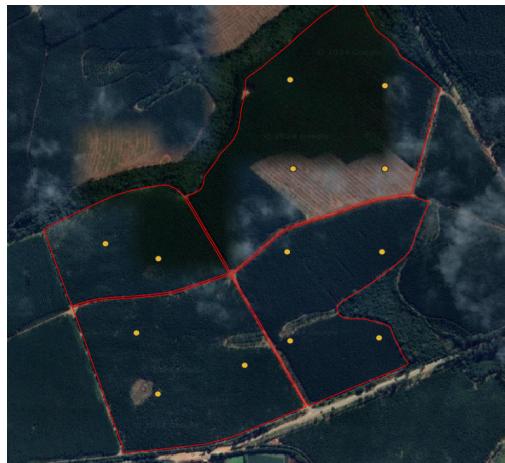


# Descrição das etapas de processamento de dados de inventário realizado com o auxílio da tecnologia LiDAR

Otávio Magalhães Silva Souza



Piracicaba, SP – Data de Emissão: 02 de setembro de 2024

# Contents

<b>1 Pacotes utilizados no R (colocar breve descrição - já tem uma descriçaozinha no R passado em aula)</b>	<b>3</b>
1.1 Tidyverse ( <a href="https://livro.curso-r.com/4-2-tidyverse.html">https://livro.curso-r.com/4-2-tidyverse.html</a> ) . . . . .	3
1.2 Sf . . . . .	3
1.3 Tidyterra . . . . .	4
1.4 Terra . . . . .	4
1.5 Stars . . . . .	4
1.6 Tools . . . . .	4
1.7 RColorBrewer . . . . .	4
1.8 Progress . . . . .	4
1.9 Reshape2 . . . . .	4
1.10 Mapview . . . . .	4
1.11 LidR . . . . .	4
1.12 RCSF . . . . .	4
1.13 Future . . . . .	4
<b>2 Descrição da área</b>	<b>5</b>
<b>3 Grid e parcelas já inventariadas</b>	<b>6</b>
<b>4 Processamento da nuvem LiDAR</b>	<b>7</b>
4.1 Fluxograma e plots relacionados à Dupla Amostragem proposta em Inventário Florestal . . .	8
<b>5 Dupla Amostragem</b>	<b>23</b>
<b>6 Fluxograma e etapas Tripla amostragem</b>	<b>24</b>
<b>7 Regressão linear simples</b>	<b>25</b>
<b>8 Arborimetria preditiva</b>	<b>26</b>
<b>9 Silvimetria com medidas auxiliares</b>	<b>28</b>
<b>10 R utilizando database grisons</b>	<b>29</b>
<b>11 Resultado</b>	<b>31</b>
<b>12 Estatística de confiança (ACS)</b>	<b>31</b>

# 1 Pacotes utilizados no R (colocar breve descrição - já tem uma descriçaozinha no R passado em aula)

## 1.1 Tidyverse (<https://livro.curso-r.com/4-2-tidyverse.html>)

O Tidyverse é um pacote guarda-chuva e contém diversas funções úteis para garantir o dinamismo no script, visualização, processamento e análise dos dados, modelagem etc.

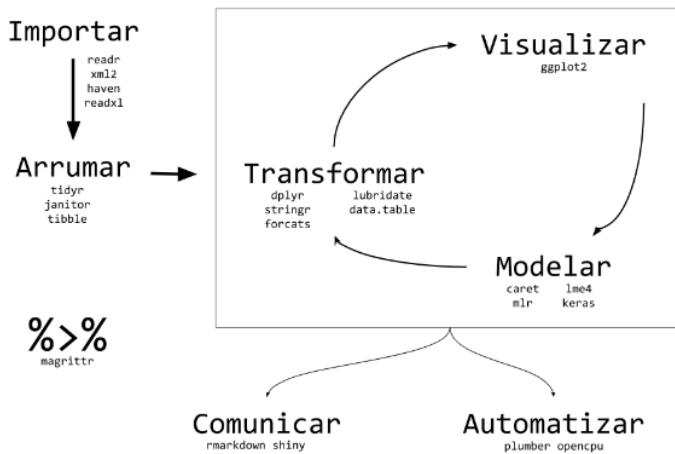


Figure 1: Tidyverse

## 1.2 Sf

Pacote utilizado para manipulação de objetos do mundo real. Descreve a forma com que esses objetos podem ser armazenados e importados e quais operações geométricas podem ser definidas por eles.

**1.3 Tidyterra**

**1.4 Terra**

**1.5 Stars**

**1.6 Tools**

**1.7 RColorBrewer**

**1.8 Progress**

**1.9 Reshape2**

**1.10 Mapview**

**1.11 LidR**

**1.12 RCSF**

**1.13 Future**

## 2 Descrição da área

A área a ser estudada como “Fazenda Modelo” localiza-se no município de São Miguel Arcanjo (SP), pode ser identificada pelas coordenadas (-23.86707°, -47.87772°) e possui 129,784 ha, que dividem-se em 4 subtalhões: 301a (18,933 ha), 301d (34,468 ha), 302a (47,602 ha) e 302c (28,781 ha).

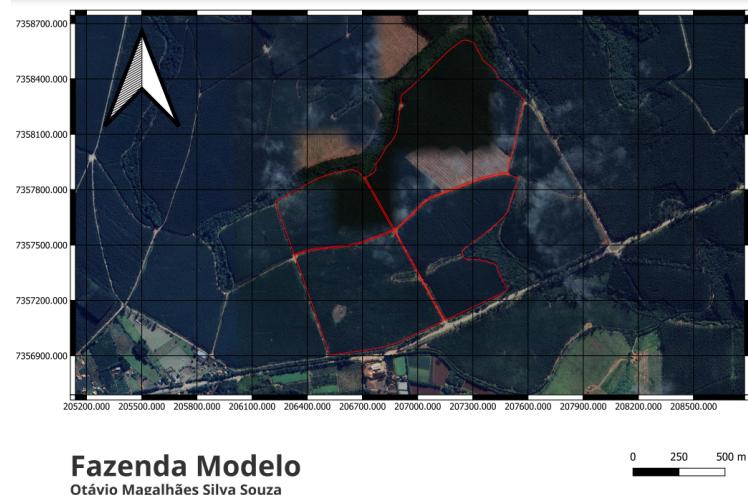


Figure 2: Mapa da propriedade

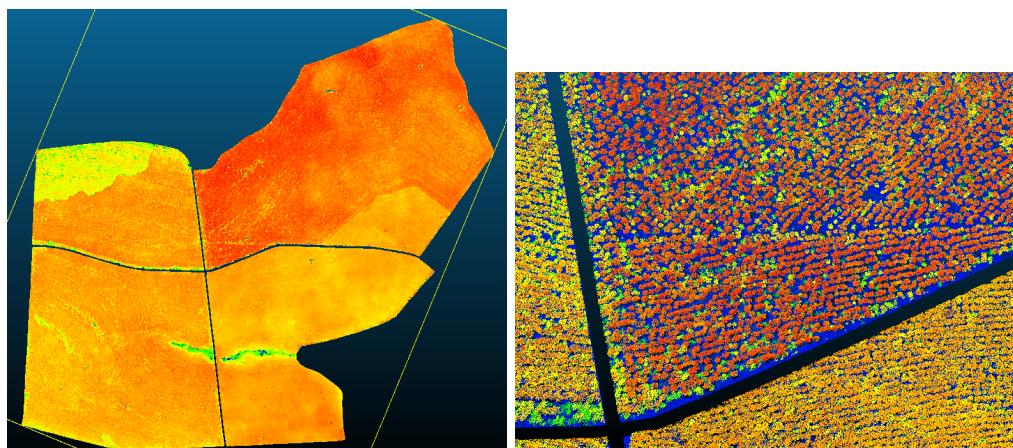


Figure 3: Nuvens LiDAR normalizadas

### 3 Grid e parcelas já inventariadas

A região foi dividida em 3454 parcelas, onde 2960 delas possuem 400m<sup>2</sup>, enquanto as outras são menores por estarem na borda e abrangerem áreas além da área de interesse. Além disso, 13 das parcelas possuem dados de inventário florestal e podem ser identificadas pelos seguintes Id's: 993, 1526, 1770, 1881, 3165, 3628, 3660, 3730, 5052, 5091, 5106 e 5122.



