

Relacao Entre Velocidade e Distancia de Frenagem para Carros de Passeio

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

^a Departamento de Estatistica - UFRN; ^b Consultor; ^c Outro Departamento - UFRN; ^d Consulfente; ^e Orientacao

2 de Agosto de 2019

Este trabalho estuda a relacao entre a velocidade de carros (mph) e a distancia (pes) que eles levaram para parar completamente. Utilizamos o modelo de regressao linear simples para determinar se existe relacao entre estas duas variaveis.

regressao linear | automobilismo | seguranca | transito

1. Objetivos

Diversos autores ja publicaram estudos referentes a seguranca no transito. McKenna *et al.* (1991), por exemplo, estuda a relacao entre as habilidades dos motoristas e a percepcao que eles possuem sobre estas habilidades. Alem desta caracteristica, existem diversas outras que, se estudadas, podem aumentar a seguranca no transito. Uma destas caracteristicas e a distancia minima necessaria para que um carro pare completamente apos seus freios serem acionados.

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

Alem de verificar se ha correlacao entre estas variaveis, desejamos obter uma relacao capaz de prever o quanto uma variavel varia em relacao a outra. Ou seja, gostariamos de poder estimar a distancia necessaria para um carro parar completamente se soubermos qual a sua velocidade de trafego no momento em que os freios foram acionados.

2. Metodologia

Os dados analisados neste trabalho foram obtidos a partir de uma amostra de 50 carros. As medicoes foram realizadas na decada de 1920 e disponibilizadas originalmente por Ezekiel (1930). Nao ha informacoes a respeito dos modelos dos carros utilizados neste experimento.

Utilizaremos um metodo estatistico chamado regressao linear a fim de verificar se ha relacao entre a distancia necessaria para um carro parar completamente e sua velocidade. Este e um metodo bastante popular, capaz de descrever com bastante precisao a relacao entre as variaveis que nos interessam.

Sejam x_1, x_2, \dots, x_n as observacoes referentes a velocidade dos carros em questao. Considere y_1, y_2, \dots, y_n as observacoes referentes a distancia necessaria para os carros pararem. De acordo com Kutner *et al.* (2004), podemos expressar a dependencia entre y e x atraves da equacao

