CE310 - Modelos de Regressão Linear

Regressão linear com variáveis categóricas

Cesar Augusto Taconeli

28 de maio, 2025

Introdução

Introdução

- Neste módulo vamos tratar da inclusão de variáveis explicativas categóricas (fatores) em modelos de regressão linear.

Introdução

- Neste módulo vamos tratar da inclusão de variáveis explicativas categóricas (fatores) em modelos de regressão linear.

- Alguns exemplos de variáveis categóricas:

- Neste módulo vamos tratar da inclusão de variáveis explicativas categóricas (fatores) em modelos de regressão linear.

- Alguns exemplos de variáveis categóricas:
    - Sexo (masculino ou feminino);

Introdução

- Neste módulo vamos tratar da inclusão de variáveis explicativas categóricas (fatores) em modelos de regressão linear.

- Alguns exemplos de variáveis categóricas:
  - Sexo (masculino ou feminino);
  - Estação do ano (primavera, verão, outono ou inverno);

Introdução

- Neste módulo vamos tratar da inclusão de variáveis explicativas categóricas (fatores) em modelos de regressão linear.

- Alguns exemplos de variáveis categóricas:
  - Sexo (masculino ou feminino);
  - Estação do ano (primavera, verão, outono ou inverno);
  - Categoria de cliente de banco (*platinum*, *gold*, *silver*,...);

Introdução

- Neste módulo vamos tratar da inclusão de variáveis explicativas categóricas (fatores) em modelos de regressão linear.

- Alguns exemplos de variáveis categóricas:
  - Sexo (masculino ou feminino);
  - Estação do ano (primavera, verão, outono ou inverno);
  - Categoria de cliente de banco (*platinum*, *gold*, *silver*,. . .);
  - Escolaridade (sem escolaridade, ensino primário, ensino secundário,. . .).

Introdução

- A forma usual de incorporar variáveis categóricas em um modelo de regressão é através de variáveis indicadoras (variáveis *dummy*).

Introdução

- A forma usual de incorporar variáveis categóricas em um modelo de regressão é através de variáveis indicadoras (variáveis *dummy*).

- Importante reforçar que vamos estudar a inclusão de **variáveis explicativas** categóricas no modelo de regressão linear.

Introdução

- A forma usual de incorporar variáveis categóricas em um modelo de regressão é através de variáveis indicadoras (variáveis *dummy*).

- Importante reforçar que vamos estudar a inclusão de **variáveis explicativas** categóricas no modelo de regressão linear.

- Modelos de regressão com variável resposta categórica vão além do escopo da regressão linear, sendo contemplados, por exemplo, pela classe de modelos lineares generalizados.

Regressão linear com variáveis categóricas

Regressão linear com variáveis categóricas

- Considere uma variável categórica (fator) com dois níveis, A e B;

Regressão linear com variáveis categóricas

- Considere uma variável categórica (fator) com dois níveis, A e B;
- A inclusão dessa variável ao modelo de regressão requer a incorporação de uma única variável indicadora:

$$x = \begin{cases} 0, & \text{se categoria A} \\ 1, & \text{se categoria B} \end{cases}$$

Regressão linear com variáveis categóricas

- Considere uma variável categórica (fator) com dois níveis, A e B;

- A inclusão dessa variável ao modelo de regressão requer a incorporação de uma única variável indicadora:

$$x = \begin{cases} 0, & \text{se categoria A} \\ 1, & \text{se categoria B} \end{cases}$$

- Supondo que essa seja a única variável na análise, o modelo de regressão linear ficaria definido por:

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

Regressão linear com variáveis categóricas

- Sob as suposições de uma regressão linear temos que:

$$E(y|x) = \begin{cases} \beta_0 & \text{se categoria A} \\ \beta_0 + \beta_1 & \text{se categoria B} \end{cases}$$

Regressão linear com variáveis categóricas

- Sob as suposições de uma regressão linear temos que:

$$E(y|x) = \begin{cases} \beta_0 & \text{se categoria A} \\ \beta_0 + \beta_1 & \text{se categoria B} \end{cases}$$

- Desta forma, β_0 corresponde à média de y para indivíduos da categoria A.

Regressão linear com variáveis categóricas

- Sob as suposições de uma regressão linear temos que:

$$E(y|x) = \begin{cases} \beta_0 & \text{se categoria A} \\ \beta_0 + \beta_1 & \text{se categoria B} \end{cases}$$

- Desta forma, β_0 corresponde à média de y para indivíduos da categoria A.

- Ainda, β_1 corresponde à diferença na média de y dos indivíduos da categoria B para os indivíduos da categoria A.

Regressão linear com variáveis categóricas

- Sob as suposições de uma regressão linear temos que:

$$E(y|x) = \begin{cases} \beta_0 & \text{se categoria A} \\ \beta_0 + \beta_1 & \text{se categoria B} \end{cases}$$

- Desta forma, β_0 corresponde à média de y para indivíduos da categoria A.

- Ainda, β_1 corresponde à diferença na média de y dos indivíduos da categoria B para os indivíduos da categoria A.

- Se no modelo houver outras variáveis (categóricas ou numéricas), então β_0 será a média de y quando todas as variáveis do modelo forem zero e a interpretação de β_1 se mantém, mas mantendo fixos os valores das demais variáveis.

Regressão linear com variáveis categóricas

- Por que não adicionar ao modelo uma variável indicadora para cada categoria da variável?

Regressão linear com variáveis categóricas

- Por que não adicionar ao modelo uma variável indicadora para cada categoria da variável?
- Seja x_1 a variável indicadora referente à categoria A e x_2 a variável indicadora referente a B;

Regressão linear com variáveis categóricas

- Por que não adicionar ao modelo uma variável indicadora para cada categoria da variável?
- Seja x_1 a variável indicadora referente à categoria A e x_2 a variável indicadora referente a B;
- A matriz do modelo ficaria da seguinte forma:

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

Regressão linear com variáveis categóricas

- Observe que ao inserir uma variável indicadora para cada categoria da variável, além do intercepto, a soma da segunda e da terceira coluna de X é igual à primeira coluna (vetor de uns).

Regressão linear com variáveis categóricas

- Observe que ao inserir uma variável indicadora para cada categoria da variável, além do intercepto, a soma da segunda e da terceira coluna de X é igual à primeira coluna (vetor de uns).

- Desta forma a matriz X não teria rank completo, de forma que haveria infinitas soluções para as equações de mínimos quadrados (os parâmetros do modelo não seriam identificáveis).

- Observe que ao inserir uma variável indicadora para cada categoria da variável, além do intercepto, a soma da segunda e da terceira coluna de X é igual à primeira coluna (vetor de uns).

- Desta forma a matriz X não teria rank completo, de forma que haveria infinitas soluções para as equações de mínimos quadrados (os parâmetros do modelo não seriam identificáveis).

- Como a matriz do modelo não tem rank completo, a solução seria excluir uma de suas colunas (o intercepto ou uma das indicadoras das categorias de x).

Regressão linear com variáveis categóricas

- Considere agora que a variável categórica (fator) tenha k níveis $(A_1, A_2, ..., A_k)$.

Regressão linear com variáveis categóricas

- Considere agora que a variável categórica (fator) tenha k níveis $(A_1, A_2, ..., A_k)$.

- Se incluíssemos no modelo k variáveis indicadoras, além do intercepto, então novamente a soma dessas variáveis resultaria num vetor de uns (dependência linear, matriz X de rank incompleto).

Regressão linear com variáveis categóricas

- Considere agora que a variável categórica (fator) tenha k níveis $(A_1, A_2, ..., A_k)$.

- Se incluíssemos no modelo k variáveis indicadoras, além do intercepto, então novamente a soma dessas variáveis resultaria num vetor de uns (dependência linear, matriz X de rank incompleto).

- Uma solução seria excluir da matriz do modelo uma das variáveis indicadoras (a categoria correspondente fica sendo a *categoria de referência*), ou o intercepto.

Regressão linear com variáveis categóricas

- Considere agora que a variável categórica (fator) tenha k níveis $(A_1, A_2, ..., A_k)$.

- Se incluíssemos no modelo k variáveis indicadoras, além do intercepto, então novamente a soma dessas variáveis resultaria num vetor de uns (dependência linear, matriz X de rank incompleto).

