Aula 1: Estatística e Probabilidade

Análise Quantitativa de Dados Ambientais

Thiago S. F. Silva - tsfsilva@rc.unesp.br

Programa de Pós Graduação em Geografia - IGCE/UNESP

8 de Abril de 2019

Outline

Probabilidade

Probabilidade

- A base de toda a estatística
- Conceitualmente simples...
- ...mas que rapidamente se torna **bem complexa**.
- A probabilidade mede as “chances” de um determinado evento ocorrer

Ex.: Qual a probabilidade de um inseto ser capturado por uma planta carnívora?

Probabilidade

Para falar de probabilidade, precisamos definir alguns termos:

- **Evento (A):** um processo probabilístico (Ex.: $A =$ tentativa de captura)

Probabilidade

Para falar de probabilidade, precisamos definir alguns termos:

- **Evento (A):** um processo probabilístico (Ex.: A = tentativa de captura)

- **Resultado (<u>outcome, A_i</u>):** resultado observado do evento (Ex.: A_1 = hove captura)

Probabilidade

Para falar de probabilidade, precisamos definir alguns termos:

- **Evento (A):** um processo probabilístico (Ex.: A = tentativa de captura)

- **Resultado (outcome, A_i):** resultado observado do evento (Ex.: A_1 = hove captura)

- **Espaço (Universo) Amostral ($S = A_i, ..., A_n$):** todos os resultados possíveis de um evento (Ex: $S_{captura} = \{HouveCaptura, NoHouveCaptura\}$)

Probabilidade

Para falar de probabilidade, precisamos definir alguns termos:

- **Evento (A):** um processo probabilístico (Ex.: A = tentativa de captura)
- **Resultado (<u>outcome, A_i</u>):** resultado observado do evento (Ex.: A_1 = hove captura)
- **Espaço (Universo) Amostral ($S = A_i, ..., A_n$):** todos os resultados possíveis de um evento (Ex: $S_{captura} = \{HouveCaptura, NoHouveCaptura\}$)
- Neste exemplo, o espaço amostral é discreto

Probabilidade em 15 minutos

Axioma 1: A probabilidade de qualquer evento dentro do espaço amostral é um número real positivo

$$P(A) \in \mathbb{R}, P(A) \geq 0 \quad \forall A \in S$$

Axioma 2: A soma das probabilidades de todos os resultados dentro do espaço amostral é igual a 1

$$\sum_{i=1}^{n} P(A_i) = 1$$

Probabilidade em 15 minutos

Regra da Subtração: A probabilidade de observar um determinado resultado é complementar à probabilidade deste resultado não ser observado

$$P(A) = 1 - P(A^c)$$

Ex.: Qual a probabilidade de tirarmos 5 em um dado?

$$P(A) = \frac{1}{6} = 1 - P(A^c) = 1 - \frac{5}{6} = \frac{1}{6}$$

Probabilidade em 15 minutos

Regra da Multiplicação: Se dois eventos são **independentes**, a probabilidade de que os dois ocorram juntos é o **produto** da probabilidade de cada evento (**interseção das probabilidades**, $\cap$)

$$P(A \cap B) = P(A) \times P(B)$$

Ex.: Qual a probabilidade de tirarmos um 5 **e** um 6 em dois dados?

$$P(A \cap B) = P(A) \times P(B) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

Probabilidade em 15 minutos

Regra da Adição: Se dois eventos são **mutuamente exclusivos (disjuntos)**, a probabilidade de que algum deles ocorra é a **soma** da probabilidade de cada evento (**união das probabilidades**, $\cup$)

$$P(A \cup B) = P(A) + P(B)$$

Ex.: Qual a probabilidade de tirarmos um 5 **ou** um 6 em um dado?

$$P(A \cup B) = P(A) + P(B) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6}$$

Probabilidade em 15 minutos

Se dois eventos **não** são mutuamente exclusivos, usamos:
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
Ex.: Qual a probabilidade de sortearmos um 7 (A) **ou** uma carta de espadas (B) de um baralho com 52 cartas?

$$P(A = 7) = \frac{4}{52} = 0.077, P(B = \text{espadas}) = \frac{13}{52} = 0.25$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.077 + 0.25 - (0.077 \times 0.25) = 0.308$$

Por que subtrair $P(A \cap B)$?

