CE310 - Modelos de Regressão Linear

Seleção de variáveis

Cesar Augusto Taconeli

26 de maio, 2025

Introdução

Introdução

Princípio da navalha de Occam

Dentre as várias explicações possíveis para um fenômeno, a mais simples é a melhor

Introdução

Princípio da navalha de Occam

Dentre as várias explicações possíveis para um fenômeno, a mais simples é a melhor

Fuechsel, técnico da IBM

Garbage in, garbage out

Introdução

- Neste módulo vamos tratar da seleção de variáveis explicativas para o ajuste de modelos de regressão linear.

Introdução

- Neste módulo vamos tratar da seleção de variáveis explicativas para o ajuste de modelos de regressão linear.

- O objetivo é identificar um modelo parcimonioso, capaz de proporcionar bom ajuste com a menor quantidade possível de parâmetros.

Introdução

- Neste módulo vamos tratar da seleção de variáveis explicativas para o ajuste de modelos de regressão linear.

- O objetivo é identificar um modelo parcimonioso, capaz de proporcionar bom ajuste com a menor quantidade possível de parâmetros.

- Diferentes métodos podem ser aplicados na seleção de um subconjunto "ótimo" de variáveis.

Introdução

- Neste módulo vamos tratar da seleção de variáveis explicativas para o ajuste de modelos de regressão linear.

- O objetivo é identificar um modelo parcimonioso, capaz de proporcionar bom ajuste com a menor quantidade possível de parâmetros.

- Diferentes métodos podem ser aplicados na seleção de um subconjunto "ótimo" de variáveis.

- Importante ter em mente que diferentes métodos de seleção podem produzir diferentes modelos (lembre-se: *"All models are wrong but some are useful"*).

Introdução

- Por que não incluir todas as variáveis no modelo?

Introdução

- Por que não incluir todas as variáveis no modelo?

1. Um dos objetivos principais da análise de regressão é explicar a relação entre as variáveis de maneira simples e interpretável;

Introdução

- Por que não incluir todas as variáveis no modelo?

1. Um dos objetivos principais da análise de regressão é explicar a relação entre as variáveis de maneira simples e interpretável;

2. Quanto maior o número de parâmetros no modelo, menos graus de liberdade para os resíduos, menor precisão para as inferências;

Introdução

- Por que não incluir todas as variáveis no modelo?

1. Um dos objetivos principais da análise de regressão é explicar a relação entre as variáveis de maneira simples e interpretável;

2. Quanto maior o número de parâmetros no modelo, menos graus de liberdade para os resíduos, menor precisão para as inferências;

3. Quanto maior o número de variáveis incluídas no modelo, maior a possibilidade de multicolinearidade;

Introdução

- Por que não incluir todas as variáveis no modelo?

1. Um dos objetivos principais da análise de regressão é explicar a relação entre as variáveis de maneira simples e interpretável;

2. Quanto maior o número de parâmetros no modelo, menos graus de liberdade para os resíduos, menor precisão para as inferências;

3. Quanto maior o número de variáveis incluídas no modelo, maior a possibilidade de multicolinearidade;

4. Quanto mais complexo (parametrizado) o modelo, melhor o ajuste da amostra, mas menor seu poder de generalização (e menor o poder preditivo).

Como proceder a seleção do modelo?

Como proceder a seleção do modelo?

- Antes de aplicar qualquer método analítico para seleção de variáveis, é recomendável fazer uma pré-triagem de variáveis, buscando eliminar variáveis que:

Como proceder a seleção do modelo?

- Antes de aplicar qualquer método analítico para seleção de variáveis, é recomendável fazer uma pré-triagem de variáveis, buscando eliminar variáveis que:
  - Sejam redundantes;

Como proceder a seleção do modelo?

- Antes de aplicar qualquer método analítico para seleção de variáveis, é recomendável fazer uma pré-triagem de variáveis, buscando eliminar variáveis que:
    - Sejam redundantes;
    - Apresentem inconsistências;

Como proceder a seleção do modelo?

- Antes de aplicar qualquer método analítico para seleção de variáveis, é recomendável fazer uma pré-triagem de variáveis, buscando eliminar variáveis que:
  - Sejam redundantes;
  - Apresentem inconsistências;
  - Estejam fora do contexto do estudo;

Como proceder a seleção do modelo?

- Antes de aplicar qualquer método analítico para seleção de variáveis, é recomendável fazer uma pré-triagem de variáveis, buscando eliminar variáveis que:
    - Sejam redundantes;
    - Apresentem inconsistências;
    - Estejam fora do contexto do estudo;
    - Apresentem elevada taxa de dados perdidos (*missing data*)...

Seleção de variáveis - princípio hierárquico

- Certos modelos apresentam alguma hierarquia natural quanto às variáveis que os compõem.

