

Pedologia Quantitativa feita Simples

Alessandro Samuel-Rosa

2016-09-16

Contents

1	Prerequisites	5
2	Introdução	7
3	Infraestrutura	11
3.1	Hardware	11
3.2	Software	11
3.3	Gestão de dados	11
4	Métodos	13
5	Amostragem	15
5.1	Tipos de amostragem	15
5.2	Sumário	18
5.3	Bibliografia consultada	18
6	Final Words	19

Chapter 1

Prerequisites

This is a *sample* book written in **Markdown**. You can use anything that Pandoc's Markdown supports, e.g., a math equation $a^2 + b^2 = c^2$.

For now, you have to install the development versions of **bookdown** from Github:

```
devtools::install_github("rstudio/bookdown")
```

Remember each Rmd file contains one and only one chapter, and a chapter is defined by the first-level heading #.

Chapter 2

Introdução

You can label chapter and section titles using `{#label}` after them, e.g., we can reference Chapter 2. If you do not manually label them, there will be automatic labels anyway, e.g., Chapter ??.

Figures and tables with captions will be placed in **figure** and **table** environments, respectively.

```
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19)
```

Reference a figure by its code chunk label with the **fig:** prefix, e.g., see Figure 2.1. Similarly, you can reference tables generated from `knitr::kable()`, e.g., see Table 2.1.

```
knitr::kable(
  head(iris, 20), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

Table 2.1: Here is a nice table!

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa

You can write citations, too. For example, we are using the **bookdown** package (Xie, 2016) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015). ddd aaa

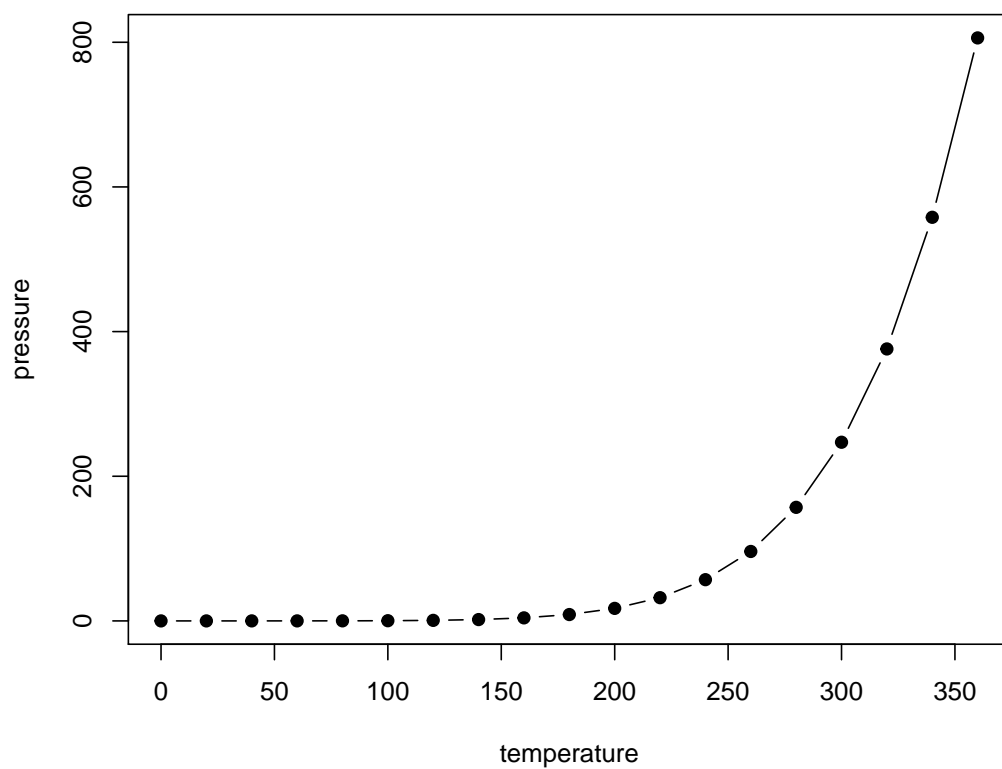


Figure 2.1: Here is a nice figure!

Chapter 3

Infraestrutura

3.1 Hardware

3.2 Software

3.3 Gestão de dados

Um importante passo consiste na criação do nosso diretório de trabalho. Para isso, acesse seu gerenciador de diretórios e crie a seguinte estrutura de diretórios em seu local favorito de trabalho:

```
projeto
|- code/      # qualquer código de programação
|  |- R/      # código de programação em R
|
|- data
|  |- grid/   # dados matriciais
|  |- R/      # dados no formato *.rda
|  |- vector/ # dados vetoriais
|
|- doc/       # arquivos usados para redigir o relatório
|  |- fig/    # figuras usadas no relatório
```

Note que, dentro do nosso diretório de trabalho principal (**projeto**) existem três subdiretórios: **code**, **data** e **doc**. O primeiro deles, **code**, serve para armazenarmos os arquivos contendo código de programação em qualquer linguagem. Para cada linguagem criamos um subsubdiretório específico. Nesse exemplo, como usaremos apenas o R, criaremos apenas um subsubdiretório chamado **R**. Ali dentro serão armazenados os *scripts* com código de programação do escritos em R.

O segundo subdiretório será utilizado para armazenarmos os dados usados no projeto. Nesse exemplo, são três os tipos de dados que utilizaremos, cada um armazenado em um subsubdiretório específico. No subsubdiretório **grid** ficarão armazenados os dados matriciais, ou seja, os dados das covariáveis e os resultados das predições espaciais. Já no subsubdiretório **vector** ficarão armazenados os dados vetoriais, ou seja, aqueles cuja forma de representação espacial pode ser a de pontos, linhas e polígonos. Isso inclui os dados de solo e dos limites da área de estudo. Por fim, o subsubdiretório **R** será usado para armazenar dados diversos produzidos durante o processamento no R, os quais serão salvos usando o formato **rda**.

O terceiro e último subdiretório de nosso diretório de trabalho **projeto**, aqui denominado **doc**, será usado para armazenar os arquivos usados para redigir os documentos resultantes do projeto.

Chapter 4

Métodos

We describe our methods in this chapter.

Chapter 5

Amostragem

5.1 Tipos de amostragem

A amostragem é um dos maiores contribuintes para os custos da modelagem espacial do solo, seja pelo método chamado “tradicional” ou usando modelos estatísticos. Assim, o adequado planejamento amostral é essencial para reduzir a necessidade de recursos (financeiros, psicológicos, humanos, entre outros) e maximizar o número de observações possíveis, sejam elas tradagens, amostras superficiais, ou descrições de perfis completos.

A *situação ideal* é aquela em que os recursos disponíveis não impõe quaisquer limitações à amostragem. Nesse caso seria possível fazer observações e coletar amostras do solo em diversas etapas. Por exemplo, iniciariamos com um levantamento exploratório para identificar a estrutura da variação espacial do solo. De posse desse novo conhecimento, executaríamos uma nova etapa amostral a fim de atender a algum critério como, por exemplo, obter uma cobertura espacial aproximadamente uniforme da área sendo modelada. Caso os resultados ainda sejam insatisfatórios, uma nova etapa amostral poderia ser executada para, por exemplo, fazer observações em locais específicos cujas condições ambientais estejam sub-representadas na base de dados. Por fim, calibrado o modelo espacial do solo e feitas as predições espaciais, coletariamos amostras para validação do modelo preditivo em número suficiente para garantir um nível de confiança pré-determinado.

Mas a *situação ideal* está longe de ser o que acontece na prática. Em geral temos que fazer todas as observações em uma única fase, coletando todo o material possível, inclusive para as amostras de validação. Isso requer que o tipo de amostragem mais apropriado seja utilizado a fim de otimizar o uso dos recursos disponíveis e obter o melhor modelo preditivo possível. E qual seria o tipo de amostragem mais apropriado? Uma resposta universal à essa pergunta continua desconhecida. A melhor estratégia costuma ser avaliar os diferentes tipos de amostragem frente (1) os objetivos do projeto, (2) os recursos disponíveis, e (3) as dificuldades operacionais encontradas na área sendo modelada.

Podemos dizer que existem dois tipos fundamentais de amostragem:

- probabilística, e
- não-probabilística.

5.1.1 Amostragem probabilística

A característica fundamental da amostragem probabilística é que a chance de um determinado local ser amostrado (probabilidade de inclusão) é conhecida e maior do que zero. Em outras palavras, todo e qualquer local possui alguma chance de ser amostrado, mesmo que alguns tenham maior chance do que outros. Um local que não pode ser amostrado tem probabilidade de inclusão igual a zero.

