CE310 - Modelos de Regressão Linear

Introdução

Cesar Augusto Taconeli

18 de março, 2025

Motivação e contexto histórico

Motivação e contexto histórico

Análise de regressão

Conjunto de técnicas aplicadas na análise e modelagem da relação estatística entre variáveis.

Motivação e contexto histórico

All models are wrong but some are useful

George Box

Motivação e contexto histórico

All models are wrong but some are useful

George Box

No matter how beautiful your theory, no matter how clever you are or what your name is, if it disagrees with experiment, it's wrong.

Richard Feynman

Motivação e contexto histórico

All models are wrong but some are useful

George Box

No matter how beautiful your theory, no matter how clever you are or what your name is, if it disagrees with experiment, it's wrong.

Richard Feynman

Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise.

John W. Tukey

Motivação e contexto histórico

- O termo regressão (*regression*) deve-se ao estatístico inglês Francis Galton (século XIX).

Motivação e contexto histórico

- O termo regressão (*regression*) deve-se ao estatístico inglês Francis Galton (século XIX).

- Em seus estudos, Galton levantou dados sobre alturas de casais e respectivos descendentes.

Motivação e contexto histórico

- O termo regressão (*regression*) deve-se ao estatístico inglês Francis Galton (século XIX).

- Em seus estudos, Galton levantou dados sobre alturas de casais e respectivos descendentes.

- Os dados revelaram relação crescente entre as alturas de pais e filhos.

Motivação e contexto histórico

- O termo regressão (*regression*) deve-se ao estatístico inglês Francis Galton (século XIX).
- Em seus estudos, Galton levantou dados sobre alturas de casais e respectivos descendentes.
- Os dados revelaram relação crescente entre as alturas de pais e filhos.
- No entanto, Galton observou que casais muito altos geravam filhos também altos, mas em geral de menor estatura (mais próximos a uma altura média). O mesmo ocorria para casais que se destacavam por uma baixa estatura.

Motivação e contexto histórico

- O termo regressão (*regression*) deve-se ao estatístico inglês Francis Galton (século XIX).
- Em seus estudos, Galton levantou dados sobre alturas de casais e respectivos descendentes.
- Os dados revelaram relação crescente entre as alturas de pais e filhos.
- No entanto, Galton observou que casais muito altos geravam filhos também altos, mas em geral de menor estatura (mais próximos a uma altura média). O mesmo ocorria para casais que se destacavam por uma baixa estatura.
- Galton denominou este fenômeno como **regressão à média**, ou simplesmente **regressão**.

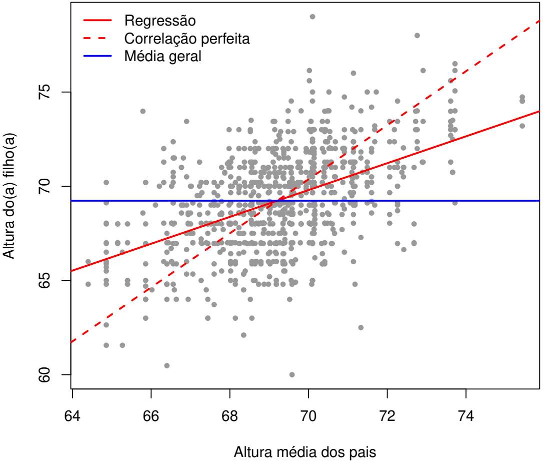

Figura 1: Ilustração - origem do termo regressão

Motivação e contexto histórico

- Século 19: Desenvolvimento do **método de mínimos quadrados**, que serve de base para o ajuste de modelos de regressão linear.

Motivação e contexto histórico

- Século 19: Desenvolvimento do **método de mínimos quadrados**, que serve de base para o ajuste de modelos de regressão linear.

- A teoria de mínimos quadrados teve origem na Física, motivada por problemas na área navegação (século 18).

Motivação e contexto histórico

- Século 19: Desenvolvimento do **método de mínimos quadrados**, que serve de base para o ajuste de modelos de regressão linear.

- A teoria de mínimos quadrados teve origem na Física, motivada por problemas na área navegação (século 18).

- Ao longo do século 19, regressão linear e o método de mínimos quadrados passaram a ser utilizados em outras ciências, baseados em modelos pré-estabelecidos, ou apenas em evidências empíricas.

Motivação e contexto histórico

- Modelos de regressão permitem descrever a relação (não determinística) entre uma variável de interesse (resposta) e uma ou mais covariáveis (preditoras).

Motivação e contexto histórico

- Modelos de regressão permitem descrever a relação (não determinística) entre uma variável de interesse (resposta) e uma ou mais covariáveis (preditoras).

- Uma análise de regressão pode ter diferentes objetivos, que em geral estão associados a duas finalidades principais:

Motivação e contexto histórico

- Modelos de regressão permitem descrever a relação (não determinística) entre uma variável de interesse (resposta) e uma ou mais covariáveis (preditoras).

- Uma análise de regressão pode ter diferentes objetivos, que em geral estão associados a duas finalidades principais:

    1. Modelos exploratórios: identificar e quantificar as relações entre a resposta e as covariáveis;

Motivação e contexto histórico

- Modelos de regressão permitem descrever a relação (não determinística) entre uma variável de interesse (resposta) e uma ou mais covariáveis (preditoras).

- Uma análise de regressão pode ter diferentes objetivos, que em geral estão associados a duas finalidades principais:

  1. Modelos exploratórios: identificar e quantificar as relações entre a resposta e as covariáveis;

  2. Modelos preditivos: utilizar valores observados das covariáveis para predizer resultados não observados da resposta.

Motivação e contexto histórico

- Modelos de regressão permitem descrever a relação (não determinística) entre uma variável de interesse (resposta) e uma ou mais covariáveis (preditoras).

- Uma análise de regressão pode ter diferentes objetivos, que em geral estão associados a duas finalidades principais:

  1. Modelos exploratórios: identificar e quantificar as relações entre a resposta e as covariáveis;

  2. Modelos preditivos: utilizar valores observados das covariáveis para predizer resultados não observados da resposta.

- Nos slides seguintes são ilustrados modelos de regressão linear e algumas generalizações.

