MODELOS ESTATÍSTICOS APLICADOS A RELAÇÕES DENDROMÉTRICAS

Uma Viagem pela Estatística

Prof. Luan Fiorentin

Laboratório de Estatística e Geoinformação
Departamento de Estatística
Universidade Federal do Paraná

22 de fevereiro de 2021

Uma viagem pela Estatística!

- Introdução.
- Biometria Florestal.
- Modelos Lineares.
- Modelos Lineares Generalizados.
- Modelos Lineares Generalizados Mistos.
- Modelos Não Lineares.
- Modelos Não Lineares Generalizados.
- Modelos Não Lineares Generalizados Mistos.
- Modelos Multivariados.
- Outras Classes de Modelos.
- Considerações Finais.

Figura 1. Foto de Lukas no Pexels.

Introdução

- **Inventário florestal** é um procedimento para coletar informações quali-quantitativas de características da floresta.
- Logo, devemos concluir a respeito de características da floresta a partir de uma **amostra**.
- Principais **variáveis**:
  - Volume de madeira na floresta;
  - Altura das árvores;
  - Número de espécies;
  - Entre outras.

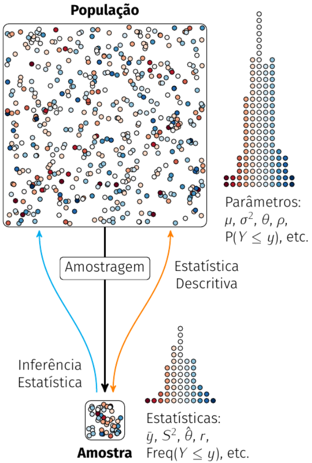

Figura 2. Extraído de Walmes Zeviani. Disponível em: http://leg.ufpr.br/ walmes/Tikz.

Biometria Florestal

- É a área da Ciência que estuda a **medição das árvores e das florestas**, com finalidade de entender o seu crescimento ao longo do tempo.
  - Volume; Altura; Diâmetro; entre outras.
- **Relações Dendrométricas** são associações desenvolvidas entre variáveis mensuradas em árvores.
  - Altura e diâmetro (relação hipsométrica);
  - Volume e diâmetro;
  - Diâmetro de copa e diâmetro;
  - Espessura de casca e idade.

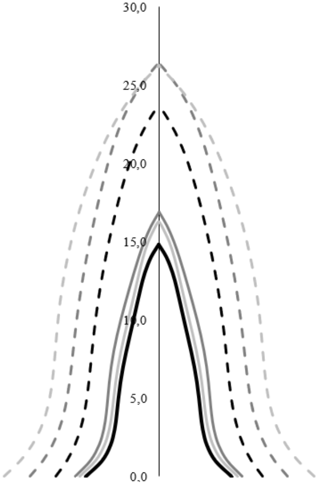

Figura 3. Dinâmica da forma da árvore.

Biometria Florestal

- Variáveis de **difícil** obtenção:
  - Volume total/individual;
  - Volume de casca de madeira;
  - Altura total/comercial;
  - Diâmetro de copa;
  - Entre outras.
- Podemos coletar informações **específicas** das árvores na floresta.
- **Relacionar** variáveis de fácil acesso com variáveis complexas.

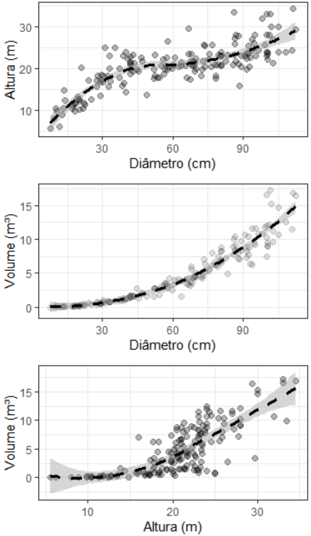

Figura 4. Extraído de Fiorentin et al. (2019).

Biometria Florestal

- Na modelagem dendrométrica, procuramos estudar a **relação** entre duas ou mais variáveis Y e X:
$$Y = f(X).$$
- Cada observação é composta por duas partes:
  - **Previsível** (ou controlada).
  - **Aleatória** (não previsível ou não controlada).

$$\text{observação} = \text{previsível} + \text{aleatória}.$$

  - Parte previsível incorpora o **conhecimento sobre a variável**, sendo expressa por uma função matemática.
  - Parte aleatória deve obedecer algum **modelo de probabilidade**.

- **Matemáticamente**, podemos descrever a relação como

$$y_i = \theta + \epsilon_i,$$

em que:

  - y_i é a observação i da variável resposta Y;
  - θ é um parâmetro de efeito fixo, comum a todas as observações;
  - ϵ_i é o erro da observação i, ou efeito resídual ou aleatório, sendo $\epsilon_i = y_i - \theta$.
  - ϵ_i pode ser resultante de outras variáveis que não foram controladas (ou não são controláveis). Logo, elas não estão explícitas no modelo.

Biometria Florestal - Exemplo 1

- **Exemplo**: Interesse em estimar a altura total de um povoamento florestal.
  - Foram mensuradas as variáveis altura, diâmetro e idade em nível de indivíduo.
  - ***Pergunta***: o que fazer com as árvores em que a altura não foi mensurada no inventário florestal?
  - ***Resposta***: estimar a altura das árvores.

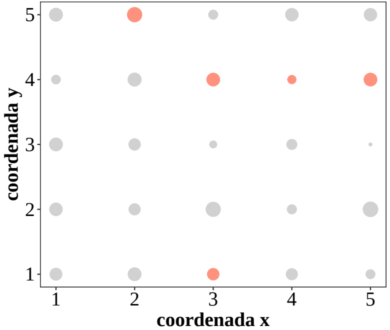

Figura 5. Árvores amostradas em colorido.

Biometria Florestal - Exemplo 1

- **Exemplo**: Amostra coletada:

altura	diâmetro
8.0	15.0
8.3	15.5
10.0	16.0
11.7	16.5
12.0	17.0

Podemos estimar a altura das árvores (y_i) de duas formas:

- Usando apenas os valores de y_i.
- Usando a relação de y_i com diâmetro das árvores (x_i).

- Modelo constante:

$$y_i = \theta + \epsilon_i$$

- Modelo linear simples:

$$y_i = \beta_0 + \beta_1 \text{diâmetro}_i + \epsilon_i$$

Biometria Florestal - Exemplo 1

- **Exemplo**: Considerando que a altura média das árvores é $\hat{\mu} = 10$ metros, então $\hat{y}_i$ pode ser descrita por:

$$\hat{y}_i = 10 + e_i,$$

em que:

- $\hat{\theta} = \hat{\mu}$;
- e_i determinará a altura de cada árvore em função de diversos fatores:

$$e_i = f(\text{diâmetro}, \text{sítio}, \text{idade}, \ldots).$$

Conforme relacionamos a altura com outras variáveis, ganhamos informação e **diminuímos o erro**.

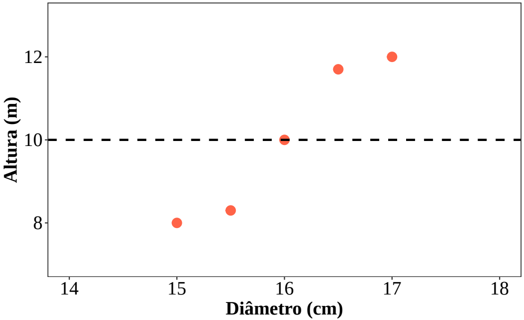

Figura 6. Modelagem da Altura.

