

Modelos, hipóteses e a lógica da ciência

Motivação

A construção de modelos matemáticos e estatísticos é fundamental para a Ciência moderna. Nesta lição vamos discutir aspectos da construção de modelos, bem como suas fundações filosóficas. No processo, vamos entender a relação entre hipóteses científicas e hipóteses estatísticas e entre hipóteses e modelos. Além disso, vamos aprender a evitar [reificação](#) e entender melhor o papel de modelos em Ciência.

Exercícios de fixação

1. **Modelos com erro de observação:** Em muitos problemas estatísticos, temos erros de medição ou observação. Em Epidemiologia, por exemplo, raramente observamos o verdadeiro número de casos de uma doença; em vez disso, temos acesso aos números de casos *notificados*. Neste exercício vamos estudar alguns aspectos de modelos de contagem com erro de observação. Suponha que X_1, X_2, \dots, X_n são contagens ($X_i \in \mathbb{N} \cup \{0\}$), assumidas i.i.d. Suponha ainda que postulamos o seguinte modelo para os dados:

$$Y_i \sim \text{Binomial-Negativa}(\mu, \phi), \quad (1)$$

$$X_i | Y_i \sim \text{Binomial}(Y_i, \theta), \quad (2)$$

para $i = 1, 2, \dots, n$ e onde a binomial negativa é parametrizada em termos de [média e dispersão](#)¹.

- a. Compute $E[X_1]$ e $\text{Var}(X_1)$ usando a [propriedade da torre](#).
- b. Calcule a probabilidade marginal conjunta $\Pr(X_1 = x_1, \dots, X_n = x_n)$.
- c. O modelo em questão é identificável?
- d. Como você modificaria $\Pr(X_1 = x_1, \dots, X_n = x_n)$ se o evento $X_i = 0$ não fosse observável?

¹Ver também [estas](#) notas do grande John Cook.

e. Suponha que o modelo fosse:

$$N \sim \text{Binomial-Negativa}(\mu, \phi), \quad (3)$$

$$X_i | N \sim \text{Binomial}(N, \theta). \quad (4)$$

Como isso mudaria suas respostas aos itens a-d?

f. Suponha agora que o processo de observação fosse: “Se pelo menos um indivíduo do cluster Y_i é observado, então todos são observados”. Como você escreveria $\Pr(X_1 = x_1, \dots, X_n = x_n)$?

2. **Todos os modelos são errados, bla bla bla:** Neste exercício vamos estudar com mais cuidado o famoso aforismo “[All models are wrong, some are useful](#)”. Para tanto, vamos nos debruçar sobre o artigo seminal “Science and statistics” (1976) do estatístico britânico [George E.P.Box](#) (1919-2013). Em particular, você deve escrever um ensaio de no máximo duas laudas sobre o paper em questão. Aqui estão algumas perguntas para lhe estimular:

- Qual a importância que Box dá à flexibilidade nos exercícios de modelagem?
- Como o autor vê o papel da parcimônia? Aliás, o que é parcimônia neste contexto? Tem relação com algum objeto perfuro-cortante?
- Discuta o ciclo teoria-prática.
- Discuta como aplicações de matemática sofisticada aparecem na prática.

Dica: considere as seções 3.6 e 3.11.

Referências

- Capítulo 1 de [Statistical Rethinking](#);
- [Box \(1976\)](#). Science and statistics. Journal of the American Statistical Association, 71(356), 791-799;
- [Nester \(1996\)](#). An applied statistician’s creed. Journal of the Royal Statistical Society: Series C (Applied Statistics), 45(4), 401-410.