Seleção de variáveis - princípio hierárquico

- Certos modelos apresentam alguma hierarquia natural quanto às variáveis que os compõem.

- Nesses casos, a seleção das variáveis a compor o modelo deve respeitar essa hierarquia.

Seleção de variáveis - princípio hierárquico

- Certos modelos apresentam alguma hierarquia natural quanto às variáveis que os compõem.

- Nesses casos, a seleção das variáveis a compor o modelo deve respeitar essa hierarquia.

- Um dos casos mais típicos é o **modelo polinomial**.

Seleção de variáveis - Modelos hierárquicos

- Considere o modelo polinomial de segunda ordem:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon.$$

Seleção de variáveis - Modelos hierárquicos

- Considere o modelo polinomial de segunda ordem:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon.$$

- Em modelos polinomiais, a remoção das variáveis sempre deve iniciar pelos termos de maior ordem.

Seleção de variáveis - Modelos hierárquicos

- Considere o modelo polinomial de segunda ordem:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon.$$

- Em modelos polinomiais, a remoção das variáveis sempre deve iniciar pelos termos de maior ordem.

- Para fins de ilustração, considere o modelo quadrático com a remoção do termo linear:

$$y = \beta_0 + \beta_2 x^2 + \epsilon.$$

Seleção de variáveis - Modelos hierárquicos

- Suponha que seja feita uma mudança de escala, trocando x por $x + c$. O modelo ficaria dado por:

$$y = \beta_0 + \beta_2 c^2 + 2\beta_2 cx + \beta_2 x^2 + \epsilon.$$

Seleção de variáveis - Modelos hierárquicos

- Suponha que seja feita uma mudança de escala, trocando x por $x + c$. O modelo ficaria dado por:

$$y = \beta_0 + \beta_2 c^2 + 2\beta_2 c x + \beta_2 x^2 + \epsilon.$$

- Observe que o termo linear reaparece mediante simples mudança de escala.

Seleção de variáveis - Modelos hierárquicos

- Suponha que seja feita uma mudança de escala, trocando x por $x + c$. O modelo ficaria dado por:

$$y = \beta_0 + \beta_2 c^2 + 2\beta_2 c x + \beta_2 x^2 + \epsilon.$$

- Observe que o termo linear reaparece mediante simples mudança de escala.

- Em geral, é indesejado que a estrutura do modelo (e as interpretações) sejam alteradas devido à simples mudança de escala dos dados.

Seleção de variáveis - Modelos hierárquicos

- Da mesma forma o princípio hierárquico se aplica a modelos com termos de interação entre variáveis.

Seleção de variáveis - Modelos hierárquicos

- Da mesma forma o princípio hierárquico se aplica a modelos com termos de interação entre variáveis.

- Considere o modelo de regressão com interação de segunda ordem:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \epsilon.$$

Seleção de variáveis - Modelos hierárquicos

- Da mesma forma o princípio hierárquico se aplica a modelos com termos de interação entre variáveis.

- Considere o modelo de regressão com interação de segunda ordem:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \epsilon.$$

- Neste caso, não é recomendável remover o termo de interação $(x_1 x_2)$ sem a remoção dos demais termos quadráticos $(x_1^2,\, x_2^2)$;

Seleção de variáveis - Modelos hierárquicos

- Da mesma forma o princípio hierárquico se aplica a modelos com termos de interação entre variáveis.

- Considere o modelo de regressão com interação de segunda ordem:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \epsilon.$$

- Neste caso, não é recomendável remover o termo de interação ($x_1 x_2$) sem a remoção dos demais termos quadráticos (x_1^2, x_2^2);

- A remoção conjunta de $x_1 x_2$, x_1^2 e x_2^2 é válida, e tem por objetivo comparar os ajustes linear e quadrático.

Critérios para comparação de modelos

Critérios para comparação de modelos

- No processo de seleção de covariáveis, diferentes critérios podem ser usados para comparar os modelos produzidos. Alguns deles são descritos na sequência.

Critérios para comparação de modelos

- No processo de seleção de covariáveis, diferentes critérios podem ser usados para comparar os modelos produzidos. Alguns deles são descritos na sequência.

- **Coeficiente de deteminação -** O coeficiente de determinação de um modelo é definido como:

$$R^2 = \frac{\text{SQ}_{\text{Total}} - \text{SQ}_{\text{Res}}}{\text{SQ}_{\text{Total}}} = 1 - \frac{\text{SQ}_{\text{Res}}}{\text{SQ}_{\text{Total}}},$$

em que

$$\text{SQ}_{\text{Total}} = \sum_{i=1}^{n}(y_i - \bar{y})^2 \quad \text{e} \quad \text{SQ}_{\text{Res}} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

são as somas de quadrados total e atribuída aos resíduos, respectivamente.

