Estatística para a Pesquisa Médica

James R. Hunter

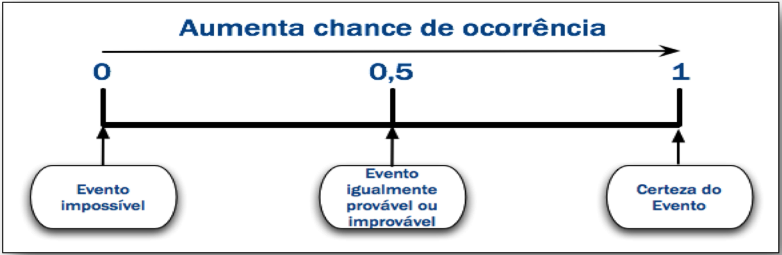
Retrovirologia - Doenças Infecciosas - Departamento de Medicina - UNIFESP

Historia

- Ponto de partida para estatística
- Assunto de interesse desde século 12
- Sempre associada com jogos de azar
- Envolve grande pensadores como
  - Fibonacci, Século 12,
  - Girolamo Cardano (*Liber de Ludo Aleae*), Sec. 15
  - Chevalier de Méré, Blaise Pascal, Pierre de Fermat, Sec. 16
  - Abraham de Moivre, Gauss, Bernoulli(s), LaPlace, Sec. 17 - 18

Escala de Probabilidade

- Em termos quantitativos, probabilidade estende de 0 até 1
- É um número puro; não tem unidade



Exemplo (Não Tão) Simples

- O que é a probabilidade de ganhar a Megasena?
- Precisa selecionar 6 números de 60
- Podemos tirar o mesmo número 2 vezes? Não – "sem reposição"
- Como calcularia você?

Exemplo Super Simples

- O que é a probabilidade que um "1" vai aparecer num dado equilibrado jogado na mesa?
- São 6 números num dado
- Só um vai aparecer
- 1 chance em 6 ou $1/6 = 16{,}67\%$

$$p[1] = p[2] = p[3] = p[4] = p[5] = p[6] = 0,1667$$

Voltamos a Megasena

- Como só tem um número que ganha no dado
- Só tem uma **combinação** que ganha na Megasena

Métodos de Contagem

- Visão frequentista da probabilidade
- Contar as possíveis soluções com o resultado desejado
- Contar todas as possíveis soluções
- Comparar as duas

$$p[E] = \frac{soluções\ com\ resultado\ desejado}{todas\ as\ soluções\ possíveis}$$

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Probabilidade de Megasena

- Numerador: Só tem uma combinação dos números que ganha
- Denominador: Todos as combinações possíveis
- **Quantas combinações de 6 números pode escolher?**

Combinações (Funções de *Choosing*)

Se tiramos `r` objetos de um conjunto de `n` objetos sem reposição e sem referência a ordem, quantas amostras diferentes são possíveis?

$$\binom{n}{r} = \frac{n!}{(n-r)!r!}$$

Nota Bene:

1. Falado: "n choose r"
2. Escrito: ou ${}_nC_r$ ou $\binom{n}{r}$

Fatoriais

$$5! = 5 * 4 * 3 * 2 * 1 = 120$$

- Fatoriais crescem muito rápido

Número	Fatorial
1	1
2	2
3	6
4	24
5	120
6	720
7	5040
8	40320
9	362880
10	3628800
11	39916800
12	479001600

Megasena, De Novo

- Cálculo de denominador

$$\binom{60}{6} = \frac{60!}{(60-6)!6!} = 50.063.860$$

- Probabilidade de Ganhar a Mega

$$\frac{1}{50063860} = 0,000000020 = 2.0E - 08$$

Caixa Concorda

PROBABILIDADE DE ACERTO NA MEGA-SENA				
Quantidade Nº Jogados	Valor de Aposta	Probabilidade de acerto (1 em ...)		
		Sena	Quina	Quadra
6	2,00	50.063.860	154.518	2.332
7	14,00	7.151.980	44.981	1.038
8	56,00	1.787.995	17.192	539
9	168,00	595.998	7.791	312
10	420,00	238.399	3.973	195
11	924,00	108.363	2.211	129
12	1.848,00	54.182	1.317	90
13	3.432,00	29.175	828	65
14	6.006,00	16.671	544	48
15	10.010,00	10.003	370	37

Compraria 2 Números???

