Métodos Estatísticos

Prof. Bruno Santos

Sala 220 (IME), brunorsantos@ufba.br

Instituto de Matemática e Estatística
Universidade Federal da Bahia

Links interessantes

Site para visualizar alguns conceitos estatísticos:

- Seeing theory
  - www.students.brown.edu/seeing-theory/

Site de notícias baseado em dados:

- FiveThirtyEight
  - www.fivethirtyeight.com

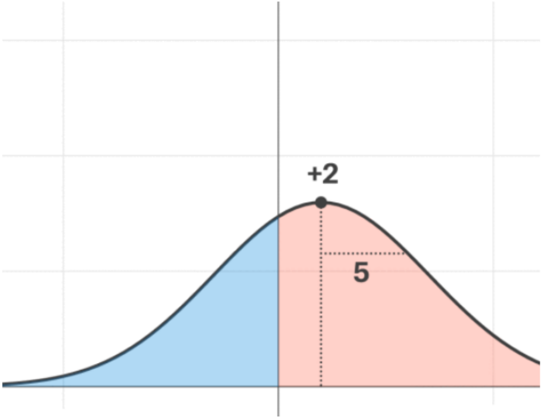
Estatística no jornalismo

SEP 17, 2014 AT 7:30 AM

How The FiveThirtyEight Senate Forecast Model Works

By Nate Silver

Filed under 2014 Midterms



Estatística na imprensa esportiva

Standings Stats Teams More

NFL season a billion times, though, a lot can go wrong (or right) over one measly 16-game stretch.

We can use the binomial distribution to estimate just how things could have gone for a 10.6-win team like the Falcons. Atlanta actually won 11 games this season, which isn't necessarily surprising, as the most likely outcome for a 10.6-win team would naturally be to win 11 games. That would occur 20.7 percent of the time, and the Falcons would win 11 or more games in 53.4 percent of simulations. And 5.6 percent of the time, a team of this caliber would win 14 or more games. On the flip side, though, a team can play excellent football and still (at least theoretically) struggle to win games. There's a 5.4 percent chance a team with Atlanta's Pythagorean expectation would post a losing record.

Hate this level of randomness? Try hoops. Teams are far less likely to put up similarly

Rating which teams are closest to being Super Bowl contenders

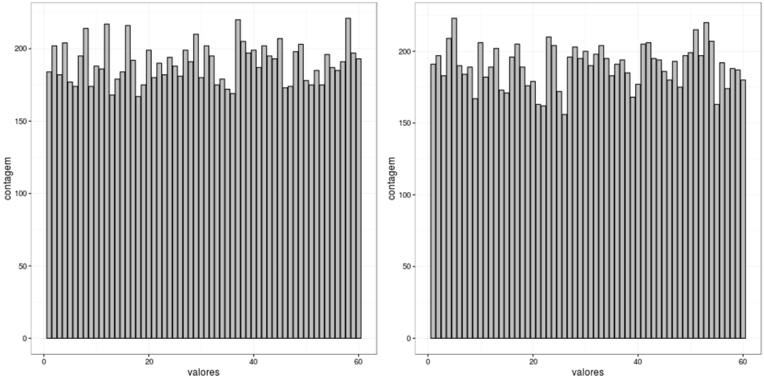
Will the Chiefs contend in 2017? Is Aaron Rodgers' title window closing? NFL Nation reporters assess 30 teams' chances of making a Super Bowl, from nowhere close to on the cusp.

Ranking every player in Super Bowl LI

From 1 to 106, we ranked every player on the active rosters for the Falcons and Patriots. Tom Brady and Matt Ryan are at the top of the list, but who is No. 1?

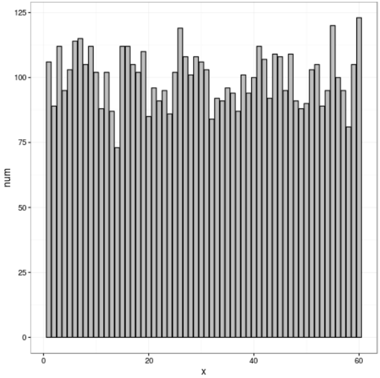
Sobre os sorteios da Mega-Sena

Você saberia qual dizer qual das figuras representa a distribuição dos números sorteados pela Mega Sena e qual foi gerado por um computador?