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

onde β_0 e β_1 sao coeficientes estimados pelas equacoes

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

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (2)$$

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 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="Distancia (pes)")
```

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

Além disso, adicionamos ao gráfico exibido na Figura 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 (1) e (2). Explicitamente, a equação representada na Figura 2 é dada por

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

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

Entretanto, precisamos testar se os coeficientes estimados e apresentados na relação (3) são, de fato, estatisticamente significativos. Para isto, testaremos as hipóteses

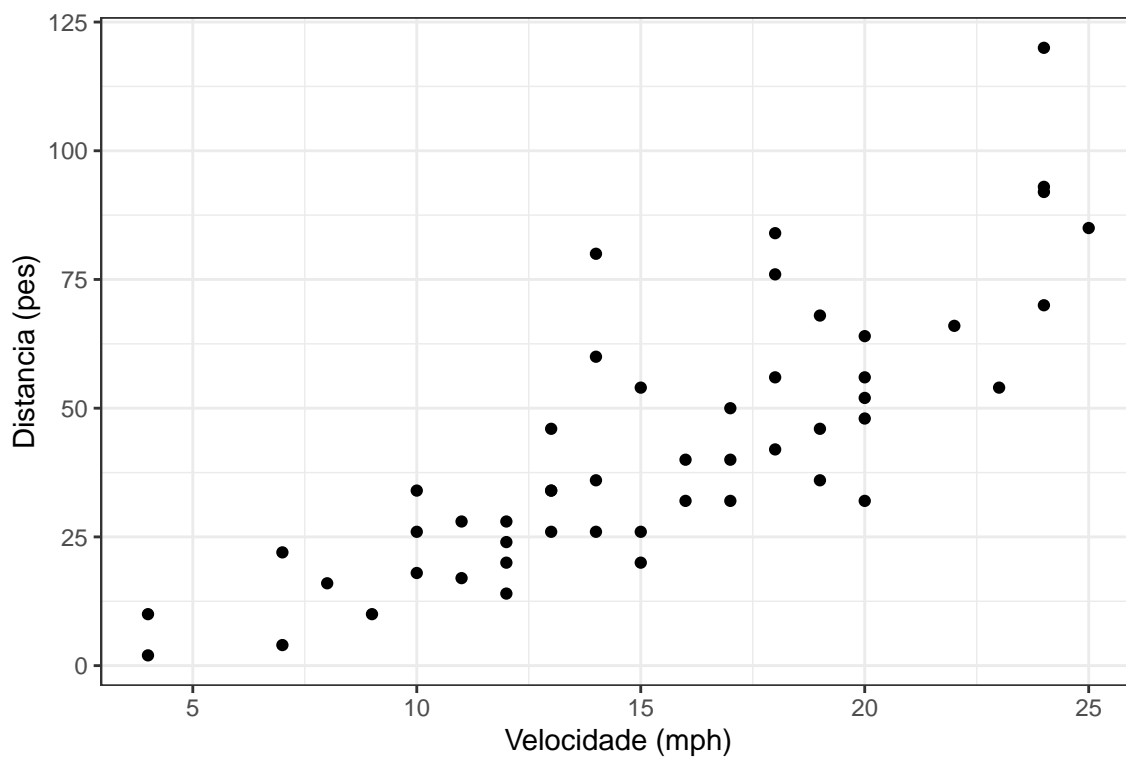


Fig. 1. Grafico de dispersao da distancia de parada completa (pes) versus velocidade (mph) dos carros.