- Se compramos 2 jogos de Megasena, temos 2 chances de ganhar.
- Dobrou a possibilidade de ganhar
- **Vale a pena?**

$$\frac{2}{50063860} = 0,000000040$$

Odds da Megasena

- Em pesquisa médica, usamos muito o conceito de "odds"
- Definição: relação entre chances de ganhar sobre as chances de perder

$$Odds = \frac{chances\ de\ ganhar}{chances\ de\ perder}$$

$$Odds\ Megasena = \frac{1}{50.063.380 - 1} = \frac{1}{50.063.379}$$

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Combinações e Permutatções

- Quantos codons diferentes são possíveis usando as 4 bases de DNA?
- $4^3 = 4 * 4 * 4 = 64$
- Permutação - ordem conta
  - GGC - glycine
  - CGC - arginine
- $(\#\, de\ resultados\ por\ evento)^{(\#\, dos\ eventos)} = n^r$
- Quando tem "com reposição" ou "ordem conta", **use poderes**

A Senhora Bebendo Vinho

- Variação no problema famosa estatística: a senhora bebendo chá
- Uma senhora afirma que pode identificar quatro vinhos diferentes só pelo sabor
- O que é a probabilidade que ela pode conseguir isso somente por hasard?
    - Ela só tem 4 taças não identificadas
    - Dica: Tem reposição????

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Vinho - Solução

- Só tem uma solução de sucesso (identificar corretamente todos os 4 vinhos)
- Quantas soluções existem?
- Número de chutes que pode fazer:
  - Taça 1: 4 escolhas
  - Taça 2: 3 escolhas restantes
  - Taça 3: 2 escolhas restantes
  - Taça 4: 1 escolha restante

$$p(sucesso) = \frac{1}{4 * 3 * 2 * 1} = \frac{1}{4!} = \frac{1}{24} = 0,04167$$

Novo Jogo com Dados

- Vamos jogar 1 dado 6 vezes
- Em cada jogo, tem uma chance em 6 de jogar um 1
- Evento "E" tem probabilidade de 1/6 $P(E) = {}^{1}\!/_{6}$
- Em 6 jogos, quantos "1" pode esperar?
- Experimento empírico
  - Vamos fazer 500 experimentos (trials) de 6 jogadas do dado
  - Medir quantos "1" resultam

Tabela dos Resultados

- Primeiro 10 experimentos

exp	uns	freqrel	freqrelcum
1	0	0.0000000	0.0000000
2	3	0.5000000	0.2500000
3	2	0.3333333	0.2777778
4	0	0.0000000	0.2083333
5	0	0.0000000	0.1666667
6	0	0.0000000	0.1388889
7	2	0.3333333	0.1666667
8	2	0.3333333	0.1875000
9	2	0.3333333	0.2037037
10	1	0.1666667	0.2000000

Gráfico da Freqüência Relativa

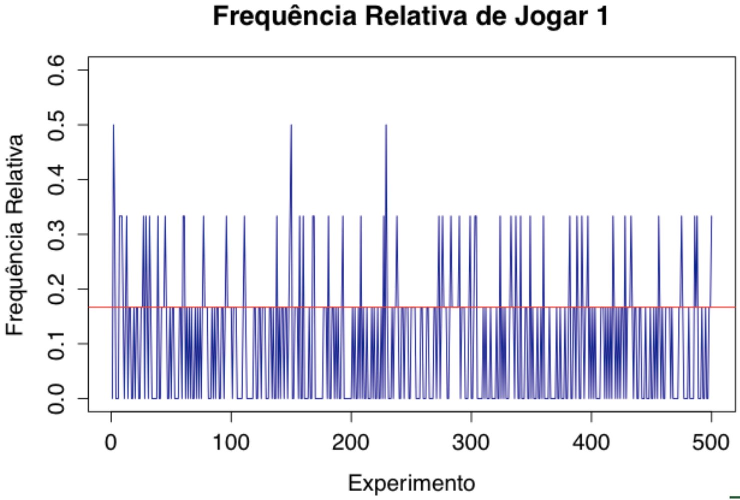
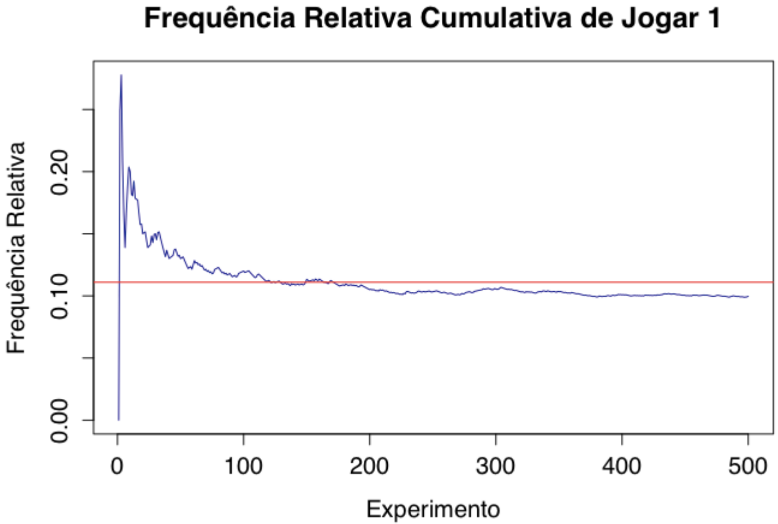


