ACA228 - Modelos de Regressão e Previsão

Regressão Linear Simples (RLS)

Prof. Carlos Trucíos
carlos.trucios@facc.ufrj.br
ctruciosm.github.io

Faculdade de Administração e Ciências Contábeis,
Universidade Federal do Rio de Janeiro

Aula 3

Introdução

Introdução

Algumas razões para estudar analise de dados:

- A análise de dados é muito importante para as grandes empresas.
- Oportunidades de trabalho em alta
- Salario competitivo
- Oportunidades de trabalho em diversos setores
- As tomadas de decisão nas empresas são influenciadas pelos dados (cultura *Data-Driven*).

Introdução

A modelagem entra em cena quando:

- Temos uma teoria econômica para testar
- Temos em mente uma relação que apresenta alguma importância na tomada de decisão
- Queremos explicar determinados fenômenos
- Queremos saber como o aumento/diminuição em uma variavel influencia em outra.
- Queremos fazer previsão

Introdução

Suponha que estamos interessados em entender/explicar/prever Y em função de um conjunto de p características (variáveis) $X_1, X_2, \ldots, X_p$.

Introdução

Suponha que estamos interessados em entender/explicar/prever Y em função de um conjunto de p características (variáveis) $X_1, X_2, \ldots, X_p$.

Vamos supor também que a relação entre Y e $X_1, X_2, \ldots, X_p$ é da forma

$$Y = f(X_1, X_2, \ldots, X_p) + u$$

em que u é uma perturbação aleatoria.

Introdução

Suponha que estamos interessados em entender/explicar/prever Y em função de um conjunto de p características (variáveis) $X_1, X_2, \ldots, X_p$.

Vamos supor também que a relação entre Y e $X_1, X_2, \ldots, X_p$ é da forma

$$Y = f(X_1, X_2, \ldots, X_p) + u$$

em que u é uma perturbação aleatoria.

- Na prática nunca conhecemos $f(\cdot)$,

Introdução

Suponha que estamos interessados em entender/explicar/prever Y em função de um conjunto de p características (variáveis) $X_1, X_2, \ldots, X_p$.

Vamos supor também que a relação entre Y e $X_1, X_2, \ldots, X_p$ é da forma

$$Y = f(X_1, X_2, \ldots, X_p) + u$$

em que u é uma perturbação aleatoria.

- Na prática nunca conhecemos $f(\cdot)$,
- mas utilizando os dados vamos estimar $f(\cdot)$ por $\widehat{f}(\cdot)$

Introdução

Suponha que estamos interessados em entender/explicar/prever Y em função de um conjunto de p características (variáveis) $X_1, X_2, \ldots, X_p$.

Vamos supor também que a relação entre Y e $X_1, X_2, \ldots, X_p$ é da forma

$$Y = f(X_1, X_2, \ldots, X_p) + u$$

em que u é uma perturbação aleatoria.

- Na prática nunca conhecemos $f(\cdot)$,
- mas utilizando os dados vamos estimar $f(\cdot)$ por $\widehat{f}(\cdot)$
- Assim, $\widehat{Y} = \widehat{f}(X_1, X_2, \ldots, X_p)$

Introdução

Suponha que estamos interessados em entender/explicar/prever Y em função de um conjunto de p características (variáveis) $X_1, X_2, \dots, X_p$.

Vamos supor também que a relação entre Y e $X_1, X_2, \dots, X_p$ é da forma

$$Y = f(X_1, X_2, \dots, X_p) + u$$

em que u é uma perturbação aleatoria.

- Na prática nunca conhecemos $f(\cdot)$,
- mas utilizando os dados vamos estimar $f(\cdot)$ por $\widehat{f}(\cdot)$
- Assim, $\widehat{Y} = \widehat{f}(X_1, X_2, \dots, X_p)$

Introdução

Suponha que estamos interessados em entender/explicar/prever Y em função de um conjunto de p características (variáveis) $X_1, X_2, \dots, X_p$.

Vamos supor também que a relação entre Y e $X_1, X_2, \dots, X_p$ é da forma

$$Y = f(X_1, X_2, \dots, X_p) + u$$

em que u é uma perturbação aleatoria.

- Na prática nunca conhecemos $f(\cdot)$,
- mas utilizando os dados vamos estimar $f(\cdot)$ por $\widehat{f}(\cdot)$
- Assim, $\widehat{Y} = \widehat{f}(X_1, X_2, \dots, X_p)$

Nota: O *chapeuzinho* significa "estimado"

Introdução

Nesta disciplina focaremos no caso em que

$$Y = \underbrace{\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p}_{f(X_1,\ldots,X_p)} + u$$

Introdução

Nesta disciplina focaremos no caso em que

$$Y = \underbrace{\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p}_{f(X_1,\ldots,X_p)} + u$$

Y	X
variável dependente	variável independente
variável explicada	variável explicativa
variável resposta	variável de controle
variável prevista	variável previsora
regressando	regressor
variável *target*	covariável
	feature

Introdução

Estrutura de dados

No processo de modelagem, lidamos com diversos tipos de dados, eles podem ser classificados em:

- Corte transversal
- Séries temporais
- Corte transversal agrupados
- Painel (ou longitudinais)

Para cada tipo de dado, teremos uma abordagem de modelagem diferente que nos permitirá explotar a informação contida nos dados.

Introdução

Corte transversal

Consiste em uma amostra de indivídios* (consumidores, empresas, cidades, paises, etc) tomadas em determino período no tempo. Podemos pensar nesse conjunto de dados como quando tiramos uma **foto** (*panoramica*).

Introdução

Corte transversal

Consiste em uma amostra de indivídios* (consumidores, empresas, cidades, paises, etc) tomadas em determino período no tempo. Podemos pensar nesse conjunto de dados como quando tiramos uma **foto** (*panoramica*).

Séries temporais

Consiste em observações sobre uma (ou várias) variaveis ao **longo do tempo**. A diferença dos dados e corte transversal, dados de séries temporais são ordenados de forma cronológica.

