

**MBA  
USP  
ESALQ**

# **Sinistralidade automobilística no Brasil explicada por meio de Regressão Logística Multinomial**

Tarciso Cavalcanti de Mello Filho  
Walter Mesquita Filho

INTRODUÇÃO



```
graph TD; A[INTRODUÇÃO] --> B[MATERIAL E MÉTODOS]; B --> C[RESULTADOS E DISCUSSÃO]; C --> D[CONSIDERAÇÕES FINAIS];
```

MATERIAL E MÉTODOS

RESULTADOS E DISCUSSÃO

CONSIDERAÇÕES FINAIS

# Contexto Histórico

- 1808 – **Companhia de Seguros Boa Fé:** Seguros Marítimos; Intituição do Tribunal do Comércio e da Provedoria de Seguros;
- 1822 – **Independência do Brasil:** Autonomia; Decreto de Lei em 1931; Criação do Código Comercial Brasileiro em 1950
- 1889-1914 – **Crescimento do Ramo de Seguros:** Introdução de Seguradoras (Nacionais/Internacionais); Seguros Marítimos, Terrestres e de Vida; Modelo Mercantil → Modelo Financeiro
- 1966 – **Decreto de Lei nº 73:** Criação da CNSP, IRB e SUSEP.
- Presente – **SUSEP:** Regulação e Fiscalização de **Operações de Seguros**, Previdência Privada (VGBl) e Capitalização

# Problema e Objetivo

- **Cenário (2023):** R\$36,72 Bilhões em Prêmios; 21,04 Bilhões em Indenização; Sinistralidade de 57,3%
- **Problema:** Quais são os fatores que impactam a sinistralidade automobilística no Brasil?
- **Objetivo:** Analisar o efeito de variáveis financeiras, biométricas e patrimoniais sobre a sinistralidade ocorrida no Brasil, através do uso de modelos de Regressão Logística Multinomial.

INTRODUÇÃO



MATERIAL E MÉTODOS

RESULTADOS E DISCUSSÃO

CONSIDERAÇÕES FINAIS



# Tipologia, Coleta e Amostra

- **Tipologia da Pesquisa:** Essência → Quantitativa; Natureza → Aplicada; Procedimento → Ex-post-facto; Objetivo → Explicativa; Temporalidade → Transversal
- **Coleta de Dados:** Dados públicos; Sistema Autoseg ; 2º semestre 2019
- **Amostra:** Grupos de apólices de seguros CASCO; 3.210.981 → Data Wrangling → 2.268.204
- **Data Wrangling:** Remoção NA's; Remoção inconsistências; Remoção Sexo (PJ); Cálculo da Sinistralidade →  $SINISTRALIDADE = \frac{PRÊMIO}{INDENIZAÇÃO} \times 100$ ; Criação da SAC (Sinistralidade Automobilística Categorizada)

# Variáveis Utilizadas

- Categoria Tarifária
- Ano Modelo
- Sexo
- Faixa Etária
- Exposição
- Prêmio
- Importância Segurada
- Quantidade de Sinistros (I-V)
- Indenização (I-V)

- SAC I – Roubo ou Furto
- SAC II – Colisão Parcial
- SAC III – Colisão Perda Total
- SAC IV – Incêndio
- SAC V – Outros

| CATEGORIA  | CRITÉRIO                            | SAC (Y) |
|------------|-------------------------------------|---------|
| NULA       | $SINISTRALIDADE = 0$                | 0       |
| BAIXA      | $0 < SINISTRALIDADE \leq 50\%$      | 1       |
| MÉDIA      | $50\% < SINISTRALIDADE \leq 100\%$  | 2       |
| ALTA       | $100\% < SINISTRALIDADE \leq 200\%$ | 3       |
| MUITO ALTA | $SINISTRALIDADE > 200\%$            | 4       |

# Análise Exploratória

- **Variáveis Qualitativas:** Gráficos Pizza
- **Variáveis Quantitativas:** Medidas de Posição; Medidas de Dispersão
- **Sinistralidade:** Medidas de Posição; Medidas de Dispersão; Gráficos de Dispersão; Histogramas
- **SAC:** Gráficos Pizza com categoria NULA; Gráficos Pizza sem categoria NULA



# Regressão Logística Multinomial

- 3 ou + Categorias
- Distribuição Binomial
- Estimação do efeito das Variáveis Independentes sobre a Variável Dependente (SAC)

$$g_k(x) = \ln \left[ \frac{P(Y = k|x)}{P(Y = 0|x)} \right]$$

$$g_k(x) = \beta_{k0} + \beta_{k1}x_1 + \beta_{k2}x_2 + \dots + \beta_{kr}x_r$$

$$g_k(x) = x'\beta_k$$

- Melhor critério AIC x BIC

$$AIC = -\frac{2}{N} \times LL + 2 \times \frac{k}{N}$$

$$BIC = -2 \times LL + \log N \times k$$

- Criação 511 Modelos

$$Qntd\ Modelos = 2^r - 1$$

- Variação % odds

$$Var\ \% odd = (e^{Estimate} - 1) \times 100$$

INTRODUÇÃO



```
graph TD; A[INTRODUÇÃO] --> B[MATERIAL E MÉTODOS]; B --> C[RESULTADOS E DISCUSSÃO]; C --> D[CONSIDERAÇÕES FINAIS];
```

