

# Relação Entre Velocidade e Distância de Frenagem para Carros de Passeio

Aluno Consultor 1<sup>a,b</sup>, Aluno Consultor 2<sup>a,b</sup>, Consulente<sup>c,d</sup> e Marcus A. Nunes<sup>a,e</sup>

<sup>a</sup>Departamento de Estatística - UFRN; <sup>b</sup>Consultor; <sup>c</sup>Outro Departamento - UFRN; <sup>d</sup>Consulente; <sup>e</sup>Orientação

13 de Março de 2018

**Este trabalho estuda a relação entre a velocidade de carros (mph) e a distância (pés) que eles levaram para parar completamente. Utilizamos regressão linear simples para determinar se há relação entre estas duas variáveis.**

regressão linear | automobilismo | segurança | trânsito

## 1. Objetivos

Diversos autores já publicaram estudos referentes à segurança no trânsito. McKenna *et al.* (1991), por exemplo, estuda a relação entre as habilidades dos motoristas e a percepção que eles possuem sobre estas habilidades. Além desta característica, existem diversas outras que, se estudadas, podem aumentar a segurança no trânsito. Uma destas características é a distância mínima necessária para que um carro pare completamente após seus freios serem acionados.

Neste trabalho estamos interessados em verificar qual é a relação que existe entre a velocidade de um carro (em milhas por hora) e a distância que ele levou para parar completamente (em pés). Este conjunto de dados foi fornecido pelo programa R: A Language and Environment for Statistical Computing (R Core Team (2017)). A hipótese com a qual trabalhamos é a de que existe uma relação positiva entre estas variáveis. Isto é, quanto mais rápido um carro estiver trafegando, maior vai ser a distância necessária para que este carro pare completamente.

Além de verificar se há correlação entre estas variáveis, desejamos obter uma relação capaz de prever o quanto uma variável varia em relação a outra. Ou seja, gostaríamos de poder estimar a distância necessária para um carro parar completamente se soubermos qual a sua velocidade de tráfego no momento em que os freios foram acionados.

## 2. Metodologia

Os dados aqui analisados foram obtidos a partir de uma amostra de 50 carros. As medições foram realizadas na década de 1920 e disponibilizadas originalmente por Ezekiel (1930). Não há informações a respeito dos modelos dos carros utilizados neste experimento.

Utilizaremos um método estatístico chamado regressão linear a fim de verificar se há relação entre a distância necessária para um carro parar completamente e sua velocidade. Este é um método bastante popular, capaz de descrever com bastante precisão a relação entre as variáveis que nos interessam.

Sejam  $x_1, x_2, \dots, x_n$  as observações referentes à velocidade dos carros em questão. Considere  $y_1, y_2, \dots, y_n$  as observações referentes à distância necessária para os carros pararem. De acordo com Kutner *et al.* (2004), podemos expressar a dependência entre  $y$  e  $x$  através da equação

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

onde  $\beta_0$  e  $\beta_1$  são coeficientes estimados pelas equações

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (2.1)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (2.2)$$

As quantidades  $\bar{x}$  e  $\bar{y}$  são, respectivamente, as médias amostrais de  $x_1, x_2, \dots, x_n$  e  $y_1, y_2, \dots, y_n$ . Estas médias amostrais são dadas por

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

Determinamos se o coeficiente  $\beta_1$  é estatisticamente significativo através de um teste  $t$ . Sob a hipótese nula, assumimos que o estimador possui distribuição  $t$  com  $n - 1$  graus de liberdade.