Introdução

Corte transversal agrupados

Agrupar várias amostras de corte transversal (cada uma tomada em diferentes períodos de tempo)

Introdução

Corte transversal agrupados

Agrupar várias amostras de corte transversal (cada uma tomada em diferentes períodos de tempo)

Painel (ou longitudinais)

Consiste em uma série temporal para cada observação de corte transversal. A diferença dos dados de *corte transversal agrupados*, nos dados de painel as **mesmas unidades** são acompanhadas ao longo do tempo

Introdução

Corte transversal agrupados

Agrupar várias amostras de corte transversal (cada uma tomada em diferentes períodos de tempo)

Painel (ou longitudinais)

Consiste em uma série temporal para cada observação de corte transversal. A diferença dos dados de *corte transversal agrupados*, nos dados de painel as **mesmas unidades** são acompanhadas ao longo do tempo

Nesta disciplina, estudaremos dados de corte transversal e de séries temporais

Modelo de Regressão Linear Simples

Modelo de Regressão Linear Simples

Sejam X e Y duas variáveis e suponha que queremos **explicar** Y **em termos de** X, **estudar como varia** Y **com variações de** X, **predizer** Y **utilizando** X.

Modelo de Regressão Linear Simples

Sejam X e Y duas variáveis e suponha que queremos **explicar** Y **em termos de** X, **estudar como varia** Y **com variações de** X, **predizer** Y **utilizando** X.

MRL Simples

$$Y = \beta_0 + \beta_1 X + u$$

em que

- Y variável dependente,
- X variável independente,
- $\beta = [\beta_0 \quad \beta_1]'$, β_0 é parâmetro de intercepto e β_1 é o parâmetro de inclinação e
- u é o termo de erro ou perturbação (representa outros fatores, além de X, que afetam Y)

Modelo de Regressão Linear Simples

Exemplos

- Salario e educação

$$\text{Salario} = \beta_0 + \beta_1 \text{Educação} + u$$

Modelo de Regressão Linear Simples

Exemplos

- Salario e educação

$$\text{Salario} = \beta_0 + \beta_1 \text{Educação} + u$$

- Vendas do Galaxy Note 20 e gasto em publicidade

$$\text{Vendas} = \beta_0 + \beta_1 \text{Gasto em publicidade} + u$$

Modelo de Regressão Linear Simples

Exemplos

- Salario e educação

$$\text{Salario} = \beta_0 + \beta_1 \text{Educação} + u$$

- Vendas do Galaxy Note 20 e gasto em publicidade

$$\text{Vendas} = \beta_0 + \beta_1 \text{Gasto em publicidade} + u$$

- Pressão arterial e dosagem de um determinada medicamento

$$\text{Pressão} = \beta_0 + \beta_1 \text{Dosagem} + u$$

Modelo de Regressão Linear Simples

Exemplos

- Salario e educação

$$\text{Salario} = \beta_0 + \beta_1 \text{Educação} + u$$

- Vendas do Galaxy Note 20 e gasto em publicidade

$$\text{Vendas} = \beta_0 + \beta_1 \text{Gasto em publicidade} + u$$

- Pressão arterial e dosagem de um determinada medicamento

$$\text{Pressão} = \beta_0 + \beta_1 \text{Dosagem} + u$$

- Notas do ENEM e horas semanais de estudo

$$\text{Nota do ENEM} = \beta_0 + \beta_1 \text{Horas de estudo} + u$$

Modelo de Regressão Linear Simples

Exemplos

- Salario e educação

$$\text{Salario} = \beta_0 + \beta_1 \text{Educação} + u$$

- Vendas do Galaxy Note 20 e gasto em publicidade

$$\text{Vendas} = \beta_0 + \beta_1 \text{Gasto em publicidade} + u$$

- Pressão arterial e dosagem de um determinada medicamento

$$\text{Pressão} = \beta_0 + \beta_1 \text{Dosagem} + u$$

- Notas do ENEM e horas semanais de estudo

$$\text{Nota do ENEM} = \beta_0 + \beta_1 \text{Horas de estudo} + u$$

- Satisfação dos trabalhadores e horas em reuniões virtuais

$$\text{Satisfação} = \beta_0 + \beta_1 \text{Horas de reunião} + u$$

Modelo de Regressão Linear Simples

- Nos exemplos anteriores, estamos interessados em saber como X afeta/explica/ajuda a predizer Y.

Modelo de Regressão Linear Simples

- Nos exemplos anteriores, estamos interessados em saber como X afeta/explica/ajuda a predizer Y.
- Um problema de bastante interesse é saber qual é o efeito de X sobre Y *mantendo fixos os outros fatores*

Modelo de Regressão Linear Simples

- Nos exemplos anteriores, estamos interessados em saber como X afeta/explica/ajuda a predizer Y.
- Um problema de bastante interesse é saber qual é o efeito de X sobre Y *mantendo fixos os outros fatores*

Modelo de Regressão Linear Simples

- Nos exemplos anteriores, estamos interessados em saber como X afeta/explica/ajuda a predizer Y.
- Um problema de bastante interesse é saber qual é o efeito de X sobre Y *mantendo fixos os outros fatores*

Exemplo

- A cada 10K em publicidade, são vendidos, em média 22 *Galaxy Note 20*

Modelo de Regressão Linear Simples

- Nos exemplos anteriores, estamos interessados em saber como X afeta/explica/ajuda a predizer Y.
- Um problema de bastante interesse é saber qual é o efeito de X sobre Y *mantendo fixos os outros fatores*

Exemplo

- A cada 10K em publicidade, são vendidos, em média 22 *Galaxy Note 20*
- A cada hora adicional de reunões virtuais por semana, em média, o indice de satisfação do trabalhadores cai em 5%

Modelo de Regressão Linear Simples

- Nos exemplos anteriores, estamos interessados em saber como X afeta/explica/ajuda a predizer Y.
- Um problema de bastante interesse é saber qual é o efeito de X sobre Y *mantendo fixos os outros fatores*

Exemplo

- A cada 10K em publicidade, são vendidos, em média 22 *Galaxy Note 20*
- A cada hora adicional de reunões virtuais por semana, em média, o indice de satisfação do trabalhadores cai em 5%
- A cada hora de estudo semanal adicional, em média, o aluno obtem 1 ponto adicional na nota do ENEM.