Figure 4: Parcelas com dados de inventário

## 4 Processamento da nuvem LiDAR

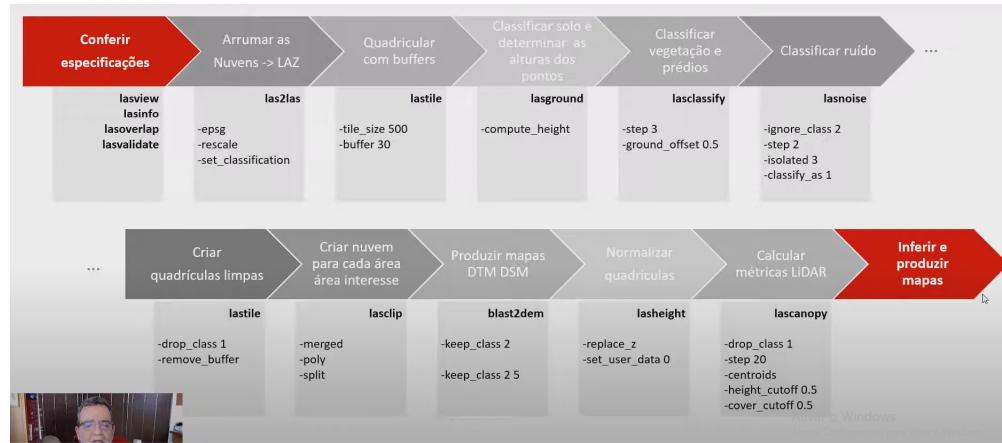


Figure 5: Fluxograma do processamento da nuvem LiDAR

- 1. Conferir especificações: Verificar se a nuvem gerada se enquadra no que é esperado: densidade de pontos, área escaneada, classificação de pontos etc;
- 2. Arrumar nuvens: Verificação do sistema de coordenadas utilizado e atributos de latitude, longitude e elevação;
- 3. Quadruplicar: Repartição da nuvem em pedaços menores a fim de ganhar eficiência no processamento em paralelo;
- 4. Definir solo e determinar altura dos pontos: Assegura e certifica a classificação do solo por diferentes métodos;
- 5. Classificar outros pontos e ruídos: Separar os pontos restantes por sua natureza (edificações, corpos d'água etc);
- 6. Quadrículas limpas: Remoção dos ruídos após classificação;
- 7. Produção de mapas DTM (elevação do solo), DSM (elevação do solo + vegetação) e DEM (elevação da vegetação);
- 8. Normalização da nuvem: Joga tudo pro mesmo plano;
- 9. Cálculo de métricas LiDAR: O p90 (percentil 90) representa a altura em que 90% dos pontos da nuvem LiDAR se encontram abaixo desta. Cada quadrícula possui o seu próprio p90 e ele tende a apresentar boa correlação com a realidade de campo. A intenção de existirem diversas métricas LiDAR é a de ver quais delas possuem boa correlação com os atributos de interesse.

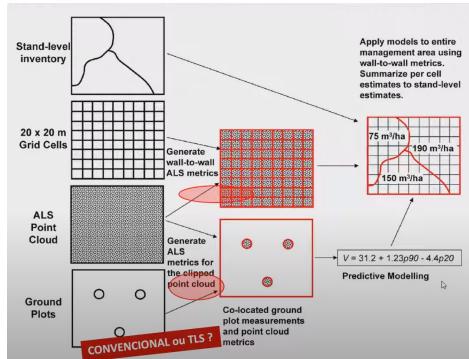


Figure 6: Esquematização do processo de inventário com ALS e informação terrestre

#### 4.1 Fluxograma e plots relacionados à Dupla Amostragem proposta em Inventário Florestal

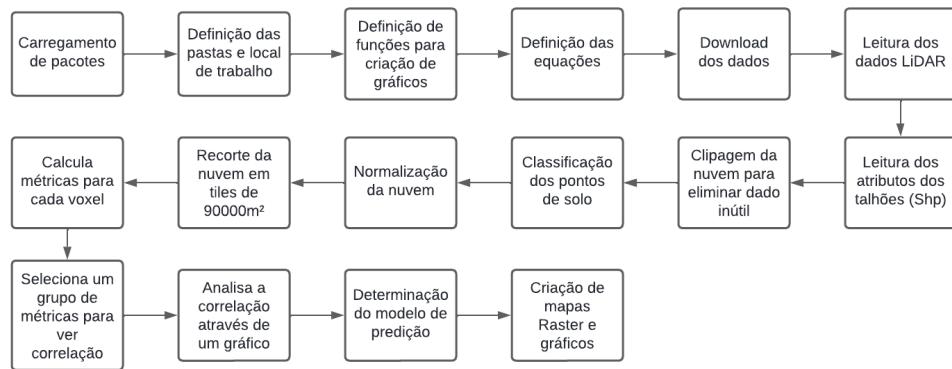


Figure 7: Fluxograma das etapas de processamento de dados LiDAR para fins de inventário florestal

1. Carregamento dos pacotes
  - i. Diversos são os pacotes carregados. Os nomes e a utilidade de cada um estão descritos na primeira seção do documento.

## 2. Definição das pastas e local de trabalho

- GitHub
  - i. C: - pasta raiz
  - ii. GitRepo - diretório em que estão agrupados os arquivos a serem upados no GitHub
  - iii. PRJ\_FAZENDAMODELO - pasta do projeto
  - iv. RMD - código e arquivos utilizados na redação do presente documento
  - v. RESULTADOS - plot da matriz de correlação
  - vi. BATCHR - arquivo tipo R com os scripts utilizados no pré e pós-processamento dos dados
  - vii. SAIDASSIG - arquivos gerados no QGIS
  - viii. SHAPES - arquivos de entrada para uso no SIG
- LiDAR
  - i. C: - pasta raiz
  - ii. LiDAR - agrupa todos as nuvens de pontos utilizadas no script
  - iii. PRJ\_FAZENDAMODELO - pasta do projeto
  - iv. NUVENS - onde se localizam as nuvens de pontos
  - v. A13 - reúne as nuvens do ano de 2013
  - vi. TALHOES - nuvem segregada por talhão
  - vii. NoNORM - nuvens com solo classificado
  - viii. SiNORM - nuvens com solo classificado e normalizadas
  - ix. RSTR\_qua - raster apresentando a estimativa da variável de interesse para cada talhão