Possíveis aplicações de regressão linear

Aplicação 1: Preço de Imóveis

- **Variável Resposta:** Preço dos imóveis de certa capital
- **Variáveis Explicativas:** Tamanho do imóvel (m^2), número de quartos, localização (bairro), idade do imóvel, presença de garagem.

Possíveis aplicações de regressão linear

Aplicação 1: Preço de Imóveis

- **Variável Resposta:** Preço dos imóveis de certa capital
- **Variáveis Explicativas:** Tamanho do imóvel (m^2), número de quartos, localização (bairro), idade do imóvel, presença de garagem.

Aplicação 2: Desempenho Escolar

- **Variável Resposta:** Nota de alunos num exame de conhecimentos gerais
- **Variáveis Explicativas:** Tipo de escola que o aluno frequenta (pública ou privada), renda familiar mensal, escolaridade dos pais (anos de ensino), localização da escola (rural ou urbana), se a família recebe algum auxílio financeiro do governo.

Possíveis aplicações de regressão linear

Aplicação 3: Vendas de produtos

- **Variável Resposta:** Vendas mensais de diferentes produtos
- **Variáveis Explicativas:** Preço do produto, gastos em marketing, número de vendedores, promoções realizadas, forma de comercialização.

Possíveis aplicações de regressão linear

Aplicação 3: Vendas de produtos

- **Variável Resposta:** Vendas mensais de diferentes produtos
- **Variáveis Explicativas:** Preço do produto, gastos em marketing, número de vendedores, promoções realizadas, forma de comercialização.

Aplicação 4: Consumo de Energia

- **Variável Resposta:** Consumo de energia elétrica por domicílios de certa região (kWh)
- **Variáveis Explicativas:** Tamanho da residência (m^2), número de moradores, uso de eletrodomésticos, temperatura média na região, tipo de aquecimento.

Aplicação 5: Rendimento de Investimentos

- **Variável Resposta:** Rendimento de diferentes carteiras de investimentos
- **Variáveis Explicativas:** Taxa de juros, valor inicial mínimo investido, taxa de administração, duração do investimento.

Modelo de regressão linear

- Vamos começar pelo modelo de regressão linear para uma variável explicativa (regressão linear simples).

Modelo de regressão linear

- Vamos começar pelo modelo de regressão linear para uma variável explicativa (regressão linear simples).

- Seja y a variável resposta e x a variável explicativa. O modelo de regressão linear simples fica definido por:

$$y = \beta_0 + \beta_1 x + \epsilon,$$

em que β_0 e β_1 são os parâmetros do modelo, definindo a reta de regressão (β_0 é o intercepto e β_1 é a inclinação da reta), e ϵ representa o erro aleatório (parte da variação de y que não é explicada pela reta de regressão).

Modelo de regressão linear

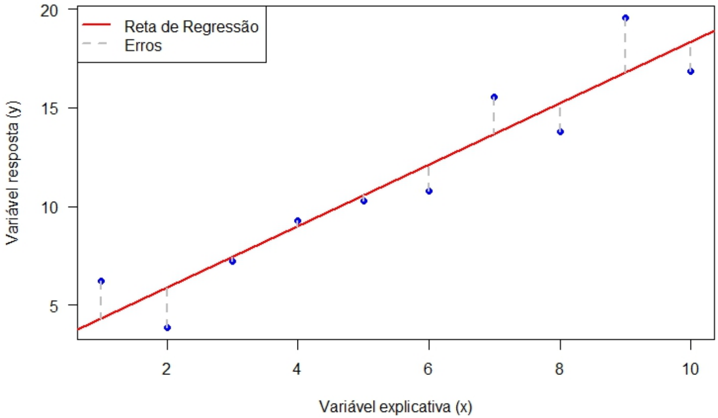

Figura 2: Regressão linear simples

Modelo de regressão linear

- No caso de múltiplas variáveis explicativas (regressão linear múltipla), o modelo inclui os termos lineares das demais variáveis.

Modelo de regressão linear

- No caso de múltiplas variáveis explicativas (regressão linear múltipla), o modelo inclui os termos lineares das demais variáveis.

- Seja y a variável resposta e $x_1, x_2, ..., x_k$ os k preditores. O modelo de regressão linear fica definido por:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \epsilon,$$

em que $\beta_0, \beta_1, ..., \beta_k$ são os parâmetros do modelo, definindo o modelo de regressão (β_0 é o intercepto e $\beta_1, ..., \beta_k$ os parâmetros de inclinação), e ϵ representa o erro aleatório.

Modelo de regressão linear

- No caso de múltiplas variáveis explicativas (regressão linear múltipla), o modelo inclui os termos lineares das demais variáveis.

- Seja y a variável resposta e $x_1, x_2, ..., x_k$ os k preditores. O modelo de regressão linear fica definido por:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \epsilon,$$

em que $\beta_0, \beta_1, ..., \beta_k$ são os parâmetros do modelo, definindo o modelo de regressão (β_0 é o intercepto e $\beta_1, ..., \beta_k$ os parâmetros de inclinação), e ϵ representa o erro aleatório.

- Numa análise de regressão, as variáveis explicativas são consideradas fixas (não aleatórias). A variável resposta é a variável aleatória.

Modelo de regressão linear

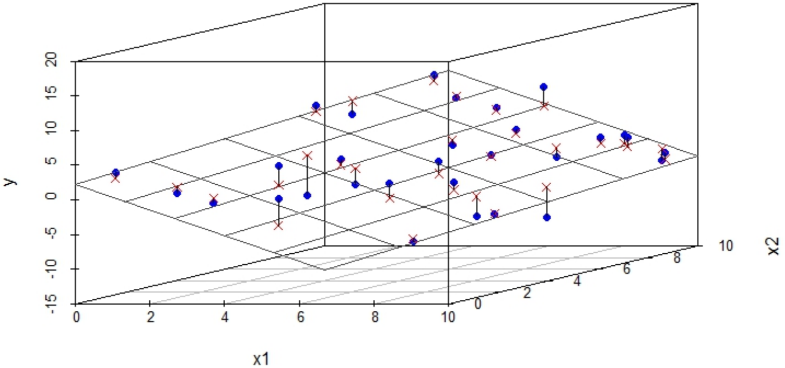

Figura 3: Regressão linear múltipla para duas variáveis explicativas

Modelo de regressão linear

- Embora sejam não observáveis e não explicados pelo modelo, na regressão linear assumimos as seguintes propriedades para os erros:

Modelo de regressão linear

- Embora sejam não observáveis e não explicados pelo modelo, na regressão linear assumimos as seguintes propriedades para os erros:
  - Os erros têm média zero;

Modelo de regressão linear

- Embora sejam não observáveis e não explicados pelo modelo, na regressão linear assumimos as seguintes propriedades para os erros:
  - Os erros têm média zero;
  - Os erros têm variância constante;

Modelo de regressão linear

- Embora sejam não observáveis e não explicados pelo modelo, na regressão linear assumimos as seguintes propriedades para os erros:
  - Os erros têm média zero;
  - Os erros têm variância constante;
  - Os erros para dois indivíduos quaisquer são não correlacionados;

Modelo de regressão linear

- Embora sejam não observáveis e não explicados pelo modelo, na regressão linear assumimos as seguintes propriedades para os erros:
  - Os erros têm média zero;
  - Os erros têm variância constante;
  - Os erros para dois indivíduos quaisquer são não correlacionados;
  - Os erros têm distribuição normal.

Modelo de regressão linear

- Embora sejam não observáveis e não explicados pelo modelo, na regressão linear assumimos as seguintes propriedades para os erros:
  - Os erros têm média zero;
  - Os erros têm variância constante;
  - Os erros para dois indivíduos quaisquer são não correlacionados;
  - Os erros têm distribuição normal.

- Este conjunto de suposições é usualmente denotado por $\epsilon \sim \mathrm{N}(0, \sigma^2)$.

Modelo de regressão linear

- Voltando ao modelo de regressão linear simples, para fins de ilustração:

$$y = \beta_0 + \beta_1 x + \epsilon, \quad \epsilon \sim \mathrm{N}(0, \sigma^2)$$

Modelo de regressão linear

- Voltando ao modelo de regressão linear simples, para fins de ilustração:

$$y = \beta_0 + \beta_1 x + \epsilon, \quad \epsilon \sim \mathrm{N}(0, \sigma^2)$$

- Como consequência das suposições assumidas para os erros, temos que, dados x, segue a seguinte distribuição para y:

$$y|x \sim \mathrm{N}(\mu_x, \sigma^2)$$
$$\mu_x = \beta_0 + \beta_1 x$$

Modelo de regressão linear

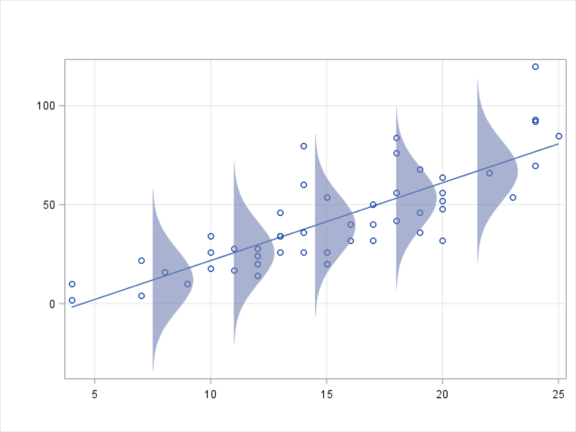

Figura 4: Regressão linear (1)

Modelo de regressão linear

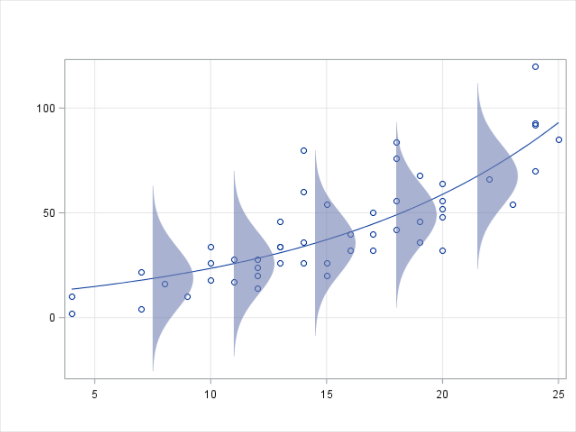

Figura 5: Regressão linear (2)

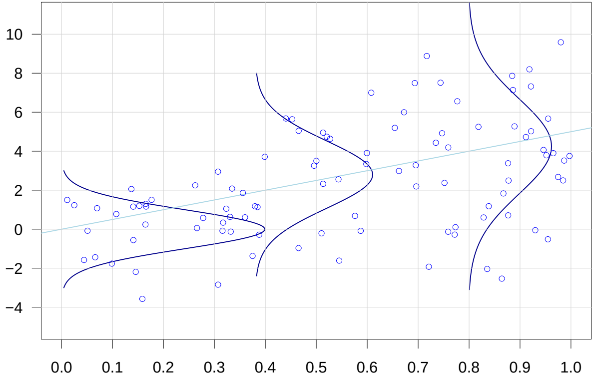

Figura 6: Regressão com erros heterocedásticos (variância não constante)