Biometria Florestal - Exemplo 1

- **Exemplo**: Considerando que a altura média das árvores é uma f(diâmetro), então $\hat{y}_i$ pode ser descrita por:

$$\hat{y}_i = -26.48 + 2.28\text{diâmetro}_i + e_i,$$

em que:

  - $\hat{\theta} = \hat{\mu}$;
  - e_i determinará a altura de cada árvore em função de diversos fatores:

$$e_i = f(\text{sítio}, \text{idade}, \ldots).$$

Conforme relacionamos a altura com outras variáveis, ganhamos informação e **diminuímos o erro**.

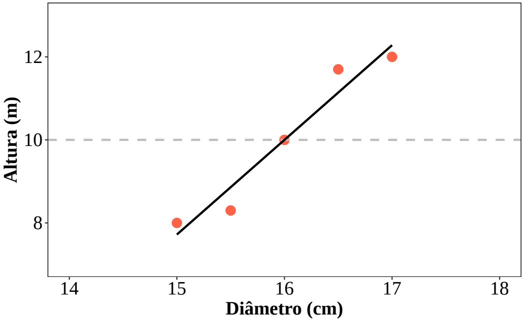

Figura 7. Modelagem da Altura dado o Diâmetro.

Modelos Lineares

- É uma classe de modelos onde o **parâmetro** ocorre de forma **linear**.
- Casos **especiais**:
  - Análise de variância (ANOVA).
  - Análise de regressão linear.
- **Análise de variância**: compara simultâneamente k médias.
- **Análise de regressão linear**: estuda a relação entre variáveis Y e X:
  - Regressão linear simples.
  - Regressão linear múltipla.

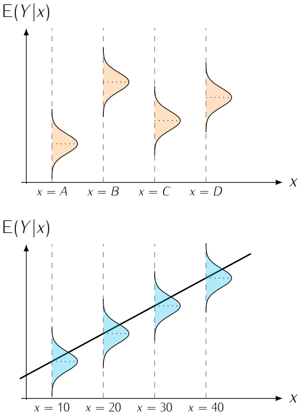

Figura 8. Representações esquemáticas dos modelos de ANOVA e regressão linear simples.

Modelos Lineares - Regressão Linear Simples

- **Modelo de regressão linear simples** é definido por uma reta que estabelece a relação entre uma variável resposta (y_i) e uma explicativa (x_i):

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

  em que:
  - i é o índice da observação;
  - β_0 é o intercepto da reta;
  - β_1 é o coeficiente angular da reta;
  - ϵ é o erro aleatório.
- Definimos que os erros seguem uma **distribuição normal** de média zero e variância constante, ou seja, $\epsilon \sim N(0, \sigma^2)$.
- Ainda, supomos que erros associados a diferentes observações sejam **não correlacionados**, então $Cov(\epsilon_i, \epsilon_i^{'}) = 0$.

Modelos Lineares - Regressão Linear Simples

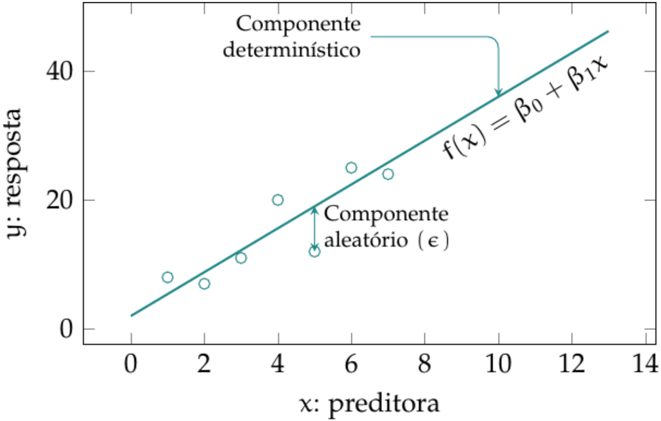

Figura 9. Componentes do modelo. Extraído de Walmes Zeviani. Disponível em: http://leg.ufpr.br/ walmes/Tikz.

Modelos Lineares - Regressão Linear Simples

- **Média** de y_i:

$$E(y_i|x_i) = \beta_0 + \beta_1 x_i.$$

- **Variância** de y_i:

$$Var(y_i|x_i) = \sigma^2.$$

- **Interpretação** dos parâmetros:
  - β_1 expressa a alteração no valor esperado de y_i associado ao acréscimo de uma unidade em x_i.
  - β_0 é o valor esperado de y_i quando $x_i = 0$.

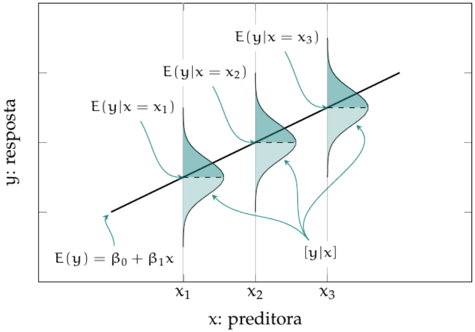

Figura 10. Valor esperado da resposta. Extraído de Walmes Zeviani. Disponível em: http://leg.ufpr.br/ walmes/Tikz.

Modelos Lineares - Regressão Linear Simples

- **Valores preditos**: A predição de valores estimados para a resposta depende de valores da variável preditora:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i.$$

- **Resíduos** são erros de predição ao se utilizar um modelo para descrever a relação entre variáveis:

$$e_i = y_i - \hat{y}_i.$$

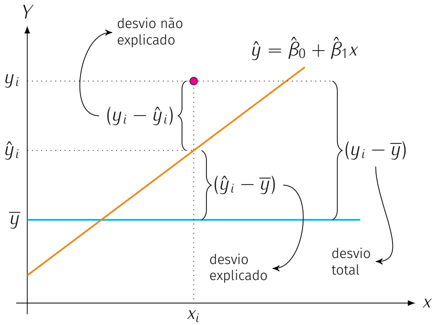

Figura 11. Resíduos. Extraído de Walmes Zeviani. Disponível em: http://leg.ufpr.br/ walmes/Tikz.

Modelos Lineares - Regressão Linear Simples

- **Estimação dos parâmetros**: Note que o resíduo (ϵ_i) é função dos parâmetros desconhecidos β_0 e β_1.
- Precisamos de um **critério** para definir valores ótimos de parâmetros.
- Critério de **mínimos quadrados** consiste em minimizar a soma de quadrados dos resíduos (SSE):

$$SSE(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2.$$

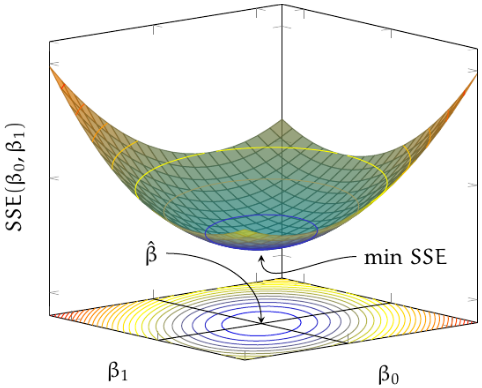

Figura 12. Minimização da SSE. Extraído de Walmes Zeviani. Disponível em: http://leg.ufpr.br/ walmes/Tikz.

