

Relação Entre Velocidade e Distância de Frenagem para Carros de Passeio

Aluno Consultor 1^{a,b}, Aluno Consultor 2^{a,b}, Consulente^{c,d}, e Marcus A. Nunes^{a,e}

^aDepartamento de Estatística - UFRN; ^bConsultor; ^cOutro Departamento - UFRN; ^dConsulente; ^eOrientação

Data: 3 de Fevereiro 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

Neste trabalho estamos interessados em verificar que existe alguma relação entre a velocidade de um carro (em milhas por hora) e a distância que o carro levou para parar (em pés). 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 viajando, maior vai ser a distância necessária para que este carro pare completamente.

Além de verificar se há uma relação entre estas variáveis, desejamos uma relação capaz de prever o quanto uma variável varia em relação a outra. Assim, gostaríamos de poder estimar a distância necessária para um carro parar se soubermos qual a sua velocidade.

2. Metodologia

Os dados foram obtidos a partir de 50 carros. As medições foram realizadas na década de 1920 e disponibilizadas no pacote estatístico R. 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. Podemos expressar a dependência entre y e x através da equação

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

onde β_0 e β_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 β_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)")
```

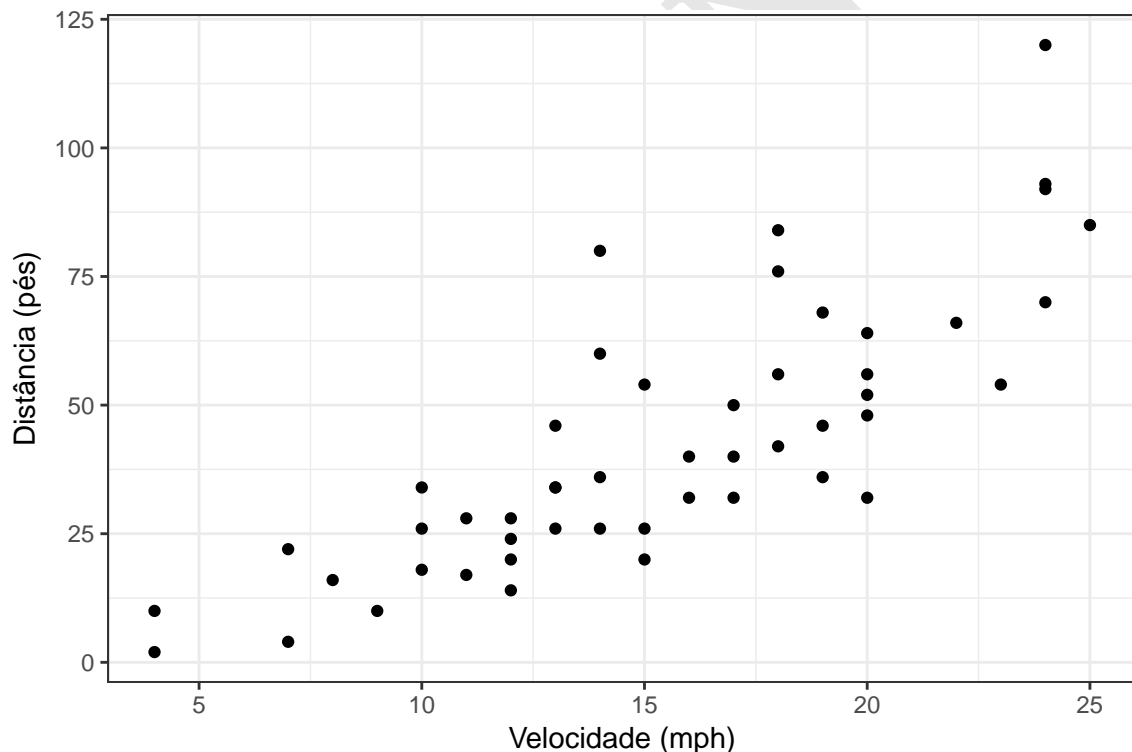


Fig. 3.1. Gráfico de dispersão da distância de parada completa (pés) versus velocidade (mph) dos carros.

Além disso, adicionamos ao gráfico exibido na Figura 3.1 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 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

$$H_0 : \beta_0 = 0$$

$$H_1 : \beta_0 \neq 0$$

e

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

Os resultados destes testes estão apresentados na Tabela 3.1.

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

Coefficiente	Estimativa	Erro Padrão	t	p-valor
β_0	-17,5791	6,7584	-2,601	0,0123
β_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 β_0 e β_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 tal, exibimos a análise de resíduos na Figura 3.2.

```
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

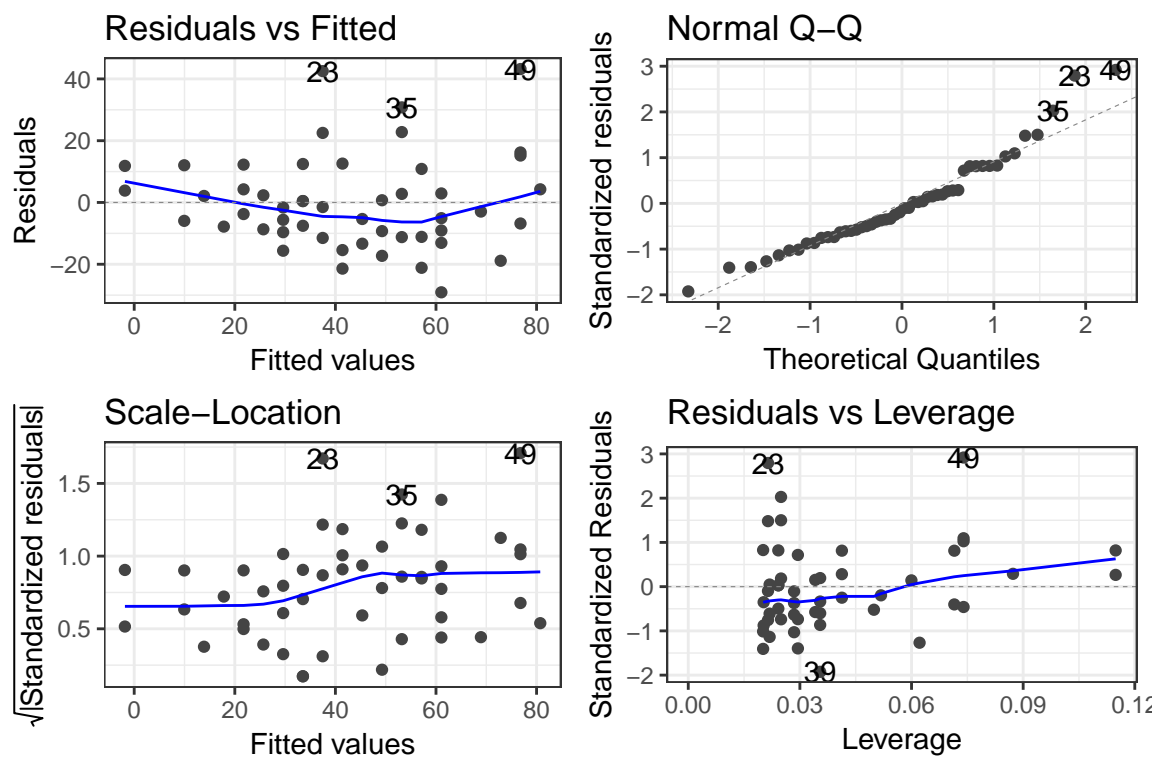


Fig. 3.2. Análise de resíduos da regressão linear realizada.