Modelo de regressão linear

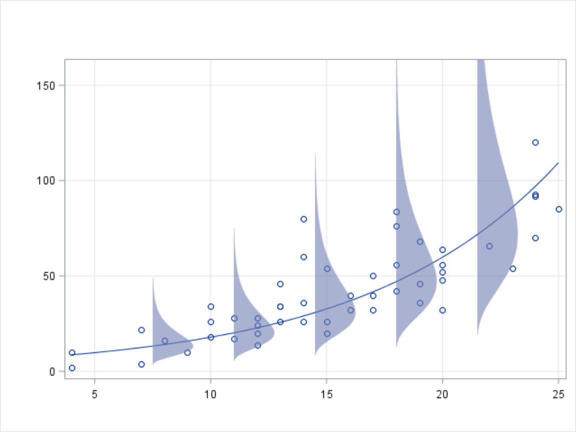

Figura 7: Regressão para dados com distribuição assimétrica

Modelo de regressão linear

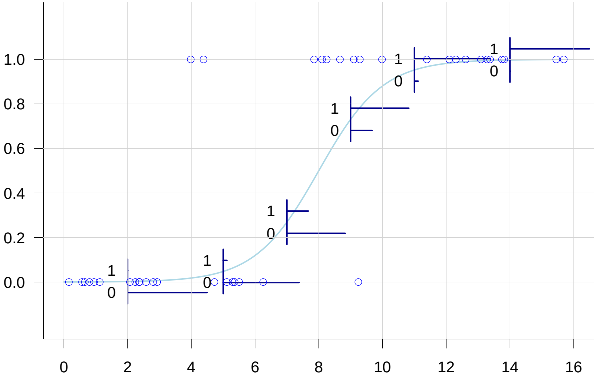

Figura 8: Regressão para dados binários

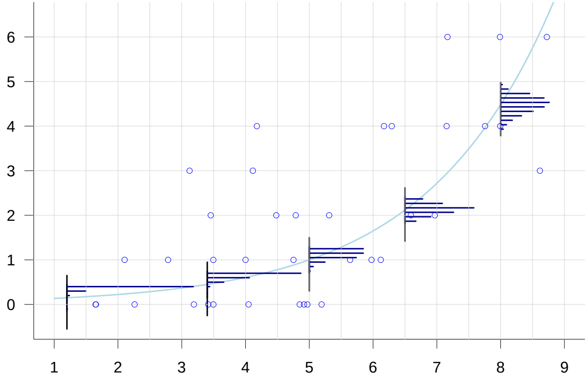

Figura 9: Regressão para dados de contagens

Exemplo- Consumo de combustível

Exemplo- Consumo de combustível

- Neste exemplo vamos analisar os dados da base `Auto`, disponíveis na biblioteca `ISLR` do `R`.

Exemplo- Consumo de combustível

- Neste exemplo vamos analisar os dados da base `Auto`, disponíveis na biblioteca `ISLR` do `R`.

- A base de dados dispõe de informações técnicas de 397 modelos de automóveis das décadas de 1970 e 1980, como consumo de combustível. potência do motor, dimensões dentre outras.

Exemplo- Consumo de combustível

- Neste exemplo vamos analisar os dados da base `Auto`, disponíveis na biblioteca `ISLR` do `R`.

- A base de dados dispõe de informações técnicas de 397 modelos de automóveis das décadas de 1970 e 1980, como consumo de combustível. potência do motor, dimensões dentre outras.

- O objetivo aqui é *ajustar* modelos de regressão linear que expliquem o consumo de combustível com base em características do modelo.

Exemplo- Consumo de combustível

- Neste exemplo vamos analisar os dados da base `Auto`, disponíveis na biblioteca `ISLR` do `R`.

- A base de dados dispõe de informações técnicas de 397 modelos de automóveis das décadas de 1970 e 1980, como consumo de combustível. potência do motor, dimensões dentre outras.

- O objetivo aqui é *ajustar* modelos de regressão linear que expliquem o consumo de combustível com base em características do modelo.

- Neste primeiro momento, para fins ilustrativos vamos considerar apenas duas variáveis de cada vez, embora o mais usual seja considerar múltiplas variáveis conjuntamente numa análise de regressão.

Exemplo- Consumo de combustível

Tabela 1: Extrato da base de dados: ano do modelo e consumo de combustível

	year	mpg
1	70	18
2	70	15
3	70	18
4	70	16
...	...	...
394	82	44
395	82	32
396	82	28
397	82	31

Exemplo- Consumo de combustível

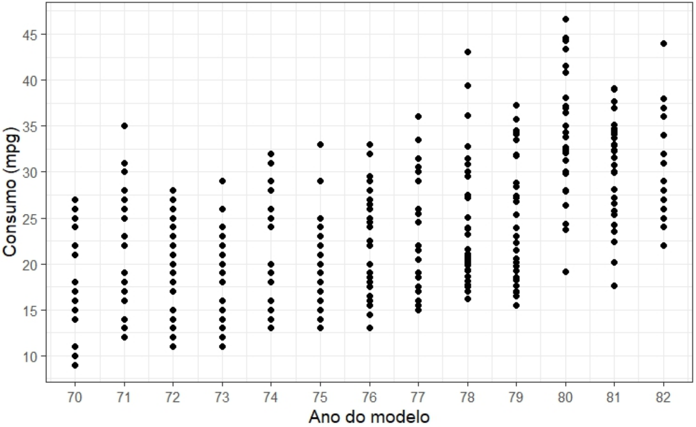

Figura 10: Consumo de combustível *vs* ano do modelo.

Exemplo- Consumo de combustível

- Inicialmente, vamos ajustar um modelo de regressão linear que permita explicar o consumo de combustível (`mpg`- variável resposta) em função do ano de lançamento do modelo (`year`- variável explicativa).

Exemplo- Consumo de combustível

- Inicialmente, vamos ajustar um modelo de regressão linear que permita explicar o consumo de combustível (`mpg`- variável resposta) em função do ano de lançamento do modelo (`year`- variável explicativa).

- A Figura 10 sugere relação crescente entre o consumo de combustível e o ano de lançamento do modelo.

Exemplo- Consumo de combustível

- Inicialmente, vamos ajustar um modelo de regressão linear que permita explicar o consumo de combustível (`mpg`- variável resposta) em função do ano de lançamento do modelo (`year`- variável explicativa).

- A Figura 10 sugere relação crescente entre o consumo de combustível e o ano de lançamento do modelo.

- O modelo de regressão linear para esse par de variáveis fica especificado por:

$$\texttt{mpg} = \beta_0 + \beta_1 \times \texttt{year} + \epsilon,$$

onde β_0 e β_1 são os parâmetros do modelo (intercepto e inclinação da reta de regressão) e ϵ representa os erros aleatórios.

Exemplo- Consumo de combustível

- O ajuste da regressão linear consiste na estimação dos parâmetros do modelo (β_0 e β_1), com base nos dados amostrais, que produzem a reta de regressão que melhor se ajusta aos dados.

Exemplo- Consumo de combustível

- O ajuste da regressão linear consiste na estimação dos parâmetros do modelo (β_0 e β_1), com base nos dados amostrais, que produzem a reta de regressão que melhor se ajusta aos dados.

- O método usual de estimação dos parâmetros de uma regressão linear é o método de mínimos quadrados, que será estudado adiante.

Exemplo- Consumo de combustível

- O ajuste da regressão linear consiste na estimação dos parâmetros do modelo (β_0 e β_1), com base nos dados amostrais, que produzem a reta de regressão que melhor se ajusta aos dados.