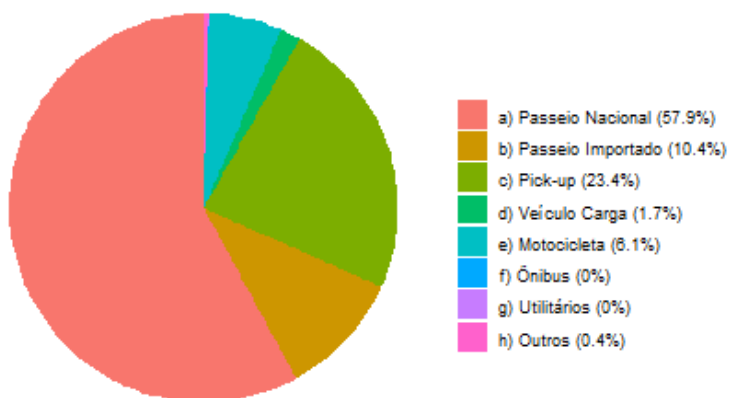
MATERIAL E MÉTODOS

RESULTADOS E DISCUSSÃO

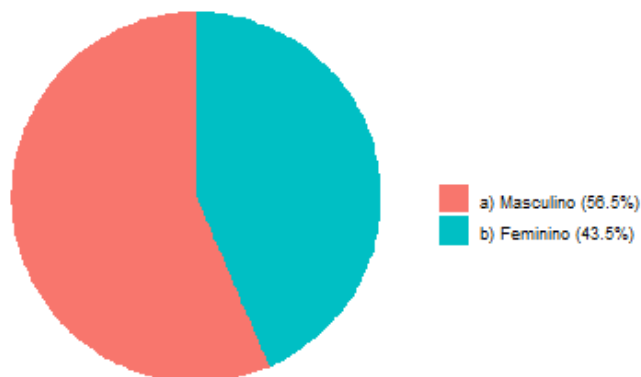
CONSIDERAÇÕES FINAIS

# Análise Exploratória – Variáveis Qualitativas

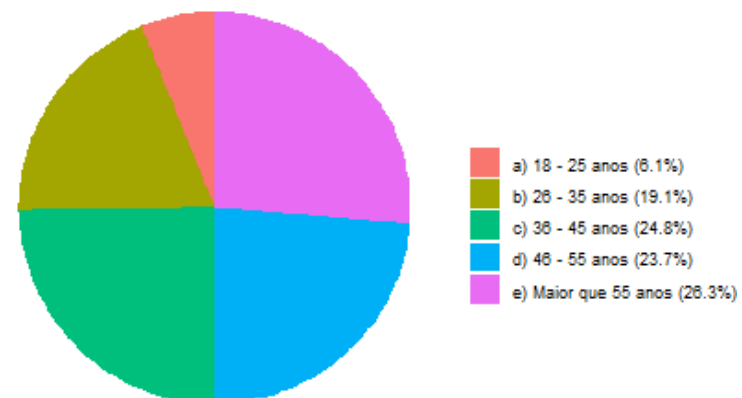
CATEGORIA TARIFÁRIA



SEXO



FAIXA ETÁRIA



# Análise Exploratória – Variáveis Quantitativas

| VARIÁVEL         | MÍN  | 1ºQUARTIL | MEDIAN  | 3ºQUARTIL | MÁX         | MEAN      |
|------------------|------|-----------|---------|-----------|-------------|-----------|
| ANO_MODELO       | 1950 | 2011      | 2014    | 2017      | 2021        | 2013      |
| EXPOSIÇÃO        | 0    | 0,5       | 0,82    | 2,21      | 916,34      | 3,047     |
| PRÊMIO           | 0    | 533,8     | 1.240,4 | 3.355     | 1.453.984,6 | 4.356,3   |
| IMP. SEGURADA    | 0    | 25.632    | 40.634  | 65.629    | 5.469.652   | 54.153    |
| QD SINISTROS I   | 0    | 0         | 0       | 0         | 43          | 0,02182   |
| INDENIZAÇÃO I    | 0    | 0         | 0       | 0         | 715.720     | 733,9     |
| QD SINISTROS II  | 0    | 0         | 0       | 0         | 54          | 0,1837    |
| INDENIZAÇÃO II   | 0    | 0         | 0       | 0         | 392.335     | 1.147     |
| QD SINISTROS III | 0    | 0         | 0       | 0         | 13          | 0,2185    |
| INDENIZAÇÃO III  | 0    | 0         | 0       | 0         | 768.709     | 898,3     |
| QD SINISTROS IV  | 0    | 0         | 0       | 0         | 3           | 0,0007746 |
| INDENIZAÇÃO IV   | 0    | 0         | 0       | 0         | 422.874     | 22,2      |
| QD SINISTROS V   | 0    | 0         | 0       | 0         | 381         | 0,7968    |
| INDENIZAÇÃO V    | 0    | 0         | 0       | 0         | 261.575     | 427,5     |