Modelos Lineares - Regressão Linear Simples

- Abordagem padrão é usar **cálculo diferencial**:
  - Obter **vetor gradiente** (vetor de derivadas parciais) de β_0 e β_1;
  - Resolver o **sistema de equações lineares** (igualar a zero e isolar β_0 e β_1):

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x},$$

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

sendo $\bar{x}$ e $\bar{y}$ a média de cada variável. Já a variância é

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}.$$

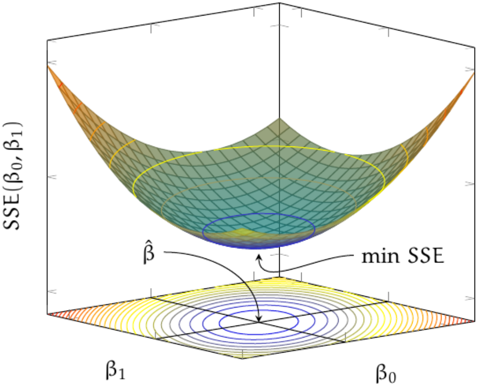

Figura 13. Minimização da SSE. Extraído de Walmes Zeviani. Disponível em: http://leg.ufpr.br/ walmes/Tikz.

Modelos Lineares - Regressão Linear Simples - Exemplo 2

- O problema consiste em **estimar a altura total** de árvores de um inventário florestal realizado em determinada idade.
- **Dados** disponíveis:
  - Altura (H) $\rightarrow$ Resposta.
  - Diâmetro (D) $\rightarrow$ Preditora.
- Modelo **teórico**:

$$H_i = \beta_0 + \beta_1 D_i + \epsilon_i.$$

- **Valor esperado**:

$$E(H_i) = \hat{\beta}_0 + \hat{\beta}_1 D_i.$$

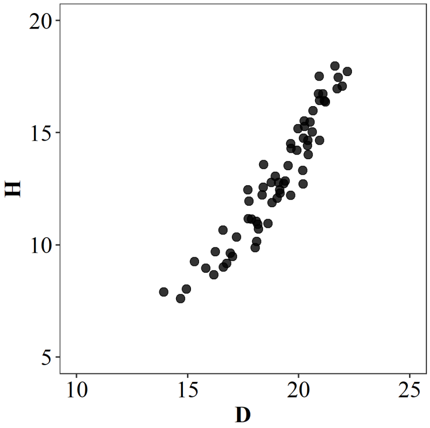

Figura 14. Exemplo 2

Modelos Lineares - Regressão Linear Simples - Exemplo 2

- O problema consiste em **estimar a altura total** de árvores de um inventário florestal realizado em determinada idade.
- **Dados** disponíveis:
  - Altura (H) $\rightarrow$ Resposta.
  - Diâmetro (D) $\rightarrow$ Preditora.
- Modelo **teórico**:

$$H_i = \beta_0 + \beta_1 D_i + \epsilon_i.$$

- **Valor esperado**:

$$E(H_i) = \hat{\beta}_0 + \hat{\beta}_1 D_i.$$

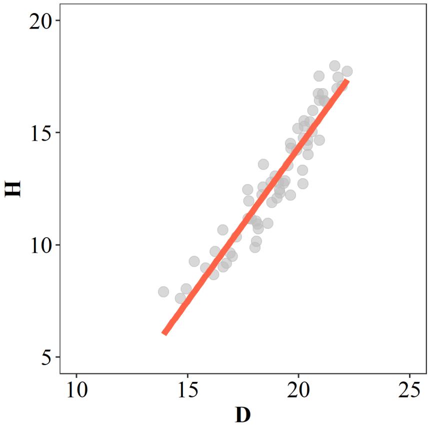

Figura 15. Exemplo 2

Modelos Lineares - Regressão Linear Múltiplo

- **Modelo de regressão linear múltiplo** é definido por função que estabelece a relação entre uma variável resposta (y_i) e múltiplas explicativas ($x_{i1}, \ldots, x_{ip}$):

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \epsilon_i,$$

  em que:
  - i é o índice da observação;
  - β_0 é o intercepto;
  - β_j é o parâmetro, sendo $j = 1, \ldots, p$;
  - x_j é a variável preditora, sendo $j = 1, \ldots, p$;
  - ϵ é o erro aleatório.
- Definimos que os erros seguem uma **distribuição normal** de média zero e variância constante, ou seja, $\epsilon \sim N(0, \sigma^2)$.
- Ainda, supomos que erros associados a diferentes observações sejam **não correlacionados**, então $Cov(\epsilon_i, \epsilon_i^{'}) = 0$.

Modelos Lineares - Regressão Linear Múltiplo

- **Média** de y_i:
$$E(y_i|\mathbf{x_i} = (x_{i1}, \ldots, x_{ip})^{'})) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}.$$
- **Variância** de y_i:
$$Var(y_i|\mathbf{x_i} = (x_{i1}, \ldots, x_{ip})^{'})) = \sigma^2.$$
- **Interpretação** dos parâmetros:
  - β_j representa a alteração esperada na resposta (y_i) para uma unidade a mais em x_{ij}, quando as demais preditoras são mantidas fixas.

Modelos Lineares - Regressão Linear Múltiplo

- **Valores preditos**: A predição de valores estimados para a resposta depende de valores das variáveis preditoras:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_p x_{ip}.$$

- **Resíduos** são erros de predição ao se utilizar um modelo para descrever a relação entre variáveis:

$$e_i = y_i - \hat{y}_i.$$

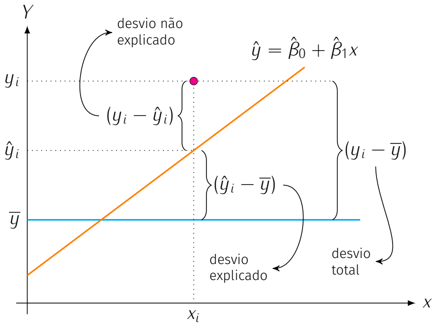

Figura 16. Resíduos. Extraído de Walmes Zeviani. Disponível em: http://leg.ufpr.br/ walmes/Tikz.

Modelos Lineares - Regressão Linear Múltiplo

- **Notação matricial**: Modelo de regressão linear múltiplo é dado por

$$\boldsymbol{y} = \boldsymbol{x}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

em que:

- $\boldsymbol{y} = (y_1, \ldots, y_n)'$ de dimensão $n \times 1$;
- $\boldsymbol{x} = (1_i, x_{i1}, \ldots, x_{np})'$ de dimensão $n \times (p+1)$;
- $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$ de dimensão $(p+1) \times 1$;
- $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)'$ de dimensão $n \times 1$.

Modelos Lineares - Regressão Linear Múltiplo

- **Estimação dos parâmetros**: Estimação de mínimos quadrados baseia-se, novamente, na determinação de $\beta_0, \beta_1, \ldots, \beta_p$ que minimizem a soma de quadrados de resíduos dada por

$$SSE(\beta_0, \beta_1, \ldots, \beta_p) = \sum_{i=1}^{n}(y_i - (\beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}))^2.$$

  Em notação matricial: $SSE(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{x}\boldsymbol{\beta})^{'}(\boldsymbol{y} - \boldsymbol{x}\boldsymbol{\beta})$.