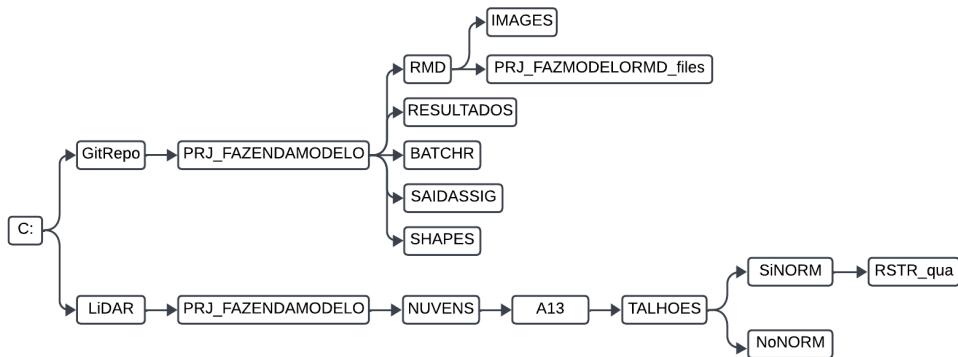


Figure 8: Organização dos diretórios

### 3. Definição das funções para criação dos gráficos

4. Definição das equações (estudar quais são)

5. Download e leitura dos dados LiDAR
  - i. Ao todo foram baixadas 6 nuvens de pontos LiDAR, que antes do processamento encontravam-se da seguinte maneira:

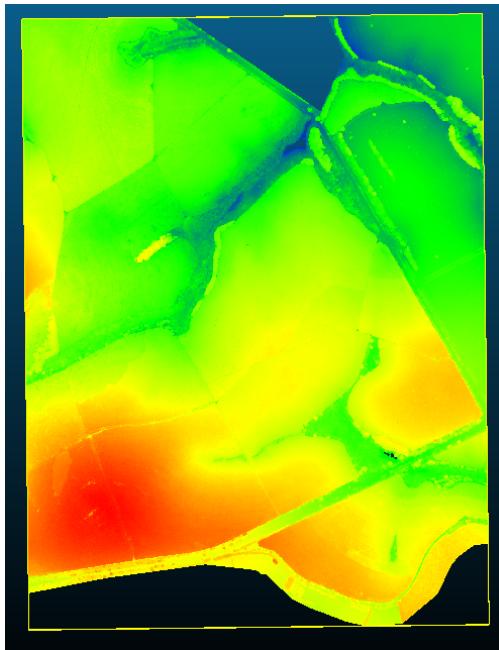


Figure 9: Nuvens de pontos LiDAR pré-processadas

- ii. As nuvens foram baixadas pelo seguinte link: [https://github.com/FlorestaR/dados/blob/main/5\\_LIDARF/Modelo/CLOUDS/](https://github.com/FlorestaR/dados/blob/main/5_LIDARF/Modelo/CLOUDS/)

6. Download e leitura dos dados em Shapefile

- i. Os shapes foram baixados pelo seguinte link: [https://github.com/FlorestaR/dados/blob/main/5\\_LIDARF/Modelo/SHAPES](https://github.com/FlorestaR/dados/blob/main/5_LIDARF/Modelo/SHAPES)

Fazenda	A
ID_Talhao	301
CODTALHAO	52I214
CHAVE	52I214_6
CHAVE2	P301410062
NUMPARCELA	6
NUMMEDICAO	2
ANOREF	2013
DATAREALIZ	4/18/2013
IDINV	3,70000
AREAPARCEL	399,73000
LAT_UTM23S	206632,09800
LONG_UTM23	7357127,11200
MHDOM	19,37000
VTCC	199,85000
VCCC	196,85000
SUBTALHAO	301d

Figure 10: Dados contidos nas parcelas inventariadas

7. Clipagem da nuvem para eliminação de dados indesejados (mostrar nuvem antes e depois)

Antes da clipagem: 68mi pontos

Após clipagem: 18mi pontos

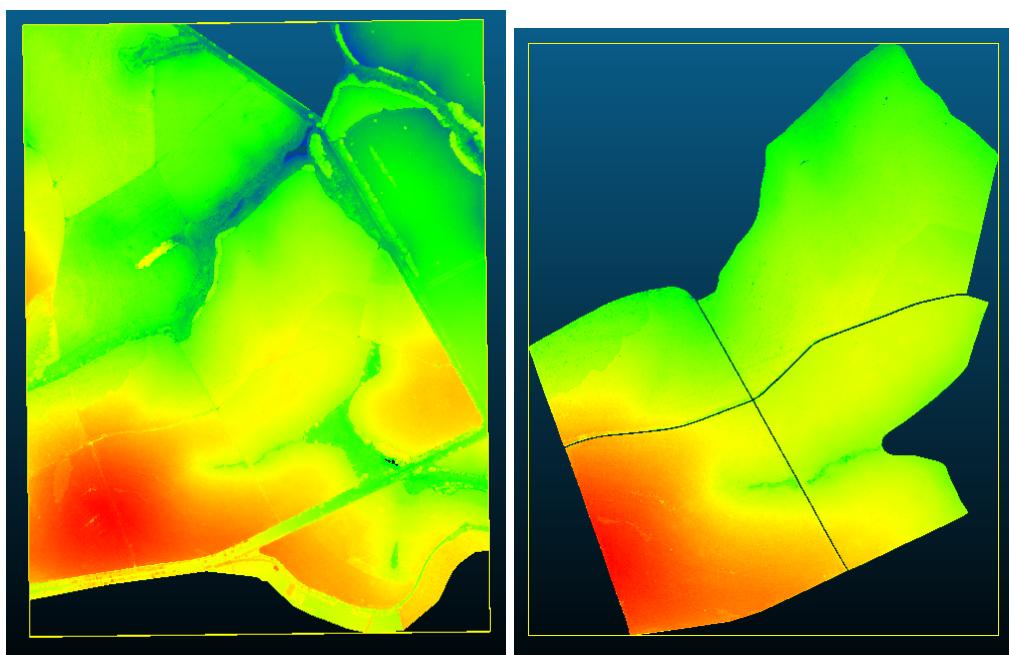


Figure 11: Comparativo entre as nuvens de pontos antes e após a clipagem

## 8. Classificação

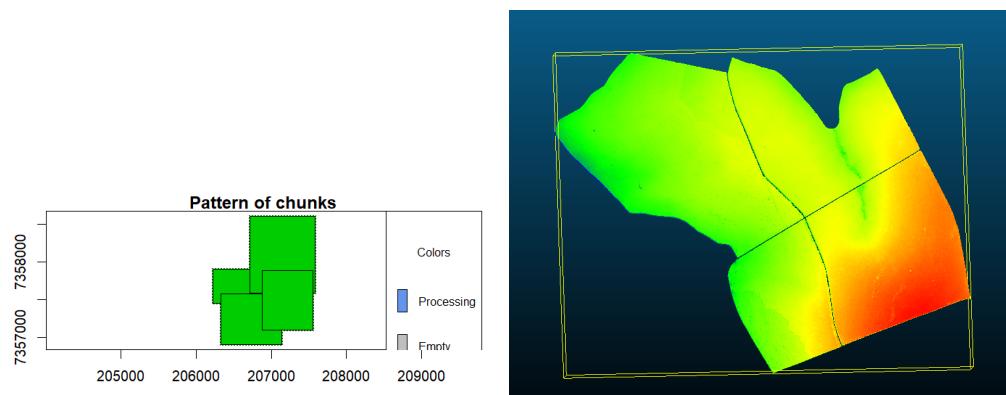


Figure 12: Processamento e resultado da classificação de solo das nuvens LiDAR

## 9. Normalização

- i. A etapa de normalização tem por finalidade nivelar toda a nuvem e é um passo que está diretamente correlacionado à classificação do solo.