Probabilidade em 15 minutos

Probabilidade condicional: é a probabilidade de que um evento ocorra, dado que outro evento relacionado **já ocorreu**:
$P(A|B) = P(A \cap B)/P(B)$
Ex.: Qual a probabilidade de uma carta sorteada ser um 7 (A), sabendo que a carta é de espadas (B)?

$$P(A = 7) = \frac{4}{52} = 0.077, P(B = \text{espadas}) = \frac{13}{52} = 0.25$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.077 \times 0.25}{0.25} = 0.077$$

Se já sabemos que a carta é de espadas, a probabilidade de obter um 7 é 1/13, que equivale a 4/52

Probabilidade em 15 minutos

Multiplicação para eventos dependentes: Se dois eventos são **dependentes**, a probabilidade de que os dois ocorram juntos pode ser obtida pela relação anterior:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad P(A \cap B) = P(B) \times P(A|B)$$

Ex.: Em Rio Claro, a chance de ser picado por Aedes egyptii (C) é de 70% por dia. Assumindo que a chance de um mosquito transmitir o vírus (T) é de 50% , qual a probabilidade de um aluno de estatística pegar dengue hoje?

Probabilidade em 15 minutos

A probabilida de transmissão é condicional à picada. Se houve picada, $P(A|B) = 0.5$. Se não houve picada, $P(A|B) = 0$.

$$P(A \cap B) = P(B) \times P(A|B)$$

$$P(T \cap C) = P(C) \times P(T|C) = 0.7 \times 0.5 = 0.35$$

Exercício: Gotellli & Ellison 2^a Ed. Ing. pp. 15-17

Plantas vs. lagartas

Em uma paisagem, temos manchas de dois fenótipos de uma planta: R é resistente à herbivoria por lagartas, enquanto R^c não é. Os fenótipos nunca ocorrem juntos na mesma mancha, e fenótipos resistentes ocorrem na paisagem com frequência de 20%.
A probabilidade de uma mancha ser invadida por lagartas (C) é de 0.7, independente da variedade.

Assumindo que as lagartas se dispersam igualmente para todas as manchas, e que somente populações resistentes sobrevivem à chegada das lagartas, qual a probabilidade de que uma mancha desapareça devido à herbivoria?

Dica: Primeiro calcule as probabilidades de ocorrência das quatro combinações possíveis de resultados.

Exercício: Gotellli & Ellison 2^a Ed. Ing. pp. 15-17

Primeiro Passo: organizando as informações:

Resistente ou suscetível são resultados mutuamente exclusivos:
$P(R) = 0.2, \quad P(R^c) = P(1 - R) = 0.8$

Presença e ausência de lagartas são resultados mutuamente exclusivos:
$P(L) = 0.7, \quad P(L^c) = P(1 - L) = 0.3$

$S = \{R^cL^c, RL^c, RL, R^cL, RL\}$

Exercício: Gotellli & Ellison 2nd Ed. Ing. pp. 15-17

Segundo Passo: Expressando as probabilidades. Resistência e invasão por lagartas são eventos independentes:

$$P(R^cL^c) = P(R^c) \times P(L^c)$$

$$P(RL^c) = P(R) \times P(L^c)$$

$$P(R^cL) = P(R^c) \times P(L)$$

$$P(RL) = P(R) \times P(L)$$

Multiplicamos porque os eventos são independentes.

Exercício: Gotellli & Ellison 2^a Ed. Ing. pp. 15-17

Terceiro Passo: Calculando as probabilidades:

$$P(R^c \cap L^c) = P(R^c) \times P(L^c) = 0.8 \times 0.3 = 0.24$$

$$P(R \cap L^c) = P(R) \times P(L^c) = 0.2 \times 0.3 = 0.06$$

$$P(R^c \cap L) = P(R^c) \times P(L) = 0.8 \times 0.7 = 0.56$$

$$P(R \cap L) = P(R) \times P(L) = 0.2 \times 0.7 = 0.14$$

Exercício: Gotellli & Ellison 2^a Ed. Ing. pp. 15-17

Quarto Passo: Combinando as probabilidades:

$P(R^c \cap L^c) = P(R^c) \times P(L^c) = 0.8 \times 0.3 = 0.24$: Planta permanece

$P(R \cap L^c) = P(R) \times P(L^c) = 0.2 \times 0.3 = 0.06$: Planta permanece

$P(R^c \cap L) = P(R^c) \times P(L) = 0.8 \times 0.7 = 0.56$: **Planta desaparece**

$P(R \cap L) = P(R) \times P(L) = 0.2 \times 0.7 = 0.14$: Planta permanece

$\mathbf{P}(\textbf{Planta desaparece}) = \mathbf{0.56}$
$P(\text{Planta permanece}) = P((R^c \cap L^c) \cup (R \cap L^c) \cup (R \cap L)) = 0.44$

Expandindo o Exercício

Plantas vs. lagartas vs. invasoras
Mesmo nos fenótipos resistentes, a herbivoria diminui a capacidade competitiva da planta estudada, facilitando o estabelecimento (I) de uma espécie invasora. Se há presença da lagarta, a invasão tem uma taxa de sucesso de 60%, e se não há plantas, o sucesso é garantido (100%).