Modelo de Regressão Linear Simples

- Nos exemplos anteriores, estamos interessados em saber como X afeta/explica/ajuda a predizer Y.
- Um problema de bastante interesse é saber qual é o efeito de X sobre Y *mantendo fixos os outros fatores*

Exemplo

- A cada 10K em publicidade, são vendidos, em média 22 *Galaxy Note 20*
- A cada hora adicional de reunões virtuais por semana, em média, o indice de satisfação do trabalhadores cai em 5%
- A cada hora de estudo semanal adicional, em média, o aluno obtem 1 ponto adicional na nota do ENEM.
- etc

Modelo de Regressão Linear Simples

- Nos exemplos anteriores, estamos interessados em saber como X afeta/explica/ajuda a predizer Y.
- Um problema de bastante interesse é saber qual é o efeito de X sobre Y *mantendo fixos os outros fatores*

Exemplo

- A cada 10K em publicidade, são vendidos, em média 22 *Galaxy Note 20*
- A cada hora adicional de reunões virtuais por semana, em média, o indice de satisfação do trabalhadores cai em 5%
- A cada hora de estudo semanal adicional, em média, o aluno obtem 1 ponto adicional na nota do ENEM.
- etc

Modelo de Regressão Linear Simples

- Nos exemplos anteriores, estamos interessados em saber como X afeta/explica/ajuda a predizer Y.
- Um problema de bastante interesse é saber qual é o efeito de X sobre Y *mantendo fixos os outros fatores*

Exemplo

- A cada 10K em publicidade, são vendidos, em média 22 *Galaxy Note 20*
- A cada hora adicional de reunões virtuais por semana, em média, o indice de satisfação do trabalhadores cai em 5%
- A cada hora de estudo semanal adicional, em média, o aluno obtem 1 ponto adicional na nota do ENEM.
- etc

Nota: *mantendo fixos os outros fatores* é conhecido como efeito *Ceteris paribus*.

Modelo de Regressão Linear Simples

No modelo

$$Y = \beta_0 + \beta_1 X + u$$

se todos os outros fatores são mantidos fixos (de modo que $\Delta u = 0$), temos

$$\Delta Y = \beta_1 \Delta X$$

Modelo de Regressão Linear Simples

No modelo

$$Y = \beta_0 + \beta_1 X + u$$

se todos os outros fatores são mantidos fixos (de modo que $\Delta u = 0$), temos

$$\Delta Y = \beta_1 \Delta X$$

Se X varia em uma unidade, Y varia em β_1 unidades.

Modelo de Regressão Linear Simples

No modelo

$$Y = \beta_0 + \beta_1 X + u$$

se todos os outros fatores são mantidos fixos (de modo que $\Delta u = 0$), temos

$$\Delta Y = \beta_1 \Delta X$$

Se X varia em uma unidade, Y varia em β_1 unidades.

Mas... como podemos manter fixos *todos os outros fatores* quando na verdade estamos ignorando eles?

Modelo de Regressão Linear Simples

No modelo

$$Y = \beta_0 + \beta_1 X + u$$

se todos os outros fatores são mantidos fixos (de modo que $\Delta u = 0$), temos

$$\Delta Y = \beta_1 \Delta X$$

Se X varia em uma unidade, Y varia em β_1 unidades.

Mas... como podemos manter fixos *todos os outros fatores* quando na verdade estamos ignorando eles?

Na verdade, somente podemos obter estimadores *confiáveis* de β_0 y β_1 quando fazemos hipotesis sobre u e como se relaciona com X

Modelo de Regressão Linear Simples

No modelo

$$Y = \beta_0 + \beta_1 X + u$$

se todos os outros fatores são mantidos fixos (de modo que $\Delta u = 0$), temos

$$\Delta Y = \beta_1 \Delta X$$

Se X varia em uma unidade, Y varia em β_1 unidades.

Mas... como podemos manter fixos *todos os outros fatores* quando na verdade estamos ignorando eles?

Na verdade, somente podemos obter estimadores *confiáveis* de β_0 y β_1 quando fazemos hipotesis sobre u e como se relaciona com X

Nota: Não assumimos hipóteses, estabelecemos hipóteses e as verificamos.

Modelo de Regressão Linear Simples
Hipóteses

- $\mathbb{E}(u) = 0$

Modelo de Regressão Linear Simples

Hipóteses

- $\mathbb{E}(u) = 0$
- $\mathbb{E}(u|X) = \mathbb{E}(u)$

Modelo de Regressão Linear Simples
Hipóteses

- $\mathbb{E}(u) = 0$
- $\mathbb{E}(u|X) = \mathbb{E}(u)$

Modelo de Regressão Linear Simples

Hipóteses

- $\mathbb{E}(u) = 0$
- $\mathbb{E}(u|X) = \mathbb{E}(u)$

O qué implicam essas hipoteses?

Modelo de Regressão Linear Simples
Hipóteses

- $\mathbb{E}(u) = 0$
- $\mathbb{E}(u|X) = \mathbb{E}(u)$

O qué implicam essas hipoteses?

1. No modelo com intercepto (ou seja com β_0), sem perda de generalidade, sempre podemos assumir que que o valor médio de u é zero ($\mathbb{E}(u) = 0$)

Modelo de Regressão Linear Simples

Hipóteses

- $\mathbb{E}(u) = 0$
- $\mathbb{E}(u|X) = \mathbb{E}(u)$

O qué implicam essas hipoteses?

1. No modelo com intercepto (ou seja com β_0), sem perda de generalidade, sempre podemos assumir que que o valor médio de u é zero ($\mathbb{E}(u) = 0$)
2. $\mathbb{E}(u|X) = \mathbb{E}(u)$ diz que o valor médio dos fatores não observáveis (u) é o mesmo para todo valor de X e que é igual à media de u.

Modelo de Regressão Linear Simples

Hipóteses

- $\mathbb{E}(u) = 0$
- $\mathbb{E}(u|X) = \mathbb{E}(u)$

O qué implicam essas hipoteses?

1. No modelo com intercepto (ou seja com β_0), sem perda de generalidade, sempre podemos assumir que que o valor médio de u é zero ($\mathbb{E}(u) = 0$)
2. $\mathbb{E}(u|X) = \mathbb{E}(u)$ diz que o valor médio dos fatores não observáveis (u) é o mesmo para todo valor de X e que é igual à media de u.
3. No MRLS, se aplicarnos $\mathbb{E}(\cdot|X)$, temos que

$$\mathbb{E}(Y|X) = \mathbb{E}(\beta_0 + \beta_1 X + u|X) = \beta_0 + \beta_1 \underbrace{\mathbb{E}(X|X)}_{X} + \underbrace{\mathbb{E}(u|X)}_{0} = \beta_0 + \beta_1 X$$

Modelo de Regressão Linear Simples

$$\mathbb{E}(Y|X) = \beta_0 + \beta_1 X$$

Modelo de Regressão Linear Simples

$$\mathbb{E}(Y|X) = \beta_0 + \beta_1 X$$

a média de Y aumenta em β_1 por unidade em X.