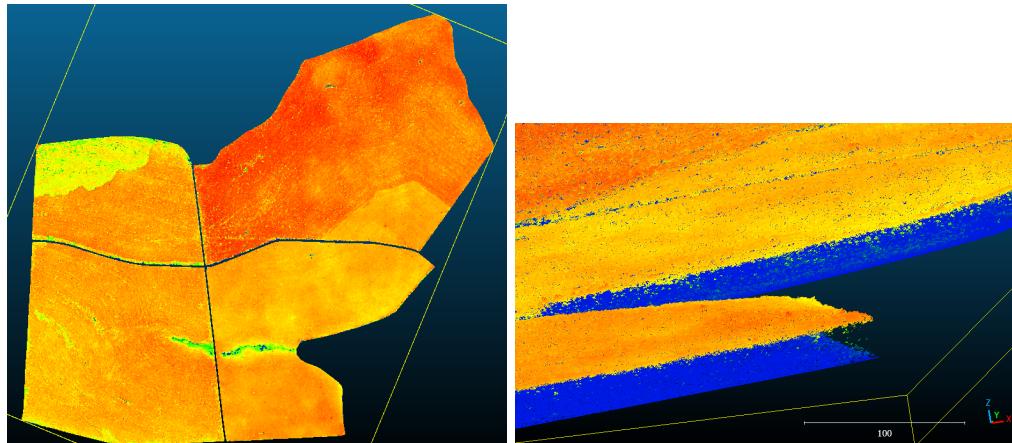


Figure 13: Resultado da normalização das nuvens

10. Recorte da nuvem em tiles 300x300m O recorte da nuvem em tiles menores tme a função de facilitar o processamento dos dados, tornando-o mais rápido e dinâmico

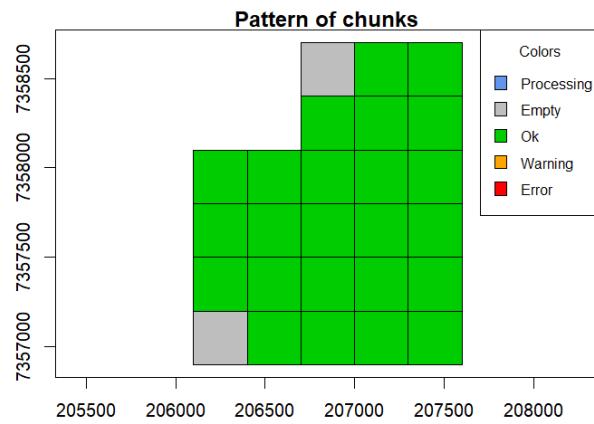


Figure 14: Retile da nuvem em quadrados de 300x300m

11. Cálculo de métricas para cada voxel (explicar voxel)

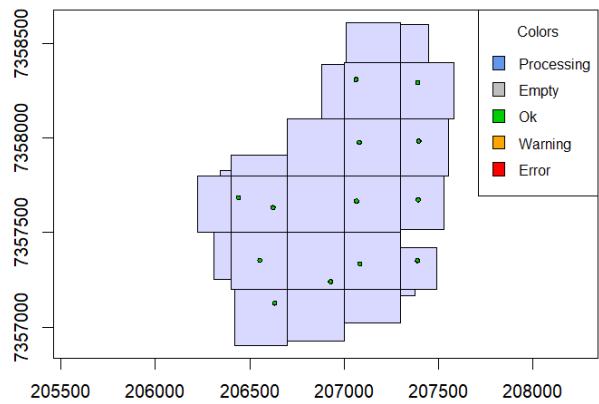


Figure 15: Cálculo das métricas para cada voxel

## 12. Seleção de um grupo de métricas para estudo de correlação

	VTCC	MHDOM	IDINV	zmean	zq45	zq75	zq95	zpcum2	zpcum4	zpcum6	pzabovezmean	pzabove2
	<db>	<db>	<db>	<db>	<db>	<db>	<db>	<db>	<db>	<db>	<db>	<db>
1	247.	22.2	5.2	12.6	13.1	23.2	24.2	1.08	1.08	1.55	55.1	55.2
2	204.	19.9	5.2	11.8	2.98	21.9	22.7	8.81	8.94	9.65	54.0	56.3
3	285.	27.4	5.2	13.2	19.8	22.1	23.7	1.14	1.34	1.59	61.0	61.3
4	249	24.5	5.2	14.1	20.4	22.7	24.2	3.21	3.21	3.94	63.8	64.7
5	335.	32.7	5.2	12.5	0	26.0	27.2	7.19	8.62	16.8	49.3	51.5
6	307.	28.1	5.2	12.3	0.37	24.8	25.9	9.33	9.50	10.3	49.9	51.0
7	200.	19.4	3.7	13.2	19.8	22.1	23.7	1.14	1.34	1.59	61.0	61.3
8	219.	19.9	3.7	14.1	20.4	22.7	24.2	3.21	3.21	3.94	63.8	64.7
9	155.	14.8	3.7	10.7	15.3	18.8	20.8	2.47	3.20	3.52	57.9	59.3
10	254.	23.4	3.7	14.7	21.4	23.6	25.5	4.86	5.56	5.56	63.1	65.3
11	247.	22.1	3.7	13.7	20.3	23.0	24.5	10.6	11.0	11.2	60.1	63.8
12	238.	23.9	5.2	14.0	21.2	22.9	24.2	2.04	2.19	2.55	62.0	62.8
13	276.	25.8	5.2	15.9	23.2	24.4	25.2	2.55	2.55	2.55	65.9	66.4