- Uma solução seria excluir da matriz do modelo uma das variáveis indicadoras (a categoria correspondente fica sendo a *categoria de referência*), ou o intercepto.

- Seja uma variável categórica com categorias A, B, C, D e E. Assumindo a categoria A como referência, as seguintes variáveis indicadoras seriam incluídas no modelo:

$$x_1 = \begin{cases} 1 & \text{se B} \\ 0 & \text{se outra} \end{cases}, \quad x_2 = \begin{cases} 1 & \text{se C} \\ 0 & \text{se outra} \end{cases}, \quad x_3 = \begin{cases} 1 & \text{se D} \\ 0 & \text{se outra} \end{cases} \quad x_4 = \begin{cases} 1 & \text{se E} \\ 0 & \text{se outra} \end{cases}$$

Regressão linear com variáveis categóricas

- Considerando que não haja outras variáveis no modelo, o modelo fica definido por:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \epsilon, \quad \epsilon \sim \text{Normal}(0, \sigma^2),$$

Regressão linear com variáveis categóricas

- Considerando que não haja outras variáveis no modelo, o modelo fica definido por:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \epsilon, \quad \epsilon \sim \text{Normal}(0, \sigma^2),$$

ou, de maneira equivalente:

$$y|\boldsymbol{x} \sim \text{Normal}(\mu_{\boldsymbol{x}}, \sigma^2)$$
$$\mu_{\boldsymbol{x}} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$$

Regressão linear com variáveis categóricas

- Desta forma:

$$E(y|x) = \begin{cases} \beta_0 & \text{se categoria A} \\ \beta_0 + \beta_1 & \text{se categoria B} \\ \beta_0 + \beta_2 & \text{se categoria C} \\ \beta_0 + \beta_3 & \text{se categoria D} \\ \beta_0 + \beta_4 & \text{se categoria E} \end{cases}$$

- Neste caso β_0 é a média de y para a categoria de referência (A).

Regressão linear com variáveis categóricas

- Desta forma:

$$E(y|x) = \begin{cases} \beta_0 & \text{se categoria A} \\ \beta_0 + \beta_1 & \text{se categoria B} \\ \beta_0 + \beta_2 & \text{se categoria C} \\ \beta_0 + \beta_3 & \text{se categoria D} \\ \beta_0 + \beta_4 & \text{se categoria E} \end{cases}$$

- Neste caso β_0 é a média de y para a categoria de referência (A).
- β_1 é a diferença na média de y da categoria B para a categoria A; β_2 da categoria C para a categoria A; β_3 da categoria D para a categoria A e β_4 da categoria E para a categoria A.

Regressão linear com variáveis categóricas

- Desta forma:

$$E(y|x) = \begin{cases} \beta_0 & \text{se categoria A} \\ \beta_0 + \beta_1 & \text{se categoria B} \\ \beta_0 + \beta_2 & \text{se categoria C} \\ \beta_0 + \beta_3 & \text{se categoria D} \\ \beta_0 + \beta_4 & \text{se categoria E} \end{cases}$$

- Neste caso β_0 é a média de y para a categoria de referência (A).
- β_1 é a diferença na média de y da categoria B para a categoria A; β_2 da categoria C para a categoria A; β_3 da categoria D para a categoria A e β_4 da categoria E para a categoria A.
- $\beta_2 - \beta_1$ é a diferença na média de y da categoria C para a categoria B; $\beta_4 - \beta_2$ da categoria E para a categoria C, e assim por diante.

Regressão linear com variáveis categóricas

- Se o modelo tiver ainda outras variáveis, então:

Regressão linear com variáveis categóricas

- Se o modelo tiver ainda outras variáveis, então:
    - β_0 é a média de y quando todas as variáveis do modelo forem zero.

Regressão linear com variáveis categóricas

- Se o modelo tiver ainda outras variáveis, então:
    - β_0 é a média de y quando todas as variáveis do modelo forem zero.
    - Os demais $\beta's$ associados a essa variável correspondem à diferença na média de y de uma particular categoria da variável para a categoria de referência, fixados os valores das demais variáveis.

Regressão linear com variáveis categóricas

- Se o modelo tiver ainda outras variáveis, então:
  - β_0 é a média de y quando todas as variáveis do modelo forem zero.
  - Os demais $\beta's$ associados a essa variável correspondem à diferença na média de y de uma particular categoria da variável para a categoria de referência, fixados os valores das demais variáveis.
  - As diferenças entre os $\beta's$ associados a essa variável correspondem às diferenças na média de y para duas categorias da variável (exceto a de referência) fixados os valores das demais variáveis.

Regressão linear com variáveis categóricas

- A categoria a ser designada como referência fica a critério do analista.

Regressão linear com variáveis categóricas

- A categoria a ser designada como referência fica a critério do analista.

- Independente da escolha, o modelo ajustado é o mesmo, apenas as interpretações dos parâmetros mudam.

Regressão linear com variáveis categóricas

- A categoria a ser designada como referência fica a critério do analista.
- Independente da escolha, o modelo ajustado é o mesmo, apenas as interpretações dos parâmetros mudam.
- No R, se você não especificar a categoria de referência, por *default* ele utiliza o primeiro nível do fator.

Regressão linear com variáveis categóricas

- A categoria a ser designada como referência fica a critério do analista.
- Independente da escolha, o modelo ajustado é o mesmo, apenas as interpretações dos parâmetros mudam.
- No R, se você não especificar a categoria de referência, por *default* ele utiliza o primeiro nível do fator.
- Há outras formas de incorporar variáveis categóricas a modelos de regressão. Cuidado ao usar outros softwares, consulte sempre a documentação!

Regressão linear com variáveis categóricas

- Ainda para o caso da variável categórica com cinco níveis, caso se opte por remover o intercepto do modelo, teremos cinco variáveis indicadoras a serem incluídas:

$$x_1 = \begin{cases} 1 & \text{se } A \\ 0 & \text{se outra} \end{cases}, \quad x_2 = \begin{cases} 1 & \text{se } B \\ 0 & \text{se outra} \end{cases}, \quad x_3 = \begin{cases} 1 & \text{se } C \\ 0 & \text{se outra} \end{cases} \quad x_4 = \begin{cases} 1 & \text{se } D \\ 0 & \text{se outra} \end{cases} \quad x_5 = \begin{cases} 1 & \text{se } E \\ 0 & \text{se outra} \end{cases}$$

Regressão linear com variáveis categóricas

- Ainda para o caso da variável categórica com cinco níveis, caso se opte por remover o intercepto do modelo, teremos cinco variáveis indicadoras a serem incluídas:

$$x_1 = \begin{cases} 1 & \text{se } A \\ 0 & \text{se outra} \end{cases}, \; x_2 = \begin{cases} 1 & \text{se } B \\ 0 & \text{se outra} \end{cases}, \; x_3 = \begin{cases} 1 & \text{se } C \\ 0 & \text{se outra} \end{cases} \; x_4 = \begin{cases} 1 & \text{se } D \\ 0 & \text{se outra} \end{cases} \; x_5 = \begin{cases} 1 & \text{se } E \\ 0 & \text{se outra} \end{cases}$$

- O modelo de regressão linear ficará definido da seguinte forma:

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \epsilon, \quad \epsilon \sim \text{Normal}(0, \sigma^2),$$

- Como consequência:

$$E(y|x) = \begin{cases} \beta_1 & \text{se categoria } A \\ \beta_2 & \text{se categoria } B \\ \beta_3 & \text{se categoria } C \\ \beta_4 & \text{se categoria } D \\ \beta_5 & \text{se categoria } E \end{cases}$$

Regressão linear com variáveis categóricas

- Como consequência:

$$E(y|x) = \begin{cases} \beta_1 & \text{se categoria } A \\ \beta_2 & \text{se categoria } B \\ \beta_3 & \text{se categoria } C \\ \beta_4 & \text{se categoria } D \\ \beta_5 & \text{se categoria } E \end{cases}$$

- Neste caso, cada parâmetro do modelo corresponde à média de y para uma das categorias da variável explicativa.

Exemplo- Crescimento de plantas

Exemplo- Crescimento de plantas

- Nesta aplicação vamos analisar os dados disponíveis na base `PlantGrowth` do `R`.