- O método usual de estimação dos parâmetros de uma regressão linear é o método de mínimos quadrados, que será estudado adiante.

- Aplicando o método de mínimos quadrados, obtemos os parâmetros estimados $\hat{\beta}_0 = -70.01$ e $\hat{\beta}_1 = 1.23$, produzindo a seguinte reta de regressão ajustada:

$$\widehat{\texttt{mpg}} = -70.01 + 1.23 \times \texttt{year}$$

Exemplo- Consumo de combustível

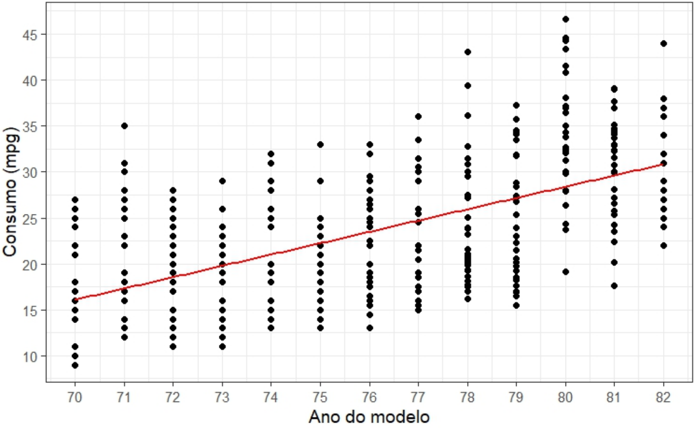

Figura 11: Consumo de combustível *vs* ano do modelo com ajuste da regressão linear.

Exemplo- Consumo de combustível

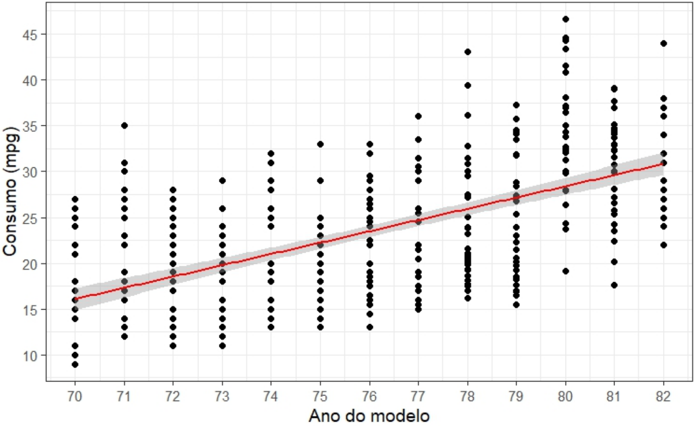

Figura 12: Consumo de combustível *vs* ano do modelo com ajuste da regressão linear e bandas de confiança.

Exemplo- Consumo de combustível

- Com base no modelo ajustado, estimamos um aumento de 1.23 mpg para cada ano a mais no lançamento do modelo (modelos mais novos fazem mais `mpg`).

Exemplo- Consumo de combustível

- Com base no modelo ajustado, estimamos um aumento de 1.23 mpg para cada ano a mais no lançamento do modelo (modelos mais novos fazem mais `mpg`).
- Podemos usar o modelo ajustado para estimar (predizer) o consumo de combustível de veículos. Por exemplo, para modelos de 1972 (`year=72`), temos a seguinte estimativa de consumo:

$$\widehat{\texttt{mpg}} = -70.01 + 1.23 \times 72 = 18.55 \text{ mpg}$$

Exemplo- Consumo de combustível

- Com base no modelo ajustado, estimamos um aumento de 1.23 mpg para cada ano a mais no lançamento do modelo (modelos mais novos fazem mais `mpg`).

- Podemos usar o modelo ajustado para estimar (predizer) o consumo de combustível de veículos. Por exemplo, para modelos de 1972 (`year=72`), temos a seguinte estimativa de consumo:

$$\widehat{\texttt{mpg}} = -70.01 + 1.23 \times 72 = 18.55 \text{ mpg}$$

- Já para veículos lançados em 1978 (`year=78`) a estimativa de consumo é igual a:

$$\widehat{\texttt{mpg}} = -70.01 + 1.23 \times 78 = 25.93 \text{ mpg}$$

Exemplo- Consumo de combustível

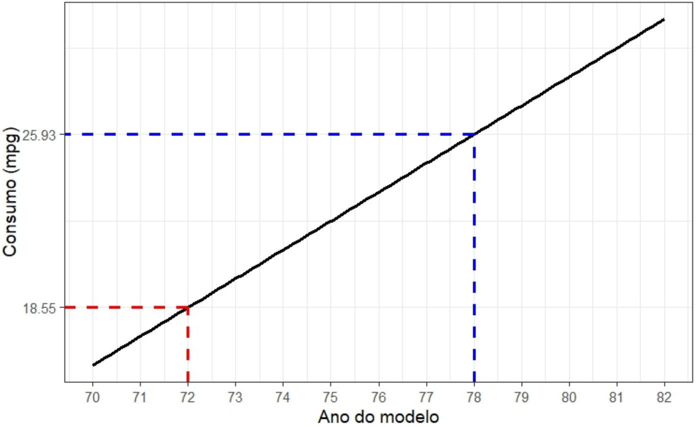

Figura 13: Consumo de combustível *vs* ano do modelo- predições.

Exemplo- Consumo de combustível

- Ainda a título de ilustração, vamos limitar nossa análise agora às variáveis `mpg` e `horsepower` (potência do motor).

Exemplo- Consumo de combustível

- Ainda a título de ilustração, vamos limitar nossa análise agora às variáveis `mpg` e `horsepower` (potência do motor).

- A figura 14 indica relação não linear decrescente entre `mpg` e a potência do motor.

Exemplo- Consumo de combustível

- Ainda a título de ilustração, vamos limitar nossa análise agora às variáveis `mpg` e `horsepower` (potência do motor).

- A figura 14 indica relação não linear decrescente entre `mpg` e a potência do motor.

- Podemos observar na Figura 15 que a reta de regressão obtida por mínimos quadrados para este par de variáveis claramente não se ajusta bem aos dados.

Exemplo- Consumo de combustível

- Ainda a título de ilustração, vamos limitar nossa análise agora às variáveis `mpg` e `horsepower` (potência do motor).