# Análise Exploratória – Variáveis Quantitativas

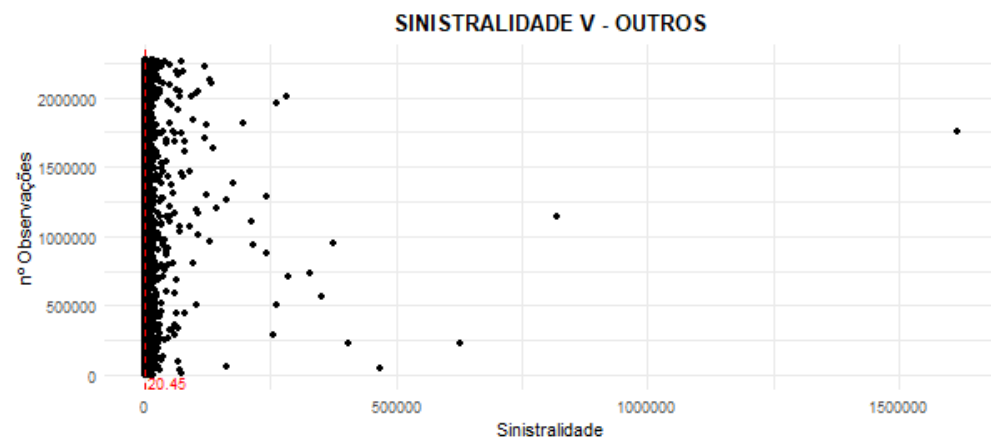
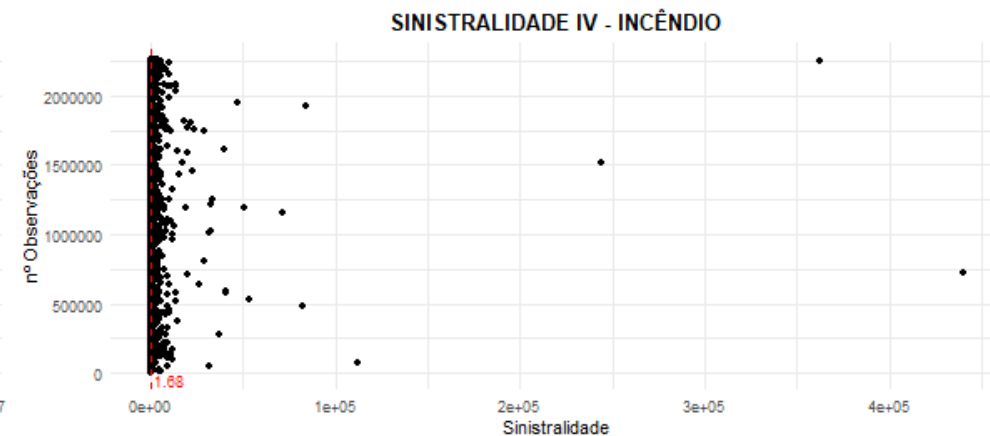
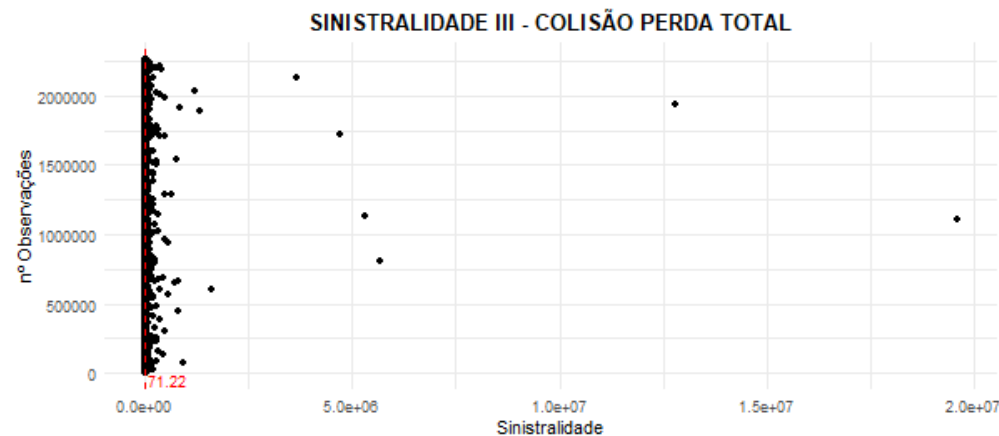
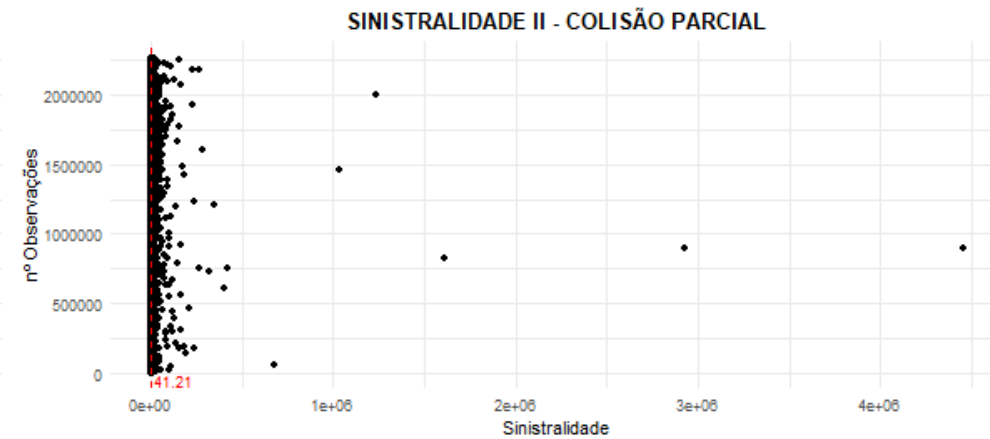
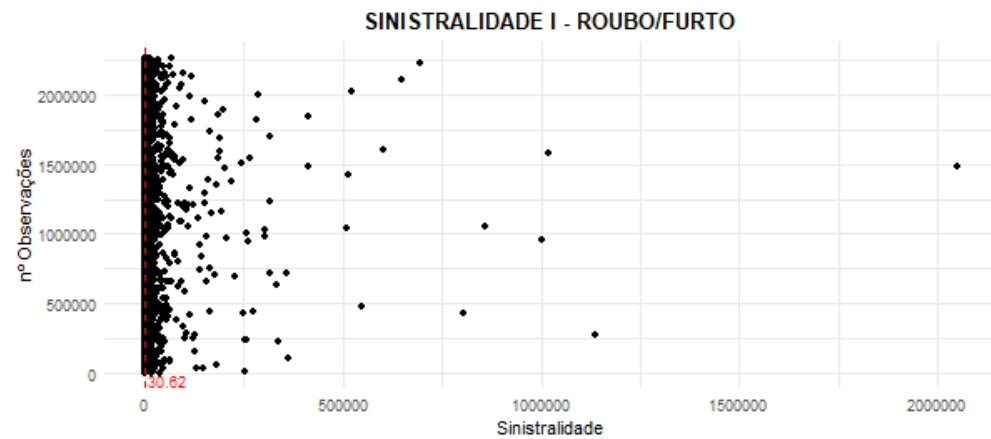
| VARIÁVEL         | RANGE        | SD        | SKEW   | KURTOSIS  | SE    |
|------------------|--------------|-----------|--------|-----------|-------|
| ANO_MODELO       | 71           | 5,16      | -1,29  | 2,83      | 0,00  |
| EXPOSIÇÃO        | 916,34       | 10,04     | 17,06  | 566,39    | 0,01  |
| PRÊMIO           | 1.453.984,54 | 14.056,83 | 18,42  | 727,84    | 9,33  |
| IMP. SEGURADA    | 5.469.651,77 | 51.639,05 | 6,46   | 188,45    | 34,29 |
| QD SINISTROS I   | 43           | 0,22      | 34,24  | 3.250,06  | 0,00  |
| INDENIZAÇÃO I    | 715.720      | 7.812,82  | 19,65  | 633,88    | 5,19  |
| QD SINISTROS II  | 54           | 0,78      | 13,15  | 369,65    | 0,00  |
| INDENIZAÇÃO II   | 392.335      | 5.930,56  | 13,79  | 353,98    | 3,94  |
| QD SINISTROS III | 13           | 0,16      | 9,22   | 137,60    | 0,00  |
| INDENIZAÇÃO III  | 768.709      | 8.228,12  | 17,07  | 523,59    | 5,46  |
| QD SINISTROS IV  | 3            | 0,03      | 42,56  | 2.091,84  | 0,00  |
| INDENIZAÇÃO IV   | 422.874      | 1.234,78  | 119,36 | 22.840,66 | 0,82  |
| QD SINISTROS V   | 381          | 3,92      | 22,48  | 967,97    | 0,00  |
| INDENIZAÇÃO V    | 261575       | 3.345,56  | 20,65  | 636,35    | 2,22  |