- Estimador de mínimos quadrados em notação matricial é dado por

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{x}^{'}\boldsymbol{x})^{-1}\boldsymbol{x}^{'}\boldsymbol{y}.$$

- Estimador de variância é dado por

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n - p}.$$

Modelos Lineares - Regressão Linear Múltiplo - Exemplo 3

- O problema consiste em **estimar o volume total** de árvores de um inventário florestal realizado em determinada idade.
- **Dados** disponíveis:
  - Volume (H) $\rightarrow$ Resposta.
  - Diâmetro (D) $\rightarrow$ Preditora.
- Modelo **teórico**:

$$V_i = \beta_0 + \beta_1 D_i + \beta_2 D_i^2 + \epsilon_i.$$

- **Valor esperado**:

$$E(H_i) = \hat{\beta}_0 + \hat{\beta}_1 D_i + \hat{\beta}_2 D_i^2.$$

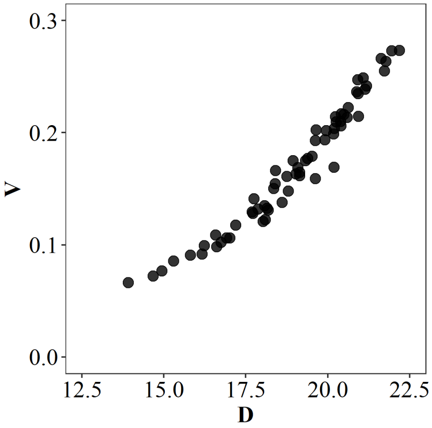

Figura 17. Exemplo 3

Modelos Lineares - Regressão Linear Múltiplo - Exemplo 3

- O problema consiste em **estimar o volume total** de árvores de um inventário florestal realizado em determinada idade.
- **Dados** disponíveis:
  - Volume (H) $\rightarrow$ Resposta.
  - Diâmetro (D) $\rightarrow$ Preditora.
- Modelo **teórico**:

$$V_i = \beta_0 + \beta_1 D_i + \beta_2 D_i^2 + \epsilon_i.$$

- **Valor esperado**:

$$E(H_i) = \hat{\beta}_0 + \hat{\beta}_1 D_i + \hat{\beta}_2 D_i^2.$$

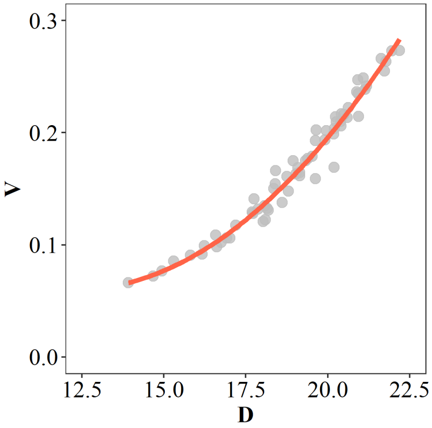

Figura 18. Exemplo 3

Modelos Lineares Generalizados

- O comportamento da resposta pode ser estudado por meio de modelos **probabilísticos**.
- Modelos disponíveis para variáveis:
  - Discreta $\rightarrow$ Poisson;
  - Contínua $\rightarrow$ Normal, Gamma;
  - Binária $\rightarrow$ Binomial, Hipergeométrica;
  - Limitada em intervalo $\rightarrow$ Beta.
- Escolha do modelo depende da **natureza** da resposta e análises **gráficas**.

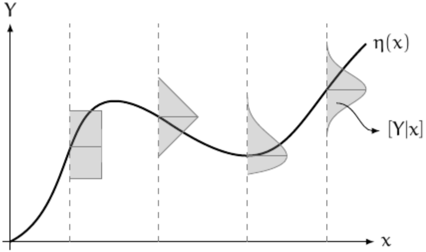

Figura 19. Modelo linear generalizado. Extraído de Walmes Zeviani. Disponível em: http://leg.ufpr.br/ walmes/Tikz.

Modelos Lineares Generalizados

- Definido pela **especificação** de três componentes:
  - **Componente aleatório**: conjunto de variáveis aleatórias independetes (y_i) com distribuição pertencente à família exponencial, e vetor de parâmetros $\boldsymbol{\theta}$, $\boldsymbol{\phi}$:

$$f_e(y_i; \boldsymbol{\theta}, \boldsymbol{\phi}) = exp\left\{\frac{y_i\boldsymbol{\theta} - b(\boldsymbol{\theta})}{a(\boldsymbol{\phi})} + c(y_i; \boldsymbol{\phi})\right\},$$

    em que a, b, c são funções adequadas.
  - **Componente sistemático**: é o preditor linear do modelo, em que as variáveis são inseridas por meio de uma combinação linear de parâmetros:

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}.$$

  - **Função de ligação**: É uma função real, monótona e diferenciável, que conecta o componente sistemático ao aleatório:

$$g(\mu_i) = \eta_i \quad \text{e} \quad \mu_i = g^{-1}(\eta_i),$$

    em que $\mu_i = E(y_i|\boldsymbol{x_i} = (x_{i1}, \ldots, x_{ip})')$.

Modelos Lineares Generalizados - Exemplo 4

- O problema consiste em **estimar o volume total** de árvores de um inventário florestal realizado em determinadas idades.
- **Dados** disponíveis:
  - Volume (V) $\rightarrow$ Resposta.
  - Idade (I) $\rightarrow$ Preditora.
- Modelo **teórico** com distribuição Gamma e função de ligação logarítmica:

$$V_i = exp(\beta_0 + \beta_1 I_i + \epsilon_i).$$

- **Valor esperado**:

$$E(V_i) = exp(\hat{\beta}_0 + \hat{\beta}_1 I_i).$$

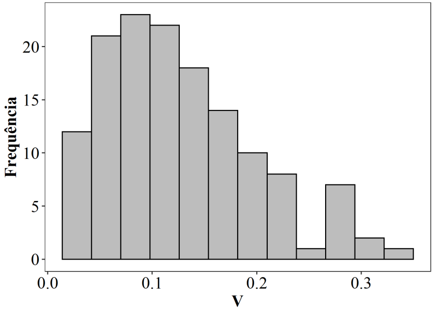

Figura 20. Exemplo 4

Modelos Lineares Generalizados - Exemplo 4

- O problema consiste em **estimar o volume total** de árvores de um inventário florestal realizado em determinadas idades.
- **Dados** disponíveis:
  - Volume (V) $\rightarrow$ Resposta.
  - Idade (I) $\rightarrow$ Preditora.
- Modelo **teórico** com distribuição Gamma e função de ligação logarítmica:

$$V_i = exp(\beta_0 + \beta_1 I_i + \epsilon_i).$$

- **Valor esperado**:

$$E(V_i) = exp(\hat{\beta}_0 + \hat{\beta}_1 I_i).$$

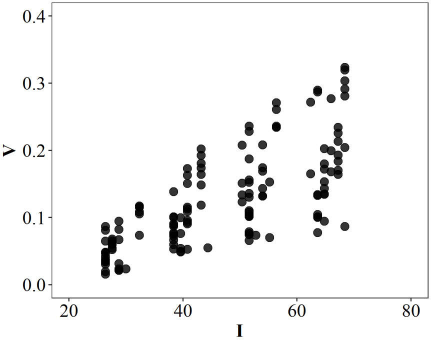

Figura 21. Exemplo 4. Dados extraídos do pacote forestmangr no R.

Modelos Lineares Generalizados - Exemplo 4

- O problema consiste em **estimar o volume total** de árvores de um inventário florestal realizado em determinadas idades.
- **Dados** disponíveis:
  - Volume (V) $\rightarrow$ Resposta.
  - Idade (I) $\rightarrow$ Preditora.
- Modelo **teórico** com distribuição Gamma e função de ligação logarítmica:

$$V_i = exp(\beta_0 + \beta_1 I_i + \epsilon_i).$$

- **Valor esperado**:

$$E(V_i) = exp(\hat{\beta}_0 + \hat{\beta}_1 I_i).$$

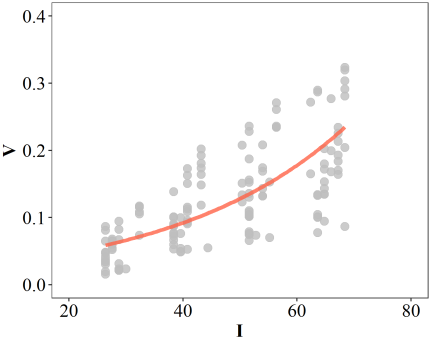

Figura 22. Exemplo 4. Dados extraídos do pacote forestmangr no R.