Modelo de Regressão Linear Simples

$$\mathbb{E}(Y|X) = \beta_0 + \beta_1 X$$

a média de Y aumenta em β_1 por unidade em X.

Exemplos

(lembre-se que u, o termo aleatório, representa outros fatores ($\neq X$) que também afetam Y)

- Suponha que u seja **aptidão** e X seja **educação** (em anos), então $\mathbb{E}(\text{aptidão}|\text{educação} = 5)$ representa a aptidão média para o grupo de pessoas com 5 anos de educação e $\mathbb{E}(\text{aptidão}|\text{educação} = 12)$ a aptidão média para o grupo de pessoas com 12 anos de educação.

Modelo de Regressão Linear Simples

$$\mathbb{E}(Y|X) = \beta_0 + \beta_1 X$$

a média de Y aumenta em β_1 por unidade em X.

Exemplos

(lembre-se que u, o termo aleatório, representa outros fatores ($\neq X$) que também afetam Y)

- Suponha que u seja **aptidão** e X seja **educação** (em anos), então $\mathbb{E}(\text{aptidão}|\text{educação} = 5)$ representa a aptidão média para o grupo de pessoas com 5 anos de educação e $\mathbb{E}(\text{aptidão}|\text{educação} = 12)$ a aptidão média para o grupo de pessoas com 12 anos de educação.
- $\mathbb{E}(u|X) = \mathbb{E}(u)$ implica que ambas as médias devem ser as mesmas.

Modelo de Regressão Linear Simples

$$\mathbb{E}(Y|X) = \beta_0 + \beta_1 X$$

a média de Y aumenta em β_1 por unidade em X.

Exemplos

(lembre-se que u, o termo aleatório, representa outros fatores ($\neq X$) que também afetam Y)

- Suponha que u seja **aptidão** e X seja **educação** (em anos), então $\mathbb{E}(\text{aptidão}|\text{educação} = 5)$ representa a aptidão média para o grupo de pessoas com 5 anos de educação e $\mathbb{E}(\text{aptidão}|\text{educação} = 12)$ a aptidão média para o grupo de pessoas com 12 anos de educação.
- $\mathbb{E}(u|X) = \mathbb{E}(u)$ implica que ambas as médias devem ser as mesmas.
- Mas se entendermos que a média da aptidão aumenta com os anos de educação formal, então $\mathbb{E}(u|X) \neq \mathbb{E}(u)$

Modelo de Regressão Linear Simples

Suponha que

$$Y = 1.05 + 0.5X + u$$

com $\mathbb{E}(u|x) = \mathbb{E}(u) = 0$. Então

$$\mathbb{E}(Y|X) = \underbrace{1.05}_{\beta_0} + \underbrace{0.5}_{\beta_1} X$$

(a média de Y aumenta em 0.5 por unidade em X).

Modelo de Regressão Linear Simples

Suponha que

$$Y = 1.05 + 0.5X + u$$

com $\mathbb{E}(u|x) = \mathbb{E}(u) = 0$. Então

$$\mathbb{E}(Y|X) = \underbrace{1.05}_{\beta_0} + \underbrace{0.5}_{\beta_1} X$$

(a média de Y aumenta em 0.5 por unidade em X).

Qual o problema com a equação acima?

Modelo de Regressão Linear Simples

Suponha que

$$Y = 1.05 + 0.5X + u$$

com $\mathbb{E}(u|x) = \mathbb{E}(u) = 0$. Então

$$\mathbb{E}(Y|X) = \underbrace{1.05}_{\beta_0} + \underbrace{0.5}_{\beta_1} X$$

(a média de Y aumenta em 0.5 por unidade em X).

Qual o problema com a equação acima?

Na prática, nunca conhecemos os valores de β_0 e β_1 e devemos estimá-los.

Estimação

Estimação

Por que preciso estimar os β's?

Na prática, nunca conhecemos os valores de β_0, β_1 então precisamos estima-los utilizando os dados. Estes valores estimados serão denotados por $\hat{\beta}_0$ e $\hat{\beta}_1$, respectivamente.

Estimação

Por que preciso estimar os β's?

Na prática, nunca conhecemos os valores de β_0, β_1 então precisamos estima-los utilizando os dados. Estes valores estimados serão denotados por $\hat{\beta}_0$ e $\hat{\beta}_1$, respectivamente.

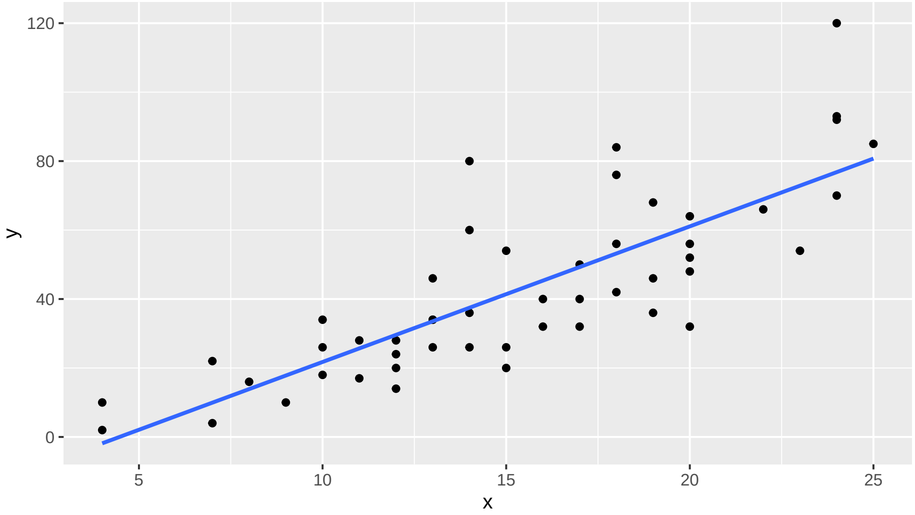
Sejam $(x_1, y_1), \ldots, (x_n, y_n)$ uma amostra aleatória (a.a.) de tamanho n da população, então

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

onde i indica a i-ésima observação.

Estimação

Estamos interessados em uma reta do tipo:



Estimação

Como vamos a obter essa reta?