- O coeficiente de determinação expressa a proporção da variabilidade total explicada pelo modelo ajustado.

Critérios para comparação de modelos

- O coeficiente de determinação expressa a proporção da variabilidade total explicada pelo modelo ajustado.

- O valor de R^2 nunca decresce à medida que novas covariáveis são incluídas no modelo.

Critérios para comparação de modelos

- O coeficiente de determinação expressa a proporção da variabilidade total explicada pelo modelo ajustado.
- O valor de R^2 nunca decresce à medida que novas covariáveis são incluídas no modelo.
- Assim, o modelo que produz maior valor de R^2 incluirá, necessariamente, todas as variáveis disponíveis.

Critérios para comparação de modelos

- **Coeficiente de deteminação ajustado -** O coeficiente de determinação ajustado (ou simplesmente R^2 ajustado) é definido por:

$$R^2_{Aj} = 1 - \left(\frac{n-1}{n-p}\right)(1 - R^2),$$

em que n e p são o número de observações e o número de parâmetros do modelo.

Critérios para comparação de modelos

- **Coeficiente de deteminação ajustado -** O coeficiente de determinação ajustado (ou simplesmente R^2 ajustado) é definido por:

$$R^2_{Aj} = 1 - \left(\frac{n-1}{n-p}\right)(1 - R^2),$$

em que n e p são o número de observações e o número de parâmetros do modelo.

- Diferentemente do que ocorre para R^2, o valor de R^2_{Aj} pode não aumentar mediante inclusão de novas variáveis ao modelo. Neste caso, deve-se optar por modelos com maiores valores de R^2_{Aj}.

Critérios para comparação de modelos

- **Quadrado médio de resíduos**, definido por:

$$\text{QM}_{\text{Res}} = \frac{\text{SQ}_{\text{Res}}}{n-p} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-p}.$$

também pode ser usado para comparação e seleção de modelos de regressão.

Critérios para comparação de modelos

- **Quadrado médio de resíduos**, definido por:

$$\mathrm{QM}_{\mathrm{Res}} = \frac{\mathrm{SQ}_{\mathrm{Res}}}{n-p} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-p}.$$

também pode ser usado para comparação e seleção de modelos de regressão.

- Deve-se optar por modelos com menores valores para $\mathrm{QM}_{\mathrm{Res}}$.

Critérios para comparação de modelos

- **Quadrado médio de resíduos**, definido por:

$$\text{QM}_{\text{Res}} = \frac{\text{SQ}_{\text{Res}}}{n-p} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-p}.$$

também pode ser usado para comparação e seleção de modelos de regressão.

- Deve-se optar por modelos com menores valores para QM_{Res}.

- Pode-se mostrar que minimizar QM_{Res} é equivalente a maximizar R^2_{Aj}, de forma que os dois critérios conduzem à seleção do mesmo conjunto de variáveis.

Critérios para comparação de modelos

- A **estatística PRESS** permite avaliar a qualidade preditiva dos modelos de regressão, sendo definida por:

$$\text{PRESS} = \sum_{i=1}^{n}\left[y_i - \hat{y}_{(i)}\right]^2 = \sum_{i=1}^{n}\left(\frac{r_i}{1-h_{ii}}\right)^2,$$

em que $\hat{y}_{(i)}$ é obtido com base no modelo ajustado apenas excluindo a $i-$ésima observação, ou seja, com as demais $n-1$ observações $(i = 1, 2, ..., n)$.

Critérios para comparação de modelos

- A **estatística PRESS** permite avaliar a qualidade preditiva dos modelos de regressão, sendo definida por:

$$\text{PRESS} = \sum_{i=1}^{n} \left[y_i - \hat{y}_{(i)}\right]^2 = \sum_{i=1}^{n} \left(\frac{r_i}{1 - h_{ii}}\right)^2,$$

em que $\hat{y}_{(i)}$ é obtido com base no modelo ajustado apenas excluindo a $i-$ésima observação, ou seja, com as demais $n-1$ observações $(i = 1, 2, ..., n)$.

- Menores valores da estatística PRESS indicam modelos com melhor poder preditivo.

Critérios para comparação de modelos

- O critério de informação de Akaike (*Akaike Information Criterion*), ou simplesmente AIC, é definido por:

$$\text{AIC} = -2l(\hat{\theta}) + 2p,$$

em que $l(\hat{\theta})$ é a log-verossimilhança maximizada do modelo (calculada com base nos estimadores de máxima verossimilhança dos parâmetros) e p o número de parâmetros do modelo.