Figure 16: Tabela com as métricas escolhidas

13. Análise da correlação por meio de gráfico (falar do gráfico)

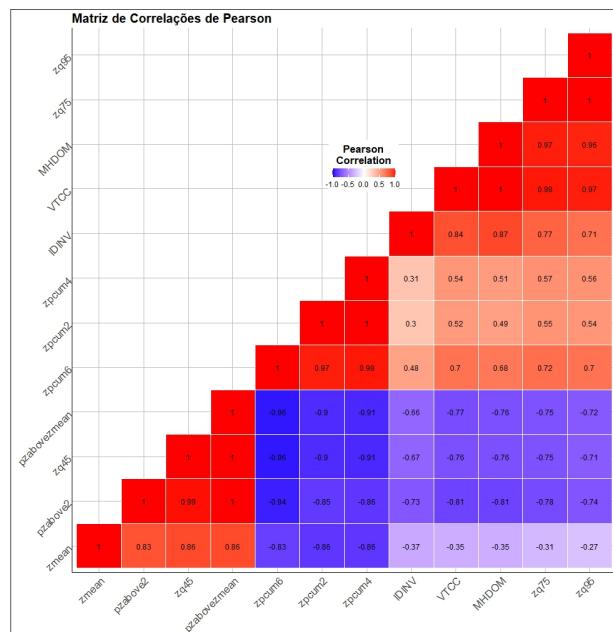


Figure 17: Resultado da análise de correlação

14. Determinação do modelo de predição

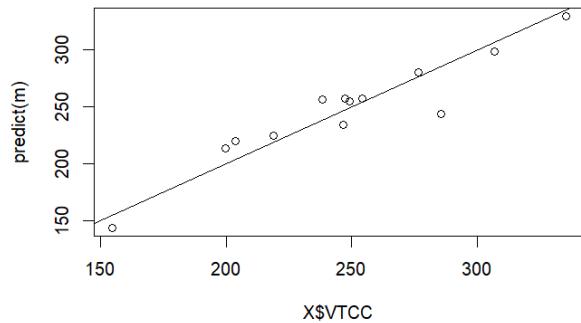


Figure 18: nao sei

## 15. Criação de mapas raster e gráficos

## 5 Dupla Amostragem

A dupla amostragem é composta por duas fases (s1 e s2) - LiDAR

- s1: comprehende uma gama de variáveis explanatórias para cada ponto pertencente a s1. As variáveis explanatórias derivam de informações auxiliares disponíveis em grande quantidade ao longo da área florestal;
- s2: constitui o inventário terrestre feito num número limitado de subamostras, onde todas pertencem a s1 e fornecem o valor das variáveis de interesse, ex. densidade local.
- É importante que s1 tenha uma grande intensidade amostral, de modo a que a média ou o total amostral da medida auxiliar possa ser estimado com precisão, sem erros amostrais. É uma variável barata e que pode ser encontrada em abundância na floresta. Os estimadores de razão e regressão utilizam a relação entre essa variável e a variável de interesse (cara e de difícil medição - VTCC) para garantir melhor precisão ao cálculo.
- O processo: busca-se estimar a média populacional da medida de interesse (ex. em média, quantos m<sup>3</sup> de madeira há por hectare?) e obter um indicativo da confiabilidade do cálculo (variância). Para aumentar a precisão da estimativa da média (diminuir a variância), utiliza-se a variável auxiliar que é de fácil observação e está associada à variável de interesse (aqui entra estimador de razão/regressão). Este movimento de redução da variância da média da medida de interesse ocorre de formas diferentes nas DA de *fase 2 de subamostra (onde todas as parcelas pertencem à fase 1) e fase 2 independente (há sorteio, então nem todas as parcelas inventariadas de forma intensiva fazem parte da fase 1)* TA CERTO ISSO??? APARENTEMENTE OS DOIS TIPOS DE FASE 2 SE DIVIDEM EM: SORTEIO ALEATORIO DAS AMOSTRAS E ESTRATIFICACAO
- Exemplo prático e definições (<https://www.ipef.br/publicacoes/scientia/nr108/cap09.pdf>);
- O erro amostral (E%) é calculado a partir do desvio padrão S<sup>2</sup>, que representa a variação de uma série de médias retiradas da população;
- A alta correlação entre métricas LiDAR e parâmetros biofísicos da floresta justificam a adoção da tecnologia na primeira fase da DA, além de reduzir a intensidade amostra (custo e trabalho). A amostragem dupla ganha maior precisão pela adoção dos estimadores de razão e regressão e para isso a média ou o total populacional da variável auxiliar deve ser conhecido e sem erro amostral. ;

$$VTCC = \beta_0 + \beta_1 P90 + \beta_2 ARMFR + \varepsilon$$

Figure 19: Modelo para cálculo da VTCC

## 6 Fluxograma e etapas Tripla amostragem

Composta por 3 fases, s0, s1 e s2:

- O princípio básico é o de que as variáveis explanatórias derivadas das informações auxiliares estão disponíveis em duas freqüências diferentes. A fase s0 fornece informações sobre toda a área, enquanto s1 possui dados adicionais de amostras de s0. Logo, a partir da informação terrestre coletada de um número x de parcelas de campo (1 camada de informação) é possível aferir sobre informações adicionais para outras parcelas a partir do uso de preditores (2 camadas de informação) **É ISSO MESMO?** **A CAMADA S1 É COMPOSTA POR DADOS ESTIMADOS?** e, por fim, o LiDAR coleta dados sobre todas as parcelas (3 camadas de informação).
- Logo, a motivação por trás da TA é a de que a gama de informações de s1 adiciona alto poder preditivo às variáveis disponíveis para todas as parcelas da área (s0)

*Journal of Statistical Software*

5

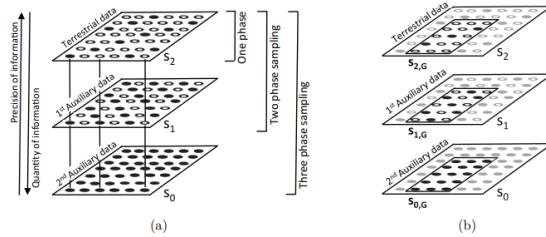


Figure 2: (a) Concept of multiphase sampling. The square represents the forest area for which an inventory is being conducted. The points denote the sample locations  $x$ . Filled points indicate available information. (b) Illustration of the small area estimation problem.

Figure 20: Esquema TA com Lidar

### Estimativas de pequenas áreas (small area estimation)

- Para sub áreas da floresta onde há pouca informação terrestre, como em G na figura acima. Para essas áreas, o uso da amostragem multifásica pode ser mais eficiente, já que utilizam um número reduzido de parcelas em campo para chegar à mesma precisão da ACE e ACS. Por outro lado, a sub área em questão pode ser pequena demais para justificar a adoção de um modelo de regressão separado, uma vez que este pode resultar em um intervalo de confiança indesejadamente abrangente;
- A ideia, então, é a de utilizar toda a riqueza de informações presente nas amostras de S2 para ajustar o modelo (equação utilizada para o cálculo de uma variável de interesse) e aplicá-lo para a sub área em questão.
- O potencial viés que surge da aplicação do modelo na sub área é corrigido pelo uso de modelos residuais empíricos derivados da área amostrada em campo. (não entendi direito essa parte) - trecho no texto: The potential bias of applying that model in G is then corrected for by using the empirical model residuals derived from that small area.
- Caso não existam parcelas de campo na sub área, então deve-se aceitar o viés na estimativa e no modelo. São essas as estimativas sintéticas, mas apesar do viés é possível calcular a sua variação.

### DESIGN-BASED VS MODEL-DEPENDENT APPROACH (ver dps)

- Model-dependent approach: as parcelas amostrais são fixas e as observações retiradas desses locais são assumidas como variáveis aleatórias, assim como a floresta assume o papel de ser o meio realizador

desse processo estocástico (?????? o que isso significa?). Embora os locais das parcelas possam ser arbitrariamente escolhidos, o modelo deve escrever adequadamente o processo estocástico, a fim de garantir resultados parciais. (????)

- Os estimadores do forestinventory se baseiam na design-based approach. Design-based approach baseia-se na randomização dos locais de amostra de forma uniforme e independente. Logo, a floresta por si só e qualquer valor de densidade local de x (pertencente a) F (floresta) são fixos e não um resultado de um processo estocástico (estudar). (????)
- Na estrutura do Model-dependent approach as parcelas são fixas e resultam em um processo estocástico GPT: what is the difference between design-based approach and model-dependent approach?