Estimação

Como vamos a obter essa reta?

- De forma que minimize a soma de quadrados dos resíduos $\hat{u}_i := y_i - \hat{y}_i$ em que $\hat{y}_i = b_0 - b_1 x_i$

Estimação

Como vamos a obter essa reta?

- De forma que minimize a soma de quadrados dos resíduos $\hat{u}_i := y_i - \hat{y}_i$ em que $\hat{y}_i = b_0 - b_1 x_i$
- Seja $SQR = \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} \left(y_i - b_0 - b_1 x_i\right)^2$, então

$$\hat{\beta}_0, \hat{\beta}_1 = \underset{b_0, b_1}{\operatorname{argmin}}\, SQR$$

Estimação

Como vamos a obter essa reta?

- De forma que minimize a soma de quadrados dos resíduos $\hat{u}_i := y_i - \hat{y}_i$ em que $\hat{y}_i = b_0 - b_1 x_i$
- Seja $SQR = \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} \left(y_i - b_0 - b_1 x_i\right)^2$, então

$$\hat{\beta}_0, \hat{\beta}_1 = \underset{b_0, b_1}{\operatorname{argmin}} \, SQR$$

Estimação

Como vamos a obter essa reta?

- De forma que minimize a soma de quadrados dos resíduos $\hat{u}_i := y_i - \hat{y}_i$ em que $\hat{y}_i = b_0 - b_1 x_i$
- Seja $SQR = \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} \left(y_i - b_0 - b_1 x_i\right)^2$, então

$$\hat{\beta}_0, \hat{\beta}_1 = \underset{b_0, b_1}{\operatorname{argmin}} \, SQR$$

Para minimizar SQR, igualamos a primeira derivada a zero (para obter os candidatos) e depois pelo critério da segunda derivada verificamos se é ponto de mínimo.

Estimação

Como vamos a obter essa reta?

- De forma que minimize a soma de quadrados dos resíduos $\hat{u}_i := y_i - \hat{y}_i$ em que $\hat{y}_i = b_0 - b_1 x_i$
- Seja $SQR = \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} \left(y_i - b_0 - b_1 x_i\right)^2$, então

$$\hat{\beta}_0, \hat{\beta}_1 = \underset{b_0, b_1}{\operatorname{argmin}}\, SQR$$

Para minimizar SQR, igualamos a primeira derivada a zero (para obter os candidatos) e depois pelo critério da segunda derivada verificamos se é ponto de mínimo.

- Derivando w.r.t b_0 e b_1 temos

$$\frac{\partial SQR}{\partial b_0} = -2 \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i); \quad \frac{\partial SQR}{\partial b_1} = -2 \sum_{i=1}^{n} x_i (y_i - b_0 - b_1 x_i)$$

Estimação

Fazendo $\frac{\partial SQR}{\partial b_0} = 0$ e $\frac{\partial SQR}{\partial b_1} = 0$, chegamos ao sistema:

$$-2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0; \quad -2\sum_{i=1}^{n}(x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)) = 0 \qquad (1)$$

Estimação

Fazendo $\frac{\partial SQR}{\partial b_0} = 0$ e $\frac{\partial SQR}{\partial b_1} = 0$, chegamos ao sistema:

$$-2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0; \quad -2\sum_{i=1}^{n}(x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)) = 0 \qquad (1)$$

Após um pouco de matemática, temos que

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \text{ e } \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Estimação

Fazendo $\frac{\partial SQR}{\partial b_0} = 0$ e $\frac{\partial SQR}{\partial b_1} = 0$, chegamos ao sistema:

$$-2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0; \quad -2\sum_{i=1}^{n}(x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)) = 0 \tag{1}$$

Após um pouco de matemática, temos que

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \text{ e } \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Importante: As equações em (1) são conhecidas como condições de primeira ordem. De (1) temos que $\overline{\hat{u}} = 0$ e $\overline{x\hat{u}} = 0$.

Estimação

Como sabemos que de fato são os valores que minimizam SQR?

Estimação

Como sabemos que de fato são os valores que minimizam SQR?

Precisamos utilizar o critério da segunda derivada!!!

Estimação

Como sabemos que de fato são os valores que minimizam SQR?

Precisamos utilizar o critério da segunda derivada!!!

A matriz

$$\begin{bmatrix} \frac{\partial^2 SQR}{\partial b_0^2} & \frac{\partial^2 SQR}{\partial b_0 \partial b_1} \\ \frac{\partial^2 SQR}{\partial b_1 \partial b_0} & \frac{\partial^2 SQR}{\partial b_1^2} \end{bmatrix}$$

deve ser definida positiva (**todos os autovalores devem ser positivos**)

Estimação

Como sabemos que de fato são os valores que minimizam SQR?

Precisamos utilizar o critério da segunda derivada!!!

A matriz

$$\begin{bmatrix} \dfrac{\partial^2 SQR}{\partial b_0^2} & \dfrac{\partial^2 SQR}{\partial b_0 \partial b_1} \\ \dfrac{\partial^2 SQR}{\partial b_1 \partial b_0} & \dfrac{\partial^2 SQR}{\partial b_1^2} \end{bmatrix}$$

deve ser definida positiva (**todos os autovalores devem ser positivos**)

O método descrito anteriormente e conhecido como o **método de mínimos quadrados ordinários (MQO)** e será amplamente utilizado na disciplina.

Métodos de estimação: MQO

Estimadores de MQO

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \quad \text{e} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

Métodos de estimação: MQO

Estimadores de MQO

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \quad \text{e} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

Note que

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} = \frac{\overbrace{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})/(n-1)}^{Cov(x,y)}}{\underbrace{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2/(n-1)}_{\hat{\sigma}_x^2}} = \frac{Cov(x,y)}{\hat{\sigma}_x^2}$$

Métodos de estimação: MQO

Por outro lado, $\hat{\rho}_{xy} := Cor(x, y) = \dfrac{Cov(x, y)}{\hat{\sigma}_x \hat{\sigma}_y}$, então

$$\hat{\beta}_1 = \frac{Cov(x, y)}{\hat{\sigma}_x^2} = \frac{\hat{\rho}_{xy} \hat{\sigma}_x \hat{\sigma}_y}{\hat{\sigma}_x^2} = \hat{\rho}_{xy} \frac{\hat{\sigma}_y}{\hat{\sigma}_x}$$