Exemplo- Crescimento de plantas

- Nesta aplicação vamos analisar os dados disponíveis na base `PlantGrowth` do `R`.

- O objetivo é comparar a produção de plantas (`weight`, peso da massa seca) sob três diferentes condições (`group`, fator com três níveis: controle (`ctrl`), tratamento 1 (`trt1`) e tratamento 2 (`trt2`)).

Exemplo- Crescimento de plantas

- Nesta aplicação vamos analisar os dados disponíveis na base `PlantGrowth` do `R`.

- O objetivo é comparar a produção de plantas (`weight`, peso da massa seca) sob três diferentes condições (`group`, fator com três níveis: controle (`ctrl`), tratamento 1 (`trt1`) e tratamento 2 (`trt2`)).

- Para cada condição experimental, $n = 10$ plantas tiveram sua massa seca medida (10 réplicas por tratamento).

Exemplo- Crescimento de plantas

- Nesta aplicação vamos analisar os dados disponíveis na base `PlantGrowth` do `R`.

- O objetivo é comparar a produção de plantas (`weight`, peso da massa seca) sob três diferentes condições (`group`, fator com três níveis: controle (`ctrl`), tratamento 1 (`trt1`) e tratamento 2 (`trt2`)).

- Para cada condição experimental, $n = 10$ plantas tiveram sua massa seca medida (10 réplicas por tratamento).

- Na sequência apresentamos uma representação gráfica dos resultados, permitindo avaliar possíveis diferenças entre tratamentos.

Exemplo- Crescimento de plantas

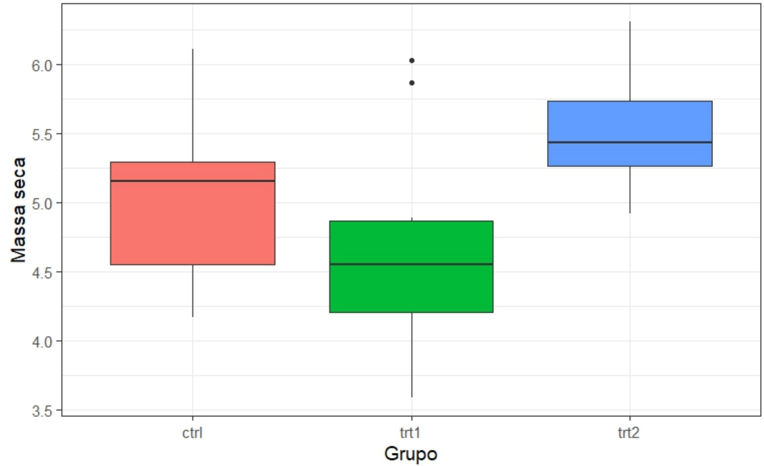

Figura 1: Dados sobre produção de massa seca em experimento agrícola

Exemplo- Crescimento de plantas

- Inicialmente, vamos considerar o fator `group` apenas com dois níveis: `ctrl` e `trt1`.

Exemplo- Crescimento de plantas

- Inicialmente, vamos considerar o fator `group` apenas com dois níveis: `ctrl` e `trt1`.

- Neste caso, definimos uma variável indicadora para `trt1`, tal que `trt1=0` para as observações do grupo controle, e `trt1=1` para aquelas do tratamento 1.

Exemplo- Crescimento de plantas

- Inicialmente, vamos considerar o fator `group` apenas com dois níveis: `ctrl` e `trt1`.

- Neste caso, definimos uma variável indicadora para `trt1`, tal que `trt1=0` para as observações do grupo controle, e `trt1=1` para aquelas do tratamento 1.

- Desta forma, o modelo de regressão linear fica especificado por:

$$\texttt{weight} = \beta_0 + \beta_1 \texttt{trt1} + \epsilon, \quad \epsilon \sim \mathrm{N}(0, \sigma^2),$$

Exemplo- Crescimento de plantas

- Inicialmente, vamos considerar o fator `group` apenas com dois níveis: `ctrl` e `trt1`.
- Neste caso, definimos uma variável indicadora para `trt1`, tal que `trt1=0` para as observações do grupo controle, e `trt1=1` para aquelas do tratamento 1.
- Desta forma, o modelo de regressão linear fica especificado por:

$$\texttt{weight} = \beta_0 + \beta_1 \texttt{trt1} + \epsilon, \quad \epsilon \sim \mathrm{N}(0, \sigma^2),$$

ou, de maneira equivalente,

$$\texttt{weight}|\texttt{grupo} \sim \mathrm{N}(\mu_{\texttt{grupo}}, \sigma^2)$$

$$\mu_{\texttt{grupo}} = \beta_0 + \beta_1 \texttt{trt1}$$

Exemplo- Crescimento de plantas

- Com base nos dados amostrais, o modelo ajustado pelo método de mínimos quadrados fica dado por:

$$\widehat{\texttt{weight}} = 5.032 - 0.371 \times \texttt{trt1}$$

Exemplo- Crescimento de plantas

- Com base nos dados amostrais, o modelo ajustado pelo método de mínimos quadrados fica dado por:

$$\widehat{\texttt{weight}} = 5.032 - 0.371 \times \texttt{trt1}$$

- A quantidade média de massa seca produzida, estimada pelo modelo, é dada por:

$$\widehat{\mu}_{\texttt{ctrl}} = 5.032 - 0.371 \times 0 = 5.032, \text{ para o grupo controle,}$$

Exemplo- Crescimento de plantas

- Com base nos dados amostrais, o modelo ajustado pelo método de mínimos quadrados fica dado por:

$$\widehat{\texttt{weight}} = 5.032 - 0.371 \times \texttt{trt1}$$

- A quantidade média de massa seca produzida, estimada pelo modelo, é dada por:

$$\widehat{\mu}_{\texttt{ctrl}} = 5.032 - 0.371 \times 0 = 5.032, \text{ para o grupo controle,}$$

$$\widehat{\mu}_{\texttt{trt1}} = 5.032 - 0.371 \times 1 = 4.661, \text{ para o tratamento 1}$$

Exemplo- Crescimento de plantas

- Desta forma, $\hat{\beta}_1 = -0.371$ representa a redução média na produção de massa seca da condição controle para o tratamento 1.

Exemplo- Crescimento de plantas

- Desta forma, $\hat{\beta}_1 = -0.371$ representa a redução média na produção de massa seca da condição controle para o tratamento 1.

- Seguindo, vamos incluir agora também o terceiro grupo (`trt2`) na análise.

Exemplo- Crescimento de plantas

- Desta forma, $\hat{\beta}_1 = -0.371$ representa a redução média na produção de massa seca da condição controle para o tratamento 1.

- Seguindo, vamos incluir agora também o terceiro grupo (`trt2`) na análise.

- Para isso, precisamos definir duas variáveis indicadoras para o fator `grupo` com três níveis:
  - `trt1` $= 1$, se a observação pertence ao tratamento 1, e `trt1` $= 0$, caso contrário;
  - `trt2` $= 1$, se a observação pertence ao tratamento 2, e `trt2` $= 0$, caso contrário.