- A figura 14 indica relação não linear decrescente entre `mpg` e a potência do motor.

- Podemos observar na Figura 15 que a reta de regressão obtida por mínimos quadrados para este par de variáveis claramente não se ajusta bem aos dados.

- No entanto, a Figura 16 indica linearidade na relação dessas variáveis quando ambas têm seus valores log-transformados.

Exemplo- Consumo de combustível

Tabela 2: Extrato da base de dados: potência do motor, consumo de combustível e variáveis log-transformadas

	horsepower	mpg	log_horsepower	log_mpg
1	130	18	4.87	2.89
2	165	15	5.11	2.71
3	150	18	5.01	2.89
4	150	16	5.01	2.77
...	...	...	...	...
389	52	44	3.95	3.78
390	84	32	4.43	3.47
391	79	28	4.37	3.33
392	82	31	4.41	3.43

Exemplo- Consumo de combustível

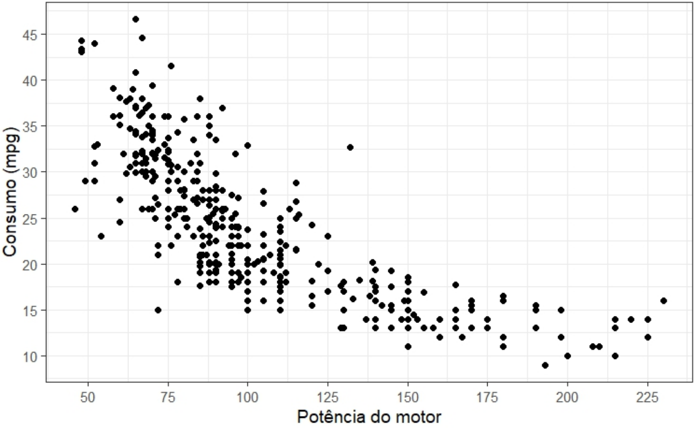

Figura 14: Consumo de combustível *vs* potência do motor.

Exemplo- Consumo de combustível

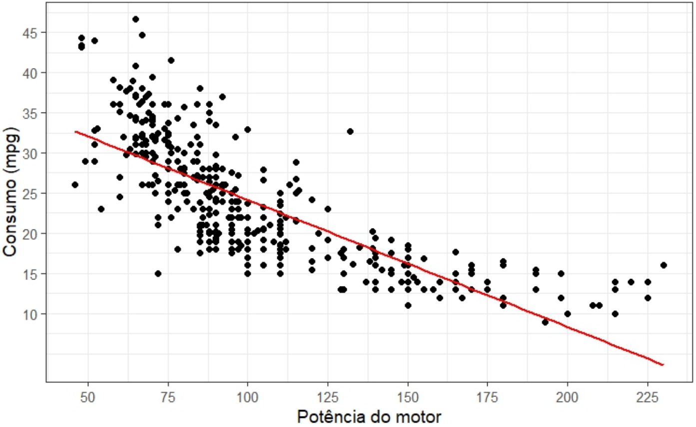

Figura 15: Consumo de combustível *vs* potência do motor com regressão linear ajustada.

Exemplo- Consumo de combustível

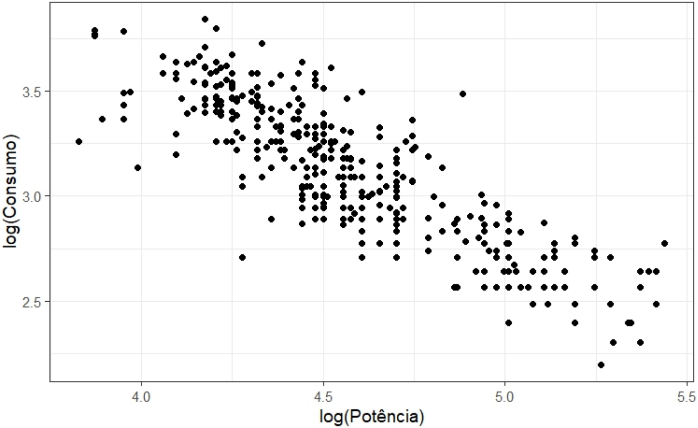

Figura 16: Consumo de combustível *vs* potência do motor com as variáveis (log) transformadas.

Exemplo- Consumo de combustível

- Na análise de regressão linear, é comum transformar a variável resposta e/ou a(s) variável(eis) explicativa(s) para linearizar a relação entre elas.

Exemplo- Consumo de combustível

- Na análise de regressão linear, é comum transformar a variável resposta e/ou a(s) variável(eis) explicativa(s) para linearizar a relação entre elas.

- Neste caso, o modelo de regressão fica especificado por:

$$\log(\texttt{mpg}) = \beta_0 + \beta_1 \times \log(\texttt{horsepower}) + \epsilon$$

Exemplo- Consumo de combustível

- Na análise de regressão linear, é comum transformar a variável resposta e/ou a(s) variável(eis) explicativa(s) para linearizar a relação entre elas.

- Neste caso, o modelo de regressão fica especificado por:

$$\log(\texttt{mpg}) = \beta_0 + \beta_1 \times \log(\texttt{horsepower}) + \epsilon$$

- O modelo ajustado pelo método de mínimos quadrados fica dado por:

$$\widehat{\log(\texttt{mpg})} = 6.96 - 0.84 \times \log(\texttt{horsepower})$$

Exemplo- Consumo de combustível

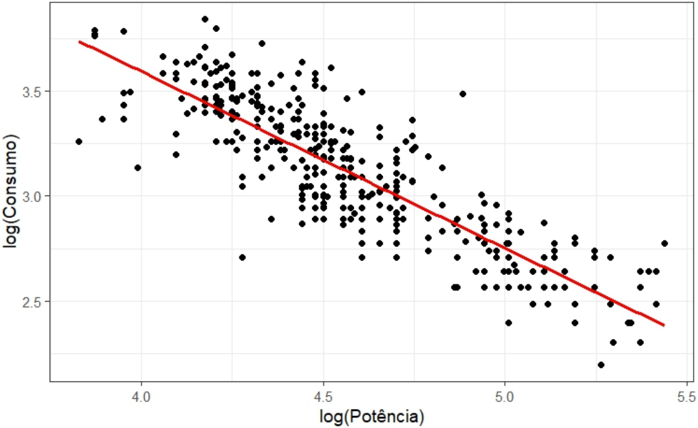

Figura 17: Consumo de combustível *vs* potência do motor com regressão linear ajustada para as variáveis (log) transformadas.