$$H_0 : \beta_0 = 0$$

$$H_1 : \beta_0 \neq 0$$

e

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

Os resultados destes testes estao apresentados na Tabela 1.s

Tabela 1. Resultados dos testes de hipoteses realizados para a analise de regressao.

Coefficiente	Estimativa	Erro Padrao	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 e inferior a $\alpha = 0,05$. Portanto, podemos rejeitar ambas as hipoteses nulas e β_0 e β_1 sao estatisticamente diferentes de zero.

Para finalizar a analise, devemos verificar se o modelo ajustado nao viola as hipoteses do modelo de regressao linear. Para verificar isto, exibimos a analise de residuos na Figura 3.

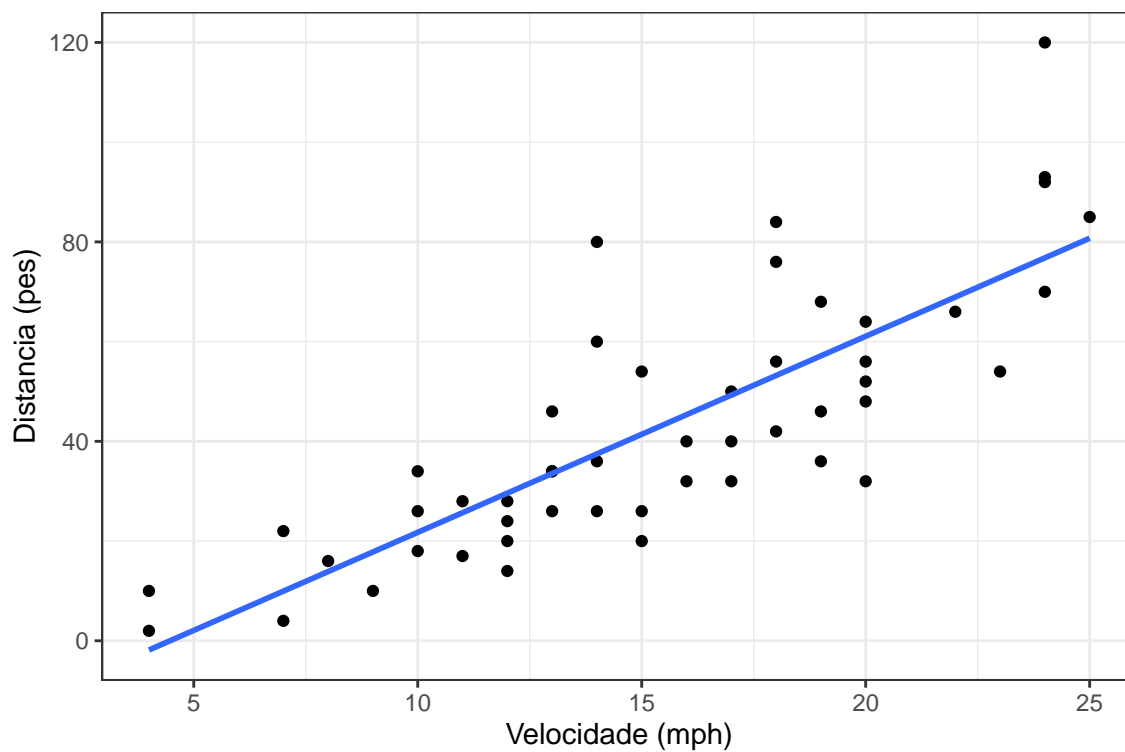


Fig. 2. Grafico de dispersao da distancia de parada completa (pes) versus velocidade (mph) dos carros com a reta que melhor se ajusta a estes dados.

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/>.

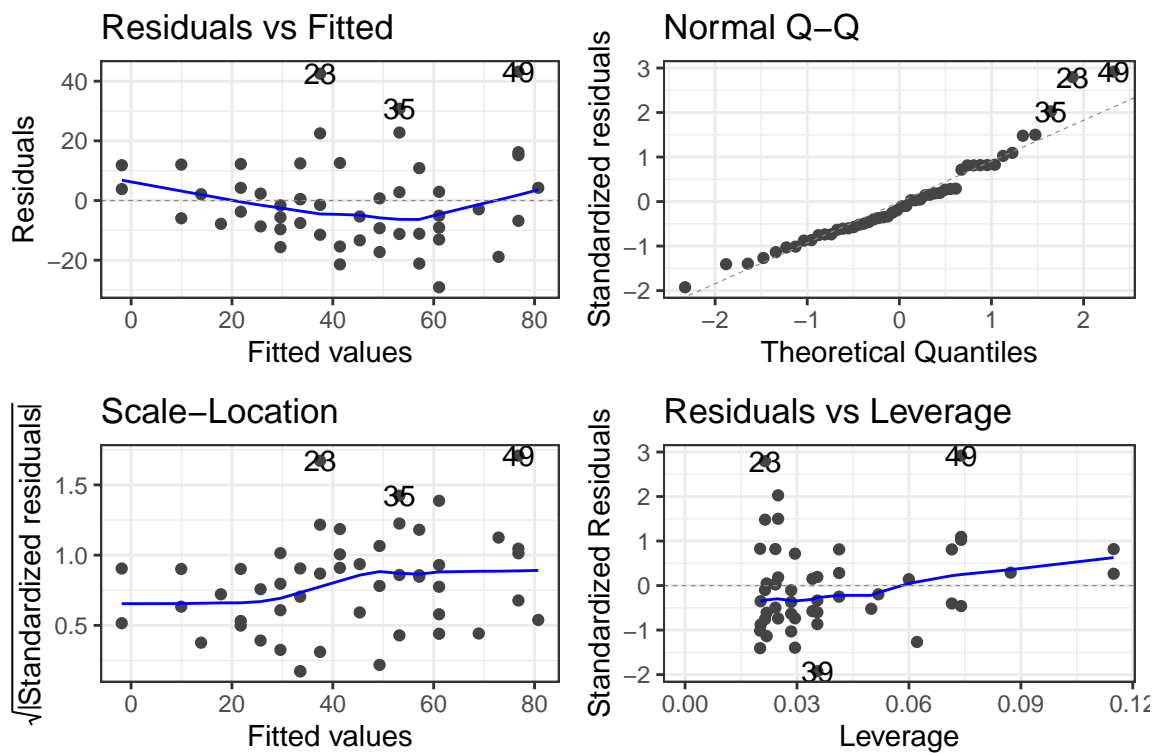


Fig. 3. Análise de resíduos do modelo de regressão linear ajustado.