## 7 Regressão linear simples

Modelar a relação entre duas variáveis e se expressa na forma de uma reta no plano cartesiano.

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

Figure 21: Equação base da regressão linear simples

- $Y$  = variável resposta ou variável dependente
- $X$  = variável independente, regressora ou explicativa
- $\epsilon$  = erro. Pode ser que a  $Y$  seja explicada por  $X$ , mas falte algo que explique essa variabilidade. É a flutuação aleatória que ocorre ao tentar explicar  $Y$  por  $X$ . Seja por imperfeição do modelo, erros na mensuração ou outras variáveis fora de controle.

## 8 Arborimetria preditiva

- Em florestas plantadas há grande relação entre o DAP e a altura da árvore. O procedimento de medição de campo é o seguinte: 1 - DAP de todas as árvores da parcela; 2 - altura total de uma amostra da parcela; 3 - estima-se a relação entre o DAP e a altura total da amostra; 4 - utiliza essa relação para predizer a altura de todas as árvores da parcela.
  - A seleção das árvores da amostra pode ser feita de forma aleatória ou sistemática. A seleção aleatória garante maior confiabilidade, enquanto a sistemática é mais simplificada (ex. 1 árvore a cada 15).
  - A **\*relação hipsométrica** é a relação média entre o DAP e a altura das árvores individuais. A curva sigmoidal é o melhor modelo para a relação média entre o DAP e a altura.

<p>que se somam. Em cada termo, tem-se a medida da DAP, ou uma transformação dela, sendo multiplicada por um</p>
<b>Quadro 6.1</b> Modelos não lineares de relação hipsométrica
<b>Modelo</b>
<b>Forma preditiva e de estimativa</b>
<b>Chapman-Richards</b>
$h = \beta_0 [1 - \exp(-\beta_1 \cdot d)]^{\beta_2} + \varepsilon$
<b>Weibull</b>
$h = \beta_0 [1 - \exp(-\beta_1 \cdot d^{\beta_2})]^{\beta_2} + \varepsilon$
<b>Monomolecular</b>
$h = \beta_0 [1 - \beta_1 \cdot \exp(-\beta_2 \cdot d)] + \varepsilon$
<b>Comportz</b>
$h = \beta_0 \exp[-\beta_1 \cdot \exp(-\beta_2 \cdot d)] + \varepsilon$
<b>Logístico</b>
$h = \beta_0 / [1 + \beta_1 \cdot \exp(-\beta_2 \cdot d)] + \varepsilon$
<i>h</i> - altura total das árvores individuais; <i>d</i> - diâmetro à altura do peito ( <i>DAP</i> ); $\beta_0, \beta_1, \beta_2$ - parâmetros a serem estimados, $\beta'_0 = \exp(\beta_0)$ , $\varepsilon$ - componente estocástico do modelo.

Figure 22: Quadro de modelos não lineares (esquerda) e lineares (direita)

- Um modelo é linear quando cada termo é uma constante ou o produto de um parâmetro e uma variável preditora. Uma equação linear é construída adicionando os resultados para cada termo. Modelos lineares seguem esse padrão: Resposta = constante + parâmetro \* preditor + ... + parâmetro \* preditor. Caso contrário a equação é não-linear.
  - O postulado de Cotta define que o volume de uma árvore depende do seu diâmetro, altura e forma. Todas as árvores de mesmo diâmetro, altura e forma possuem o mesmo volume. Isso serve para o lenho de tronco único. O fator forma pode ser visto como um índice quantitativo que representa a razão entre o volume sólido e cilíndrico de uma árvore. O volume sólido de uma amostra da parcela pode ser medido por cubagem rigorosa. O fator forma é igual ao somatório dos volumes sólidos dividido pelo somatório dos volumes cilíndricos ( $[\pi/4] * d^2 * h$ ).
  - O volume de uma dada árvore é o resultado da multiplicação entre o fator de forma e o volume cilíndrico do indivíduo. A seleção das árvores deve ser aleatória, sem preocupação em selecionar árvores de todas as classes de DAP e altura. **Vamos usar o LiDAR TLS para estimar o diâmetro e a forma?**

Quantificação de recursos florestais		
Modos: lineares de relação homométrica. A forma de estimativa se refere à maneira utilizada para estimar os parâmetros dos modelos, enquanto que a forma preditiva se refere à maneira como o modelo é aplicado para se prever a altura		
Modelo		
Polinômios		
1	$h = \beta_0 + \beta_1 \cdot d + \varepsilon$	—
2	$h = \beta_0 + \beta_1 \cdot d + \beta_2 \cdot d^2 + \varepsilon$	—
Hiperbólicos		
1	$h = \beta_0 + \beta_1 \cdot (d/d') + \varepsilon$	$\hat{h} = \frac{\beta_0 \cdot d' - \beta_1}{d - d'}$
2	$1/\sqrt{h} = \beta_0 + \beta_1 \cdot (1/d') + \varepsilon$	$\hat{h} = \left[ \frac{\beta_0 \cdot d'}{\beta_1} \right]^2$
3	$1/h = \beta_0 + \beta_1 \cdot (d/d') + \varepsilon$	$\hat{h} = \frac{\beta_0}{\beta_1} \cdot d' - \beta_1 \cdot d'/d^2$
4	$d^2/h = \beta_0 + \beta_1 \cdot d + \beta_2 \cdot d^2 + \varepsilon$	$\hat{h} = \frac{\beta_0}{\beta_2} \cdot d^2 - \beta_1 \cdot d^2 - \beta_2$
5	$d/\sqrt{h} = \beta_0 + \beta_1 \cdot d + \beta_2 \cdot d^2 + \varepsilon$	$\hat{h} = \frac{(\beta_0 + \beta_1 \cdot d)^2 - \beta_2 \cdot d^2}{(\beta_0 + \beta_1 \cdot d)^2}$
Potência		
1	$\ln(h) = \beta_0 + \beta_1 \cdot \ln(d) + \varepsilon$	$\hat{h} = e^{\beta_0 + \beta_1 \cdot d}$
2	$\ln(1/h) = \beta_0 + \beta_1 \cdot \ln(d) + \varepsilon$	$\hat{h} = \frac{1}{e^{\beta_0 + \beta_1 \cdot d}}$
3	$\ln(h) = \beta_0 + \beta_1 \cdot \ln(d/(1+d)) + \varepsilon$	$\hat{h} = e^{\beta_0 + \beta_1 \cdot \ln(d/(1+d))}$
Exponencial (Schumacher)		
1	$\ln(h) = \beta_0 + \beta_1 \cdot \ln(d) + \varepsilon$	$\hat{h} = e^{\beta_0 + \beta_1 \cdot \ln(d)}$
Semiexponencial		
1	$h = \beta_0 + \beta_1 \cdot \ln(d) + \varepsilon$	$\hat{h} = \beta_0 + \beta_1 \cdot \ln(d) \Rightarrow d = \exp\left(\frac{h - \beta_0}{\beta_1}\right)$

$h$  - medida da altura total das árvores individuais.  
 $\hat{h}$  - altura total predita.  
 $d$  - diâmetro a altura ou perito ( $DAP$ ).  
 $\beta_0, \beta_1$  - parâmetros a serem estimados.  
 $\beta_0 = \exp(\beta_{00})$   
 $\varepsilon$  - componente estatístico do modelo.  
 $\ln$  - logaritmo natural ou neperiano.