Métodos de estimação: MQO

Por outro lado, $\hat{\rho}_{xy} := Cor(x,y) = \dfrac{Cov(x,y)}{\hat{\sigma}_x \hat{\sigma}_y}$, então

$$\hat{\beta}_1 = \frac{Cov(x,y)}{\hat{\sigma}_x^2} = \frac{\hat{\rho}_{xy}\hat{\sigma}_x\hat{\sigma}_y}{\hat{\sigma}_x^2} = \hat{\rho}_{xy}\frac{\hat{\sigma}_y}{\hat{\sigma}_x}$$

Ou seja, temos três formulas equivalentes para calcular $\hat{\beta}_1$

Métodos de estimação: MQO

Por outro lado, $\hat{\rho}_{xy} := Cor(x, y) = \dfrac{Cov(x, y)}{\hat{\sigma}_x \hat{\sigma}_y}$, então

$$\hat{\beta}_1 = \frac{Cov(x, y)}{\hat{\sigma}_x^2} = \frac{\hat{\rho}_{xy}\hat{\sigma}_x\hat{\sigma}_y}{\hat{\sigma}_x^2} = \hat{\rho}_{xy}\frac{\hat{\sigma}_y}{\hat{\sigma}_x}$$

Ou seja, temos três formulas equivalentes para calcular $\hat{\beta}_1$

Nota: É comum encontrar nos livros os termos *estimador*, *estimativa* e *valor estimado* e isso pode causar um pouco de confussão.

- *Estimador* é a *fórmula*,
- Quando utilizamos os dados observados, aplicamos a fórmula e obtemos um número, esse número é a *estimativa* ou *valor estimado*

Métodos de estimação: MQO

Uma vez obtidos $\hat{\beta}_0$ e $\hat{\beta}_1$, podemos obter $\hat{Y}$ (**y chapéu**)

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

($\hat{y}$: valores ajustados/previstos da regressão)

Métodos de estimação: MQO

Uma vez obtidos $\hat{\beta}_0$ e $\hat{\beta}_1$, podemos obter $\hat{Y}$ (**y chapéu**)

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

($\hat{y}$: valores ajustados/previstos da regressão)

Veja que

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x \quad \text{ou equivalentemente} \quad \hat{\beta}_1 = \Delta\hat{y}/\Delta x$$

Métodos de estimação: MQO

Uma vez obtidos $\hat{\beta}_0$ e $\hat{\beta}_1$, podemos obter $\hat{Y}$ (**y chapéu**)

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

($\hat{y}$: valores ajustados/previstos da regressão)

Veja que

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x \quad \text{ou equivalentemente} \quad \hat{\beta}_1 = \Delta\hat{y}/\Delta x$$

$\hat{\beta}_1$ nos diz quanto varia $\hat{y}$ quando x aumenta em uma unidade

Exemplos no R

Exemplos no R: salarios CEO

Informações sobre o *salario* (anual em milhares de dólares) e o retorno médio sobre o patrimônio líquido dos últimos 3 anos, *roe* (definido como uma porcentagem do patrimônio líquido) de 209 CEOs estão disponíveis no dataset `ceosal1` do pacote do *R* `wooldridge`.

Queremos estudar a relação

$$\text{salario} = \beta_0 + \beta_1 \text{roe} + u$$

Exemplos no R: salarios CEO

```
library(wooldridge)
library(dplyr)
ceosal1 %>%  select(salary,roe) %>% head(3)
```

```
##   salary  roe
## 1   1095 14.1
## 2   1001 10.9
## 3   1122 23.5
```

Exemplos no R: salarios CEO

```
library(wooldridge)
library(dplyr)
ceosal1 %>%  select(salary,roe) %>% head(3)
```

```
##   salary  roe
## 1   1095 14.1
## 2   1001 10.9
## 3   1122 23.5
```

- Para o CEO 1, temos um salário anual de 1095.000 USD e um retorno médio sobre o patrimônio líquido de 14.1%
- Para o CEO 2, temos um salário anual de 1001.000 USD e um retorno médio sobre o patrimônio líquido de 10.9%

Exemplos no R: salarios CEO

```
ceosal1 %>% select(salary,roe) %>% summary()
```

```
##      salary            roe       
##  Min.   :  223   Min.   : 0.50  
##  1st Qu.:  736   1st Qu.:12.40  
##  Median : 1039   Median :15.50  
##  Mean   : 1281   Mean   :17.18  
##  3rd Qu.: 1407   3rd Qu.:20.00  
##  Max.   :14822   Max.   :56.30  
```

Exemplos no R: salarios CEO

$$\hat{\beta}_1 = \frac{cov(x, y)}{\hat{\sigma}_x} \quad \text{e} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

```
# betas (hat)
CEOSAL1 <- ceosal1 %>%  select(salary,roe)
b1_hat = cov(CEOSAL1)[1,2]/var(CEOSAL1$roe)
b0_hat = mean(CEOSAL1$salary) - b1_hat*mean(CEOSAL1$roe)
c(b0_hat,b1_hat)
```

```
## [1] 963.19134  18.50119
```

Ejercicio:

Calcule $\hat{\beta}_1$ utilizando a fórmula $\hat{\beta}_1 = \hat{\rho}_{xy} \frac{\hat{\sigma}_y}{\hat{\sigma}_x}$. Os números batem?

Exemplos no R: salarios CEO

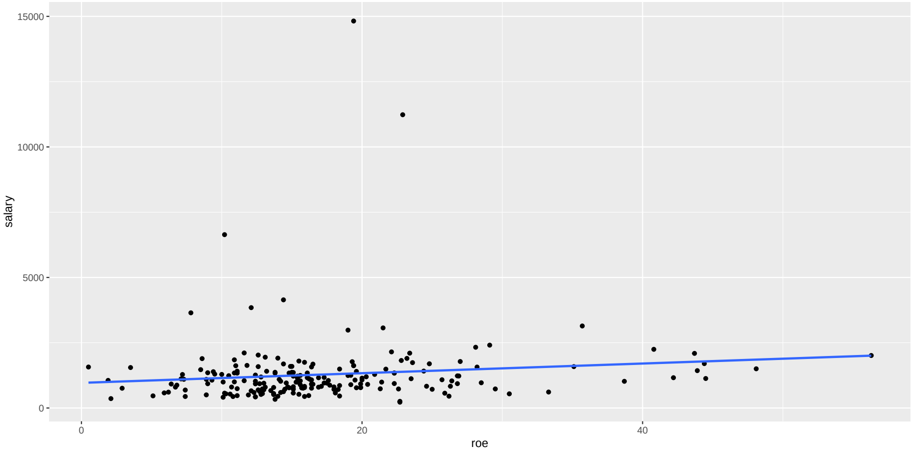
```
lm(salary~roe, data = CEOSAL1)
```

```
##
## Call:
## lm(formula = salary ~ roe, data = CEOSAL1)
##
## Coefficients:
## (Intercept)          roe
##       963.2         18.5
```

$$\widehat{\text{salario}} = 963.2 + 18.5 \text{ roe}$$

Se o *roe* aumentar 1%, espera-se que o salario anual aumente em 18.500 USD.