Critérios para comparação de modelos

- O critério de informação de Akaike (*Akaike Information Criterion*), ou simplesmente AIC, é definido por:

$$\text{AIC} = -2l(\hat{\theta}) + 2p,$$

em que $l(\hat{\theta})$ é a log-verossimilhança maximizada do modelo (calculada com base nos estimadores de máxima verossimilhança dos parâmetros) e p o número de parâmetros do modelo.

- No caso de um modelo de regressão linear temos:

$$\text{AIC} = -n\ln(\text{SQ}_{\text{Res}}/n) + 2p.$$

Critérios para comparação de modelos

- O componente $2p$, na expressão do AIC, atua como *termo de penalização* associado à complexidade (número de parâmetros) do modelo.

Critérios para comparação de modelos

- O componente $2p$, na expressão do AIC, atua como *termo de penalização* associado à complexidade (número de parâmetros) do modelo.

- Um critério alternativo ao AIC é o Critério de Informação Bayesiano (BIC), definido, para um modelo de regressão linear, por:

$$\text{BIC} = -n\ln(\text{SQ}_{\text{Res}}/n) + \ln(n)p.$$

Critérios para comparação de modelos

- O componente $2p$, na expressão do AIC, atua como *termo de penalização* associado à complexidade (número de parâmetros) do modelo.

- Um critério alternativo ao AIC é o Critério de Informação Bayesiano (BIC), definido, para um modelo de regressão linear, por:

$$\text{BIC} = -n \ln(\text{SQ}_{\text{Res}}/n) + \ln(n)p.$$

- O BIC penaliza mais fortemente a complexidade do modelo que o AIC ao substituir p por $\ln(n)$ como fator de penalização.

Critérios para comparação de modelos

- O componente $2p$, na expressão do AIC, atua como *termo de penalização* associado à complexidade (número de parâmetros) do modelo.

- Um critério alternativo ao AIC é o Critério de Informação Bayesiano (BIC), definido, para um modelo de regressão linear, por:

$$\text{BIC} = -n\ln(\text{SQ}_{\text{Res}}/n) + \ln(n)p.$$

- O BIC penaliza mais fortemente a complexidade do modelo que o AIC ao substituir p por $\ln(n)$ como fator de penalização.

- Devemos selecionar modelos com menores valores de AIC (ou BIC).

Algoritmos de seleção de variáveis

Algoritmos de seleção de variáveis

- Um primeiro algoritmo de seleção de modelos, baseados em critérios de qualidade de ajuste, é o método **todas as regressões possíveis**.

Algoritmos de seleção de variáveis

- Um primeiro algoritmo de seleção de modelos, baseados em critérios de qualidade de ajuste, é o método **todas as regressões possíveis**.

- Suponha uma análise de regressão com k variáveis.

Algoritmos de seleção de variáveis

- Um primeiro algoritmo de seleção de modelos, baseados em critérios de qualidade de ajuste, é o método **todas as regressões possíveis**.

- Suponha uma análise de regressão com k variáveis.

- Neste caso, devemos ajustar todos os possíveis modelos de regressão com $j \leq k$ variáveis explicativas, para $j = 0, 1, 2, ..., k$;

Algoritmos de seleção de variáveis

- Um primeiro algoritmo de seleção de modelos, baseados em critérios de qualidade de ajuste, é o método **todas as regressões possíveis**.

- Suponha uma análise de regressão com k variáveis.

- Neste caso, devemos ajustar todos os possíveis modelos de regressão com $j \leq k$ variáveis explicativas, para $j = 0, 1, 2, ..., k$;

- Para cada modelo de regressão ajustado calcula-se o valor do critério de seleção escolhido (AIC, BIC ou $R^2_{Aj} \dots$).

Algoritmos de seleção de variáveis

- Um primeiro algoritmo de seleção de modelos, baseados em critérios de qualidade de ajuste, é o método **todas as regressões possíveis**.

- Suponha uma análise de regressão com k variáveis.

- Neste caso, devemos ajustar todos os possíveis modelos de regressão com $j \leq k$ variáveis explicativas, para $j = 0, 1, 2, ..., k$;

- Para cada modelo de regressão ajustado calcula-se o valor do critério de seleção escolhido (AIC, BIC ou $R^2_{Aj} \dots$).

- O modelo selecionado será quele que apresentar o valor ótimo para o critério adotado (maior R^2_{Aj}, menor AIC,$\dots$).

Todas as regressões possíveis

- O método baseado em todas as regressões possíveis torna-se inviável mesmo para um número moderado de variáveis.

Todas as regressões possíveis

- O método baseado em todas as regressões possíveis torna-se inviável mesmo para um número moderado de variáveis.

- Para k variáveis o número de regressões possíveis é 2^k. Para $k = 30$, por exemplo, teríamos 1.073.741.824 modelos possíveis!