A amostragem probabilística é muito utilizada em experimentos controlados, como aqueles desenvolvidos em campos experimentais, casas de vegetação e laboratórios. No caso da modelagem espacial do solo, a amostragem probabilística costuma ser usada para a validação das predições espaciais. Entretanto, ela também pode ser usada para obter observações para a calibração dos modelos preditivos.

5.1.2 Amostragem não-probabilística

Na amostragem não-probabilística, como o próprio nome já diz, não são considerados os valores de probabilidade de inclusão para a seleção dos locais de amostragem. A escolha dos locais de amostragem depende da definição de um critério a ser atendido.

A amostragem não-probabilística costuma ser dividida em três categorias:

- casual,
- conveniente, e
- intencional.

5.1.2.1 Amostragem casual

Na amostragem casual (em inglês, *haphazard sampling*) os locais de amostragem são escolhidos, fundamentalmente, em função da subjetividade da pessoa conduzindo a amostragem. Não existe um critério claro a ser atendido. Outros locais amostrais podem ser escolhidos caso outra pessoa conduza a amostragem, mesmo que não haja justificativa plausível para isso. Assim, a amostragem casual deve ser evitada sempre que possível.

Na prática, a amostragem casual consiste em transitar pela área a ser amostrada e, aqui e acolá, definir um local para amostragem. Isso dá a impressão de que as observações são *aleatórias*, um pressuposto estatístico comum. Entretanto, supor que tais observações são aleatórias é um equívoco, dado que é impossível calcular as probabilidades de inclusão de cada amostra. É devido a esse comum equívoco que a expressão *amostragem aleatória* tem caído em desuso junto a comunidade estatística, dando-se preferência à expressão *amostragem probabilística* (veja acima).

5.1.2.2 Amostragem conveniente

A amostragem conveniente (em inglês, *convenience sampling*) está diretamente relacionada à otimização do uso dos recursos disponíveis. Ela consiste em evitar realizar observações em locais de difícil acesso como áreas densamente florestadas, distantes de estradas, terrenos íngremes, ou áreas que apresentem risco para a saúde ou à vida devido à presença de, por exemplo, animais peçonhentos. Assim sendo, o critério usado para a definição dos locais de observação é a soma dos custos financeiro e operacional. Quanto menor forem os custos financeiro e operacional, maior será o número de observações.

O resultado da amostragem conveniente é a concentração das observações em áreas usadas para agricultura, próximo de estradas, nas bordas de florestas, e em terrenos não-montanhosos. Perfis de solo descritos e amostrados em cortes de estrada são resultado de amostragem conveniente. Nesse caso, o responsável pela amostragem usa os recursos que seriam necessários para abrir uma trincheira e descrever/amostrar um perfil para descrever/amostrar vários perfis usando os cortes de estrada.

O uso da amostragem conveniente exige assumir que a estrutura da variação espacial do solo pode ser representada pelas observações obtidas, por exemplo, ao longo de estradas. Em outras palavras, qualquer estrutura de variação espacial que não apareça ao longo das estradas é descartada. Em áreas de pequena complexidade pedológica e bem servidas por estradas, a amostragem conveniente tem grande chance de ser eficiente. Do contrário, a amostragem conveniente é incapaz de capturar a maior parte da estrutura de variação espacial do solo. Isso porque estradas costumam ser construídas em posições elevadas da paisagem, com boa drenagem, como nos divisores de águas. As áreas de borda de florestas são fortemente influenciadas pelo

uso da terra subjacente. E os terrenos não-montanhosos costumam ter solo com maior desenvolvimento em profundidade. O modelo preditivo calibrado com tais observações terá bom desempenho aqui, mas não no restante da área.

5.1.2.3 Amostragem intencional

A amostragem intencional (em inglês, *purposive sampling*) é semelhante à amostragem conveniente no sentido de que em ambas os locais amostrais são definidos a fim de otimizar um critério pré-determinado. A diferença fundamental entre as duas é a natureza desse critério. Enquanto na amostragem conveniente o critério tem origem puramente econômica, na amostragem intencional o critério tem origem pedológica e/ou estatística, podendo-se agregar critérios de origem econômica.

Um critério pedológico e/ou estatístico tem origem no modelo usado para descrever a estrutura da variação espacial do solo e o método usado para fazer predições espaciais. Por exemplo, o objetivo pode ser selecionar os locais amostrais de maneira a obter a melhor cobertura espacial porque isso minimiza a incerteza das predições. O resultado disso seria uma amostra com observações equidistantes.