Qual a probilidade de que haja invasão, sabendo que as lagartas já atingiram a mancha?

Expandindo o Exercício

O primeiro impulso é calcular $P(I \cap L) = P(I|L) \times P(L)$. Mas a herbivoria leva à remoção da planta quando esta não é resistente, modificando a probabilidade de invasão.
Temos, assim, duas probabilidades condicionais:
$P(I \cap R^c L) = P(I|R^c L) \times P(R^c L) = 1 \times 0.56 = 0.56$

$P(I \cap RL) = P(I|RL) \times P(RL) = 0.6 \times 0.14 = 0.084$

RL e $R^c L$ são mutuamente exclusivos, então temos:
$P(I \cap L) = P(I \cap R^c L) \cup P(I \cap RL) = 0.56 + 0.084 = 0.644$

Estimando Probabilidades

Qual a probabilidade de uma planta carnívora capturar um inseto?

- **Como podemos estimar essa probabilidade?**

Estimando Probabilidades

Qual a probabilidade de uma planta carnívora capturar um inseto?

- **Como podemos estimar essa probabilidade?**

- Realizando uma **contagem** dos sucessos e fracassos da planta, para várias visitas de insetos.

Estimando Probabilidades

Qual a probabilidade de uma planta carnívora capturar um inseto?

- **Como podemos estimar essa probabilidade?**

- Realizando uma **contagem** dos sucessos e fracassos da planta, para várias visitas de insetos.

- Cada visita individual é uma **realização** do evento: capturado ou não. Também conhecida como **réplica** ou **observação**.

Estimando Probabilidades

Qual a probabilidade de uma planta carnívora capturar um inseto?

- **Como podemos estimar essa probabilidade?**
- Realizando uma **contagem** dos sucessos e fracassos da planta, para várias visitas de insetos.
- Cada visita individual é uma **realização** do evento: capturado ou não. Também conhecida como **réplica** ou **observação**.
- O conjunto de realizações sucessivas compreende um **experimento**.

Frequência vs. Probabilidade

Frequência de Captura:

$$F = \frac{\text{número de capturas}}{\text{número de visitas}}$$

Frequência vs. Probabilidade

Frequência de Captura:

$$F = \frac{\text{número de sucessos}}{\text{número de realizações}}$$

Frequência vs. Probabilidade

Probabilidade de Captura:

$$P(\text{captura}) \approx \lim_{n_t \to \infty} \frac{\text{número de sucessos}(n_r)}{\text{número de realizações }(n_t)}$$

Culturas Estatísticas

Estatística Frequentista:

- Associada principalmente a Sir Ronald Aymer Fisher, FRS.
- Se baseia na associação entre frequências e probabilidades.
- Ex: Jogo uma moeda 100 vezes, obtenho 45 caras e 55 coroas. Estimo minhas probabilidades como 0.45 e 0.55
- Vê a amostra como uma realização aleatória de um evento
- Parte do princípio de que se o processo fosse repetido infinitamente, seria possivel estimar as probabilidades associadas aos resultados do evento

Culturas Estatísticas

$$p < 0.05?$$

Culturas Estatísticas

$$p < 0.05?$$

$P(A|H)$ ou $P(H|A)$?

Culturas Estatísticas

$$p < 0.05?$$

$P(A|H)$ ou $P(H|A)$?

É a mesma coisa?

Culturas Estatísticas

Na visão frequentista, se avalia a probabilidade de se obter a amostra observada, dada uma determinada hipótese:

$$P(A|H)$$

Ex: Joguei uma moeda 100 vezes, e obtive 65 caras e 35 coroas. Se a minha hipótese é de que a moeda é honesta ($P(\text{cara}) = P(\text{coroa}) = 0.5$), qual chance de eu obter esse resultado, **ou um resultado mais extremo**?

$$p = 8.6 \times 10^{-4}$$

Se eu repetir esse experimento infinitas vezes (jogar 100 moedas), vou encontrar um resultado igual ou mais extremo 0.086% das vezes.