Exemplo- Crescimento de plantas

- Desta forma, o modelo de regressão linear fica especificado por:

Exemplo- Crescimento de plantas

- Desta forma, o modelo de regressão linear fica especificado por:

$$\texttt{weight} = \beta_0 + \beta_1 \texttt{trt1} + \beta_2 \texttt{trt2} + \epsilon, \quad \epsilon \sim \mathrm{N}(0, \sigma^2),$$

Exemplo- Crescimento de plantas

- Desta forma, o modelo de regressão linear fica especificado por:

$$\texttt{weight} = \beta_0 + \beta_1 \texttt{trt1} + \beta_2 \texttt{trt2} + \epsilon, \quad \epsilon \sim \mathrm{N}(0, \sigma^2),$$

ou, de maneira equivalente,

$$\texttt{weight}|\texttt{grupo} \sim \mathrm{N}(\mu_{\texttt{grupo}}, \sigma^2)$$

$$\mu_{\texttt{grupo}} = \beta_0 + \beta_1 \texttt{trt1} + \beta_2 \texttt{trt2}$$

Exemplo- Crescimento de plantas

- Com base nos dados amostrais, o modelo ajustado pelo método de mínimos quadrados fica dado por:

$$\widehat{\texttt{weight}} = 5.032 - 0.371 \times \texttt{trt1} + 0.494 \times \texttt{trt2}$$

Exemplo- Crescimento de plantas

- Com base nos dados amostrais, o modelo ajustado pelo método de mínimos quadrados fica dado por:

$$\widehat{\texttt{weight}} = 5.032 - 0.371 \times \texttt{trt1} + 0.494 \times \texttt{trt2}$$

- A quantidade média de massa seca produzida, estimada pelo modelo, é dada por:

$$\widehat{\mu}_{\texttt{ctrl}} = 5.032 - 0.371 \times 0 + 0.494 \times 0 = 5.032, \text{ para o grupo controle,}$$

Exemplo- Crescimento de plantas

- Com base nos dados amostrais, o modelo ajustado pelo método de mínimos quadrados fica dado por:

$$\widehat{\texttt{weight}} = 5.032 - 0.371 \times \texttt{trt1} + 0.494 \times \texttt{trt2}$$

- A quantidade média de massa seca produzida, estimada pelo modelo, é dada por:

$$\widehat{\mu}_{\texttt{ctrl}} = 5.032 - 0.371 \times 0 + 0.494 \times 0 = 5.032, \text{ para o grupo controle,}$$

$$\widehat{\mu}_{\texttt{trt}_1} = 5.032 - 0.371 \times 1 + 0.494 \times 0 = 4.661, \text{ para o tratamento 1,}$$

Exemplo- Crescimento de plantas

- Com base nos dados amostrais, o modelo ajustado pelo método de mínimos quadrados fica dado por:

$$\widehat{\texttt{weight}} = 5.032 - 0.371 \times \texttt{trt1} + 0.494 \times \texttt{trt2}$$

- A quantidade média de massa seca produzida, estimada pelo modelo, é dada por:

$$\widehat{\mu}_{\texttt{ctrl}} = 5.032 - 0.371 \times 0 + 0.494 \times 0 = 5.032, \text{ para o grupo controle,}$$

$$\widehat{\mu}_{\texttt{trt}_1} = 5.032 - 0.371 \times 1 + 0.494 \times 0 = 4.661, \text{ para o tratamento 1,}$$

$$\widehat{\mu}_{\texttt{trt}_2} = 5.032 - 0.371 \times 0 + 0.494 \times 1 = 5.526, \text{ para o tratamento 2.}$$

Exemplo- Crescimento de plantas

- Desta forma, $\hat{\beta}_1 = -0.371$ representa a redução média na produção de massa seca da condição controle para o tratamento 1.

Exemplo- Crescimento de plantas

- Desta forma, $\hat{\beta}_1 = -0.371$ representa a redução média na produção de massa seca da condição controle para o tratamento 1.

- Adicionalmente, $\hat{\beta}_2 = 0.494$ representa o aumento médio na produção de massa seca da condição controle para o tratamento 2.

Exemplo- Crescimento de plantas

- Desta forma, $\hat{\beta}_1 = -0.371$ representa a redução média na produção de massa seca da condição controle para o tratamento 1.

- Adicionalmente, $\hat{\beta}_2 = 0.494$ representa o aumento médio na produção de massa seca da condição controle para o tratamento 2.

- A diferença média na produção sob os tratamentos 2 e 1 é dada por $\hat{\beta}_2 - \beta_1 = 0.494 - (-0.371) = 0.865$.

Exemplo- Crescimento de plantas

- Desta forma, $\hat{\beta}_1 = -0.371$ representa a redução média na produção de massa seca da condição controle para o tratamento 1.

- Adicionalmente, $\hat{\beta}_2 = 0.494$ representa o aumento médio na produção de massa seca da condição controle para o tratamento 2.

- A diferença média na produção sob os tratamentos 2 e 1 é dada por $\hat{\beta}_2 - \beta_1 = 0.494 - (-0.371) = 0.865$.

- As demais inferências, como testes de hipóteses, intervalos de confiança e predições, são realizadas da maneira usual.

Exemplo- Teor de gordura no leite

Exemplo- Teor de gordura no leite

- Nesta aplicação vamos analisar dados de gordura butírica em 100 amostras de leite, disponíveis na base de dados `butterfat` da biblioteca `GLMsData` do `R`.

Exemplo- Teor de gordura no leite

- Nesta aplicação vamos analisar dados de gordura butírica em 100 amostras de leite, disponíveis na base de dados `butterfat` da biblioteca `GLMsData` do `R`.

- As variáveis que compõem a análise são as seguintes:

Exemplo- Teor de gordura no leite

- Nesta aplicação vamos analisar dados de gordura butírica em 100 amostras de leite, disponíveis na base de dados `butterfat` da biblioteca `GLMsData` do `R`.

- As variáveis que compõem a análise são as seguintes:
  - `Butterfat`: porcentagem de gordura butírica (variável resposta);

Exemplo- Teor de gordura no leite

- Nesta aplicação vamos analisar dados de gordura butírica em 100 amostras de leite, disponíveis na base de dados `butterfat` da biblioteca `GLMsData` do `R`.

- As variáveis que compõem a análise são as seguintes:
  - `Butterfat`: porcentagem de gordura butírica (variável resposta);
  - `Age`: idade da vaca, classificada em `2year` e `Mature`;

Exemplo- Teor de gordura no leite

- Nesta aplicação vamos analisar dados de gordura butírica em 100 amostras de leite, disponíveis na base de dados `butterfat` da biblioteca `GLMsData` do `R`.

- As variáveis que compõem a análise são as seguintes:
  - `Butterfat`: porcentagem de gordura butírica (variável resposta);
  - `Age`: idade da vaca, classificada em `2year` e `Mature`;
  - `Breed`: raça do gado, com níveis `Ayrshire`, `Canadian`, `Guernsey`, `Holstein-Fresian` e `Jersey`.

Exemplo- Teor de gordura no leite

- Tomando `Ayrshire` e `2year` como categorias de referência, o seguinte modelo de regressão linear foi especificado:

$$\texttt{Butterfat} = \beta_0 + \beta_1\texttt{Mature} + \beta_2\texttt{Canadian} + \beta_3\texttt{Guernsey}$$

$$+\beta_4\texttt{Holstein-Fresian} + \beta_5\texttt{Jersey} + \epsilon,$$

com $\epsilon \sim \text{Normal}(0, \sigma^2)$, e $\texttt{Mature} = 1$ se o animal for classificado de maduro segundo a sua idade, e $\texttt{Mature} = 0$ caso contrário, e similar para as demais variáveis.

Exemplo- Teor de gordura no leite

- O modelo de regressão linear ajustado é o seguinte:

$$\widehat{\texttt{Butterfat}} = 4.00 + 0.10 \times \texttt{Mature} + 0.37 \times \texttt{Canadian} + 0.89 \times \texttt{Guernsey}$$
$$-0.39 \times \texttt{Holstein-Fresian} + 1.23 \times \texttt{Jersey}$$

Exemplo- Teor de gordura no leite

- O modelo de regressão linear ajustado é o seguinte:

$$\widehat{\texttt{Butterfat}} = 4.00 + 0.10 \times \texttt{Mature} + 0.37 \times \texttt{Canadian} + 0.89 \times \texttt{Guernsey}$$
$$-0.39 \times \texttt{Holstein-Fresian} + 1.23 \times \texttt{Jersey}$$

- Na sequência apresentamos as interpretações das estimativas dos parâmetros de regressão (não vamos nos ater à significância dos efeitos ou à qualidade do ajuste do modelo neste momento).