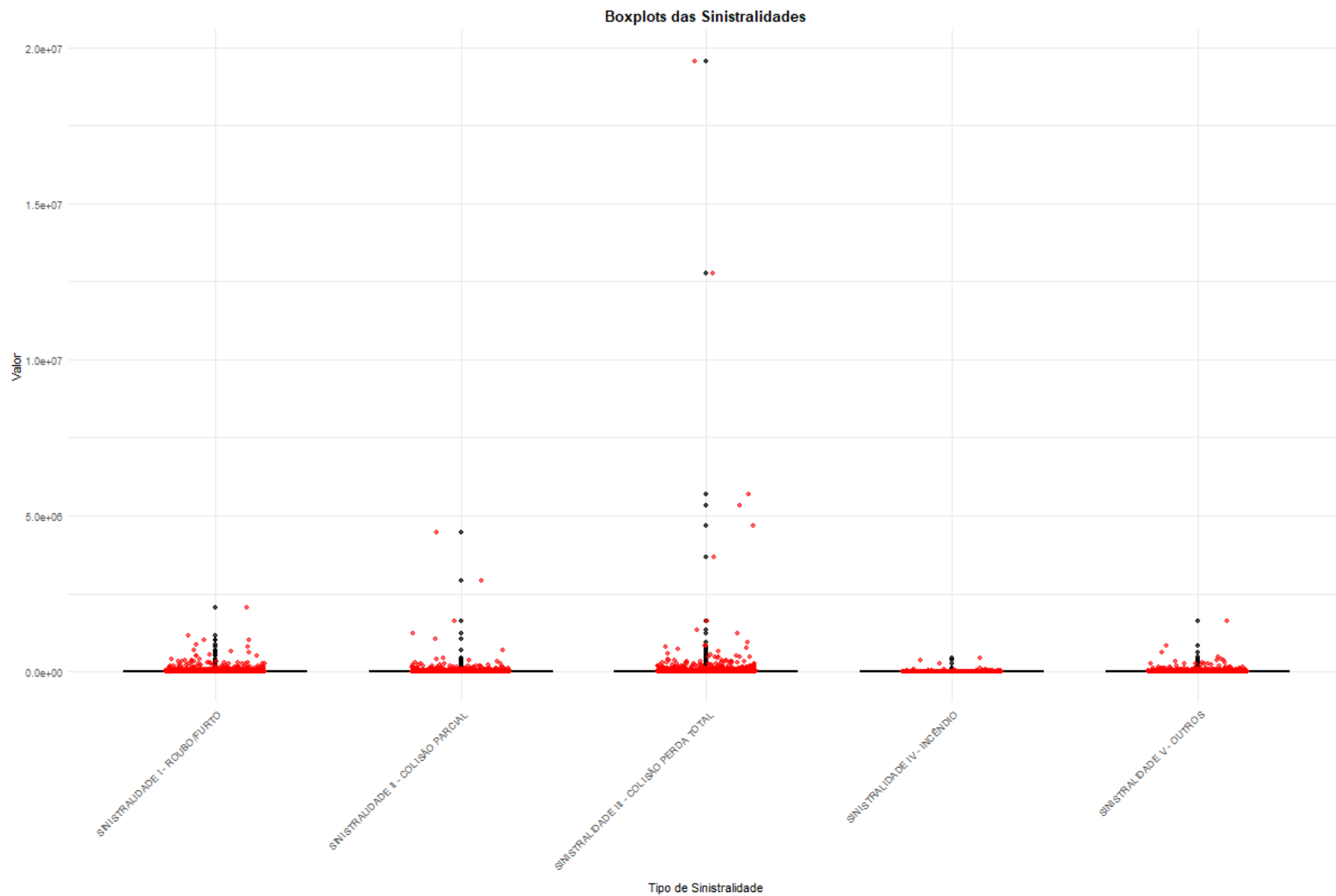
# Análise Exploratória – Sinistralidade

| VARIÁVEL           | MÍN | 1ºQUARTIL | MEDIAN | 3ºQUARTIL | MÁX         | MEAN |
|--------------------|-----|-----------|--------|-----------|-------------|------|
| SINISTRALIDADE I   | 0   | 0         | 0      | 0         | 2.048.202,2 | 30,6 |
| SINISTRALIDADE II  | 0   | 0         | 0      | 0         | 4.453.056   | 41   |
| SINISTRALIDADE III | 0   | 0         | 0      | 0         | 19.571.250  | 71   |
| SINISTRALIDADE IV  | 0   | 0         | 0      | 0         | 439.712,5   | 1,7  |
| SINISTRALIDADE V   | 0   | 0         | 0      | 0         | 1.615.379,8 | 20,4 |

| VARIÁVEL           | RANGE        | SD        | SKEW   | KURTOSIS  | SE    |
|--------------------|--------------|-----------|--------|-----------|-------|
| SINISTRALIDADE I   | 2.048.202,2  | 2.581,71  | 398,05 | 232.156,0 | 1,71  |
| SINISTRALIDADE II  | 4.453.055,6  | 4.005,25  | 832,38 | 817.349,4 | 2,66  |
| SINISTRALIDADE III | 19.571.250,0 | 17.058,98 | 895,37 | 915.551,8 | 11,33 |
| SINISTRALIDADE IV  | 439.712,5    | 445,18    | 754,36 | 655.977,4 | 0,30  |
| SINISTRALIDADE V   | 1.615.379,8  | 1.591,84  | 604,05 | 520.595,5 | 1,06  |

