0.264    0.792
```

- Modelo sem newspaper.

```
Model2 <- lm(sales ~ youtube + facebook, data = train.data)
summary(model2)
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.434458   0.408770   8.402 2.32e-14 ***
youtube     0.045582   0.001587  28.725  < 2e-16 ***
facebook    0.187877   0.009202  20.418  < 2e-16 ***
```

Exemplo: marketing

Erro preditivo:

```
predictions <- model1 %>% predict(test.data)
RMSE(predictions, test.data$sales)
1.590096
predictions <- model2 %>% predict(test.data)
RMSE(predictions, test.data$sales)
1.57911
```

Podemos melhorar o modelo para previsão? Vamos considerar o modelo com interações.

```
model.full <- lm(sales ~ youtube * facebook * newspaper, data = train.data)
summary(model.full)
```

```
Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               7.269e+00  6.552e-01  11.094  < 2e-16 ***
youtube                   2.136e-02  3.044e-03   7.015 6.81e-11 ***
facebook                  2.571e-02  1.861e-02   1.381    0.169
newspaper                 2.617e-02  1.961e-02   1.334    0.184
youtube:facebook          9.526e-04  9.089e-05  10.481  < 2e-16 ***
youtube:newspaper        -8.706e-05  8.555e-05  -1.018    0.310
facebook:newspaper       -2.132e-04  4.473e-04  -0.477    0.634
youtube:facebook:newspaper 2.865e-07 2.050e-06   0.140    0.889
```

```
RMSE(predictions, test.data$sales)
1.020196
```

Exemplo: marketing

- Erro preditivo do modelo com interações:

```
predictions <- model.full %>% predict(test.data)
RMSE(predictions, test.data$sales)
1.020196
```

- Esse modelo tem todas as variáveis e interações. Podemos ajustar um modelo mais simples?

```
step.model2 <- stepAIC(model.full, direction = "both", trace = FALSE)
summary(step.model2)
predictions <- step.model2 %>% predict(test.data)
RMSE(predictions, test.data$sales)
1.007243
```

```
Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         7.526e+00  3.608e-01  20.860   <2e-16 ***
youtube             2.099e-02  1.773e-03  11.838   <2e-16 ***
facebook            1.867e-02  1.075e-02   1.736   0.0845 .
newspaper           1.747e-02  7.368e-03   2.371   0.0190 *
youtube:facebook    9.579e-04  5.211e-05  18.382   <2e-16 ***
youtube:newspaper  -7.101e-05  3.306e-05  -2.148   0.0333 *
```

Exemplo: marketing

- Caminho do método stepwise:

```
step.model2$anova
```

```
Stepwise Model Path
Analysis of Deviance Table

Initial Model:
sales ~ youtube * facebook * newspaper

Final Model:
sales ~ youtube + facebook + newspaper + youtube:facebook + youtube:newspaper