Gráfico da Freqüência Relativa Cumulativa



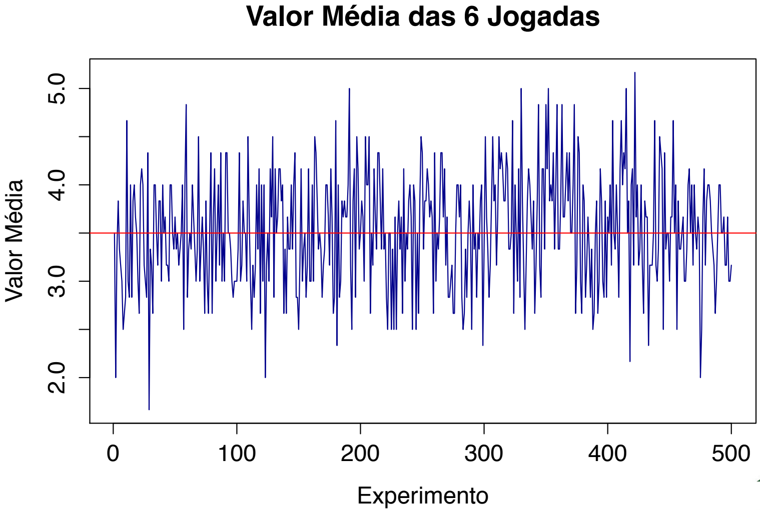
Lei de Grandes Números

- Maior o número dos experimentos
  - Mais perto seja a probabilidade empírica a probabilidade verdadeira
  - Mundo real a mundo teórico
- Aplicações
  - Seguro da vida
  - Pesquisa médica
  - Casinos – por quê?

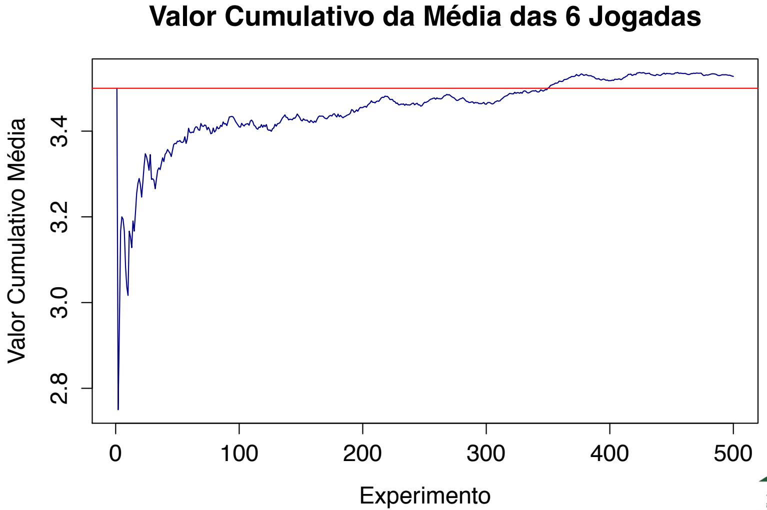
Reversão à Média

$\boxed{\text{Com mais experimentos, a média dos resultados aproximará à média teórica}}$