Culturas Estatísticas

A lógica nos diz que o mais importante é saber $P(H|S)$. Mas como?

Culturas Estatísticas

A lógica nos diz que o mais importante é saber $P(H|S)$. Mas como?

Teorema de Bayes:

$$P(H \mid A) = \frac{P(H \cap A)}{P(A)} = \frac{P(A \mid H) \times P(H)}{P(A)}$$

$P(A|H)$: probabilidade da amostra se a hipótese é verdadeira
$P(A)$: probabilidade da amostra, garante que $0 \leq P(H|A) \leq 1$
$\mathbf{P(H)}$: probabilidade da hipótese ser verdadeira. Conhecida como **priori (prior)**.

Culturas Estatísticas

Estatística Bayesiana

- Associada a Thomas Bayes
- Na visão bayesiana, a análise estatística serve para **atualizar** o conhecimento anterior
- O conhecimento prévio pode ser usado para definir uma probabilidade priori da hipótese ser verdadeira
- O resultado do experimento permite que voce atualize (melhore) essa estimativa de probabilidade, com base na amostra observada.

Ex.: Joguei uma moeda 100 vezes, e obtive 65 caras e 35 coroas. Se a minha hipótese é de que a moeda é honesta ($P(\text{cara}) = P(\text{coroa}) = 0.5$), qual a probabilidade que essa hipótese esteja correta?

H_0: A moeda é honesta
H_1: A moeda é tendenciosa

Baseado em meu conhecimento de moedas, eu poderia dizer que a probabilidade dela ser honesta é 0.9 ($P(H_0) = 0.9$), e a probabilidade dela ser tendenciosa é 0.1 ($P(H_1) = 0.1$).

Estatística Bayesiana

Para H_0:

$$P(H_0|A) = \frac{P(A|H_0) \times P(H_0)}{P(A)}$$

$$P(H_0|A) = \frac{P(A|H_0) \times P(H_0)}{P(A|H_0) \times P(H_0) + P(A|H_1) \times P(H_1)}$$

$$P(H_0|A) = \frac{P(8.6 \times 10^{-4})P(0.9)}{8.6 \times 10^{-4} \times 0.9 + 0.0834 \times 0.1}$$

$P(H_0|A) = 0.085$

Estatística Bayesiana

Para H_1:

$$P(H_1|A) = \frac{P(A|H_1) \times P(H_1)}{P(A)}$$

$$P(H_1|A) = \frac{P(A|H_1) \times P(H_1)}{P(A|H_1) \times P(H_1) + P(A|H_0) \times P(H_0)}$$

$$P(H_0|A) = \frac{P(0.0834) \times P(0.1)}{0.0834 \times 0.1 + 8.6 \times 10^{-4} \times 0.9}$$

$P(H_1|A) = 0.915$

Estatística Bayesiana

A escolha da priori afeta fortememente a posteriori:

$$\mathbf{P(H_0 = 0.5), P(H_1 = 0.5)} \rightarrow P(H_0|S) = 0.01, P(H_1|S) = 0.98$$

$$\mathbf{P(H_0 = 0.75), P(H_1 = 0.25)} \rightarrow P(H_0|S) = 0.03, P(H_1|S) = 0.97$$

$$\mathbf{P(H_0 = 0.95), P(H_1 = 0.05)} \rightarrow P(H_0|S) = 0.16, P(H_1|S) = 0.84$$

$$\mathbf{P(H_0 = 0.99), P(H_1 = 0.01)} \rightarrow P(H_0S) = 0.506, P(H_1|S) = 0.494$$

Frequentista vs. Bayesiana

Bayesianos sobre frequentistas:

- Ignoram qualquer informação a priori
- Se baseiam em experimentos fictícios

Frequentistas sobre Bayesianos:

- Podem gerar o resultado que quiserem manipulando as <u>priori</u>

Qual vamos usar?

O Curso se baseará na filosofia frequentista.

- Mais frequentemente usada (tu-dum psh).
- Mais familiar à comunidade ecológico-científica.
- É a estatística com a qual voces já tiveram algum contato prévio.
- É a que eu sei ensinar.

Contudo, tomaremos cuidado em enfatizar os maus usos e compreensões equivocadas da estatística frequentista.