### 3. Resultados

A fim de verificar visualmente se há algum tipo de relação entre as variáveis consideradas neste estudo, exibimos o gráfico de dispersão dos dados na Figura 3.1. Note que é possível perceber uma forte tendência linear positiva na relação entre estas variáveis. Quanto maior o valor da velocidade, maior a distância necessária para o carro parar completamente.

```
ggplot(cars, aes(x=speed, y=dist)) +  
  geom_point() +  
  labs(x="Velocidade (mph)", y="Distância (pés)")
```

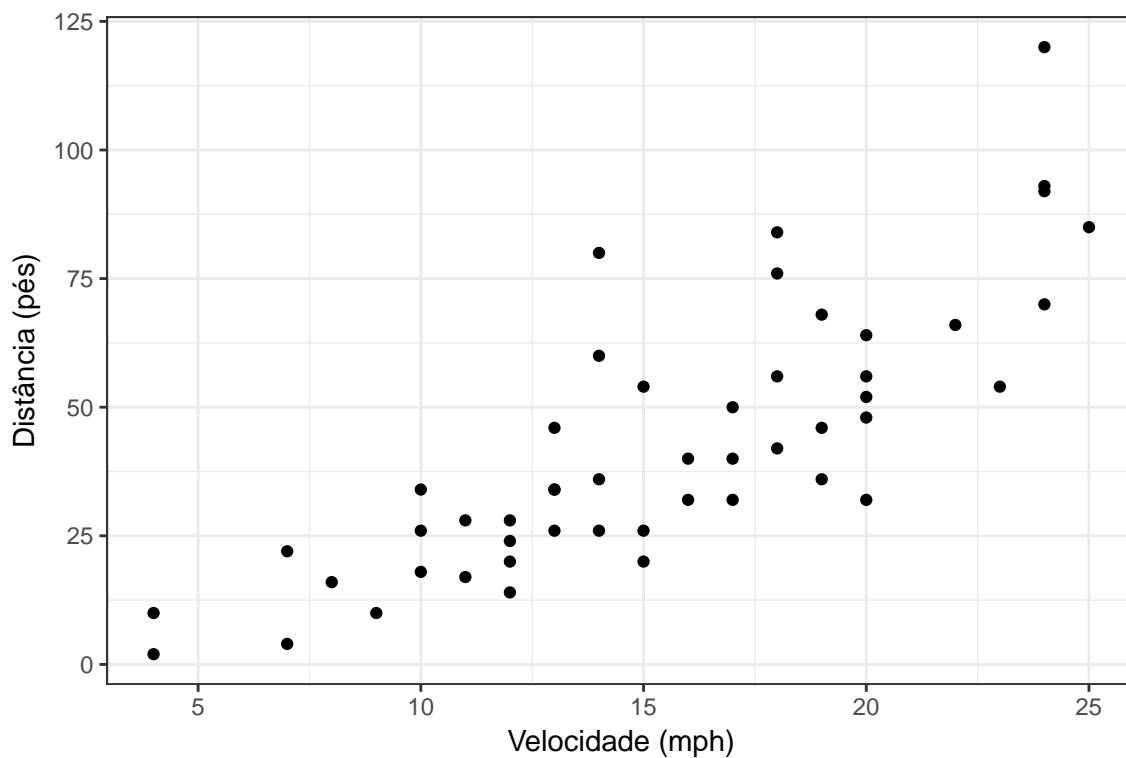
```
ggplot(cars, aes(x=speed, y=dist)) +  
  geom_point() +  
  labs(x="Velocidade (mph)", y="Distância (pés)") +  
  geom_smooth(method="lm", se=FALSE)
```

Além disso, adicionamos ao gráfico exibido na Figura 3.2 a reta que melhor descreve a relação entre estas variáveis. Esta reta foi obtida através do método descrito na seção anterior, fazendo uso das fórmulas (2.1) e (2.2). Explicitamente, a equação representada na Figura 3.2 é dada por

```
ajuste <- lm(dist ~ speed, data=cars)
```

$$\hat{y}_i = -17,5791 + 3,9324x_i. \quad (3.1)$$

Entretanto, precisamos testar se os coeficientes estimados e apresentados na relação (3.1) são, de fato, estatisticamente significantes. Para isto, testaremos as hipóteses