Valor Médio de 6 Jogadas



Valor Cumulativo Médio de 6 Jogadas



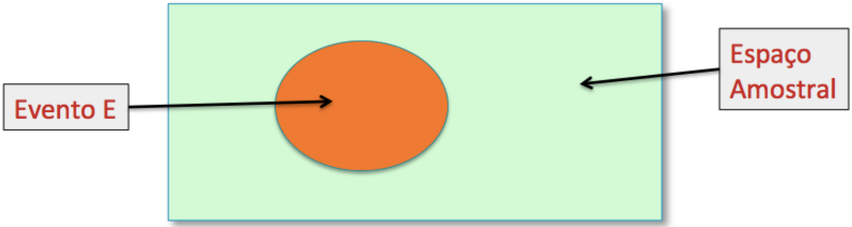
Operações Lógicas

- Eventos: E;F
- Operadores
  - União (E ou F) $[E \cup F]$
  - Interseção (E e F) $[E \cap F]$
  - Negação (não E) $[\overline{E}\ ou\ \sim E]$
    - Também chamado complemento
  - Condicionado (E se F ocorreu) $[E|F]$

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

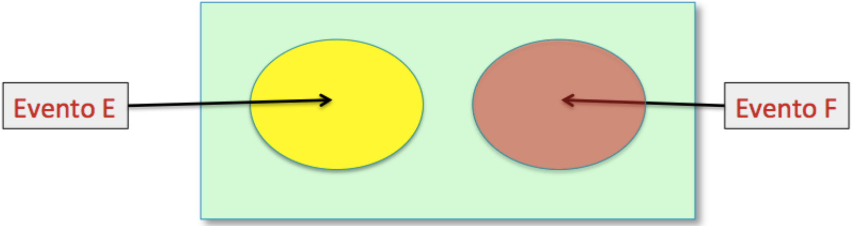
Lei de Probabilidade Total

$$P(E) + P(\sim E) = 1$$



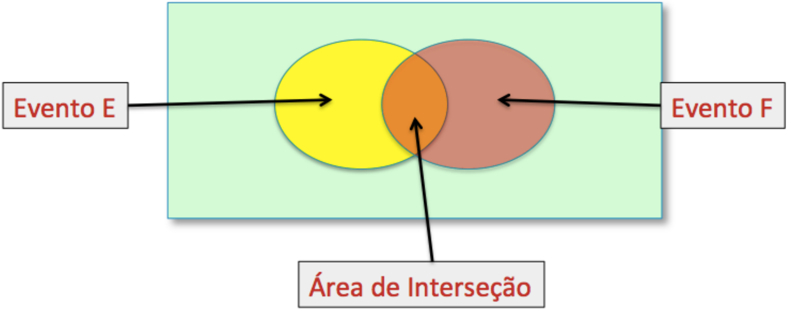
União – Eventos Mutuamente Exclusivos

$$P(E \cup F) = P(E) + P(F)$$



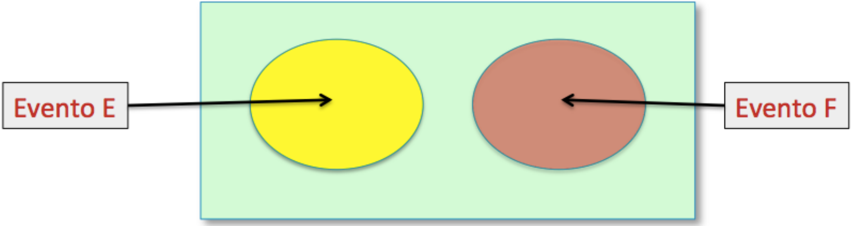
União – Lei de Adição

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$



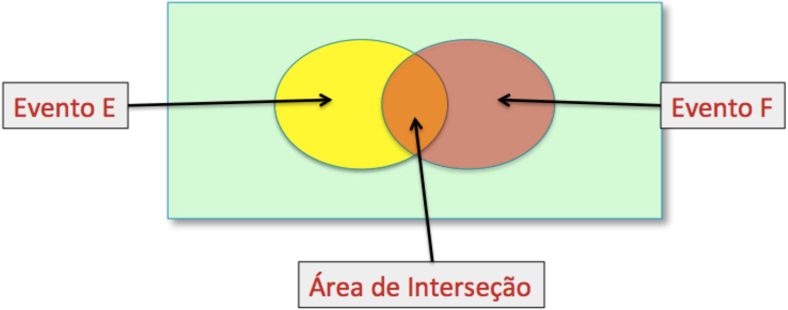
Interseção – Lei de Multiplicação - Independência