Exemplo- Consumo de combustível

- Podemos expressar o modelo de regressão ajustado diretamente em função de `mpg` na sua escala original exponenciando ambos os lados da equação:

$$\widehat{\texttt{mpg}} = e^{6.96 - 0{,}84 \times \log(\texttt{horsepower})}$$

Exemplo- Consumo de combustível

- Podemos expressar o modelo de regressão ajustado diretamente em função de `mpg` na sua escala original exponenciando ambos os lados da equação:

$$\widehat{\texttt{mpg}} = e^{6.96 - 0{,}84 \times \log(\texttt{horsepower})}$$

- A Figura 18 apresenta a regressão ajustada transformando a resposta para a escala original.

Exemplo- Consumo de combustível

- Podemos expressar o modelo de regressão ajustado diretamente em função de `mpg` na sua escala original exponenciando ambos os lados da equação:

$$\widehat{\texttt{mpg}} = e^{6.96 - 0{,}84 \times \log(\texttt{horsepower})}$$

- A Figura 18 apresenta a regressão ajustada transformando a resposta para a escala original.

- Assim como anteriormente, podemos estimar (predizer) o consumo para diferentes valores de `horsepower` usando o modelo ajustado.

Exemplo- Consumo de combustível

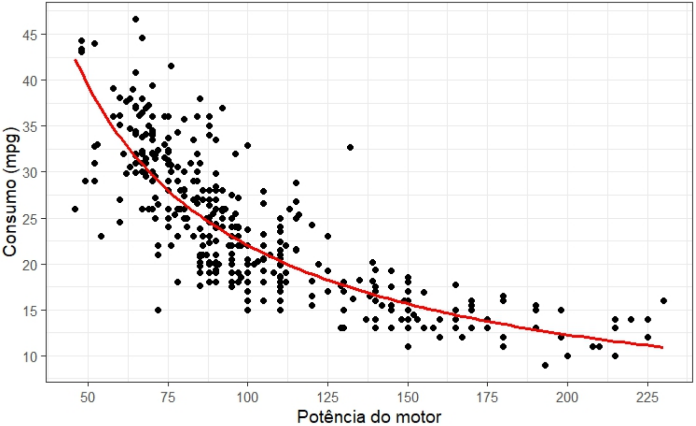

Figura 18: Consumo de combustível *vs* potência do motor com regressão linear ajustada para as variáveis na escala original.

Exemplo- Consumo de combustível

- O consumo de combustível estimado para um veículo de potência `horsepower=90` é igual a:

$$\widehat{\texttt{mpg}} = e^{6.96 - 0{,}84 \times \log(90)} = 24.05 \text{ mpg}$$

Exemplo- Consumo de combustível

- O consumo de combustível estimado para um veículo de potência `horsepower=90` é igual a:

$$\widehat{\texttt{mpg}} = e^{6.96-0,84\times\log(90)} = 24.05 \text{ mpg}$$

- Já o consumo de combustível estimado para um veículo de potência `horsepower=200` é:

$$\widehat{\texttt{mpg}} = e^{6.96-0,84\times\log(200)} = 12.30 \text{ mpg}$$

Exemplo- Consumo de combustível

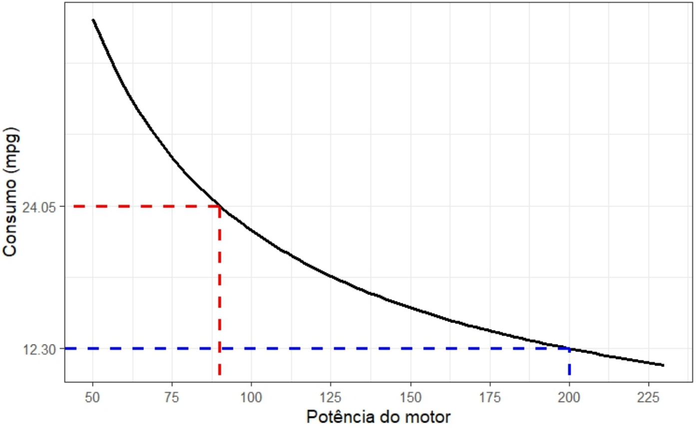

Figura 19: Predições para o consumo de combustível baseadas na regressão linear ajustada.

Exemplo- Consumo de combustível

- A base de dados `Auto` possui diversas outras variáveis que poderiam ser usadas como preditores do consumo, como:

Exemplo- Consumo de combustível

- A base de dados `Auto` possui diversas outras variáveis que poderiam ser usadas como preditores do consumo, como:
    - `cylinders`: número de cilindros;

Exemplo- Consumo de combustível

- A base de dados `Auto` possui diversas outras variáveis que poderiam ser usadas como preditores do consumo, como:
  - `cylinders`: número de cilindros;
  - `displacement`: cilindradas do motor;

Exemplo- Consumo de combustível

- A base de dados `Auto` possui diversas outras variáveis que poderiam ser usadas como preditores do consumo, como:
  - `cylinders`: número de cilindros;
  - `displacement`: cilindradas do motor;
  - `weight`: peso do veículo (em libras);

Exemplo- Consumo de combustível

- A base de dados `Auto` possui diversas outras variáveis que poderiam ser usadas como preditores do consumo, como:
  - `cylinders`: número de cilindros;
  - `displacement`: cilindradas do motor;
  - `weight`: peso do veículo (em libras);
  - `acceleration`: tempo de aceleração (de 0 a 60 mph);

Exemplo- Consumo de combustível

- A base de dados `Auto` possui diversas outras variáveis que poderiam ser usadas como preditores do consumo, como:
  - `cylinders`: número de cilindros;
  - `displacement`: cilindradas do motor;
  - `weight`: peso do veículo (em libras);
  - `acceleration`: tempo de aceleração (de 0 a 60 mph);
  - `origin`: origem do carro (1. American, 2. European, 3. Japanese).