**Fig. 3.1.** Gráfico de dispersão da distância de parada completa (pés) versus velocidade (mph) dos carros.

$$H_0 : \beta_0 = 0$$

$$H_1 : \beta_0 \neq 0$$

e

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

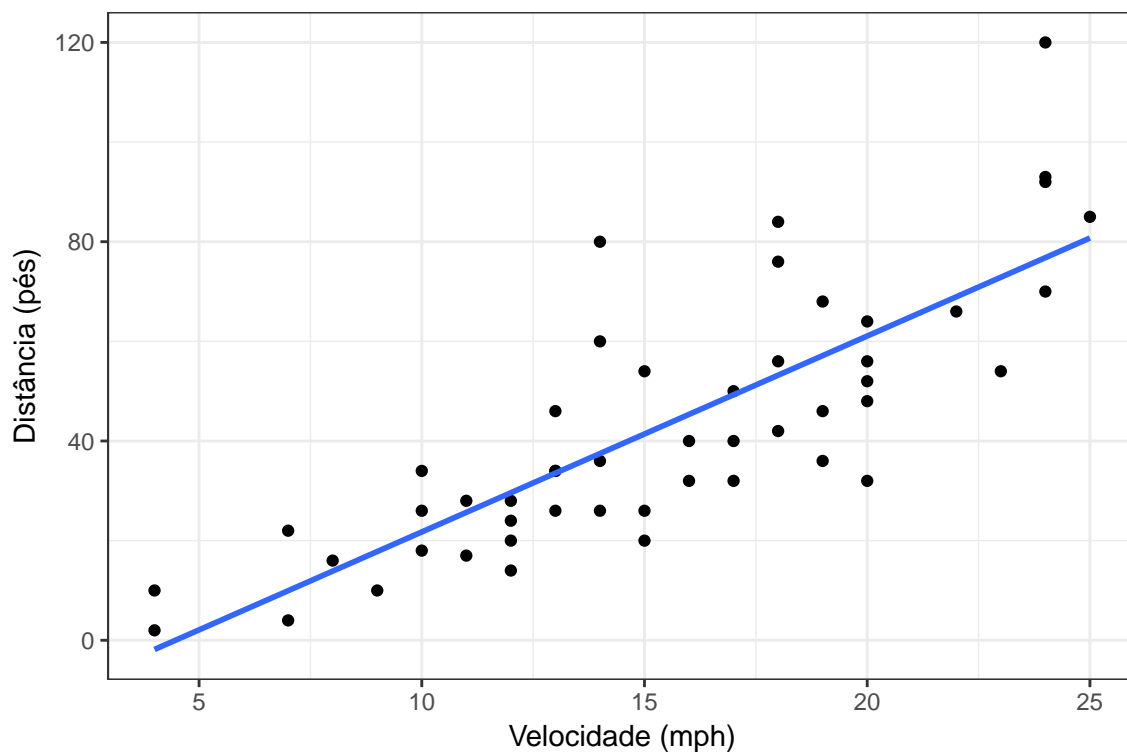
Os resultados destes testes estão apresentados na Tabela 3.1.s

**Tabela 3.1.** Resultados dos testes de hipóteses realizados para a análise de regressão.

Coeficiente	Estimativa	Erro Padrão	t	p-valor
$\beta_0$	-17,5791	6,7584	-2,601	0,0123
$\beta_1$	3,9324	0,4155	9,464	<0,0001

Note que, em ambos os casos, o p-valor encontrado é inferior a  $\alpha = 0,05$ . Portanto, podemos rejeitar ambas as hipóteses nulas e  $\beta_0$  e  $\beta_1$  são estatisticamente diferentes de zero.

Para finalizar a análise, devemos verificar se o modelo ajustado não viola as hipóteses do modelo de regressão linear. Para verificar isto, exibimos a análise de resíduos na Figura 3.3.



**Fig. 3.2.** Gráfico de dispersão da distância de parada completa (pés) versus velocidade (mph) dos carros com a reta que melhor se ajusta a estes dados.