Em geral, a amostragem intencional é o tipo de amostragem mais eficiente para a obtenção de observações de calibração para a modelagem espacial do solo. Isso se dá exatamente porque a localização das observações é definida com base no modelo usado para descrever a estrutura da variação espacial do solo. Em outras palavras, a configuração espacial das observações é otimizada para atender aos pressupostos e requerimentos do modelo de distribuição espacial que será usado. Quanto mais conhecido for o modelo de distribuição espacial, tanto mais ótima será a configuração espacial das observações.

O método do caminhar livre (em inglês, *free survey*), usado no mapeamento tradicional, utiliza-se da amostragem intencional. Aqui a localização das observações é definida com base no conhecimento tácito do responsável, seu modelo mental das relações solo-paisagem ou pedogênese [WebsterEtAl1990, Rossiter2000, BrusEtAl2007a]. O modelo mental é construído com a experiência obtida no campo e sua qualidade geralmente é diretamente proporcional ao número de anos de trabalho de campo. É este modelo aquele usado para descrever a estrutura da variação espacial do solo.

Um dos critérios utilizados para a escolha dos locais de observação no método do caminhar livre é a obtenção de uma amostra “representativa” das feições geomórficas, manchas de características do solo e usos da terra da área sendo mapeada [SamuelRosa2012]. A localização das observações também costuma ser definida de forma a testar as hipóteses postuladas pelo responsável em função do seu modelo mental de pedogênese. Isso resulta em um grande número de observações concentradas na chamadas áreas-problema [Rossiter2000]. As áreas-problema são aquelas para as quais o modelo mental de pedogênese é incompleto ou possui limitações significativas, ou seja, a variação espacial do solo é predita pobremente.

Outra forma de amostragem intencional é o uso de uma malha regular de pontos amostrais. Isso é comumente usado em trabalhos de mapeamento do solo para fins de agricultura de precisão, onde as características do solo costumam ser espacialmente homogêneas devido ao manejo agrícola. O objetivo principal não é a construção de um modelo para descrever a estrutura da variação espacial das propriedades do solo, mas sim obter predições com o menor erro possível usando métodos de interpolação como a krigagem. Quanto mais homogênea for a distribuição espacial das observações, menor será o erro de predição na área como um todo. Em outras palavras, o critério usado para otimizar a configuração espacial das observações é o erro de predição.

Um aspecto importante da amostragem intencional é que uma configuração espacial otimizada para um modelo de distribuição espacial geralmente será sub-ótimo para outro modelo de distribuição espacial. Por exemplo, uma configuração espacial otimizada para atender aos requerimentos de um modelo mental de pedogênese será sub-ótimo para um modelo estatístico, e vice-versa. O mesmo vale para dois modelos mentais de pedogênese ou dois modelos estatísticos diferentes. É nesse sentido que quanto mais conhecido for o modelo de distribuição espacial, tanto mais ótima será a configuração espacial das observações.

5.2 Sumário

Na amostragem não-probabilística, os critérios usados na definição dos locais de amostragem são:

- Amostragem casual: desconhecido - baseado na subjetividade do responsável
- Amostragem por conveniência: minimizar os custos financeiro e operacional
- Amostragem intencional: atender aos pressupostos e requerimentos do modelo de distribuição espacial

Para o mapeamento do solo, é mais adequado usar: * Amostragem intencional para calibrar o modelo de distribuição espacial * Amostragem probabilística para validar as predições espaciais

5.3 Bibliografia consultada

de Gruijter, J.J.; Brus, D.; Bierkens, M.; Knotters, M. *Sampling for natural resource monitoring*. Berlin: Springer, p. 332, 2006.

Diggle, P.J.; Ribeiro Jr, P.J. *Model-based geostatistics*. New York: Springer, p. 228, 2007.

Müller, W.G. *Collecting spatial data - optimum design of experiments for random fields*. Berlin: Springer, p. 242, 2007.

Webster, R.; Lark, R.M. *Field sampling for environmental science and management*. London: Routledge, p. 200, 2013.

Chapter 6

Final Words

We have finished a nice book.

Bibliography

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.

Xie, Y. (2016). *bookdown: Authoring Books with R Markdown*. R package version 0.0.65.