Exemplo- Consumo de combustível

- A base de dados `Auto` possui diversas outras variáveis que poderiam ser usadas como preditores do consumo, como:
  - `cylinders`: número de cilindros;
  - `displacement`: cilindradas do motor;
  - `weight`: peso do veículo (em libras);
  - `acceleration`: tempo de aceleração (de 0 a 60 mph);
  - `origin`: origem do carro (1. American, 2. European, 3. Japanese).

- Poderíamos considerar todas essas variáveis, ou parte delas, numa análise de regressão linear múltipla.

Exemplo- Consumo de combustível

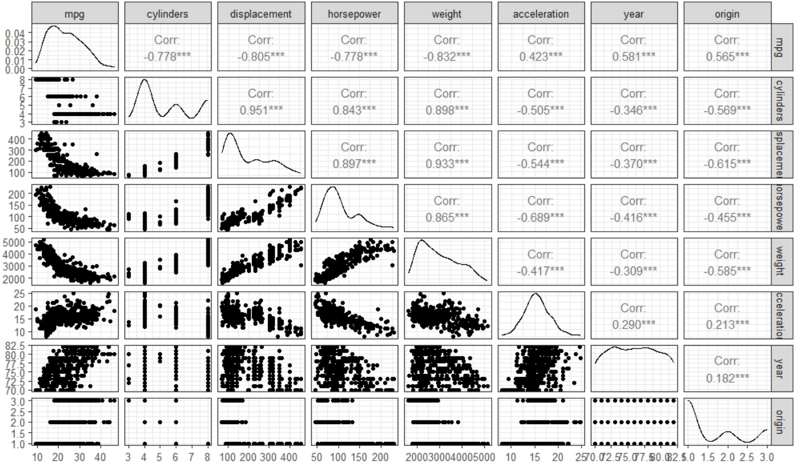

Figura 20: Gráfico de correlações para as variáveis da base de dados `Auto`.

Exemplo- Consumo de combustível

- A título de ilustração, um modelo de regressão linear múltipla com efeito linear de `cylinders`, `displacement` e `weight` em `mpg` fica definido por:

$$\texttt{mpg} = \beta_0 + \beta_1 \times \texttt{cylinders} + \beta_2 \times \texttt{displacement} + \beta_3 \times \texttt{weight} + \epsilon$$

Exemplo- Consumo de combustível

- A título de ilustração, um modelo de regressão linear múltipla com efeito linear de `cylinders`, `displacement` e `weight` em `mpg` fica definido por:

$$\texttt{mpg} = \beta_0 + \beta_1 \times \texttt{cylinders} + \beta_2 \times \texttt{displacement} + \beta_3 \times \texttt{weight} + \epsilon$$

- O modelo de regressão linear ajustado por mínimos quadrados tem a seguinte expressão:

$$\widehat{\texttt{mpg}} = 44.370 - 0.268 \times \texttt{cylinders} - 0.013 \times \texttt{displacement} - 0.006 \times \texttt{weight}$$

Exemplo- Consumo de combustível

- Diversos modelos de regressão linear podem ser definidos como variações do modelo originalmente proposto, por exemplo:

Exemplo- Consumo de combustível

- Diversos modelos de regressão linear podem ser definidos como variações do modelo originalmente proposto, por exemplo:
  - Excluindo ou adicionando variáveis explicativas;

Exemplo- Consumo de combustível

- Diversos modelos de regressão linear podem ser definidos como variações do modelo originalmente proposto, por exemplo:
  - Excluindo ou adicionando variáveis explicativas;
  - Transformando a resposta e/ou variáveis explicativas (Ex: $\sqrt{\texttt{mpg}}$, $\log(\texttt{displacement})$, $1/\texttt{cylinders}$);

Exemplo- Consumo de combustível

- Diversos modelos de regressão linear podem ser definidos como variações do modelo originalmente proposto, por exemplo:
  - Excluindo ou adicionando variáveis explicativas;
  - Transformando a resposta e/ou variáveis explicativas (Ex: $\sqrt{\texttt{mpg}}$, $\log(\texttt{displacement})$, $1/\texttt{cylinders}$);
  - Inserindo potências de variáveis explicativas ($\texttt{displacement}^2$, $\texttt{displacement}^3$);

Exemplo- Consumo de combustível

- Diversos modelos de regressão linear podem ser definidos como variações do modelo originalmente proposto, por exemplo:
  - Excluindo ou adicionando variáveis explicativas;
  - Transformando a resposta e/ou variáveis explicativas (Ex: $\sqrt{\texttt{mpg}}$, $\log(\texttt{displacement})$, $1/\texttt{cylinders}$);
  - Inserindo potências de variáveis explicativas ($\texttt{displacement}^2$, $\texttt{displacement}^3$);
  - Criando uma variável categórica que agrupe os valores de uma variável em faixas ($\texttt{weight} \leq 2000$, $2000 < \texttt{weight} \leq 4000$, $\texttt{weight} > 4000$)...

Gênese dos modelos de regressão

Mas afinal, de onde vem os modelos de regressão?

- A teoria (Física, Química, Biologia,...) pode sugerir o modelo. Por exemplo, segundo a lei de Ohm a voltagem aplicada nos terminais de um condutor é proporcional à corrente elétrica que o percorre. Logo, a relação entre as variáveis é linear, induzindo o modelo;

Mas afinal, de onde vem os modelos de regressão?

- A teoria (Física, Química, Biologia,...) pode sugerir o modelo. Por exemplo, segundo a lei de Ohm a voltagem aplicada nos terminais de um condutor é proporcional à corrente elétrica que o percorre. Logo, a relação entre as variáveis é linear, induzindo o modelo;

- Experiência empírica. Se um particular modelo proporcionou bom ajuste aos dados em estudos similares, possivelmente ele vai se ajustar bem aos dados do presente estudo;

Mas afinal, de onde vem os modelos de regressão?

- A teoria (Física, Química, Biologia,...) pode sugerir o modelo. Por exemplo, segundo a lei de Ohm a voltagem aplicada nos terminais de um condutor é proporcional à corrente elétrica que o percorre. Logo, a relação entre as variáveis é linear, induzindo o modelo;

- Experiência empírica. Se um particular modelo proporcionou bom ajuste aos dados em estudos similares, possivelmente ele vai se ajustar bem aos dados do presente estudo;

- Não há qualquer modelo indicado a priori. A escolha de um modelo pode resultar da exploração dos dados e comparação de diferentes especificações.