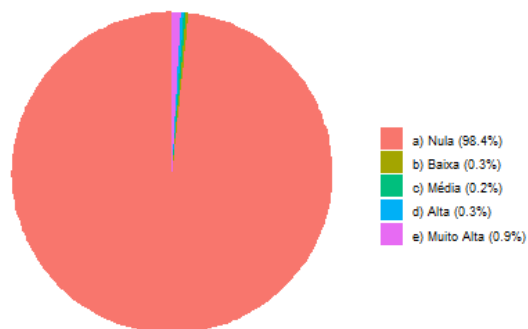




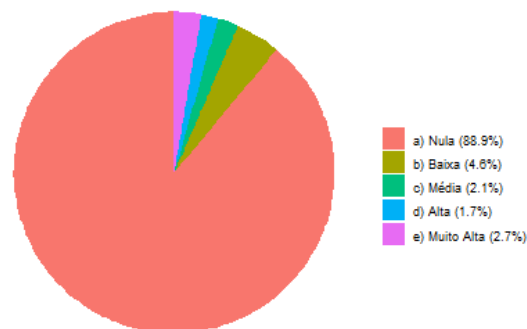


# Análise Exploratória – SAC

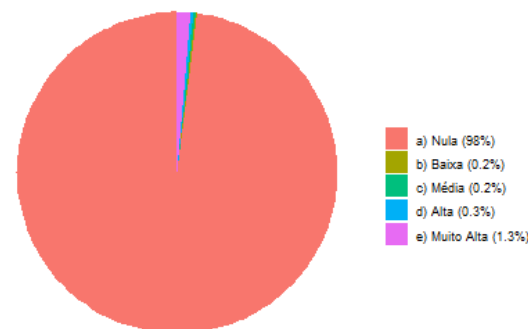
SAC I - ROUBO/FURTO



SAC II - COLISÃO PARCIAL



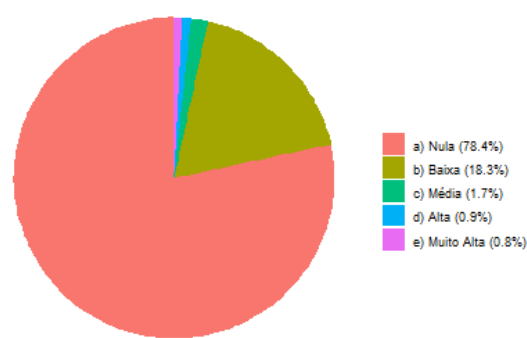
SAC III - COLISÃO PERDA TOTAL



SAC IV - INCÊNDIO

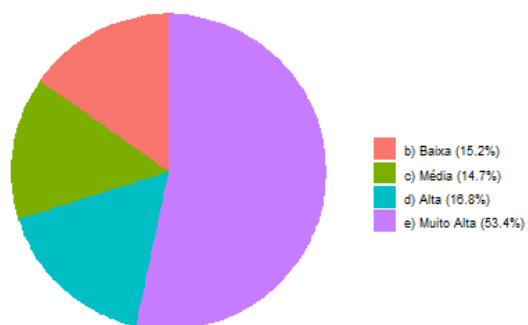


SAC V - OUTROS

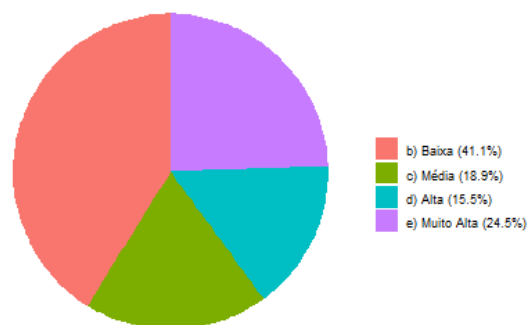


# Análise Exploratória – SAC sem classe “NULA”

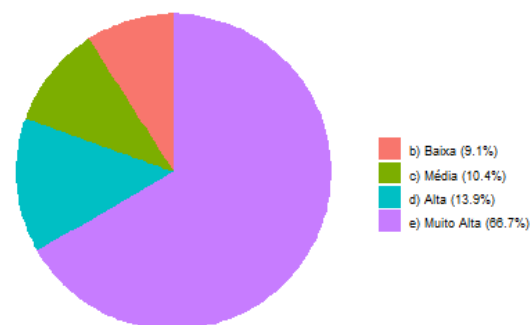
SAC I - ROUBO/FURTO



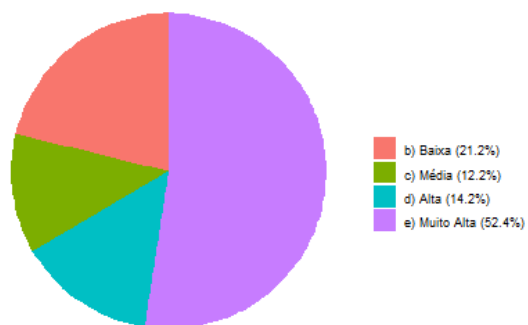
SAC II - COLISÃO PARCIAL



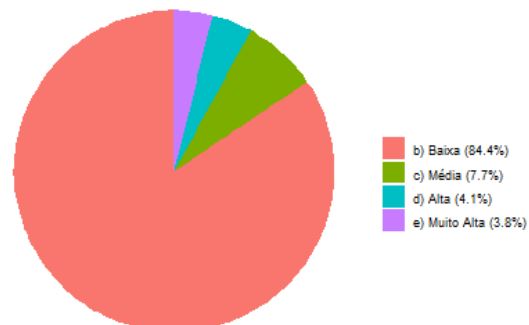
SAC III - COLISÃO PERDA TOTAL



SAC IV - INCÊNDIO



SAC V - OUTROS



# Regressão Logística Multinomial – AIC x BIC

- |   |   |
|---|---|
| • <b>AIC:</b> Prioriza a qualidade de ajuste do modelo  | • <b>BIC:</b> Prioriza a simplicidade do modelo         |
| • - severidade na penalização da complexidade do modelo | • + severidade na penalização da complexidade do modelo |
| • + chances de overfitting em grandes amostras          | • - chances de overfitting em grandes amostras          |
- 

**DECISÃO:** O BIC é eficaz em contextos em que a simplicidade do modelo é desejável, sem comprometer a qualidade do ajuste. O uso do BIC permite identificar um modelo que não apenas se ajuste bem aos dados, mas que também seja mais robusto e generalizável.

# Regressão Logística Multinomial – Modelagem






| VARIÁVEL DEPENDENTE           | MODELO  | BIC          | nº | VARIÁVEIS INDEPENDENTES  |
|-------------------------------|---------|--------------|----|--|
| SAC I – ROUBO/FURTO           | 359_I   | 36.051,44    | 5  | Ano Modelo   Exposição   Importância Segurada   Quantidade Sinistros I   Indenização I                 |
| SAC II – COLISÃO PARCIAL      | 121_II  | 1.075.324,38 | 3  | Exposição   Prêmio   Quantidade Sinistros II   |