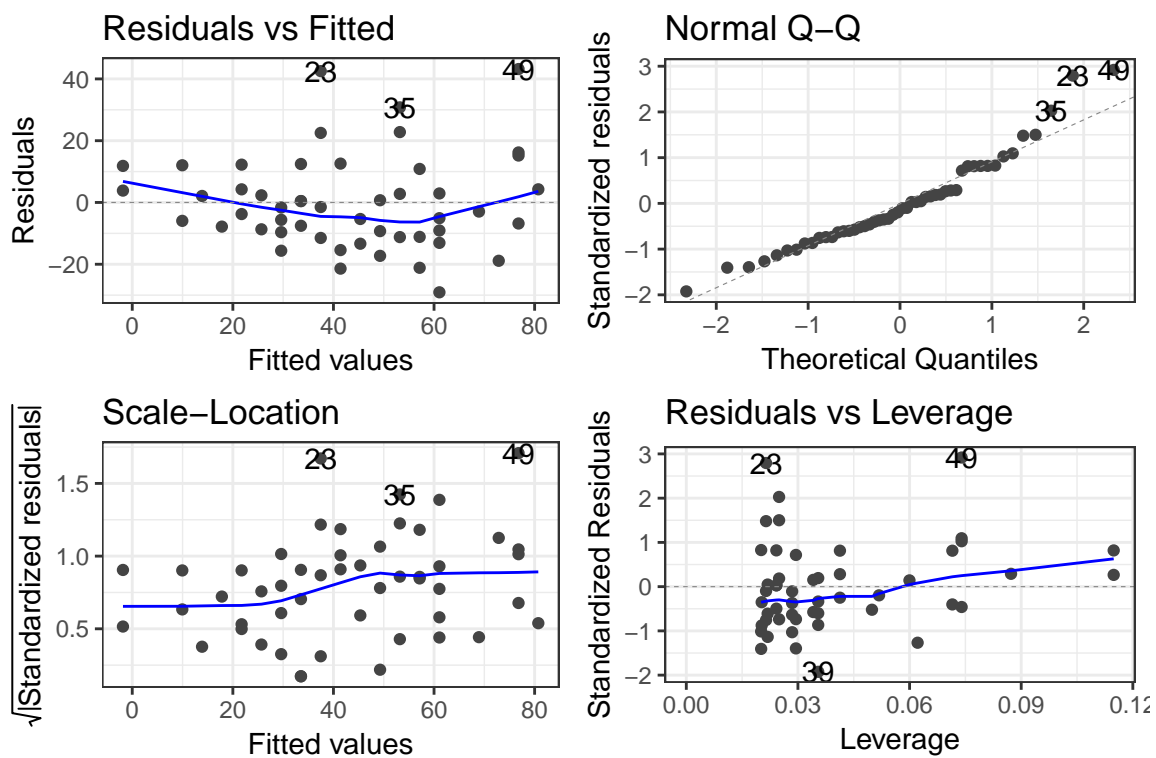
```
autoplot(ajuste)
```

Note que na parte superior esquerda da imagem, embora o gráfico dos resíduos versus valores ajustados não apresente tendência, a variância não é constante. Note que os pontos próximos de zero estão mais próximos entre si do que os pontos mais à direita no gráfico. Portanto, há uma violação das hipóteses da regressão linear neste caso.

Assim, podemos sugerir uma transformação nestes dados ou a utilização de outro método de análise, como um modelo linear generalizado.

## Referências

- Ezekiel M (1930). *Methods of Correlation Analysis*. Wiley, New York.
- Kutner M, Nachtsheim C, Neter J, Li W (2004). *Applied Linear Statistical Models - Fifth Edition*. McGraw-Hill/Irwin, New York.
- McKenna FP, Stanier RA, Lewis C (1991). "Factors underlying illusory self- assessment of driving skill in males and females." *Accident Analysis and Prevention*, **23**(1), 45–52.
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.



**Fig. 3.3.** Análise de resíduos do modelo de regressão linear ajustado.