Modelos Lineares Generalizados Mistos

- Em **dados longitudinais**, a resposta é mensurada no mesmo indivíduo em diversas ocasiões ao longo do tempo.
- Observações tomadas ao longo do tempo são **não independentes**.
- Qual a solução? Incluir **efeitos aleatórios** no modelo.
- **Modelo misto**:

$$\text{Efeitos Fixos} + \text{Efeitos Aleatórios}.$$

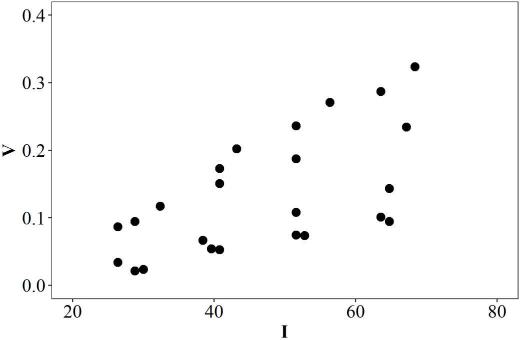

Figura 23. Observações tomadas em diferentes tempos. Dados extraídos do pacote forestmangr no R.

Modelos Lineares Generalizados Mistos

- Em **dados longitudinais**, a resposta é mensurada no mesmo indivíduo em diversas ocasiões ao longo do tempo.
- Observações tomadas ao longo do tempo são **não independentes**.
- Qual a solução? Incluir **efeitos aleatórios** no modelo.
- **Modelo misto**:

$$\text{Efeitos Fixos} + \text{Efeitos Aleatórios}.$$

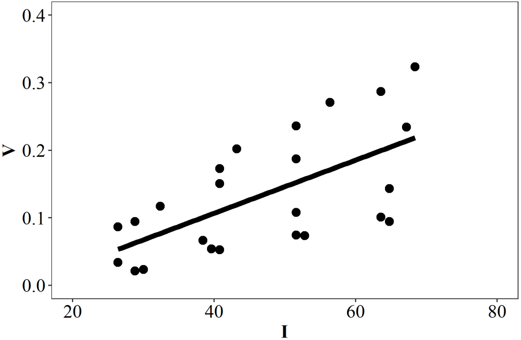

Figura 24. Observações tomadas em diferentes tempos. Dados extraídos do pacote forestmangr no R.

Modelos Lineares Generalizados Mistos

- Em **dados longitudinais**, a resposta é mensurada no mesmo indivíduo em diversas ocasiões ao longo do tempo.
- Observações tomadas ao longo do tempo são **não independentes**.
- Qual a solução? Incluir **efeitos aleatórios** no modelo.
- **Modelo misto**:

$$\text{Efeitos Fixos} + \text{Efeitos Aleatórios}.$$

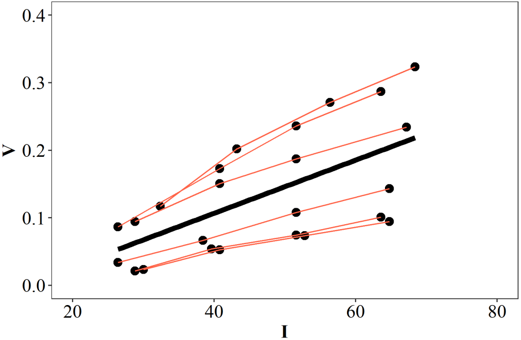

Figura 25. Observações tomadas em diferentes tempos. Dados extraídos do pacote forestmangr no R.

Modelos Lineares Generalizados Mistos

- **Independência** entre observações é obtida dado os efeitos aleatórios.
- Efeitos aleatórios induzem a uma **estrutura de correlação** entre as respostas medidas no mesmo indivíduo.
- Em geral, as correlações são positivas, mas **decrescem** ao longo do tempo.
- **Correlação positiva**: indivíduo que apresenta valor *acima da média* no tempo (t), tende a apresentar valor *acima da média* em tempo posterior ($t + u$).

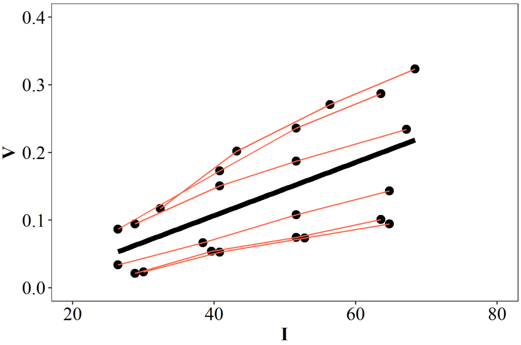

Figura 26. Observações tomadas em diferentes tempos. Dados extraídos do pacote forestmangr no R.

Modelos Lineares Generalizados Mistos

- **Notação matricial** utilizada para descrever o indivíduo i, dada por

$$\boldsymbol{y}_i = \boldsymbol{z}_i\boldsymbol{\beta} + \boldsymbol{x}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i,$$

  em que:

  - $\boldsymbol{y}_i$ é vetor de resposta;
  - $\boldsymbol{x}_i$ e $\boldsymbol{z}_i$ é matriz associada ao efeito fixo e aleatório, respectivamente;
  - $\boldsymbol{\beta}$ e $\boldsymbol{b}_i$ é vetor de efeito fixo e aleatório, respectivamente;
  - $\boldsymbol{b}_i \sim N(\mathbf{0}, \boldsymbol{G})$, sendo $\boldsymbol{G}$ a matriz de covariância dos parâmetros de efeito aleatório;
  - $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \boldsymbol{R}_i)$, sendo $\boldsymbol{R}_i = \sigma^2\boldsymbol{I}$ a matriz de covariância de resíduos com $\boldsymbol{I}$ a matriz identidade.

Modelos Lineares Generalizados Mistos

- **Efeitos fixos** descrevem características populacionais (β).
- **Efeitos aleatórios** descrevem características individuais (b_i).
- Portanto, modelos mistos permitem construir **trajetórias individuais**.
- Podemos indentificar **perfis** acima e abaixo da média populacional.

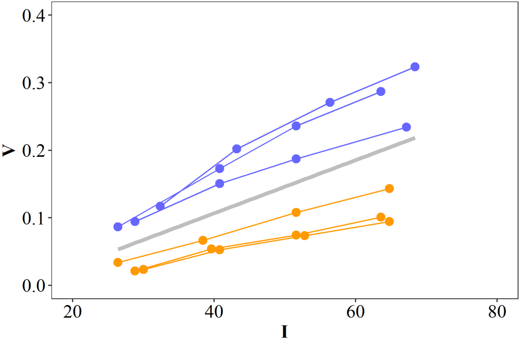

Figura 27. Observações tomadas em diferentes tempos. Dados extraídos do pacote forestmangr no R.

Modelos Lineares Generalizados Mistos - Exemplo 5

- O problema consiste em **estimar o volume total** de árvores de um inventário florestal realizado em múltiplas ocasiões.
- **Dados** disponíveis:
  - Volume (V) $\rightarrow$ Resposta.
  - Idade (I) $\rightarrow$ Preditora.