Exemplos no R: salarios CEO



Exemplos no R: vendas e gastos em publicidade

O dataset `Advertising` contém informações das vendas de um determinado produto em 200 lojas diferentes junto com o gasto em publicidade em três diferentes tipos de mídia: TV, rádio e jornal.

Exemplos no R: vendas e gastos em publicidade

O dataset `Advertising` contém informações das vendas de um determinado produto em 200 lojas diferentes junto com o gasto em publicidade em três diferentes tipos de mídia: TV, rádio e jornal.

Estamos interessados em aumentar as vendas do produto (nós não podemos diretamente aumentar as vendas, mas se existir uma relação entre vendas e gastos em publicidade, podemos construir um modelo que nos ajude a entender essa dinâmica e fazer uma melhor tomada de decisão sobre quanto gastar em publicidade em determinada mídia a fim de aumentar as vendas.)

Exemplos no R: vendas e gastos em publicidade

O dataset `Advertising` contém informações das vendas de um determinado produto em 200 lojas diferentes junto com o gasto em publicidade em três diferentes tipos de mídia: TV, rádio e jornal.

Estamos interessados em aumentar as vendas do produto (nós não podemos diretamente aumentar as vendas, mas se existir uma relação entre vendas e gastos em publicidade, podemos construir um modelo que nos ajude a entender essa dinâmica e fazer uma melhor tomada de decisão sobre quanto gastar em publicidade em determinada mídia a fim de aumentar as vendas.)

Para fins ilustrativos vamos apenas considerar o gasto em publicidade pela TV, ou seja, estamos interessados em construir um modelo da forma

$$\text{Sales} = \beta_0 + \beta_1 TV + u$$

Exemplos no R: vendas e gastos em publicidade

```r
uri <- "https://raw.githubusercontent.com/ctruciosm/ISLR/master/c
Advertising <- read.csv(uri)
head(Advertising)
```

```
##   X    TV Radio Newspaper Sales
## 1 1 230.1  37.8      69.2  22.1
## 2 2  44.5  39.3      45.1  10.4
## 3 3  17.2  45.9      69.3   9.3
## 4 4 151.5  41.3      58.5  18.5
## 5 5 180.8  10.8      58.4  12.9
## 6 6   8.7  48.9      75.0   7.2
```

Exemplos no R: vendas e gastos em publicidade

```r
uri <- "https://raw.githubusercontent.com/ctruciosm/ISLR/master/c
Advertising <- read.csv(uri)
head(Advertising)
```

```
##   X    TV Radio Newspaper Sales
## 1 1 230.1  37.8      69.2  22.1
## 2 2  44.5  39.3      45.1  10.4
## 3 3  17.2  45.9      69.3   9.3
## 4 4 151.5  41.3      58.5  18.5
## 5 5 180.8  10.8      58.4  12.9
## 6 6   8.7  48.9      75.0   7.2
```

Exemplos no R: vendas e gastos em publicidade

```
Advertising %>% select(Sales, TV) %>% summary()
```

```
##      Sales             TV        
##  Min.   : 1.60   Min.   :  0.70  
##  1st Qu.:10.38   1st Qu.: 74.38  
##  Median :12.90   Median :149.75  
##  Mean   :14.02   Mean   :147.04  
##  3rd Qu.:17.40   3rd Qu.:218.82  
##  Max.   :27.00   Max.   :296.40  
```

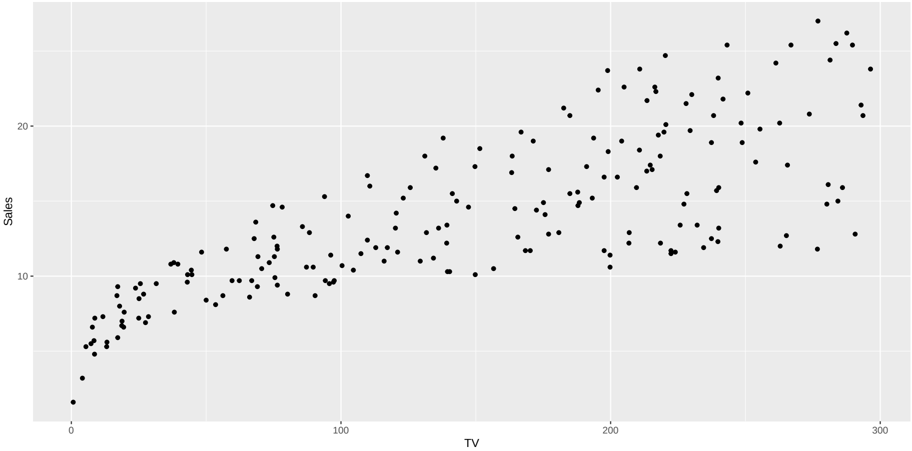
Exemplos no R: vendas e gastos em publicidade

```
Advertising %>% select(Sales, TV) %>% summary()
```

```
##      Sales             TV
##  Min.   : 1.60   Min.   :  0.70
##  1st Qu.:10.38   1st Qu.: 74.38
##  Median :12.90   Median :149.75
##  Mean   :14.02   Mean   :147.04
##  3rd Qu.:17.40   3rd Qu.:218.82
##  Max.   :27.00   Max.   :296.40
```

Faremos um grafico para analisar visualmente se existe alguma relação entre `Sales` e `TV`