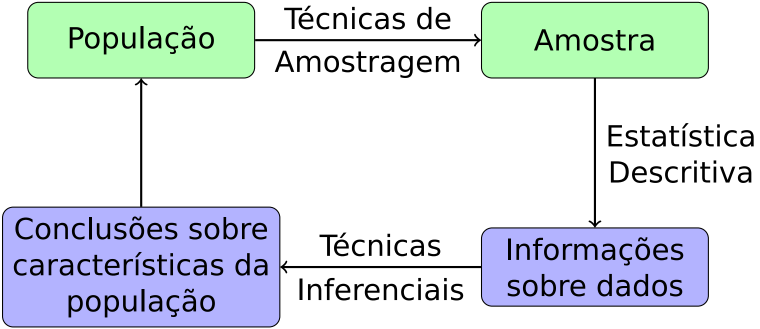


Sobre tamanhos de amostra

Evolução da distribuição de frequência dos números sorteados na Mega-Sena



Processo de análise estatística



BASE: PROBABILIDADE

Definições

EXPERIMENTO ALEATÓRIO: Todo processo em que o seu resultado não pode ser determinado devido à incerteza que está associada a esse processo.

MODELO ESTATÍSTICO: Componentes:

- **ESPAÇO AMOSTRAL (Ω)**: conjunto de todos os resultados possíveis de um experimento aleatório.
- Uma função **PROBABILIDADE**, $P(.)$, para cada ponto do espaço amostral.

EVENTO: Qualquer subconjunto do espaço amostral. Usualmente representado por uma letra maiúscula: A, B, C.

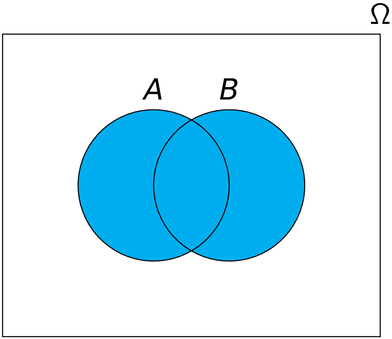
Exemplos

1. EXPERIMENTO: jogar um dado de 6 fases e observar a face superior.
   - $\Omega = \{1, 2, 3, 4, 5, 6\}$.
   - $A =$ faces maiores que 3. ($A = \{4, 5, 6\}$)
2. EXPERIMENTO: avaliar um produto recebido e verificar se é defeituoso ou não.
   - $\Omega = \{\text{DEFEITUOSO}, \text{NÃO DEFEITUOSO}\}$.
   - $\Omega = \{\text{DEFEIT.}, \text{NÃO DEFEIT.}, \text{NÃO É POSSÍVEL DIZER}\}$.
   - $A = \text{DEFEITUOSO}$

Operações com conjuntos

Dados dois eventos, A e B, de um mesmo espaço amostral Ω:

- UNIÃO de A e B ($A \cup B$) representa o evento em que pelo menos um dos eventos ocorre.



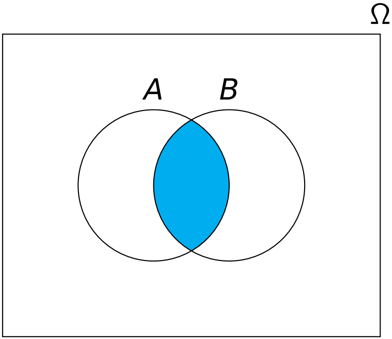
Exemplos

1. Lançamento de dados:
   - A: face par.
   - B: face maior que 3.
   - $A = \{2, 4, 6\}$.
   - $B = \{4, 5, 6\}$.
   - $A \cup B = \{2, 4, 5, 6\}$.
2. Lançamento de três moedas:
   - A = Sair três caras.
   - B = Sair pelo menos duas caras.
   - $A = \{(C, C, C)\}$.
   - $B = \{(K, C, C), (C, K, C), (C, C, K), (C, C, C)\}$
   - $A \cup B = B$

Operações com conjuntos

Dados dois eventos, A e B, de um mesmo espaço amostral Ω:

- INTERSECÇÃO de A e B ($A \cap B$) representa o evento em que os dois eventos ocorrem simultaneamente.