Todas as regressões possíveis

- O método baseado em todas as regressões possíveis torna-se inviável mesmo para um número moderado de variáveis.

- Para k variáveis o número de regressões possíveis é 2^k. Para $k = 30$, por exemplo, teríamos 1.073.741.824 modelos possíveis!

- Como alternativa ao método de todas as regressões possíveis podemos usar os algoritmos backward, forward ou stepwise.

Todas as regressões possíveis

- O método baseado em todas as regressões possíveis torna-se inviável mesmo para um número moderado de variáveis.

- Para k variáveis o número de regressões possíveis é 2^k. Para $k = 30$, por exemplo, teríamos 1.073.741.824 modelos possíveis!

- Como alternativa ao método de todas as regressões possíveis podemos usar os algoritmos backward, forward ou stepwise.

- Esses algoritmos visam determinar um modelo subótimo, obtido mediante um número consideravelmente menor de modelos ajustados e comparados.

Seleção de variáveis - Método backward

1. Ajuste o modelo de regressão com todas as k covariáveis disponíveis;

Seleção de variáveis - Método backward

1. Ajuste o modelo de regressão com todas as k covariáveis disponíveis;

2. Remova as variáveis do modelo uma por vez e calcule o valor do critério de qualidade de ajuste adotado (digamos o AIC, para ilustração) para cada modelo;

Seleção de variáveis - Método backward

1. Ajuste o modelo de regressão com todas as k covariáveis disponíveis;

2. Remova as variáveis do modelo uma por vez e calcule o valor do critério de qualidade de ajuste adotado (digamos o AIC, para ilustração) para cada modelo;

3. Remova do modelo (em definitivo) a variável cuja exclusão resultar em menor AIC;

Seleção de variáveis - Método backward

1. Ajuste o modelo de regressão com todas as k covariáveis disponíveis;

2. Remova as variáveis do modelo uma por vez e calcule o valor do critério de qualidade de ajuste adotado (digamos o AIC, para ilustração) para cada modelo;

3. Remova do modelo (em definitivo) a variável cuja exclusão resultar em menor AIC;

4. Os passos 1 a 3 são repetidos para o novo modelo (com a remoção da variável) e o processo continua até o momento em que nenhuma remoção produzir um ajuste com menor AIC que o modelo do passo anterior.

Seleção de variáveis - Método forward

1. Ajuste o modelo nulo (sem variáveis explicativas);

Seleção de variáveis - Método forward

1. Ajuste o modelo nulo (sem variáveis explicativas);
2. Adicione as variáveis ao modelo uma de cada vez e calcule o valor do critério de qualidade de ajuste adotado (digamos o AIC, para ilustração) para cada modelo;

Seleção de variáveis - Método forward

1. Ajuste o modelo nulo (sem variáveis explicativas);
2. Adicione as variáveis ao modelo uma de cada vez e calcule o valor do critério de qualidade de ajuste adotado (digamos o AIC, para ilustração) para cada modelo;
3. Inclua no modelo (em definitivo) a variável cuja exclusão resultar em menor AIC;

Seleção de variáveis - Método forward

1. Ajuste o modelo nulo (sem variáveis explicativas);
2. Adicione as variáveis ao modelo uma de cada vez e calcule o valor do critério de qualidade de ajuste adotado (digamos o AIC, para ilustração) para cada modelo;
3. Inclua no modelo (em definitivo) a variável cuja exclusão resultar em menor AIC;
4. Os passos 1 a 3 são repetidos para o novo modelo (com a inclusão da variável) e o processo continua até o momento em que nenhuma inclusão produzir um ajuste com menor AIC que o modelo do passo anterior.

Seleção de variáveis - Método combinado

1. Ajuste o modelo de regressão com todas as k covariáveis disponíveis. Proceda como descrito no algoritmo backward, mas...

Seleção de variáveis - Método combinado

1. Ajuste o modelo de regressão com todas as k covariáveis disponíveis. Proceda como descrito no algoritmo backward, mas...

2. A cada passo do algoritmo, avalie tanto a inclusão variáveis não inseridas quanto a exclusão das variáveis que compõem o modelo.

Seleção de variáveis - Método combinado

1. Ajuste o modelo de regressão com todas as k covariáveis disponíveis. Proceda como descrito no algoritmo backward, mas...

2. A cada passo do algoritmo, avalie tanto a inclusão variáveis não inseridas quanto a exclusão das variáveis que compõem o modelo.

3. Inclua (exclua) no modelo a variável cuja inclusão (exclusão) resultar em menor AIC;

Seleção de variáveis - Método combinado

1. Ajuste o modelo de regressão com todas as k covariáveis disponíveis. Proceda como descrito no algoritmo backward, mas...