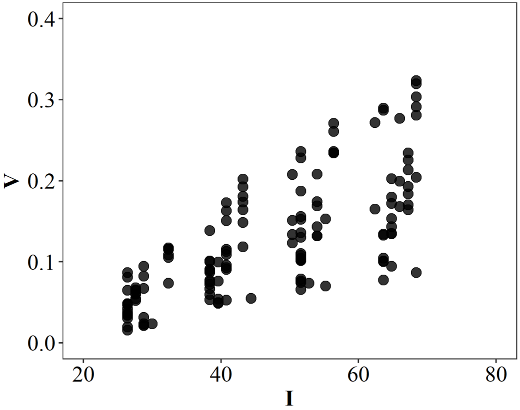

Figura 28. Exemplo 5. Dados extraídos do pacote forestmangr no R.

Modelos Lineares Generalizados Mistos - Exemplo 5

- O problema consiste em **estimar o volume total** de árvores de um inventário florestal realizado em múltiplas ocasiões.
- **Dados** disponíveis:
  - Volume (V) $\rightarrow$ Resposta.
  - Idade (I) $\rightarrow$ Preditora.

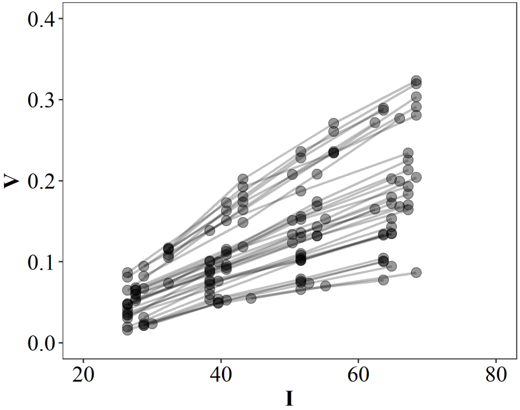

Figura 29. Exemplo 5. Dados extraídos do pacote forestmangr no R.

Modelos Lineares Generalizados Mistos - Exemplo 5

- O problema consiste em **estimar o volume total** de árvores de um inventário florestal realizado em múltiplas ocasiões.
- **Dados** disponíveis:
  - Volume (V) $\rightarrow$ Resposta.
  - Idade (I) $\rightarrow$ Preditora.
- Modelo linear misto **teórico**:

$$V_i = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})I_i + \epsilon_i.$$

- **Valor esperado**:

$$E(V_i|b_{0i},b_{1i}) = (\hat{\beta}_0 + b_{0i}) + (\hat{\beta}_1 + b_{1i}))I_i.$$

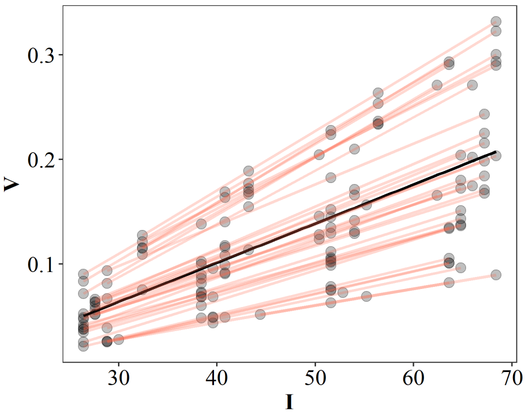

Figura 30. Exemplo 5. Dados extraídos do pacote forestmangr no R.

Modelos Não Lineares

- Descreve uma **relação não linear** entre a variável resposta (y_i) e a variável preditora (x_i).
- É uma **função não linear** do vetor de parâmetros ($\boldsymbol{\theta}$):

$$y_i = f(\boldsymbol{\theta}, x_i) + \epsilon_i.$$

- **Definição**: modelo é não linear se pelo menos uma das derivadas parciais de $f(\cdot)$ em relação a cada parâmetro do vetor $\boldsymbol{\theta}$ envolve parâmetro desconhecido.

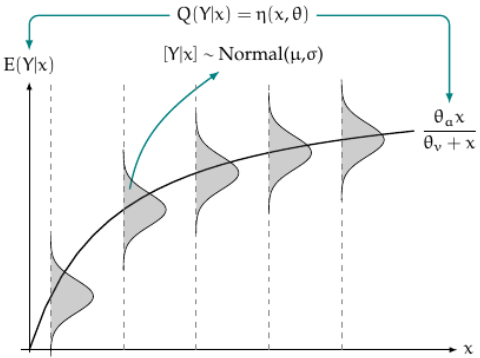

Figura 31. Valor esperado da resposta. Extraído de Walmes Zeviani. Disponível em: http://leg.ufpr.br/ walmes/Tikz.

Modelos Não Lineares

- **Vantagens**:
  - Escolha do modelo é baseada no comportamento do fenômeno;
  - Parâmetros interpretáveis;
  - Modelos mais parcimoniosos;
  - Predições mais confiáveis para fora do domínio da preditora.
- **Desvantagens**:
  - Necessidade de conhecimento técnico do fenômeno em estudo;
  - Requer procedimentos iterativos de estimação;
  - Métodos de inferências são aproximados.

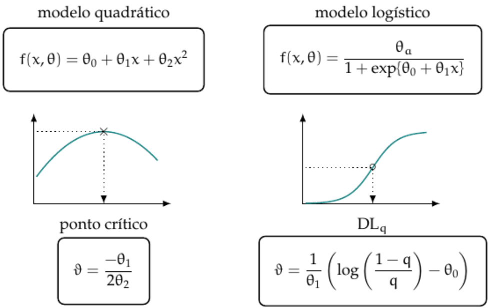

Figura 32. Modelo linear e não linear. Extraído de Walmes Zeviani. Disponível em: http://leg.ufpr.br/ walmes/Tikz.

Modelos Não Lineares Generalizados

- É uma **extensão** dos modelos lineares generalizados.
- O **valor esperado** da resposta é modelado como uma função não linear do vetor de parâmetros $\boldsymbol{\theta}$:

$$y_i = f(\boldsymbol{\theta}, x_i) + \epsilon_i,$$

em que:

  - $y_i \sim f_e(\cdot)$, ou seja, pertence a família exponencial.

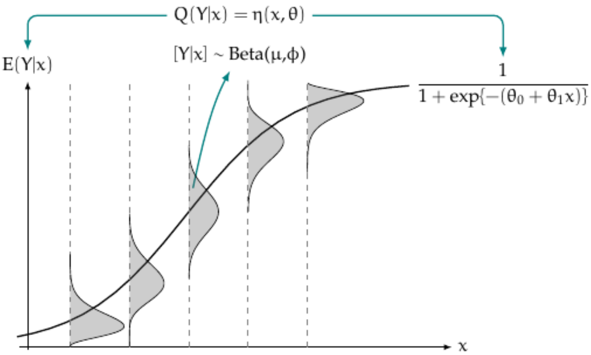

Figura 33. Modelo não linear generalizado. Extraído de Walmes Zeviani. Disponível em: http://leg.ufpr.br/ walmes/Tikz.

Modelos Não Lineares Generalizados Mistos

- É uma **extensão** dos modelos não lineares generalizados.
- O **valor esperado** da resposta é modelado como uma função não linear do vetor de parâmetros $\boldsymbol{\theta}$:

$$\boldsymbol{y}_i = f(\boldsymbol{W}_i) + \boldsymbol{\epsilon}_i,$$

em que:

  - $\boldsymbol{W}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i$;
  - $\boldsymbol{y}_i \sim f_e(\cdot)$, ou seja, pertence a família exponencial.

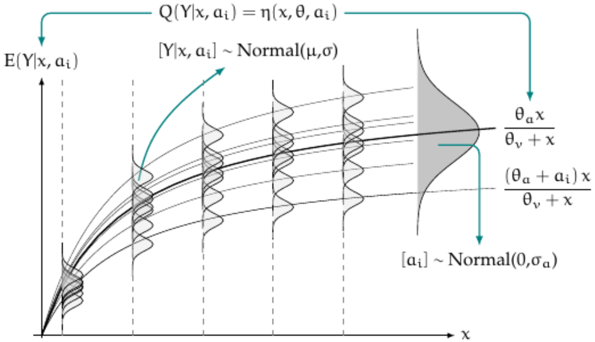

Figura 34. Modelo misto. Extraído de Walmes Zeviani. Disponível em: http://leg.ufpr.br/ walmes/Tikz.