Exemplos

1. Lançamento de dados:
   - A: face par.
   - B: face maior que 3.

   - $A = \{2, 4, 6\}$.
   - $B = \{4, 5, 6\}$.

   - $A \cap B = \{4, 6\}$.

2. Lançamento de três moedas:
   - A = Sair três caras.
   - B = Sair pelo menos duas caras.

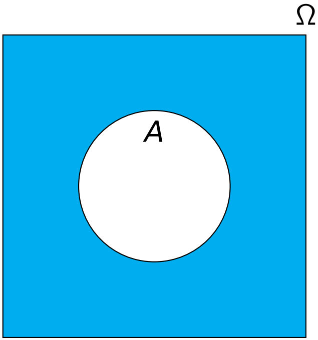
   - $A = \{(C, C, C)\}$.
   - $B = \{(K, C, C), (C, K, C), (C, C, K), (C, C, C)\}$

   - $A \cap B = A$

Operações com conjuntos

Dados um evento A de um espaço amostral Ω:

- o conjunto COMPLEMENTAR de A (A^C ou $\overline{A}$) representa o evento em que A não ocorre.



Exemplos

1. Lançamento de dados:
   - A: face par.
   - A^c: face ímpar.

   - $A = \{2, 4, 6\}$.
   - $A^c = \{1, 3, 5\}$.

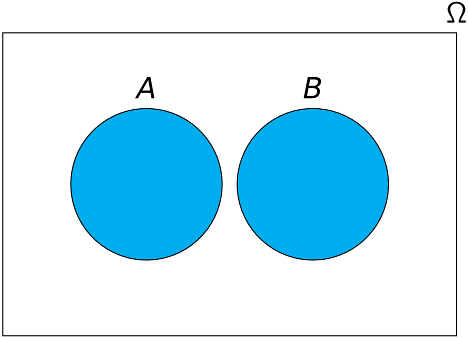
2. Lançamento de duas moedas:
   - A = Sair duas caras.
   - A^c = Não sair duas caras.

   - $A = \{(C, C)\}$.
   - $A^c = \{(K, C), (C, K), (K, K)\}$

Operações com conjuntos

Dados dois eventos A e B de um espaço amostral Ω:

- se A e B são mutuamente exclusivos ou disjuntos, então eles não podem ocorrer simultaneamente, isto é, $A \cap B = \emptyset$.



Exercícios

1. Defina o espaço amostral do experimento aleatório:
   - Lançamento de três moedas;
   - Resultado ao final do semestre em uma disciplina na UFBA;
2. Faça a representação dos seguintes conjuntos:
   - $\overline{A \cup B}$;
   - $A^c \cup B^c$;
   - $A^c \cap B^c$;

Conceitos de Probabilidade

- DEFINIÇÃO AXIOMÁTICA:
  Sejam E um experimento aleatório e Ω o espaço amostral associado a esse experimento.

  A cada evento X associamos um número real representado por $P(X)$, denominado probabilidade de X. O número $P(.)$ deve satisfazer as seguintes propriedades

  1. $P(\Omega) = 1$,
  2. $0 \leqslant P(X) \leqslant 1$, para todo $X \in \Omega$
  3. Se X e Y são disjuntos ($X \cap Y = \varnothing$), temos

     $$P(X \cup Y) = P(X) + P(Y)$$

     De forma geral, podemos dizer que para qualquer sequência de eventos disjuntos dois-a-dois, $A_1, \ldots, A_n$, temos:

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

PROPRIEDADES

1. Se $\emptyset$ é um evento impossível, então:

$$P(\emptyset) = 0.$$

2. Se $\overline{A}$ for o evento complementar de A, então:

$$P(\overline{A}) = 1 - P(A).$$

3. Se A e B são dois eventos quaisquer em Ω, então

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

4. Se A, B e C são eventos quaisquer em Ω, então

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Conceitos de probabilidade

- DEFINIÇÃO CLÁSSICA:
  Seja A um evento associado a um espaço amostral finito, Ω, em que todos os eventos são equiprováveis. A probabilidade de ocorrência do evento A é dada por

$$P(A) = \frac{\#(A)}{\#(\Omega)},$$

  em que $\#(A)$ é o número de eventos favoráveis a A.