2. A cada passo do algoritmo, avalie tanto a inclusão variáveis não inseridas quanto a exclusão das variáveis que compõem o modelo.

3. Inclua (exclua) no modelo a variável cuja inclusão (exclusão) resultar em menor AIC;

4. Os passos 1 a 3 são repetidos para o novo modelo (com a inclusão (exclusão) da variável) e o processo continua até o momento em que nenhuma inclusão (ou exclusão) produzir um ajuste com menor AIC que o modelo do passo anterior.

Exemplo- Desempenho de atletas de remo

Exemplo- Desempenho de atletas de remo

- Nesta aplicação vamos analisar dados de 90 atletas de remo, com idades entre 18 e 24 anos.

Exemplo- Desempenho de atletas de remo

- Nesta aplicação vamos analisar dados de 90 atletas de remo, com idades entre 18 e 24 anos.

- A variável resposta é `racetime`, o tempo gasto numa prova de 2km realizada num simulador.

Exemplo- Desempenho de atletas de remo

- Nesta aplicação vamos analisar dados de 90 atletas de remo, com idades entre 18 e 24 anos.
- A variável resposta é `racetime`, o tempo gasto numa prova de 2km realizada num simulador.
- As variáveis explicativas são as seguintes:

Exemplo- Desempenho de atletas de remo

- Nesta aplicação vamos analisar dados de 90 atletas de remo, com idades entre 18 e 24 anos.

- A variável resposta é `racetime`, o tempo gasto numa prova de 2km realizada num simulador.

- As variáveis explicativas são as seguintes:
  - `tall`- altura (cm)

Exemplo- Desempenho de atletas de remo

- Nesta aplicação vamos analisar dados de 90 atletas de remo, com idades entre 18 e 24 anos.

- A variável resposta é `racetime`, o tempo gasto numa prova de 2km realizada num simulador.

- As variáveis explicativas são as seguintes:
  - `tall`- altura (cm)
  - `Weight`- peso (kg)

Exemplo- Desempenho de atletas de remo

- Nesta aplicação vamos analisar dados de 90 atletas de remo, com idades entre 18 e 24 anos.

- A variável resposta é `racetime`, o tempo gasto numa prova de 2km realizada num simulador.

- As variáveis explicativas são as seguintes:
  - `tall`- altura (cm)
  - `Weight`- peso (kg)
  - `arm_span`- envergadura (cm)

Exemplo- Desempenho de atletas de remo

- Nesta aplicação vamos analisar dados de 90 atletas de remo, com idades entre 18 e 24 anos.
- A variável resposta é `racetime`, o tempo gasto numa prova de 2km realizada num simulador.
- As variáveis explicativas são as seguintes:
  - `tall`- altura (cm)
  - `Weight`- peso (kg)
  - `arm_span`- envergadura (cm)
  - `flex_arm`- circunferência do braço (cm)

Exemplo- Desempenho de atletas de remo

- Nesta aplicação vamos analisar dados de 90 atletas de remo, com idades entre 18 e 24 anos.

- A variável resposta é `racetime`, o tempo gasto numa prova de 2km realizada num simulador.

- As variáveis explicativas são as seguintes:
  - `tall`- altura (cm)
  - `Weight`- peso (kg)
  - `arm_span`- envergadura (cm)
  - `flex_arm`- circunferência do braço (cm)
  - `thigh_ci`- circunferência da coxa (cm)

Exemplo- Desempenho de atletas de remo

- Relação de variáveis explicativas (continuação)

Exemplo- Desempenho de atletas de remo

- Relação de variáveis explicativas (continuação)
  - `calf_ci`- circunferência da panturrilha (cm)

Exemplo- Desempenho de atletas de remo

- Relação de variáveis explicativas (continuação)
  - `calf_ci`- circunferência da panturrilha (cm)
  - `tricep`- prega do triceps (mm)

Exemplo- Desempenho de atletas de remo

- Relação de variáveis explicativas (continuação)
  - `calf_ci`- circunferência da panturrilha (cm)
  - `tricep`- prega do triceps (mm)
  - `biceps`- prega do biceps (mm)

Exemplo- Desempenho de atletas de remo

- Relação de variáveis explicativas (continuação)
  - `calf_ci`- circunferência da panturrilha (cm)
  - `tricep`- prega do triceps (mm)
  - `biceps`- prega do biceps (mm)
  - `thigh`- prega da coxa

Exemplo- Desempenho de atletas de remo

- Relação de variáveis explicativas (continuação)
  - `calf_ci`- circunferência da panturrilha (cm)
  - `tricep`- prega do triceps (mm)
  - `biceps`- prega do biceps (mm)
  - `thigh`- prega da coxa
  - `est_ffm`- massa livre de gordura estimada (kg)