| SAC III – COLISÃO PERDA TOTAL | 464_III | 33.132,43    | 6  | Sexo   Exposição   Prêmio   Importância Segurada   Quantidade Sinistros III   Indenização III          |
| SAC IV – INCÊNDIO             | 254_IV  | 1.498,78     | 4  | Exposição   Importância Segurada   Quantidade Sinistros IV   Indenização IV                            |
| SAC V – OUTROS                | 499_V   | 2.786.038,40 | 7  | Ano Modelo   Sexo   Exposição   Prêmio   Importância Segurada   Quantidade Sinistros V   Indenização V |



# SAC I – ROUBO/FURTO

| VARIÁVEL         | NÍVEL 1 - BAIXA | NÍVEL 2 - MÉDIA | NÍVEL 3 - ALTA | NÍVEL 4 - MUITO ALTA |
|------------------|-----------------|-----------------|----------------|----------------------|
| INTERCEPTO       | +               | -               | -              | -                    |
| ANO MODELO       | -3,7162%        | 4,1165%         | 0,5319%        | 0,7639%              |
| EXPOSIÇÃO        | 49,4214%        | 52,9590%        | 43,4620%       | 30,2910%             |
| IMP. SEGURADA    | 0,0036%         | 0,0039%         | 0,0035%        | 0,0023%              |
| QNTD SINISTROS I | -99,9999%       | -69,6690%       | -50,8159%      | -36,5552%            |
| INDENIZAÇÃO I    | -0,0151%        | -0,0197%        | -0,0113%       | -0,0060%             |

## NÍVEL DE SIGNIFICÂNCIA

-  0,1%
-  1%
-  5%
-  10%
-  Insignificante

# SAC II – COLISÃO PARCIAL

| VARIÁVEL          | NÍVEL 1 - BAIXA | NÍVEL 2 - MÉDIA | NÍVEL 3 - ALTA | NÍVEL 4 - MUITO ALTA |
|-------------------|-----------------|-----------------|----------------|----------------------|
| INTERCEPTO        | +               | +               | -              | -                    |
| EXPOSIÇÃO         | 11,8289%        | 10,6387%        | 6,0553%        | 3,3819%              |
| PRÊMIO            | 0,0027%         | 0,0028          | 0,0014         | 0,0010               |
| QNTD SINISTROS II | -96,8286%       | -48,7216%       | -26,5158%      | -23,1873%            |