$$P(E \cap F) = P(E) * P(F)$$



Interseção – Lei de Multiplicação

$$P(E \cap F) = P(F) * P(E|F)$$



Probabilidade Condicional

- $P(E \mid F)$ lida como probabilidade de E dado F
- Exemplo: Médicos num hospital
  - Promovidos ou não
  - Médicos mulheres pensam que estão sofrendo discriminação
  - As promoções favorecem os homens?

P(Promovido | Homem)

	Homens (H)	Mulheres (M)	Total
Promovidos (P)	288	36	324
Não Promovidos (N)	672	204	876
Total	960	240	1200

$$P(P|H) = \frac{288}{960} = 30\%$$

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

P(Promovido | Mulher)

	Homens (H)	Mulheres (M)	Total
Promovidos (P)	288	36	324
Não Promovidos (N)	672	204	876
Total	960	240	1200

$$P(P|M) = \frac{36}{240} = 15\%$$

- Homens: 30%; Mulheres: 15%
- Discriminação possível

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Caso de Uma Doença Nova

- Uma doença nova infeta 1 em 1000 pessoas
- Existe um teste para a doença
- Se pessoa tem doença, teste mostra positivo 99% dos casos
- Mostra positivo em 2% dos casos que pessoa não tem doença
- Você faz o teste e tem resultado positivo.
- O que é a probabilidade de ter a doença?

Diagrama de Arvore

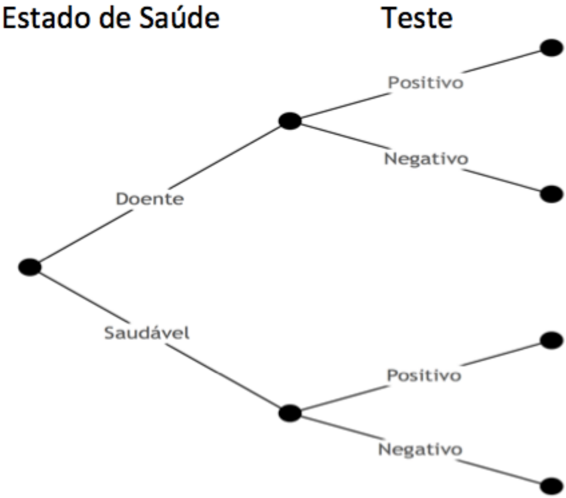
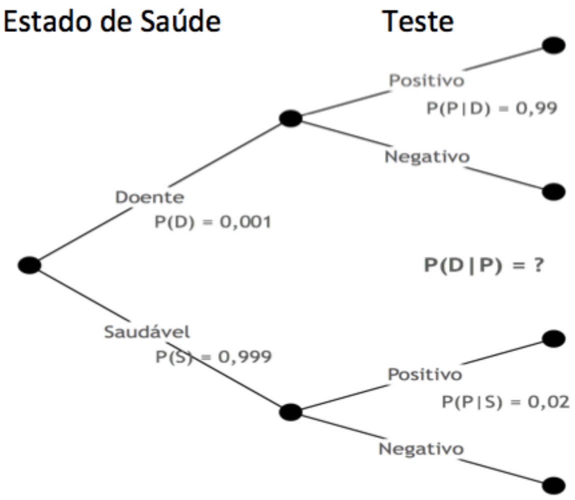


Diagrama de Arvore 2



Organizar as Probabilidades

1. Definir o espaço amostral			
	Doente (D)	Saudável (~D)	
Positivo (P)	p(D e P)	p(~D e P)	
Negativo (~P)	p(D e ~P)	p(~D e ~P)	

2. Definir probabilidades			
	Doente (D)	Saudável (~D)	
Positivo (P)	p(D e P)	p(~D e P)	p(P)
Negativo (~P)	P(D e ~P)	p(~D e ~P)	p(~P)
	p(D)	p(~D)	1