Exemplos no R: vendas e gastos em publicidade



Exemplos no R: vendas e gastos em publicidade

```r
lm(Sales~TV, data = Advertising)
```

```
##
## Call:
## lm(formula = Sales ~ TV, data = Advertising)
##
## Coefficients:
## (Intercept)           TV
##     7.03259      0.04754
```

$$\widehat{\text{Sales}} = 7.03259 + 0.04754 \text{ TV}$$

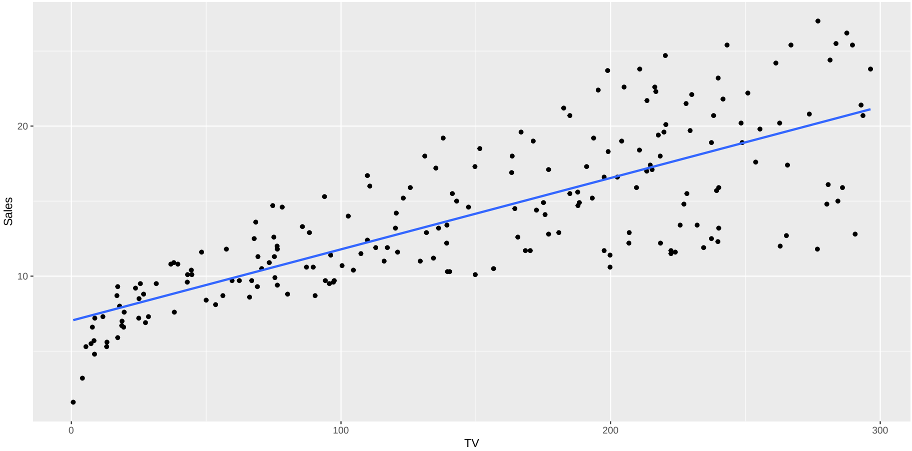
Exemplos no R: vendas e gastos em publicidade

```
lm(Sales~TV, data = Advertising)
```

```
##
## Call:
## lm(formula = Sales ~ TV, data = Advertising)
##
## Coefficients:
## (Intercept)           TV
##     7.03259      0.04754
```

$$\widehat{\text{Sales}} = 7.03259 + 0.04754\ \text{TV}$$

Exemplos no R: resultados eleitorais US



Comentários finais

- Na aula de hoje vimos o que é um modelo de regressão linear simples, por que os parâmetros β_0 e β_1 precisam ser estimados e vimos o métodod MQO que é um método para obter $\hat{\beta}_0$ e $\hat{\beta}_1$

Comentários finais

- Na aula de hoje vimos o que é um modelo de regressão linear simples, por que os parâmetros β_0 e β_1 precisam ser estimados e vimos o métodod MQO que é um método para obter $\hat{\beta}_0$ e $\hat{\beta}_1$
- Vimos dois exemplos, um deles utilizando o conjunto de dados `ceosal1` do pacote `wooldridge` e o conjunto de dados `Advertising` que foi importado ao R utilizando o endereço web dele.

Comentários finais

- Na aula de hoje vimos o que é um modelo de regressão linear simples, por que os parâmetros β_0 e β_1 precisam ser estimados e vimos o métodod MQO que é um método para obter $\hat{\beta}_0$ e $\hat{\beta}_1$
- Vimos dois exemplos, um deles utilizando o conjunto de dados `ceosal1` do pacote `wooldridge` e o conjunto de dados `Advertising` que foi importado ao R utilizando o endereço web dele.

Comentários finais

- Na aula de hoje vimos o que é um modelo de regressão linear simples, por que os parâmetros β_0 e β_1 precisam ser estimados e vimos o métodod MQO que é um método para obter $\hat{\beta}_0$ e $\hat{\beta}_1$
- Vimos dois exemplos, um deles utilizando o conjunto de dados `ceosal1` do pacote `wooldridge` e o conjunto de dados `Advertising` que foi importado ao R utilizando o endereço web dele.

Mas...

- Os modelos ajustados são bons ou ruins? como avaliamos a qualidade do ajuste dos modelos?

Comentários finais

- Na aula de hoje vimos o que é um modelo de regressão linear simples, por que os parâmetros β_0 e β_1 precisam ser estimados e vimos o métodod MQO que é um método para obter $\hat{\beta}_0$ e $\hat{\beta}_1$
- Vimos dois exemplos, um deles utilizando o conjunto de dados `ceosal1` do pacote `wooldridge` e o conjunto de dados `Advertising` que foi importado ao R utilizando o endereço web dele.

Mas...

- Os modelos ajustados são bons ou ruins? como avaliamos a qualidade do ajuste dos modelos?
- A interpretação dos $\hat{\beta}$'s é sempre igual?

Comentários finais

- Na aula de hoje vimos o que é um modelo de regressão linear simples, por que os parâmetros β_0 e β_1 precisam ser estimados e vimos o métodod MQO que é um método para obter $\hat{\beta}_0$ e $\hat{\beta}_1$
- Vimos dois exemplos, um deles utilizando o conjunto de dados `ceosal1` do pacote `wooldridge` e o conjunto de dados `Advertising` que foi importado ao R utilizando o endereço web dele.

Mas...

- Os modelos ajustados são bons ou ruins? como avaliamos a qualidade do ajuste dos modelos?
- A interpretação dos $\hat{\beta}$'s é sempre igual?
- Podemos *confiar* no valor obtido pelos $\hat{\beta}$'s?

Comentários finais

- Na aula de hoje vimos o que é um modelo de regressão linear simples, por que os parâmetros β_0 e β_1 precisam ser estimados e vimos o métodod MQO que é um método para obter $\hat{\beta}_0$ e $\hat{\beta}_1$
- Vimos dois exemplos, um deles utilizando o conjunto de dados `ceosal1` do pacote `wooldridge` e o conjunto de dados `Advertising` que foi importado ao R utilizando o endereço web dele.

Mas...

- Os modelos ajustados são bons ou ruins? como avaliamos a qualidade do ajuste dos modelos?
- A interpretação dos $\hat{\beta}$'s é sempre igual?
- Podemos *confiar* no valor obtido pelos $\hat{\beta}$'s?
- essas e outras respostas veremos nas próximas aulas

Leituras recomendadas

Leituras recomendadas

- Wooldridge, Jeffrey M. (2016). *Introdução à Econometria: Uma abordagem moderna*. Cengage Learning. – **Cap 1** e **Cap 2.1–2.3**
- James, G., Witten, D., Hastie, T., e Tibshirani, R. (2013). *An Introduction to Statistical Learning*. New York: Springer. **Chapter 3.1.1**
- Johnston, J. e Dinardo, J. (1997). *Econometric Methods*. 4ed, Mc Graw Hill. **Chapter 1.1–1.4.3**