Exemplo- Desempenho de atletas de remo

- Relação de variáveis explicativas (continuação)
  - `calf_ci`- circunferência da panturrilha (cm)
  - `tricep`- prega do triceps (mm)
  - `biceps`- prega do biceps (mm)
  - `thigh`- prega da coxa
  - `est_ffm`- massa livre de gordura estimada (kg)
  - `est_fm`- massa gorda estimada (kg)

Exemplo- Desempenho de atletas de remo

- Relação de variáveis explicativas (continuação)
  - **calf_ci**- circunferência da panturrilha (cm)
  - **tricep**- prega do triceps (mm)
  - **biceps**- prega do biceps (mm)
  - **thigh**- prega da coxa
  - **est_ffm**- massa livre de gordura estimada (kg)
  - **est_fm**- massa gorda estimada (kg)
  - **best_snr**- teste de alcance sentada (cm)

Exemplo- Desempenho de atletas de remo

- Relação de variáveis explicativas (continuação)
  - `calf_ci`- circunferência da panturrilha (cm)
  - `tricep`- prega do triceps (mm)
  - `biceps`- prega do biceps (mm)
  - `thigh`- prega da coxa
  - `est_ffm`- massa livre de gordura estimada (kg)
  - `est_fm`- massa gorda estimada (kg)
  - `best_snr`- teste de alcance sentada (cm)
  - `best_vj`- teste de salto vertical (cm)

Exemplo- Desempenho de atletas de remo

- Relação de variáveis explicativas (continuação)
  - `calf_ci`- circunferência da panturrilha (cm)
  - `tricep`- prega do triceps (mm)
  - `biceps`- prega do biceps (mm)
  - `thigh`- prega da coxa
  - `est_ffm`- massa livre de gordura estimada (kg)
  - `est_fm`- massa gorda estimada (kg)
  - `best_snr`- teste de alcance sentada (cm)
  - `best_vj`- teste de salto vertical (cm)
  - `leg_powe`- teste de pernas (kpm/min)

Exemplo- Desempenho de atletas de remo

- Relação de variáveis explicativas (continuação)
  - `calf_ci`- circunferência da panturrilha (cm)
  - `tricep`- prega do triceps (mm)
  - `biceps`- prega do biceps (mm)
  - `thigh`- prega da coxa
  - `est_ffm`- massa livre de gordura estimada (kg)
  - `est_fm`- massa gorda estimada (kg)
  - `best_snr`- teste de alcance sentada (cm)
  - `best_vj`- teste de salto vertical (cm)
  - `leg_powe`- teste de pernas (kpm/min)
  - `endo`- endomorfia

Exemplo- Desempenho de atletas de remo

- Relação de variáveis explicativas (continuação)
  - `calf_ci`- circunferência da panturrilha (cm)
  - `tricep`- prega do triceps (mm)
  - `biceps`- prega do biceps (mm)
  - `thigh`- prega da coxa
  - `est_ffm`- massa livre de gordura estimada (kg)
  - `est_fm`- massa gorda estimada (kg)
  - `best_snr`- teste de alcance sentada (cm)
  - `best_vj`- teste de salto vertical (cm)
  - `leg_powe`- teste de pernas (kpm/min)
  - `endo`- endomorfia
  - `meso`- mesomorfia

Exemplo- Desempenho de atletas de remo

- Relação de variáveis explicativas (continuação)
  - `calf_ci`- circunferência da panturrilha (cm)
  - `tricep`- prega do triceps (mm)
  - `biceps`- prega do biceps (mm)
  - `thigh`- prega da coxa
  - `est_ffm`- massa livre de gordura estimada (kg)
  - `est_fm`- massa gorda estimada (kg)
  - `best_snr`- teste de alcance sentada (cm)
  - `best_vj`- teste de salto vertical (cm)
  - `leg_powe`- teste de pernas (kpm/min)
  - `endo`- endomorfia
  - `meso`- mesomorfia
  - `ecto`- ectomorfia

Exemplo- Ergonomia de motoristas

Exemplo- Ergonomia de motoristas

- Nesta aplicação vamos analisar dados de ergonomia de motorista, apresentados na base de dados `seatpos` da biblioteca `faraway` do `R`.

Exemplo- Ergonomia de motoristas

- Nesta aplicação vamos analisar dados de ergonomia de motorista, apresentados na base de dados `seatpos` da biblioteca `faraway` do `R`.

- A variável resposta é `hipcenter`, a distância do ponto médio dos quadris a um ponto fixo no carro (em mm).

Exemplo- Ergonomia de motoristas

- Nesta aplicação vamos analisar dados de ergonomia de motorista, apresentados na base de dados `seatpos` da biblioteca `faraway` do `R`.