Calcular Probabilidades

3. Calcular probabilidades			
	Doente (D)	Saudável (~D)	
Positivo (P)	p(D e P)	p(~D e P)	p(P)
Negativo (~P)	p(D e ~P)	p(~D e ~P)	p(~P)
	p(D)	p(~D)	1

$$p(P|D) = 0,99$$

$$p(D) = 0,001$$

$$p(D \cap P) = p(D) * p(P|D)$$

$$p(D \cap P) = 0,001 * 0,99 = 0,00099$$

Calcular Mais Probabilidades

	Doente (D)	Saudável (~D)	
Positivo (P)	0,00099	p(~D e P)	p(P)
Negativo (~P)	p(D e ~P)	p(~D e ~P)	p(~P)
	0,001	0,999	1

$$p(P|\sim D) = 0,02 \qquad p(\sim D) = 0.999$$

$$p(\sim D \cap P) = p(P|\sim D) * p(\sim D)$$

$$p(\sim D \cap P) = 0,02 * 0,999 = 0,01998$$

Completar Tabela

3. Calcular probabilidades			
	Doente (D)	Saudável (~D)	
Positivo (P)	0,00099	0,01998	p(P)
Negativo (~P)	p(D e ~P)	p(~D e ~P)	p(~P)
	0,001	0,999	1

3. Calcular probabilidades			
	Doente (D)	Saudável (~D)	
Positivo (P)	0,00099	0,01998	0,02097
Negativo (~P)	0,00001	0,97902	0,97903
	0,001	0,999	1

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Para Lembrar

	Doente (D)	Saudável (~D)	
Positivo (P)	p(D e P)	p(~D e P)	p(P)
Negativo (~P)	p(D e ~P)	p(~D e ~P)	p(~P)
	p(D)	p(~D)	1

	Doente (D)	Saudável (~D)	
Positivo (P)	0,00099	0,01998	0,02097
Negativo (~P)	0,00001	0,97902	0,97903
	0,001	0,999	1

Lei de Multiplicação (de novo)

$$p(E \cap F) = p(F) * p(E|F)$$

- Com um pouco de álgebra

$$p(E|F) = \frac{p(E \cap F)}{p(F)}$$

Resposta

$$p(D|P) = \frac{p(D \cap P)}{p(P)} = \frac{0,00099}{0,02097} = 4.72\%$$

- Menos de 5% das pessoas com testes positivos têm a doença
  - 95% não têm
- O grupo de saudáveis (99,9% da população) é tão maior que os doentes
- 4.72% representa só 1 em 1000
- Uma demonstração da Teorema de Bayes (Paradoxo de Falsos Positivos)

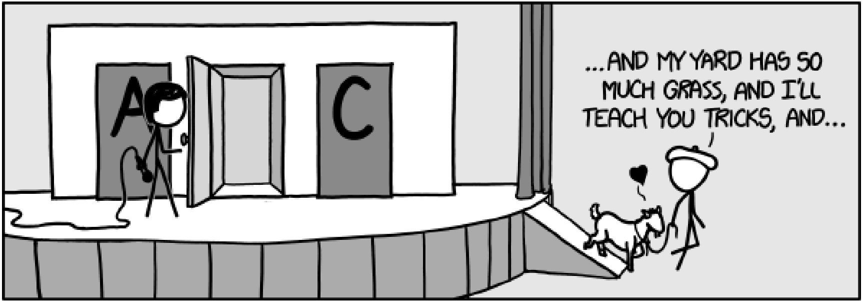
UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Teaser 1 - Os Aniversariantes

- Numa festa com 30 pessoas, o que é a probabilidade que 2 ou mais pessoas têm o mesmo aniversário?
- Dica: Lembre a lei de probabilidade total
- Dica #2: 1 - P(0) = P(pelo menos 1)

Teaser 2 - O Problema Monty Hall

- Programa de tv de premios,
- Jogador vê 3 portas
  - Atrás 1ª fica um premio muito bom (carro, casa, etc.)
  - Atrás 2ª fica um premio ruim (bode)
  - Atrás 3ª fica um premio ruim (bode)
- Monty Hall pede que você escolhe uma porta
- Você escolha porta # 1
- Ele abre porta #2 – premio ruim
  - Bom premio ainda fica no jogo
- Ele pergunta se você quer ficar com porta #1 ou trocar
- O que você deve fazer? Ficar com porta #1 ou trocar?


A
C
...AND MY YARD HAS SO MUCH GRASS, AND I'LL TEACH YOU TRICKS, AND...