                          Step Df   Deviance Resid. Df Resid. Dev      AIC
1                                                  154   209.8617 57.93407
2 - youtube:facebook:newspaper  1 0.02662091       155   209.8883 55.95462
3          - facebook:newspaper  1 0.68097906       156   210.5693 54.47938
```

Transformação da variável resposta

Suponha que desejamos transformar a variável y para obter variância constante.

Porém não sabemos qual transformação usar.

O método de Box-Cox transforma y para $g_\lambda(y)$ onde

$$g_\lambda(y) = \begin{cases} \frac{y^\lambda - 1}{\lambda} & se\ \lambda \neq 0 \\ log(y) & se\ \lambda = 0 \end{cases}$$

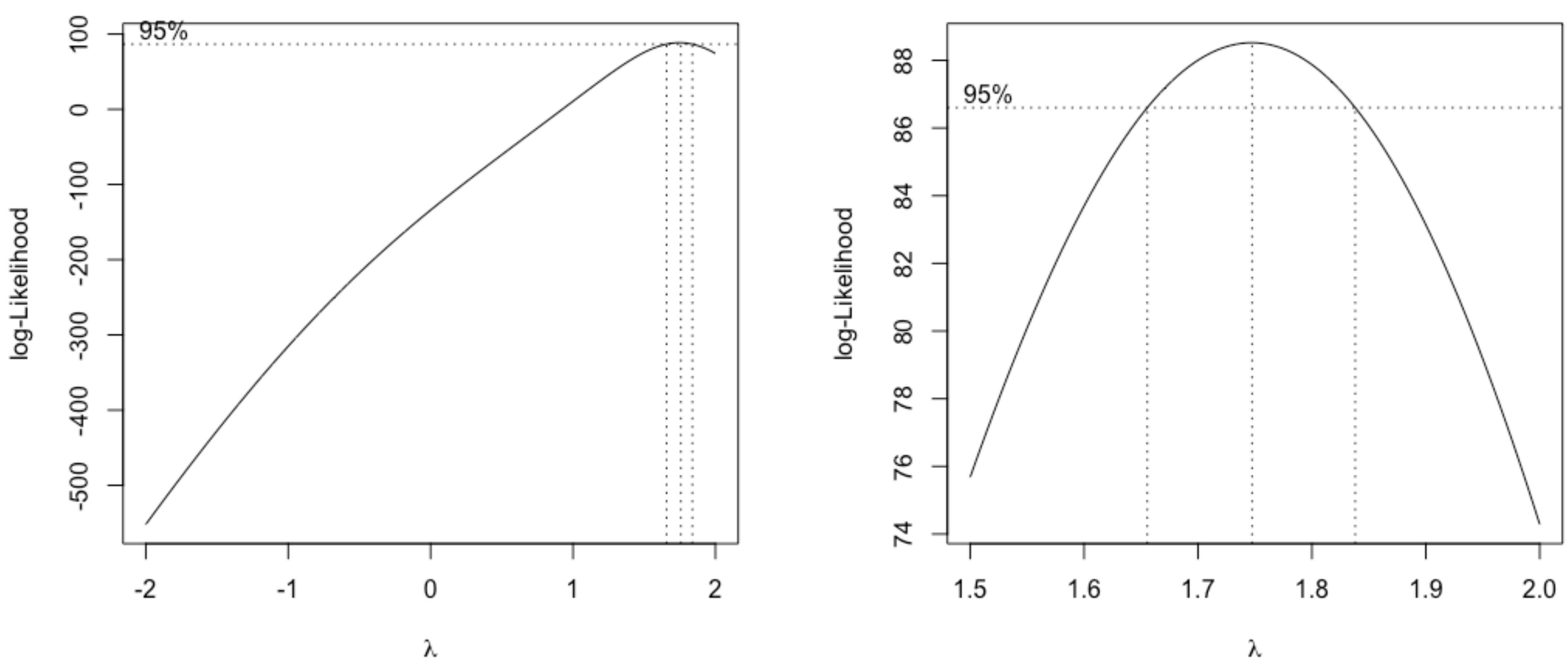
Para $y > 0$, g é contínua em λ. Podemos usar máxima verossimilança para estimar λ.

No R: usar a função *boxcox(ajuste)* da biblioteca *MASS*.

Exemplo: sales

```
############### transformação Box-Cox ###############

library(MASS)
boxcox(model.full)
transf.y = boxcox(model.full,lambda = seq(1.5, 2, 1/20)) ### refinando a busca
```



Exemplo: preço de venda

```
############### transformação Box-Cox ###############

lambda0 = transf.y$x[transf.y$y==max(transf.y$y)]
train.data$salesnew = train.data$sales^lambda0

model.full2 <- lm(salesnew ~ youtube*facebook*newspaper, data = train.data)
summary(model.full)
summary(model.full2)
```

```
                            Estimate Std. Error t value Pr(>|t|)    
(Intercept)                7.269e+00  6.552e-01  11.094  < 2e-16 ***
youtube                    2.136e-02  3.044e-03   7.015 6.81e-11 ***
facebook                   2.571e-02  1.861e-02   1.381    0.169    
newspaper                  2.617e-02  1.961e-02   1.334    0.184    
youtube:facebook           9.526e-04  9.089e-05  10.481  < 2e-16 ***
youtube:newspaper         -8.706e-05  8.555e-05  -1.018    0.310    
facebook:newspaper        -2.132e-04  4.473e-04  -0.477    0.634    
youtube:facebook:newspaper 2.865e-07  2.050e-06   0.140    0.889    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.167 on 154 degrees of freedom
Multiple R-squared:  0.9675,	Adjusted R-squared:  0.966 
F-statistic: 654.9 on 7 and 154 DF,  p-value: < 2.2e-16
```

```
                             Estimate Std. Error t value Pr(>|t|)    
(Intercept)                 3.745e+01  5.485e+00   6.828 1.87e-10 ***
youtube                     1.441e-01  2.548e-02   5.654 7.42e-08 ***
facebook                   -1.688e-01  1.558e-01  -1.084    0.280    
newspaper                   5.196e-02  1.642e-01   0.316    0.752    
youtube:facebook            1.809e-02  7.609e-04  23.777  < 2e-16 ***
youtube:newspaper          -1.885e-04  7.162e-04  -0.263    0.793    
facebook:newspaper          1.316e-03  3.744e-03   0.351    0.726    
youtube:facebook:newspaper -2.995e-06  1.716e-05  -0.175    0.862    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.772 on 154 degrees of freedom
Multiple R-squared:  0.9901,	Adjusted R-squared:  0.9896 
F-statistic:  2191 on 7 and 154 DF,  p-value: < 2.2e-16
```