- A variável resposta é `hipcenter`, a distância do ponto médio dos quadris a um ponto fixo no carro (em mm).

- As variáveis explicativas são as seguintes:

Exemplo- Ergonomia de motoristas

- Nesta aplicação vamos analisar dados de ergonomia de motorista, apresentados na base de dados `seatpos` da biblioteca `faraway` do `R`.

- A variável resposta é `hipcenter`, a distância do ponto médio dos quadris a um ponto fixo no carro (em mm).

- As variáveis explicativas são as seguintes:
  - `Age`- idade (anos)

Exemplo- Ergonomia de motoristas

- Nesta aplicação vamos analisar dados de ergonomia de motorista, apresentados na base de dados `seatpos` da biblioteca `faraway` do `R`.

- A variável resposta é `hipcenter`, a distância do ponto médio dos quadris a um ponto fixo no carro (em mm).

- As variáveis explicativas são as seguintes:
  - `Age`- idade (anos)
  - `Weight`- peso (libras)

Exemplo- Ergonomia de motoristas

- Nesta aplicação vamos analisar dados de ergonomia de motorista, apresentados na base de dados `seatpos` da biblioteca `faraway` do `R`.

- A variável resposta é `hipcenter`, a distância do ponto médio dos quadris a um ponto fixo no carro (em mm).

- As variáveis explicativas são as seguintes:
  - `Age`- idade (anos)
  - `Weight`- peso (libras)
  - `HtShoes`- altura calçado (cm)

Exemplo- Ergonomia de motoristas

- Nesta aplicação vamos analisar dados de ergonomia de motorista, apresentados na base de dados `seatpos` da biblioteca `faraway` do `R`.

- A variável resposta é `hipcenter`, a distância do ponto médio dos quadris a um ponto fixo no carro (em mm).

- As variáveis explicativas são as seguintes:
    - `Age`- idade (anos)
    - `Weight`- peso (libras)
    - `HtShoes`- altura calçado (cm)
    - `Ht`- altura descalço (cm)

Exemplo- Ergonomia de motoristas

- Nesta aplicação vamos analisar dados de ergonomia de motorista, apresentados na base de dados `seatpos` da biblioteca `faraway` do `R`.

- A variável resposta é `hipcenter`, a distância do ponto médio dos quadris a um ponto fixo no carro (em mm).

- As variáveis explicativas são as seguintes:
  - `Age`- idade (anos)
  - `Weight`- peso (libras)
  - `HtShoes`- altura calçado (cm)
  - `Ht`- altura descalço (cm)
  - `Seated`- altura sentado (cm)

Exemplo- Ergonomia de motoristas

- Nesta aplicação vamos analisar dados de ergonomia de motorista, apresentados na base de dados `seatpos` da biblioteca `faraway` do `R`.

- A variável resposta é `hipcenter`, a distância do ponto médio dos quadris a um ponto fixo no carro (em mm).

- As variáveis explicativas são as seguintes:
  - `Age`- idade (anos)
  - `Weight`- peso (libras)
  - `HtShoes`- altura calçado (cm)
  - `Ht`- altura descalço (cm)
  - `Seated`- altura sentado (cm)
  - `Arm`- comprimento do antebraço (cm)

Exemplo- Ergonomia de motoristas

- Nesta aplicação vamos analisar dados de ergonomia de motorista, apresentados na base de dados `seatpos` da biblioteca `faraway` do `R`.

- A variável resposta é `hipcenter`, a distância do ponto médio dos quadris a um ponto fixo no carro (em mm).

- As variáveis explicativas são as seguintes:
  - `Age`- idade (anos)
  - `Weight`- peso (libras)
  - `HtShoes`- altura calçado (cm)
  - `Ht`- altura descalço (cm)
  - `Seated`- altura sentado (cm)
  - `Arm`- comprimento do antebraço (cm)
  - `Thigh`- comprimento da coxa (cm)

Exemplo- Ergonomia de motoristas

- Nesta aplicação vamos analisar dados de ergonomia de motorista, apresentados na base de dados `seatpos` da biblioteca `faraway` do `R`.

- A variável resposta é `hipcenter`, a distância do ponto médio dos quadris a um ponto fixo no carro (em mm).

- As variáveis explicativas são as seguintes:
    - `Age`- idade (anos)
    - `Weight`- peso (libras)
    - `HtShoes`- altura calçado (cm)
    - `Ht`- altura descalço (cm)
    - `Seated`- altura sentado (cm)
    - `Arm`- comprimento do antebraço (cm)
    - `Thigh`- comprimento da coxa (cm)
    - `Leg`- comprimento da panturrilha (cm)

Exercícios

Exercícios

- Resolva os exercícios da lista de exercícios relativa a este módulo, disponibilizada na página da disciplina.