Exemplo- Teor de gordura no leite

- Estima-se em média 0.10 pontos percentuais a mais de gordura butírica no leite de vacas classificadas como `Mature`, em relação a `2year`, fixada a raça;

Exemplo- Teor de gordura no leite

- Estima-se em média 0.10 pontos percentuais a mais de gordura butírica no leite de vacas classificadas como `Mature`, em relação a `2year`, fixada a raça;

- Estima-se em média 0.37 pontos percentuais a mais de gordura butírica no leite de vacas da raça `Canadian`, em relação a `Ayrshire`, fixada a idade;

Exemplo- Teor de gordura no leite

- Estima-se em média 0.10 pontos percentuais a mais de gordura butírica no leite de vacas classificadas como `Mature`, em relação a `2year`, fixada a raça;

- Estima-se em média 0.37 pontos percentuais a mais de gordura butírica no leite de vacas da raça `Canadian`, em relação a `Ayrshire`, fixada a idade;

- Estima-se em média 0.89 pontos percentuais a mais de gordura butírica no leite de vacas da raça `Guernsey`, em relação a `Ayrshire`, fixada a idade;

Exemplo- Teor de gordura no leite

- Estima-se em média 0.10 pontos percentuais a mais de gordura butírica no leite de vacas classificadas como `Mature`, em relação a `2year`, fixada a raça;
- Estima-se em média 0.37 pontos percentuais a mais de gordura butírica no leite de vacas da raça `Canadian`, em relação a `Ayrshire`, fixada a idade;
- Estima-se em média 0.89 pontos percentuais a mais de gordura butírica no leite de vacas da raça `Guernsey`, em relação a `Ayrshire`, fixada a idade;
- Estima-se em média 0.39 pontos percentuais a menos de gordura butírica no leite de vacas da raça `Holstein-Fresian`, em relação a `Ayrshire`, fixada a idade;

Exemplo- Teor de gordura no leite

- Estima-se em média 0.10 pontos percentuais a mais de gordura butírica no leite de vacas classificadas como `Mature`, em relação a `2year`, fixada a raça;
- Estima-se em média 0.37 pontos percentuais a mais de gordura butírica no leite de vacas da raça `Canadian`, em relação a `Ayrshire`, fixada a idade;
- Estima-se em média 0.89 pontos percentuais a mais de gordura butírica no leite de vacas da raça `Guernsey`, em relação a `Ayrshire`, fixada a idade;
- Estima-se em média 0.39 pontos percentuais a menos de gordura butírica no leite de vacas da raça `Holstein-Fresian`, em relação a `Ayrshire`, fixada a idade;
- Estima-se em média 1.23 pontos percentuais a mais de gordura butírica no leite de vacas da raça `Jersey`, em relação a `Ayrshire`, fixada a idade.

Análise de covariância (ANCOVA)

Análise de covariância (ANCOVA)

- Vamos considerar um modelo de regressão com uma variável quantitativa (x) e uma variável categórica com k níveis.

Análise de covariância (ANCOVA)

- Vamos considerar um modelo de regressão com uma variável quantitativa (x) e uma variável categórica com k níveis.

- Para simplificar a notação, vamos considerar D_2, D_3, $\ldots$, D_k as variáveis indicadoras para os níveis 2, 3,$\ldots$, k da variável categórica.

Análise de covariância (ANCOVA)

- Vamos considerar um modelo de regressão com uma variável quantitativa (x) e uma variável categórica com k níveis.
- Para simplificar a notação, vamos considerar D_2, D_3, $\ldots$, D_k as variáveis indicadoras para os níveis 2, 3,$\ldots$, k da variável categórica.
- Um primeiro modelo a ser considerado é o de efeitos aditivos (sem interação):

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 d_{i2} + \beta_3 d_{i3} + ... + \beta_k d_{ik} + \epsilon_i,$$

em que $d_{ij} = 1$, se o indivíduo pertence à categoria j, e $d_{ij} = 0$, caso contrário.

Análise de covariância (ANCOVA)

- Para modelos com variáveis categóricas e quantitativas é sempre útil escrever a equação do modelo para cada nível da variável categórica.

Análise de covariância (ANCOVA)

- Para modelos com variáveis categóricas e quantitativas é sempre útil escrever a equação do modelo para cada nível da variável categórica.

- O modelo de regressão com efeitos aditivos pode ser expresso, para cada nível da variável categórica, por:

$$E(y|\mathrm{x}) = \begin{cases} \beta_0 + \beta_1 x_1 & \text{se categoria } A_1 \\ (\beta_0 + \beta_2) + \beta_1 x_1 & \text{se categoria } A_2 \\ (\beta_0 + \beta_3) + \beta_1 x_1 & \text{se categoria } A_3 \\ \vdots & \\ (\beta_0 + \beta_k) + \beta_1 x_1 & \text{se categoria } A_k \end{cases}$$

Análise de covariância (ANCOVA)

- Observe que para o modelo com efeitos aditivos, as retas de regressão para os k níveis da variável categórica apresentam somente interceptos diferentes. As inclinações são as mesmas.

Análise de covariância (ANCOVA)

- Observe que para o modelo com efeitos aditivos, as retas de regressão para os k níveis da variável categórica apresentam somente interceptos diferentes. As inclinações são as mesmas.

- Para o modelo com efeitos aditivos, β_j corresponde à diferença na média de y da categoria j em relação à categoria 1 (referência) da variável categórica, fixado o valor de x_1.

Análise de covariância (ANCOVA)

- O modelo com efeitos multiplicativos (com interação), por sua vez, é definido por:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 d_{i2} + ... + \beta_k d_{ik} + \beta_{k+1} d_{i2} x_{i1} + ... + \beta_p d_{ik} x_{i1} + \epsilon_i,$$

podendo ser expresso, para cada nível da variável categórica, por:

$$E(y|\mathrm{x}) = \begin{cases} \beta_0 + \beta_1 x_1 & \text{se categoria } A_1 \\ (\beta_0 + \beta_2) + (\beta_1 + \beta_{k+1}) x_1 & \text{se categoria } A_2 \\ (\beta_0 + \beta_3) + (\beta_1 + \beta_{k+2}) x_1 & \text{se categoria } A_3 \\ \vdots & \\ (\beta_0 + \beta_k) + (\beta_1 + \beta_{2k-1}) x_1 & \text{se categoria } A_k \end{cases}$$

Análise de covariância (ANCOVA)

- Observe que para o modelo com efeitos multiplicativos as retas de regressão têm interceptos e inclinações diferentes para cada nível da variável categórica.

Análise de covariância (ANCOVA)

- Observe que para o modelo com efeitos multiplicativos as retas de regressão têm interceptos e inclinações diferentes para cada nível da variável categórica.

- Para o modelo com efeitos multiplicativos, β_2, β_3, $\dots$, β_k expressam as diferenças de intercepto das retas de regressão para os níveis 2, 3,.., k da variável categórica em relação ao intercepto para o nível de referência.

Análise de covariância (ANCOVA)

- Observe que para o modelo com efeitos multiplicativos as retas de regressão têm interceptos e inclinações diferentes para cada nível da variável categórica.

- Para o modelo com efeitos multiplicativos, $\beta_2, \beta_3, \ldots, \beta_k$ expressam as diferenças de intercepto das retas de regressão para os níveis 2, 3,.., k da variável categórica em relação ao intercepto para o nível de referência.

- Já $\beta_{k+1}, \beta_{k+2}, \ldots, \beta_p$ expressam as diferenças das inclinações das retas de regressão para os níveis 2, 3,.., k da variável categórica em relação à inclinação para o nível de referência.

Análise de covariância (ANCOVA)

- Outra forma de explorar os efeitos das variáveis em modelos com variáveis categóricas e quantitativas é através de gráficos de efeitos.

Análise de covariância (ANCOVA)

- Outra forma de explorar os efeitos das variáveis em modelos com variáveis categóricas e quantitativas é através de gráficos de efeitos.

- Para avaliar o efeito de uma variável quantitativa x_1, plota-se o gráfico da estimativa de $E(y|\mathrm{x})$ ao longo de x_1, separado para cada nível da variável categórica.

Análise de covariância (ANCOVA)

- Outra forma de explorar os efeitos das variáveis em modelos com variáveis categóricas e quantitativas é através de gráficos de efeitos.

- Para avaliar o efeito de uma variável quantitativa x_1, plota-se o gráfico da estimativa de $E(y|\mathrm{x})$ ao longo de x_1, separado para cada nível da variável categórica.

- Se houver mais variáveis no modelo, elas podem ser fixadas em valores típicos (por exemplo em suas médias).

Análise de covariância (ANCOVA)

- A seleção de modelos quando se tem efeito de interação deve obedecer ao princípio hierárquico.

Análise de covariância (ANCOVA)

- A seleção de modelos quando se tem efeito de interação deve obedecer ao princípio hierárquico.