Modelos Multivariados

- Modelo de regressão **multivariado** relaciona $r = 1, \ldots, R$ variáveis:

$$y_{ir} = \beta_{0r} + \beta_{1r}x_{ijr} + \ldots + \beta_{pr}x_{ipr} + \epsilon_{ir}$$

Modelos Multivariados

- Modelo de regressão **multivariado** relaciona $r = 1, \ldots, R$ variáveis:

$$y_{ir} = \beta_{0r} + \beta_{1r}x_{ijr} + \ldots + \beta_{pr}x_{ipr} + \epsilon_{ir}$$

$$y_{ir} = \beta_{0r} + \underbrace{\beta_{1r}x_{ijr}}_{\text{observações correlacionadas}} + \ldots + \beta_{pr}x_{ipr} + \epsilon_{ir}$$

Modelos Multivariados

- Modelo de regressão **multivariado** relaciona $r = 1, \dots, R$ variáveis:

$$y_{ir} = \beta_{0r} + \beta_{1r} x_{ijr} + \ldots + \beta_{pr} x_{ipr} + \epsilon_{ir}$$

$$y_{ir} = \beta_{0r} + \underbrace{\beta_{1r} x_{ijr}}_{\text{observações correlacionadas}} + \ldots + \beta_{pr} x_{ipr} + \epsilon_{ir}$$

$$y_{ir} = \beta_{0r} + \underbrace{\beta_{1r} x_{ijr} + \ldots + \beta_{pr} x_{ipr}}_{\text{covariáveis correlacionadas}} + \epsilon_{ir}$$

Modelos Multivariados

- Modelo de regressão **multivariado** relaciona $r = 1, \ldots, R$ variáveis:

$$y_{ir} = \beta_{0r} + \beta_{1r}x_{ijr} + \ldots + \beta_{pr}x_{ipr} + \epsilon_{ir}$$

$$y_{ir} = \beta_{0r} + \underbrace{\beta_{1r}x_{ijr}}_{\text{observações correlacionadas}} + \ldots + \beta_{pr}x_{ipr} + \epsilon_{ir}$$

$$y_{ir} = \beta_{0r} + \underbrace{\beta_{1r}x_{ijr} + \ldots + \beta_{pr}x_{ipr}}_{\text{covariáveis correlacionadas}} + \epsilon_{ir}$$

$$\underbrace{y_{ir}}_{\text{respostas correlacionadas}} = \beta_{0r} + \beta_{1r}x_{ijr} + \ldots + \beta_{pr}X_{ip} + \epsilon_{ir}$$

Modelos Multivariados

- Modelo multivariado modela **múltiplas respostas** simultaneamente.
- A potencial **correlação** entre respostas é levada em consideração.
- Medidas de **incertezas** podem ser estimadas de forma adequada.

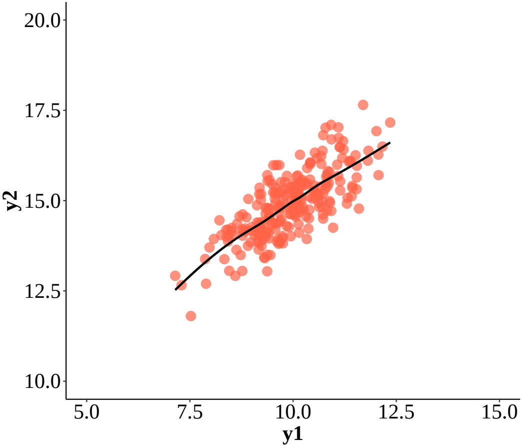

Figura 35. Variáveis respostas.

Modelos Multivariados - MCGLM

- BONAT e JØRGENSEN desenvolveram a classe ***Multivariate Covariance Generalized Linear Models*** (MCGLM):
  - Bastante flexível para modelar dados correlacionados univariados e multivariados;
  - Considerando respostas de diferentes naturezas;
  - E permite definir diversas estruturas de covariância.
- MCGLM é baseado em suposições de **segunda ordem**: média e covariância.
- **Covariâncias** são introduzidas usando uma combinação linear de matrizes conhecidas.
- Portanto, é uma classe de modelos com **grande potencial** no manejo florestal.

Modelos Multivariados - MCGLM

- **Formulação genérica** para um MCGLM é dada por

$$E(Y) = M = \{g_1^{-1}(X_1\boldsymbol{\beta}_1), \ldots, g_R^{-1}(X_R\boldsymbol{\beta}_R)\},$$

$$Var(\mathrm{Y}) = C = \boldsymbol{\Sigma}_R \otimes \boldsymbol{\Sigma}_b.$$

A matriz de covariância $\boldsymbol{\Sigma}_r$ para cada resposta é dada por

$$\boldsymbol{\Sigma}_r = V(\boldsymbol{\mu}_r; \boldsymbol{p}_r)^{1/2} h\{\Omega(\boldsymbol{\tau}_r)\} V(\boldsymbol{\mu}_r; \boldsymbol{p}_r)^{1/2},$$

$\boldsymbol{\Sigma}_R$ é uma $N \times N$ matriz de covariância dentro da resposta $r = 1,\ldots R$; $\boldsymbol{\Sigma}_b$ é uma matriz de correlação entre respostas; $V(\boldsymbol{\mu}_r; \boldsymbol{p}_r)$ é uma matriz diagonal, cujas entradas principais denotam a função de variância; $\boldsymbol{p}_r$ é um vetor de parâmetros de potência; $h\{\Omega(\tau_r)\} = \tau_0 Z_0 + \ldots + \tau_D Z_D$; e h é uma função de ligação de covariância.

Modelos Multivariados - Exemplo 6

- **Preditor linear**:
  - Resposta: altura (H) e volume (V).
  - Covariável: diâmetro (D).

$$E(H_i) = \hat{\beta}_0 + \hat{\beta}_1 D_i + \hat{\beta}_2 D_i^2 + \hat{\beta}_3 D_i^3$$

$$E\,(V_i) = exp(\hat{\beta}_0 + \hat{\beta}_1 D_i + \hat{\beta}_2 D_i^2)$$

- **Preditor linear de matriz**:
  - Modelagem da variância é uma função de D.
  - Função de variância inserida para V.

$$\boldsymbol{\Omega}(\boldsymbol{\tau}_r) = \hat{\tau}_0 I$$

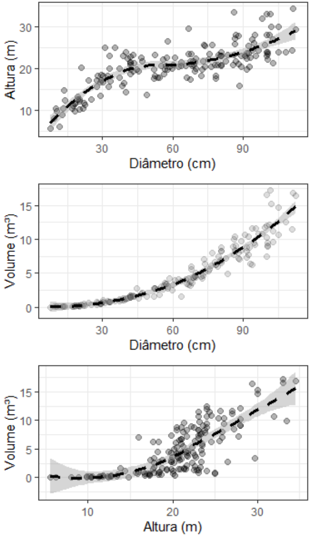

Figura 36. Extraído de Fiorentin et al. (2019).