Exemplo: Lançar um dado equilibrado e observar o número da face voltada pra cima. Qual é a probabilidade de sair um número par?

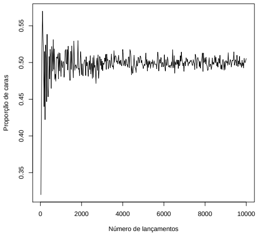
$$\#(A) = 3, \; \#(\Omega) = 6 \Rightarrow P(A) = \frac{3}{6} = \frac{1}{2}$$

Conceitos de probabilidade

- DEFINIÇÃO FREQUENTISTA
  Admita que n_A seja o número de vezes que o evento A ocorre em n repetições independentes do experimento. Então,

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n}$$

  Exemplo: Fazendo o lançamento de uma moeda um determinado número de vezes, o seguinte comportamento pode ser observado



Exemplo - Definição frequentista

Em uma pesquisa realizada numa faculdade:

- 1.162 afirmaram que “colaram” nos exames
- 2.468 afirmaram que nunca “colaram” em exames

Selecionando aleatoriamente um estudante, determine a probabilidade dessa pessoa ter colado em um exame.

Resposta: A = Pessoa “colar” no exame.

$$P(A) = \frac{1.162}{1.162 + 2.468} \approx 0,32$$

Exercícios - pg. 26, Q31

Dois processadores do tipo A e B são colocados em teste por 50 mil horas. A probabilidade de que um erro de cálculo aconteça com um processador do tipo A é de 1/30, no tipo B é 1/80 e em ambos 1/1000.

Qual é a probabilidade:

a) pelo menos um dos processadores tenha problema?

b) Nenhum dos processadores tenha apresentado erro?

c) Apenas o processador tenha A tenha apresentado erro?

PROBABILIDADE CONDICIONAL

Exemplo: Experimento aleatório: retirar uma carta do baralho completo com 52 cartas e observar o "valor" e o naipe.

$$\Omega = \{2, 3, \ldots, 10, J, Q, K, A\} \times \{\diamondsuit \spadesuit \heartsuit \clubsuit\}$$

$$P(\{\omega_i\}) = \frac{1}{52} \quad (\Omega \text{ é equiprovável}),$$

ω_i qualquer elemento do espaço amostral.

- A = { a carta é um ás }.
- B = { o naipe é vermelho }.
- C = { a carta é um número }.
- D = { a carta é um ás de copas }.
- E = { naipe é copas }.
- F = { naipe é espadas }.

Cálculo das probabilidades

$$P(\{A = \text{a carta é um ás}\}) = \frac{4}{52} = \frac{1}{13}$$

$$P(\{B = \text{o naipe é vermelho}\}) = \frac{26}{52} = \frac{1}{2}$$

$$P(\{C = \text{a carta é um número}\}) = \frac{36}{52} = \frac{9}{13}$$

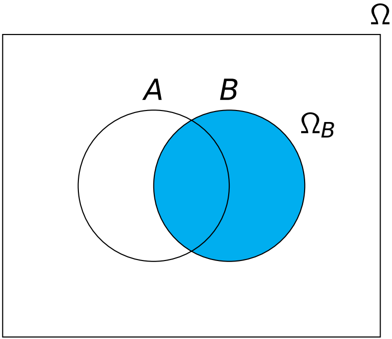
$$P(\{D = \text{a carta é um ás de copas}\}) = \frac{1}{52}$$

$$P(\{E = \text{o naipe é copas}\}) = \frac{13}{52} = P(F)$$

Alterando o universo

Suponha que sabemos que, a priori, que a carta selecionada é de naipe vermelho

Dado que B ocorreu, temos uma restrição no espaço amostral



$$\Omega \rightarrow \Omega_B$$

$$P \rightarrow P_B$$

Intuição da função probabilidade condicional, $P_B(.)$

Considerando essa nova função, $P_B(.)$:

$$P_B(A) = \frac{2}{26} = \frac{1}{13}, \quad P_B(C) = \frac{18}{26} = \frac{9}{13},$$

$$P_B(D) = \frac{1}{26}, \qquad P_B(E) = \frac{13}{26}$$

$$P_B(F) = 0$$

Observações

Note que

$$A = (A \cap B) \cup (A \cap \overline{B})$$

$$P(A) = P(A \cap B) + P(A \cap \overline{B})$$

Dado que B ocorreu, $A \cap \overline{B} = \emptyset$. Logo

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

Observações

Considere que para todo evento A, o seguinte evento é verdade

$$P_B(A) = k\,P(A) = k\,P(A \cap B)$$

Como obter k?