Figure 23: Quadro de modelos não lineares (esquerda) e lineares (direita)

- O volume pode ser calculado em função de diferentes variáveis preditoras, como a altura, DAP, altura até a primeira ramificação ou outro. Equações de dupla entrada ou equações padrão são aquelas que utilizam o DAP e a altura para cálculo do volume.
- A relação entre o volume sólido e o volume cilíndrico tende à uma relação de proporcionalidade (a curva que os representa passa pela origem). Com a adoção do volume comercial, são consideradas as árvores a partir de um tal diâmetro, o que afasta a curva da origem e, portanto, não evidencia uma relação proporcional.
- É possível modelar o fator de forma como uma função do DAP ao invés de encontrá-lo a partir da relação entre o DAP e a altura. É adequado para quando a forma da árvore muda de acordo com o DAP.

**Quadro 6.3** Modelos de equação padrão de volume ou equação de dupla entrada. Modelos marcados com asterisco (\*) são modelos não lineares. Todos os modelos são apresentados na sua forma de estimação

Modelos de variável combinada		
VC-1	Spurr	$v_i = \beta_0 + \beta_1(d_i^2 \cdot h_i) + \varepsilon_i$
VC-2	log-Spurr	$\ln(v_i) = \beta_0 + \beta_1 \cdot \ln(d_i^2 \cdot h_i) + \varepsilon_i$
VC-3*		$v_i = \beta_0 + \beta_1(d_i^2 \cdot h_i)^{\beta_2} + \varepsilon_i$
Modelos de potência variável		
PV-1	Schumacher-Hall	$\ln(v_i) = \beta_0 + \beta_1 \cdot \ln(d_i) + \beta_2 \cdot \ln(h_i) + \varepsilon_i$
PV-2*		$v_i = \beta_0 \cdot d_i^{\beta_1} \cdot h_i^{\beta_2} + \varepsilon_i$
PV-3*		$v_i = \beta_{00} + \beta_0 \cdot d_i^{\beta_1} \cdot h_i^{\beta_2} + \varepsilon_i$
PV-4	Baden-Württemberg	$\ln(v_i) = \beta_0 + \beta_1 \cdot \ln(d_i) + \beta_2 \cdot \ln^2(d_i) + \beta_3 \cdot \ln(h_i) + \beta_4 \cdot \ln^2(h_i) + \varepsilon_i$
PV-5*		$v_i = \beta_0 \cdot d_i^{\beta_1+\beta_2 \cdot \ln(d_i)} \cdot h_i^{\beta_3+\beta_4 \cdot \ln(h_i)} + \varepsilon_i$
Modelos de harmonização DAP-altura		
H-1	Stoate	$v_i = \beta_0 + \beta_1 \cdot d_i^2 + \beta_2(d_i^2 \cdot h_i) + \beta_3 \cdot h_i + \varepsilon_i$
H-2	Meyer	$v_i = \beta_0 + \beta_1 \cdot d_i + \beta_2 \cdot d_i^2 + \beta_3(d_i \cdot h_i) + \beta_4(d_i^2 \cdot h_i) + \beta_5 \cdot h_i + \varepsilon_i$
H-3	Meyer modificada	$v_i = \beta_0 + \beta_1 \cdot d_i + \beta_2 \cdot d_i^2 + \beta_3(d_i \cdot h_i) + \beta_4(d_i^2 \cdot h_i) + \varepsilon_i$
H-4	Näslund	$v_i = \beta_0 + \beta_1 \cdot d_i^2 + \beta_2(d_i^2 \cdot h_i) + \beta_3(d_i \cdot h_i^2) + \beta_4 \cdot h_i^2 + \varepsilon_i$
Modelos de fator forma		
FF-1*	Takata	$v_i = (d_i^2 \cdot h_i) / [\beta_0 + \beta_1 \cdot d_i] + \varepsilon_i$
FF-2*	Honer	$v_i = d_i^2 / [\beta_0 + \beta_1/h_i] + \varepsilon_i$

*i* – índice que indica as árvores na amostra destrutiva ( $i = 1, 2, \dots, n$ );  
*V<sub>i</sub>* – volume sólido (total ou comercial);  
*d<sub>i</sub>* – DAP;  
*h<sub>i</sub>* – altura total ou comercial;  
 $\ln$  – logaritmo natural ou neperiano;  
 $\varepsilon_i$  – componente estocástico do modelo.

Figure 24: Equação base da regressão linear simples

## 9 Silvimetria com medidas auxiliares

- Medidas de fácil obtenção: densidade de estande e área basal e possuem boa relação com as estimativas de produção (volume). Existem dois tipos de relação entre as medidas auxiliares e as de interesse e resultam em dois estimadores diferentes: estimadores de razão (razão entre a medida de interesse e a auxiliar) e o estimador de regressão (relação entre as duas medidas é uma relação linear simples).
- Estimador de razão é a razão entre a medida de interesse e a medida auxiliar. A razão pode ser em relação ao total ou às médias das medidas. Ex: saber o total das árvores em vias públicas de uma cidade (caso 1); saber a proporção das vias públicas sombreadas por árvores (caso 2). O primeiro caso se trata de um total - número de árvores na cidade -, o segundo caso envolve a razão entre a área sombreada e área total das vias públicas. O estimador de razão é uma razão das médias, e não uma média das razões.

$$R = \frac{(1/n) \sum_{i=1}^n Y_i}{(1/n) \sum_{i=1}^n X_i} = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n X_i} = \frac{TotalY}{TotalX}$$

- $R$  é o estimador de razão
- $i$  indica os arvoredos da amostra ( $i = 1, 2, \dots, n$ )
- $Y_i$  é a observação da medida de interesse no arvoredo  $i$
- $X_i$  é a observação da medida auxiliar no mesmo arvoredo  $i$
- O estimador de razão nada mais é do que a soma das médias da medida de interesse dividida pela soma das médias da medida auxiliar. é um meio para obter a estimativa da média da variável

de interesse com maior precisão. Quanto maior a relação entre a medida de interesse e a medida auxiliar, menor é a soma de quadrados da expressão  $\sum(Y_i - R * X_i)^2$  **FORMULA VARIANCIA ESTIMADOR DE RAZAO E FORMULA DA VARIANCIA POPULACIONAL - PG 296**

- O estimador de razão requer o conhecimento da *média* ou do *total populacional* da variável auxiliar, que é um atributo dos arvoredos, a exemplo da densidade de estande e da área basal, o que poderia ser obtido a partir de um censo, descartando erros, mas é caro. O estimador de razão é utilizado nas situações em que a medida de interesse (Y) é diretamente proporcional à medida auxiliar (X). Logo, deve satisfazer 3 condições:
  1. A relação média entre Y e X é linear: pode ser representada por uma reta;
  2. A reta X-Y passa pela origem do plano cartesiano;
  3. O grau de dispersão dos valores de Y (denota a variância) ao redor da reta é proporcional aos valores de X.
- O **Estimador de regressão** é utilizado na existência de uma relação linear entre a variável de interesse e a auxiliar, porém a reta que as representa não passa pela origem. Logo, o estimador de regressão possui uma abordagem mais generalista, pois é eficiente mesmo quando as medidas da variável auxiliar e de interesse estão negativamente associadas - inversamente proporcionais.