- Basicamente, para um modelo ajustado com preditor `x1 + x2 + x1:x2`, em que `x1:x2` representa o termo de interação, não é recomendável remover os termos de menor ordem (`x1` e `x2`) na presença do termo de maior ordem `x1:x2`.

Análise de covariância (ANCOVA)

- A seleção de modelos quando se tem efeito de interação deve obedecer ao princípio hierárquico.

- Basicamente, para um modelo ajustado com preditor `x1 + x2 + x1:x2`, em que `x1:x2` representa o termo de interação, não é recomendável remover os termos de menor ordem (`x1` e `x2`) na presença do termo de maior ordem `x1:x2`.

- Vamos considerar novamente que x_1 é uma variável quantitativa e x_2 uma variável categórica com k níveis.

Análise de covariância (ANCOVA)

- A seleção de modelos quando se tem efeito de interação deve obedecer ao princípio hierárquico.

- Basicamente, para um modelo ajustado com preditor `x1 + x2 + x1:x2`, em que `x1:x2` representa o termo de interação, não é recomendável remover os termos de menor ordem (`x1` e `x2`) na presença do termo de maior ordem `x1:x2`.

- Vamos considerar novamente que x_1 é uma variável quantitativa e x_2 uma variável categórica com k níveis.

- Pode-se proceder a seleção de modelos com base na sequência de modelos encaixados apresentada na sequência.

Análise de covariância (ANCOVA)

- **Modelo nulo:** sem efeito de variáveis (`y ~ 1`).

Análise de covariância (ANCOVA)

- **Modelo nulo:** sem efeito de variáveis (`y ~ 1`).

- **Modelo de retas coincidentes:** sem efeito da variável categórica (`y ~ x1`).

Análise de covariância (ANCOVA)

- **Modelo nulo:** sem efeito de variáveis (`y ~ 1`).

- **Modelo de retas coincidentes:** sem efeito da variável categórica (`y ~ x1`).

- **Modelo de retas paralelas:** modelo de efeitos aditivos, sem interação (`y ~ x1 + x2`).

Análise de covariância (ANCOVA)

- **Modelo nulo:** sem efeito de variáveis (`y ~ 1`).
- **Modelo de retas coincidentes:** sem efeito da variável categórica (`y ~ x1`).
- **Modelo de retas paralelas:** modelo de efeitos aditivos, sem interação (`y ~ x1 + x2`).
- **Modelo de retas concorrentes:** modelo de efeitos multiplicativos, com interação (`y ~ x1 * x2`).

Análise de covariância (ANCOVA)

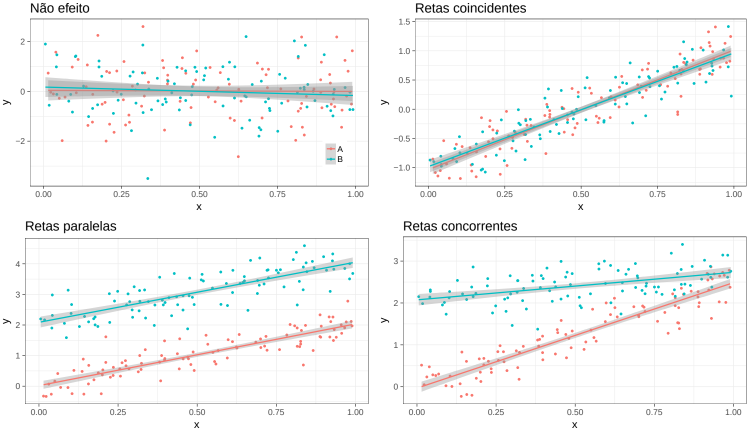

Figura 2: Ilustração para os cenários correspondentes aos quatro modelos incluindo uma variável quantitativa e outra categórica.

Análise de covariância (ANCOVA)

- Como os quatro modelos sugeridos são encaixados, podemos compará-los, sequencialmente, usando o teste F baseado na variação da soma de quadrados de resíduos;

Análise de covariância (ANCOVA)

- Como os quatro modelos sugeridos são encaixados, podemos compará-los, sequencialmente, usando o teste F baseado na variação da soma de quadrados de resíduos;

- Caso se disponha de mais variáveis para análise, pode-se aplicar algum algoritmo de seleção de variáveis. Além disso, interações com outras variáveis podem também ser consideradas;

Análise de covariância (ANCOVA)

- Como os quatro modelos sugeridos são encaixados, podemos compará-los, sequencialmente, usando o teste F baseado na variação da soma de quadrados de resíduos;
- Caso se disponha de mais variáveis para análise, pode-se aplicar algum algoritmo de seleção de variáveis. Além disso, interações com outras variáveis podem também ser consideradas;
- Outra possibilidade é avaliar modelos em que as retas de regressão são paralelas ou coincidentes apenas para algum subconjunto dos níveis da variável categórica;

Análise de covariância (ANCOVA)

- Como os quatro modelos sugeridos são encaixados, podemos compará-los, sequencialmente, usando o teste F baseado na variação da soma de quadrados de resíduos;

- Caso se disponha de mais variáveis para análise, pode-se aplicar algum algoritmo de seleção de variáveis. Além disso, interações com outras variáveis podem também ser consideradas;

- Outra possibilidade é avaliar modelos em que as retas de regressão são paralelas ou coincidentes apenas para algum subconjunto dos níveis da variável categórica;

- Como exemplo poderíamos ter retas de mesma inclinação descrevendo a relação entre índice de massa corporal e consumo calórico para adultos e idosos, mas de menor inclinação para a população de crianças.

Exemplo- Efeito da puromicina

Exemplo- Efeito da puromicina

- Nesta aplicação vamos analisar dados de um experimento químico com o objetivo de avaliar o efeito da purocimina, um particular tipo de antibiótico.

Exemplo- Efeito da puromicina

- Nesta aplicação vamos analisar dados de um experimento químico com o objetivo de avaliar o efeito da purocimina, um particular tipo de antibiótico.

- A base de dados é composta por 23 amostras de células, e as variáveis são as seguintes:
  - `rate`: velocidade da reação (resposta, em contagem/min/min);
  - `conc`: concentração de substrato (em ppm);
  - `state`: um fator com níveis `treated` e `untreated` (com puromicina).

Exemplo- Efeito da puromicina

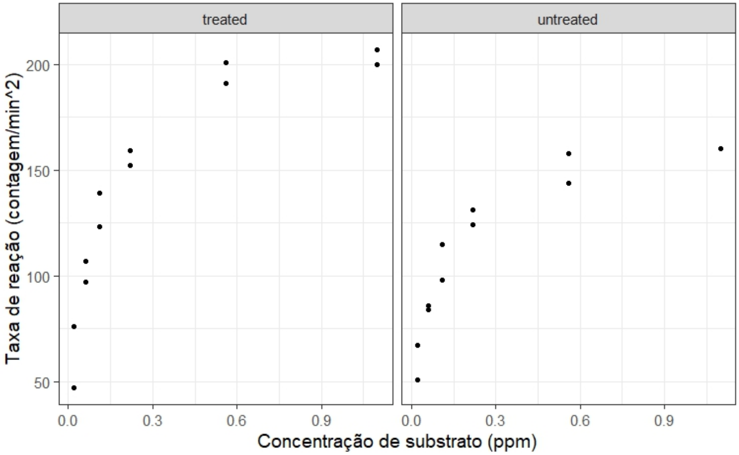

Figura 3: Dados de reação química em estudo sobre o efeito da puromicina

Exemplo- Efeito da puromicina

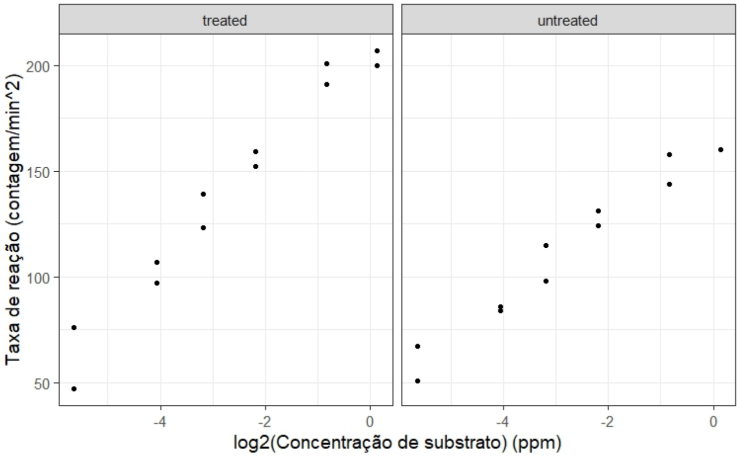

Figura 4: Dados de reação química em estudo sobre o efeito da puromicina com transformação logarítmica no substrato