## NÍVEL DE SIGNIFICÂNCIA

- 0,1%
- 1%
- 5%
- 10%
- Insignificante

# SAC III – COLISÃO PERDA TOTAL

| VARIÁVEL           | NÍVEL 1 - BAIXA | NÍVEL 2 - MÉDIA | NÍVEL 3 - ALTA | NÍVEL 4 - MUITO ALTA |
|--------------------|-----------------|-----------------|----------------|----------------------|
| INTERCEPTO         | +               | -               | -              | -                    |
| SEXO (M→F)         | -8,3030%        | -14,2042%       | -9,8504%       | -15,2530%            |
| EXPOSIÇÃO          | 11,5497%        | 13,3035%        | 11,8737%       | 9,5411%              |
| PRÊMIO             | 0,0161%         | 0,0165%         | 0,0133%        | 0,0089%              |
| IMP. SEGURADA      | 0,0004%         | 0,0006%         | 0,0005%        | 0,0006%              |
| QNTD SINISTROS III | -99,9998%       | -81,4370%       | -79,2578%      | -67,6320%            |
| INDENIZAÇÃO III    | -0,0144%        | -0,0190%        | -0,0104%       | -0,0052%             |

## NÍVEL DE SIGNIFICÂNCIA

- 0,1%
- 1%
- 5%
- 10%
- Insignificante

# SAC IV – INCÊNDIO

| VARIÁVEL          | NÍVEL 1 - BAIXA | NÍVEL 2 - MÉDIA | NÍVEL 3 - ALTA | NÍVEL 4 - MUITO ALTA |
|-------------------|-----------------|-----------------|----------------|----------------------|
| INTERCEPTO        | +               | -               | -              | -                    |
| EXPOSIÇÃO         | 122,6431%       | 125,0607%       | 92,7647%       | 53,3266%             |
| IMP. SEGURADA     | 0,0067%         | 0,0069%         | 0,0048%        | 0,0029%              |
| QNTD SINISTROS IV | -100,0000%      | -15,2615%       | -70,2696%      | -44,4118%            |
| INDENIZAÇÃO IV    | -0,0542%        | -0,0593%        | -0,0304%       | -0,0103%             |






## NÍVEL DE SIGNIFICÂNCIA

- 0,1%
- 1%
- 5%
- 10%
- Insignificante

# SAC V – OUTROS

| VARIÁVEL         | NÍVEL 1 - BAIXA | NÍVEL 2 - MÉDIA | NÍVEL 3 - ALTA | NÍVEL 4 - MUITO ALTA |
|------------------|-----------------|-----------------|----------------|----------------------|
| INTERCEPTO       | +               | -               | +              | +                    |
| ANO MODELO       | 0,01731%        | 0,7505%         | -0,0859%       | -0,0207%             |
| SEXO (M→F)       | -6,1864%        | 5,7111%         | 0,0704%        | 0,0230%              |
| EXPOSIÇÃO        | 0,0839%         | 1,9610%         | -0,3522%       | -0,0544%             |
| PRÊMIO           | -0,0006%        | 0,0021%         | 0,0001%        | 0,0002%              |
| IMP. SEGURADA    | 0,0003%         | -0,0002%        | 0,0000%        | 0,0000%              |
| QNTD SINISTROS V | -13,9722%       | 12,1649%        | 2,7481%        | 1,2589%              |
| INDENIZAÇÃO V    | -0,0122%        | -0,0106%        | -0,0047%       | -0,0046%             |

## NÍVEL DE SIGNIFICÂNCIA

-  0,1%
-  1%
-  5%
-  10%
-  Insignificante

INTRODUÇÃO



```
graph TD; A[INTRODUÇÃO] --> B[MATERIAL E MÉTODOS]; B --> C[RESULTADOS E DISCUSSÃO]; C --> D[CONSIDERAÇÕES FINAIS];
```

MATERIAL E MÉTODOS

RESULTADOS E DISCUSSÃO

CONSIDERAÇÕES FINAIS



# Sensibilidade das Variáveis

- **Categoria Tarifária:** Não sensível
- **Faixa Etária:** Não sensível
- **Ano Modelo:** Roubo/Furto e Outros
- **Sexo:** Colisão Perda Total e Outros
- **Prêmio:** Colisão Perda Total, Colisão Parcial e Outros
- **Importância Segurada:** Todas menos Colisão Parcial
- **Indenização:** Todas menos Colisão Parcial
- **Exposição:** Todas
- **Quantidade de Sinistros:** Todas

# Eficácia, Restrições e Expansão

- **Atendimento ao Objetivo:** Dados Concretos ; Identificadas as Variáveis que influenciam a Sinistralidade; Utilizável para tomadas de decisão por parte de seguradoras
- **Limitações do Estudo:** Período da base; Tempo de execução
- **Perspectivas Futuras:** Exploração de outras variáveis; Complementar o modelo com Matriz de Confusão, Curva ROC e/ou Algoritmos; Fácil reprodução; Aplicação em regiões específicas

# OBRIGADO!!

Tarciso Cavalcanti de Mello Filho  
Walter Mesquita Filho