## 10 R utilizando database grisons

- Para essa etapa será utilizado o database “grisons”. A fase s1 é composta por 306 amostras (phase\_id\_2p) escaneadas com LiDAR e s2 possui 67 subamostras de s1, de onde se calculou o volume com os dados terrestres.
- Para estimadores globais de dupla amostragem, devemos especificar:
  1. O modelo de regressão (formula);
  2. O dataframe que possui as informações de inventário;
  3. O objeto de lista (*list*) *phase-id* contendo: o argumento *phase.col* identificando o nome da coluna e especificando para s1 ou s2.
  4. Argumento *terrgrid.id*, que especifica qual valor numérico indica s2 na coluna.
  5. Opcional: coluna contendo os tamanhos de borda

```
head(grisons)
```

```
##   phase_id_2p phase_id_3p boundary_weights      mean     stddev      max
## 2             1          0       0.8133455 10.263023  7.278974 33.17151
## 3             1          0       1.0000000 23.351052 11.372732 45.22998
## 4             1          1       0.8472245  8.259801  5.958564 21.26001
## 6             1          0       1.0000000 21.563423  7.493390 32.65515
## 7             1          0       1.0000000 15.751000  8.447083 33.10999
## 8             1          1       1.0000000 21.160314 13.129320 42.31677
##       q75 smallarea tvol tvol.3p
## 2 15.10999        A    NA    NA
## 3 31.14239        A    NA    NA
## 4 13.06995        A    NA    NA
## 6 27.81149        A    NA    NA
## 7 22.83515        A    NA    NA
## 8 32.40747        A    NA    NA
```

- Cria-se um modelo para o cálculo do volume a partir da regressão múltipla da média das alturas (mean), desvio padrão dessa média (stddev), altura máxima (max) e p75 (q75). O modelo retorna a estimativa, a variância externa (variância causada por fatores externos e que não são contemplados pelo modelo - ex. ansiedade do participante ao realizar um teste cognitivo), a variância peso-g (contempla a força de um fator geral para a avaliação sobre os itens medidos. É como se houvessem pesos diferentes para os critérios - lembra uma média ponderada), a quantidade de amostras utilizadas em s1 e s2 e o R<sup>2</sup> (correlação entre as variáveis). A variância peso-g geralmente é mais preferível.

```

reg2p_nex <- twophase(formula = tvol ~ mean + stddev + max + q75,
                       data = grisons, phase_id = list(phase.col = "phase_id_2p",
                                                       terrgrid.id = 2), boundary_weights = "boundary_weights")
summary(reg2p_nex)

## 
## Two-Phase global estimation
##
## Call:
## twophase(formula = tvol ~ mean + stddev + max + q75, data = grisons,
##           phase_id = list(phase.col = "phase_id_2p", terrgrid.id = 2),
##           boundary_weights = "boundary_weights")
##
## Method used:
## Non-exhaustive global estimator
##
## Regression Model:
## tvol ~ mean + stddev + max + q75
##
## Estimation results:
## estimate ext_variance g_variance n1 n2 r.squared
## 383.5354    279.954   271.5057 306 67 0.6428771
##
## 'boundary_weight'- option was used to calculate weighted means of auxiliary variables

```

- O estimador exaustivo pode ser aplicado passando adicionalmente um vetor que contenha as exatas médias das variáveis explanatórias através do argumento *exhaustive*. Este vetor deve ser calculado previamente de tal maneira que qualquer ajuste de borda desejado já tenha sido aplicado. Enquanto um estimador exaustivo utiliza toda a informação disponível no conjunto de dados para gerar a informação, o estimador não-exaustivo, por sua vez, não utiliza toda a informação.

```
colnames(lm(formula = tvol ~ mean + stddev + max + q75, data = grisons, x = TRUE)$x)
```

```
## [1] "(Intercept)" "mean"          "stddev"        "max"          "q75"
```

*Não entendi o código acima*

- O estimador exaustivo pode ser aplicado após definir o vetor com as exatas médias, *true.means.Z*:

```

true.means.Z <- c(1, 11.39, 8.84, 32.68, 18.03)
reg2p_ex <- twophase(formula = tvol ~ mean + stddev + max + q75,
                      data = grisons, phase_id = list (phase.col = "phase_id_2p",
                                                       terrgrid.id = 2), exhaustive = true.means.Z)

summary(reg2p_ex)

```

```

## 
## Two-Phase global estimation
##
## Call:
## twophase(formula = tvol ~ mean + stddev + max + q75, data = grisons,
##           phase_id = list(phase.col = "phase_id_2p", terrgrid.id = 2),
##           exhaustive = true.means.Z)
##
## Method used:
## Exhaustive global estimator
##
## Regression Model:
## tvol ~ mean + stddev + max + q75
##
## Estimation results:
## estimate ext_variance g_variance n1 n2 r.squared
## 376.7426     202.5602    187.2787 Inf 67 0.6428771

```

- As variâncias que resultam do cálculo através dos estimador exaustivo são menores do que o valor encontrado pelos estimadores não-exaustivos, uma vez que foram trocadas as médias estimadas das variáveis explanatórias pelas médias reais.

## 11 Resultado

Two-Phase global estimation

Call: twophase(formula = VTCC ~ zq95, data = X, phase\_id = list(phase.col = “Inventario”, terrgrid.id = 2), boundary\_weights = “boundaryweights”)

Method used: Non-exhaustive global estimator

Regression Model: VTCC ~ zq95

Estimation results: estimate ext\_variance g\_variance n1 n2 r.squared 185.5156 58.22226 259.216 3453 8 0.7322413

‘boundary\_weight’- option was used to calculate weighted means of auxiliary variables

## 12 Estatística de confiança (ACS)

1. `calcCI = function(err, n, alpha=.05)`: err é o desvio padrão, n é o número de amostras e alpha é o nível de significância; essa função calcula o IC;
2. `qt(1 - alpha/2, n - 1) * err`:  $1 - \alpha/2$  traz um IC de 95% (2,5% para cauda superior e 2,5% para a inferior), n - 1 refere-se aos graus de liberdade e \* err faz com que retorne o IC;
3. `calcPars = function(df, N, alpha=.05)`: df é um data.frame com as informações de MHDOM e VTCC de cada uma das 13 parcelas, N é o número máximo possível de parcelas de campo na área e alpha diz sobre nível de significância do IC;
4. `means = apply(df, 2, mean)` calcula a média das colunas de df (MHDOM e VTCC)
5. `vars = apply(df, 2, function(y){(var(y) / length(y)) * ((N - length(y)) / N)})`: Aplica uma função para calcular as variâncias ajustadas em cada uma das colunas de df (MHDOM e VTCC); a fórmula é:  $\frac{var(y)}{length(y)} * \frac{N-length(y)}{N}$

6. `err = sqrt(vars)` calcula o desvio padrão com base nas variâncias ajustadas
7. `cis = calcCI(err, nrow(df), alpha)` dá os parâmetros para calcCI ser calculada
8. `err_pc = 100*cis/means` calcula o erro percentual
9. Organiza em um data.frame:

```
out = rbind(means, vars, err, cis, err_pc, nrow(df)) %>% as.data.frame
row.names(out) = c('media', 'var', 'dp', 'ic', 'erro', 'n')
names(out) = names(df)
return(out)
```