Exemplo- Efeito da puromicina

- Os modelos que vamos considerar são os seguintes:

Exemplo- Efeito da puromicina

- Os modelos que vamos considerar são os seguintes:

1. Modelo nulo: sem efeito de substrato e tratamento:

$$\texttt{rate} = \beta_0 + \epsilon, \quad \epsilon \sim \mathrm{N}(0, \sigma^2)$$

Exemplo- Efeito da puromicina

- Os modelos que vamos considerar são os seguintes:

1. Modelo nulo: sem efeito de substrato e tratamento:

$$\texttt{rate} = \beta_0 + \epsilon, \quad \epsilon \sim \mathrm{N}(0, \sigma^2)$$

2. Modelo de retas coincidentes: sem efeito de tratamento:

$$\texttt{rate} = \beta_0 + \beta_1 \texttt{log}_2(\texttt{conc}) + \epsilon, \quad \epsilon \sim \mathrm{N}(0, \sigma^2)$$

Exemplo- Efeito da puromicina

3. Modelo de retas paralela: diferentes interceptos, mas mesmo efeito de tratamento para tratados e não tratados (mesma inclinação):

$$\texttt{rate} = \beta_0 + \beta_1 \texttt{log}_2(\texttt{conc}) + \beta_2 \texttt{state:untreated} + \epsilon, \quad \epsilon \sim \mathrm{N}(0, \sigma^2),$$

onde `state:untreated` $= 1$ para o grupo `untreated` e `state:untreated` $= 0$ para `treated`.

Exemplo- Efeito da puromicina

4. Modelo de retas concorrentes: diferentes interceptose diferentes efeitos de tratamento para tratados e não tratados (diferentes inclinações:

$$\texttt{rate} = \beta_0 + \beta_1 \texttt{log}_2(\texttt{conc}) + \beta_2 \texttt{state:untreated} + \beta_3 \texttt{state:untreated} \times \texttt{log}_2(\texttt{conc}) + \epsilon, \quad \epsilon \sim \mathrm{N}(0, \sigma^2)$$

Exemplo- Efeito da puromicina

- Os modelos ajustados são os seguintes:

Exemplo- Efeito da puromicina

- Os modelos ajustados são os seguintes:

1. Modelo nulo:

$$\widehat{\texttt{rate}} = 126.8$$

Exemplo- Efeito da puromicina

- Os modelos ajustados são os seguintes:

1. Modelo nulo:

$$\widehat{\texttt{rate}} = 126.8$$

2. Modelo de retas coincidentes:

$$\widehat{\texttt{rate}} = 190.09 + 23.01 \times \texttt{log}_2(\texttt{conc})$$

Exemplo- Efeito da puromicina

3. Modelo de retas paralelas:

$$\widehat{\texttt{rate}} = 200.91 + 22.57 \times \texttt{log}_2(\texttt{conc}) - 25.18 \times \texttt{state:untreated},$$

Exemplo- Efeito da puromicina

3. Modelo de retas paralelas:

$$\widehat{\texttt{rate}} = 200.91 + 22.57 \times \texttt{log}_2(\texttt{conc}) - 25.18 \times \texttt{state:untreated},$$

de tal forma que:

$$\widehat{\texttt{rate}} = 200.91 + 22.57 \times \texttt{log}_2(\texttt{conc}), \text{ para o grupo tratado}$$

Exemplo- Efeito da puromicina

3. Modelo de retas paralelas:

$$\widehat{\texttt{rate}} = 200.91 + 22.57 \times \texttt{log}_2(\texttt{conc}) - 25.18 \times \texttt{state:untreated},$$

de tal forma que:

$$\widehat{\texttt{rate}} = 200.91 + 22.57 \times \texttt{log}_2(\texttt{conc}), \text{ para o grupo tratado}$$

$$\widehat{\texttt{rate}} = 175.73 + 22.57 \times \texttt{log}_2(\texttt{conc}), \text{ para o grupo não tratado}$$

Exemplo- Efeito da puromicina

4. Modelo de retas concorrentes:

$$\widehat{\texttt{rate}} = 209.19 + 25.72 \times \texttt{log}_2(\texttt{conc}) - 44.61 \times \texttt{state:untreated} - 7.02 \times \texttt{state:untreated} \times \texttt{log}_2(\texttt{conc}),$$

Exemplo- Efeito da puromicina

4. Modelo de retas concorrentes:

$$\widehat{\texttt{rate}} = 209.19 + 25.72 \times \texttt{log}_2(\texttt{conc}) - 44.61 \times \texttt{state:untreated} - 7.02 \times \texttt{state:untreated} \times \texttt{log}_2(\texttt{conc}),$$

de tal forma que:

$$\widehat{\texttt{rate}} = 209.19 + 25.72 \times \texttt{log}_2(\texttt{conc}), \text{ para o grupo tratado}$$

Exemplo- Efeito da puromicina

4. Modelo de retas concorrentes:

$$\widehat{\texttt{rate}} = 209.19 + 25.72 \times \texttt{log}_2(\texttt{conc}) - 44.61 \times \texttt{state:untreated} - 7.02 \times \texttt{state:untreated} \times \texttt{log}_2(\texttt{conc}),$$

de tal forma que:

$$\widehat{\texttt{rate}} = 209.19 + 25.72 \times \texttt{log}_2(\texttt{conc}), \text{ para o grupo tratado}$$

$$\widehat{\texttt{rate}} = 164.58 + 18.7 \times \texttt{log}_2(\texttt{conc}), \text{ para o grupo não tratado}$$

Exemplo- Efeito da puromicina

- Algumas interpretações no modelo de retas concorrentes:

Exemplo- Efeito da puromicina

- Algumas interpretações no modelo de retas concorrentes:
  - $\hat{\beta}_0 = 209.19$ é a velocidade média de reação para o grupo `treated` quando $\texttt{log}_2(\texttt{conc}) = 0$, ou seja, $\texttt{conc} = 1$;

Exemplo- Efeito da puromicina

- Algumas interpretações no modelo de retas concorrentes:
  - $\hat{\beta}_0 = 209.19$ é a velocidade média de reação para o grupo `treated` quando $\texttt{log}_2(\texttt{conc}) = 0$, ou seja, $\texttt{conc} = 1$;
  - $\hat{\beta}_1 = 25.72$ é a taxa de variação na velocidade média de reação para uma unidade a mais em $\texttt{log}_2(\texttt{conc})$ para o grupo `treated`;

Exemplo- Efeito da puromicina

- Algumas interpretações no modelo de retas concorrentes:
  - $\hat{\beta}_0 = 209.19$ é a velocidade média de reação para o grupo `treated` quando $\texttt{log}_2(\texttt{conc}) = 0$, ou seja, $\texttt{conc} = 1$;
  - $\hat{\beta}_1 = 25.72$ é a taxa de variação na velocidade média de reação para uma unidade a mais em $\texttt{log}_2(\texttt{conc})$ para o grupo `treated`;
  - $\hat{\beta}_2 = -44.61$ é o incremento (redução) no intercepto para o grupo `untreated` em relação ao grupo `treated`;

Exemplo- Efeito da puromicina

- Algumas interpretações no modelo de retas concorrentes:
  - $\hat{\beta}_0 = 209.19$ é a velocidade média de reação para o grupo `treated` quando $\texttt{log}_2(\texttt{conc}) = 0$, ou seja, $\texttt{conc} = 1$;
  - $\hat{\beta}_1 = 25.72$ é a taxa de variação na velocidade média de reação para uma unidade a mais em $\texttt{log}_2(\texttt{conc})$ para o grupo `treated`;
  - $\hat{\beta}_2 = -44.61$ é o incremento (redução) no intercepto para o grupo `untreated` em relação ao grupo `treated`;
  - $\hat{\beta}_3 = -7.02$ é o incremento (redução) na taxa de variação na velocidade média de reação para uma unidade a mais em $\texttt{log}_2(\texttt{conc})$ para o grupo `untreated` em relação ao grupo `treated`.