$P_B(B) = P_B(\Omega_B) = 1.$

$$1 = P_B(B) = kP(B \cap B) = kP(B) \Rightarrow k = \frac{1}{B}$$

Então,

$$P_B(A) = \frac{P(A \cap B)}{P(B)} = P(A|B), \forall A \in \Omega, P(B) > 0.$$

Definição Probabilidade Condicional

DEFINIÇÃO: Dado (Ω, P), seja $A, B \in \Omega$ com $P(B) > 0$. A probabilidade condicional de A dado que B ocorreu é

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

De forma análoga, temos

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

Regra do Produto

a) Sejam $A, B \in \Omega$, então

$$P(A \cap B) = P(A|B)P(B)$$
$$\text{ou } [P(B|A)P(A)]$$

b) $A_1, A_2, \ldots, A_n \in \Omega$.

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1 \cap A_2 \cap \cdots \cap A_n)$$
$$= P(A_1)P(A_2|A_1)\cdots P(A_n|A_{n-1} \cap \cdots \cap A_1)$$

Observação:

1. Notação: $P(A_1 \cap A_2 \cap A_3) = P(A_1, A_2, A_3)$.

EVENTOS INDEPENDENTES

Os eventos A e B são independentes se a informação a B não altera a probabilidade do evento A. Ou seja,

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

$$P(A|B) = P(B)$$

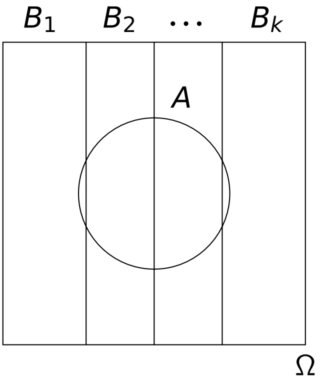
Também,

$$P(B|A) = P(B)$$

Exercícios

1. Um dado equilibrado é lançado e o número voltado para cima é observado.
   a) Se o resultado obtido é par, qual a probabilidade dele ser maior ou igual a 5?
   b) Se o resultado obtido for maior ou igual a 5, qual a probabilidade dele ser par?

Partição do espaço amostral



Dizemos que $B_1, B_2, \cdots, B_k$ formam uma partição do espaço amostral (Ω) se eles são disjuntos dois-a-dois e sua união é Ω:

$$\bigcup_{i=1}^{k} B_i = \Omega, \text{em que } B_i \cap B_j = \varnothing, \forall i \neq j.$$

Lei da probabilidade total

Suponha que os eventos $B_1, \cdots, B_k$ formam uma partição de Ω e todos com probabilidade positiva. Então, para qualquer evento A, temos

$$
\begin{aligned}
P(A) &= \sum_{i=1}^{k} P(A \cap B_i) \\
&= \sum_{i=1}^{k} P(A|B_i)P(B_i)
\end{aligned}
$$

Teorema de Bayes
Inversão de probabilidades condicionais

Vimos que

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

Pela Regra do Produto, temos que

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

E pela Lei da Probabilidade Total, podemos reescrever o resultado anterior como

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)},$$

em que $A_1, \ldots, A_k$ formam uma partição do espaço amostral

Base da inferência Bayesiana

$$\boxed{P(A_i|B)} = \frac{P(B|A_i)}{\sum_{i=1}^{k} P(B|A_i)\, P(A_i)} \quad \boxed{P(A_i)}$$

Probabilidade a posteriori   Probabilidade a priori

- O Teorema de Bayes mostra como devemos atualizar nossa incerteza diante de novas informações.
- O cérebro é bayesiano?

Exercício, Ex. 22, pg. 25

Numa faculdade 30% dos homens e 20% das mulheres estudam matemática. Além disso, 45% dos estudantes são mulheres. Se um estudante selecionado aleatoriamente qual a probabilidade de que esta pessoa seja uma mulher?

Sugestão de exercícios: 27, 28, 29.