Modelos Multivariados - Exemplo 6

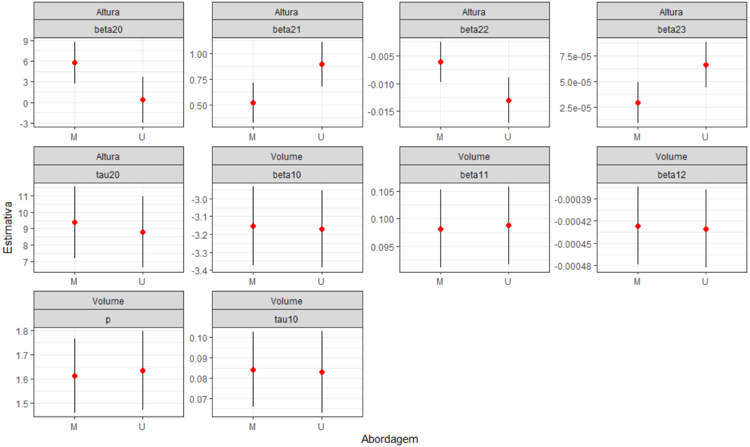

Figura 37. Extraído de Fiorentin et al. (2019).

Modelos Multivariados - Exemplo 6

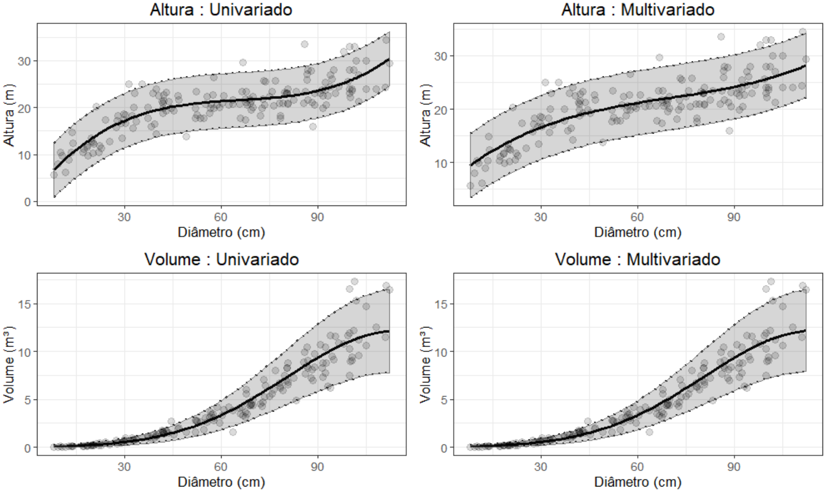

Figura 38. Extraído de Fiorentin et al. (2019).

Outras Classes de Modelos

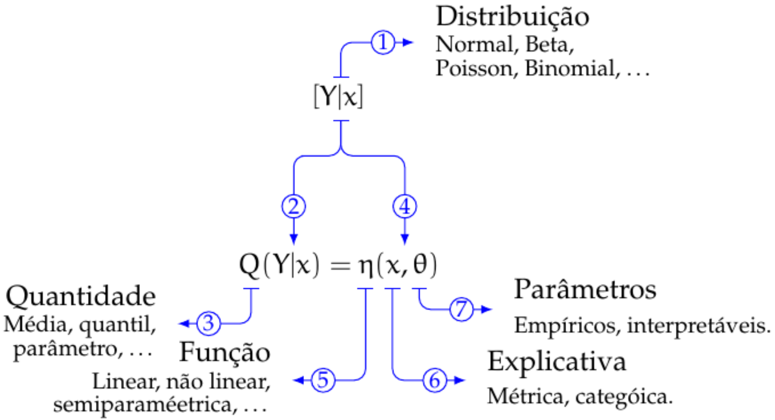

Figura 39. Componentes do modelo. Extraído de Walmes Zeviani. Disponível em: http://leg.ufpr.br/ walmes/Tikz.

Outras Classes de Modelos

- Há diversas classes de modelos:
  - Modelos Não Lineares;
  - Modelos Não Lineares Generalizados;
  - Modelos Não Lineares Generalizados Mistos;
  - Entre outros.
- Modelos de *Statistical Learning*:
  - *Neural Network*;
  - *Support Vector Machine*;
  - *Random Forest*;
  - Entre outros.
- Características dos modelos:
  1. **Interpretabilidade**.
  2. **Flexibilidade**.

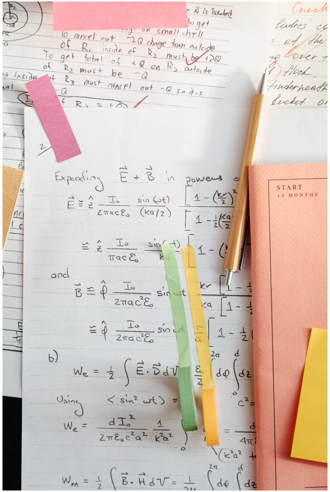

Figura 40. Foto de cottonbro no Pexels

Outras Classes de Modelos

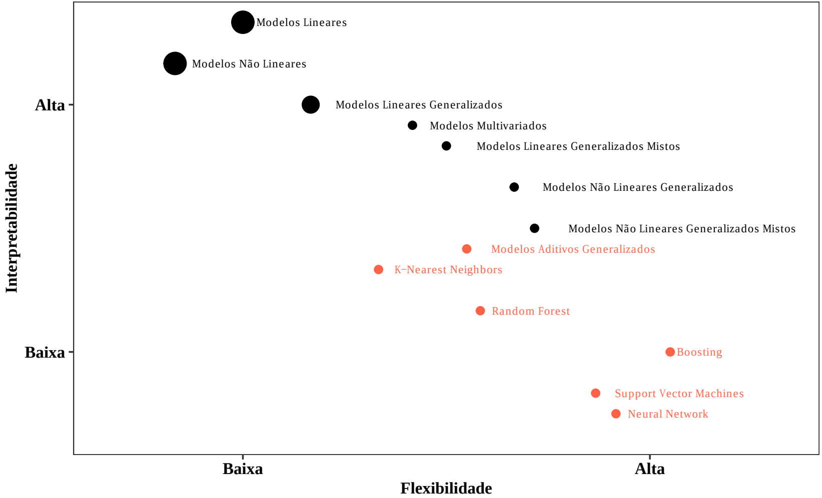

Figura 41. Relação entre interpretabilidade e flexibilidade de classes de modelos. Figura produzida pelo Autor.

Outras Classes de Modelos - *Softwares*

- Opções de ***softwares* pagos**:
  - SPSS;
  - SAS;
  - Stata;
  - Minitab;
  - Entre outros.
- Opções de ***softwares* não pagos**:
  - R.
- Linguagens de **programação**:
  - R;
  - Python;
  - Julia.

Figura 42. Logo do R.

Considerações Finais

Principais tópicos

- Modelagem de **relações dendrométricas** é fundamental na análise de dados.
- Natureza das variáveis e estrutura dos dados indica o **modelo estatístico** adequado.
- Portanto, é importante desenvolver **habilidades estatísticas**.

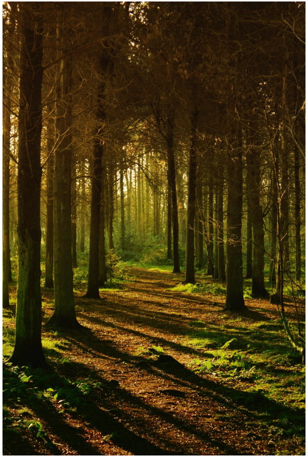

Figura 43. Foto de Anthony no Pexels.

That's all folks!!!

- **Contato**:
  - **E-mail**: luanfiorentin@hotmail.com
  - **Instagram**: luann_fiorentin
  - **Instagram** : fordata4d

Figura 44. Conheça a For Data!

That's all folks!!!

Figura 45. Foto de Pexels