Exemplo- Efeito da puromicina

Tabela 1: Análise de variância

Model	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
Nulo	22	49665.30				
Coincidentes	21	6210.03	1	43455.28	518.87	0.0000
Paralelas	20	2587.18	1	3622.85	43.26	0.0000
Concorrentes	19	1591.25	1	995.94	11.89	0.0027

Exemplo- Efeito da puromicina

- Com base nos resultados da análise de variância, podemos concluir:

Exemplo- Efeito da puromicina

- Com base nos resultados da análise de variância, podemos concluir:

1. Há uma redução significativa ($p<0.0001$) na soma de quadrados de resíduos do modelo nulo ($SQR = 49665.30$) para o modelo de retas coincidentes ($SQR=6210.03$), de tal forma que o modelo de retas coincidentes é preferível ao modelo nulo;

Exemplo- Efeito da puromicina

- Com base nos resultados da análise de variância, podemos concluir:

1. Há uma redução significativa ($p<0.0001$) na soma de quadrados de resíduos do modelo nulo (SQR = 49665.30) para o modelo de retas coincidentes (SQR=6210.03), de tal forma que o modelo de retas coincidentes é preferível ao modelo nulo;
2. Há uma redução significativa ($p<0.0001$) na soma de quadrados de resíduos do modelo de retas coincidentes (SQR = 6210.03) para o modelo de retas paralelas (SQR=2587.18), de tal forma que o modelo de retas paralelas é preferível ao modelo de retas coincidentes;

Exemplo- Efeito da puromicina

- Com base nos resultados da análise de variância, podemos concluir:

1. Há uma redução significativa ($p<0.0001$) na soma de quadrados de resíduos do modelo nulo ($SQR = 49665.30$) para o modelo de retas coincidentes ($SQR=6210.03$), de tal forma que o modelo de retas coincidentes é preferível ao modelo nulo;
2. Há uma redução significativa ($p<0.0001$) na soma de quadrados de resíduos do modelo de retas coincidentes ($SQR = 6210.03$) para o modelo de retas paralelas ($SQR=2587.18$), de tal forma que o modelo de retas paralelas é preferível ao modelo de retas coincidentes;
3. Há uma redução significativa ($p=0.0027$) na soma de quadrados de resíduos do modelo de retas paralelas ($SQR = 2587.18$) para o modelo de retas concorremtes ($SQR=1591.25$), de tal forma que o modelo de retas concorrentes é preferível ao modelo de retas paralelas.

Exemplo- Efeito da puromicina

- Desta forma, ao final do processo selecionamos o modelo mais complexo (retas concorrentes).

Exemplo- Efeito da puromicina

- Desta forma, ao final do processo selecionamos o modelo mais complexo (retas concorrentes).

- Assim, a relação entre a velocidade de reação e a (log) concentração de substrato é definida por diferentes interceptos e diferentes taxas de variação nos grupos `treated` e `untreated`.

Exemplo- Efeito da puromicina

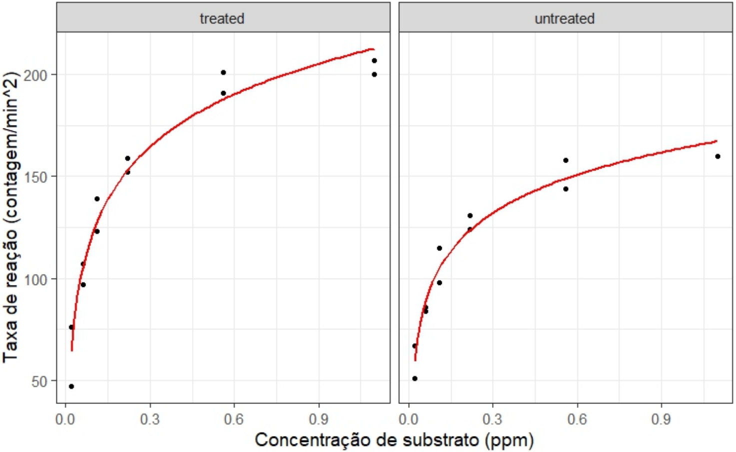

Figura 5: Dados de reação química em estudo sobre o efeito da puromicina com o modelo de regressão linear ajustado

Exemplo- Comportamento sexual de insetos

Exemplo- Comportamento sexual de insetos

- Nesta aplicação vamos considerar a base de dados `fruitfly` disponível na biblioteca `faraway` do `R`.

Exemplo- Comportamento sexual de insetos

- Nesta aplicação vamos considerar a base de dados `fruitfly` disponível na biblioteca `faraway` do `R`.

- A base de dados tem observações de 125 insetos machos em um experimento sobre seu comportamento sexual.

Exemplo- Comportamento sexual de insetos

- Nesta aplicação vamos considerar a base de dados `fruitfly` disponível na biblioteca `faraway` do `R`.

- A base de dados tem observações de 125 insetos machos em um experimento sobre seu comportamento sexual.

- As variáveis são as seguintes:

Exemplo- Comportamento sexual de insetos

- Nesta aplicação vamos considerar a base de dados `fruitfly` disponível na biblioteca `faraway` do `R`.

- A base de dados tem observações de 125 insetos machos em um experimento sobre seu comportamento sexual.

- As variáveis são as seguintes:
  - `longevity`: tempo de vida em dias (resposta);

Exemplo- Comportamento sexual de insetos

- Nesta aplicação vamos considerar a base de dados `fruitfly` disponível na biblioteca `faraway` do `R`.

- A base de dados tem observações de 125 insetos machos em um experimento sobre seu comportamento sexual.

- As variáveis são as seguintes:
  - `longevity`: tempo de vida em dias (resposta);
  - `actitivity`: um fator com cinco níveis, referente à consição sexual experimental (`isolated`, `one`, `many`, `low` e `high`);

Exemplo- Comportamento sexual de insetos

- Nesta aplicação vamos considerar a base de dados `fruitfly` disponível na biblioteca `faraway` do `R`.

- A base de dados tem observações de 125 insetos machos em um experimento sobre seu comportamento sexual.

- As variáveis são as seguintes:
  - `longevity`: tempo de vida em dias (resposta);
  - `actitivity`: um fator com cinco níveis, referente à consição sexual experimental (`isolated`, `one`, `many`, `low` e `high`);
  - `thorax`: tamanho do tórax.

Exemplo- Comportamento sexual de insetos

- O objetivo é comparar os tempos de vida nos cinco grupos experimentais, ajustando também o efeito do tamanho do tórax.

Exemplo- Comportamento sexual de insetos

- O objetivo é comparar os tempos de vida nos cinco grupos experimentais, ajustando também o efeito do tamanho do tórax.

- A análise será realizada usando os scripts `R` disponibilizados.

Exemplo- Comportamento sexual de insetos

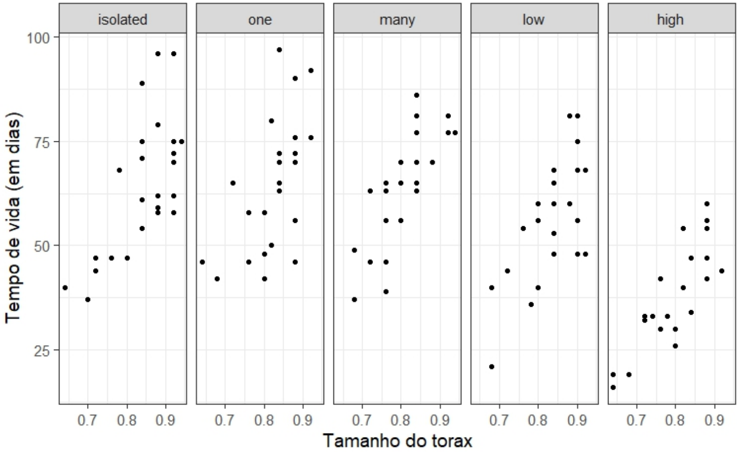

Figura 6: Dados de experimento sobre o comportamento sexual de insetos

Exercícios

Exercícios

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