

Planejamento de Experimento

Etapas de uma pesquisa

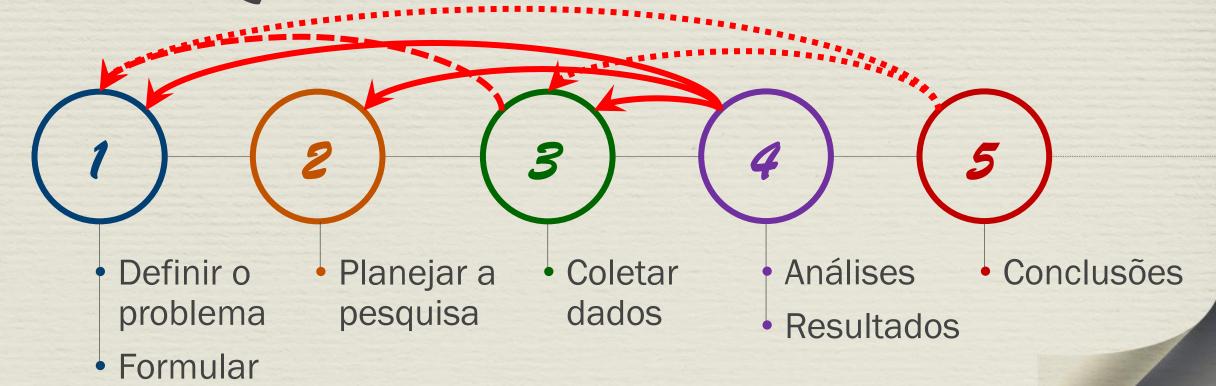
Como fazer?

Quais dados?

Quais ferramentas?

PESQUISA

objetivos







PLANEJAMENTO



Dois tipos de pesquisas:

- OBSERVAÇÃO:
 - Através de questionário ou entrevista
 - Características da amostra/população
- EXPERIMENTO
 - Aplicado a cada elemento
 - O pesquisador tem controle do processo





População e amostra:

- População-alvo:
 - conjunto de elementos que queremos abranger
- População-acessível:
 - passíveis de serem observados
- Amostra:
 - seleção de parte da população



AMOSTRAGEM





Tipos de dados:

- Dados **primários** são observados diretamente;
- Dados **secundários** são obtidos de outras fontes (publicação, banco de dados e etc);

Variáveis:

- São características que podem ser observadas (medida) em cada elemento da amostra/população;
- Quantitativas são dados quantitativos;
- Qualitativas representam qualidade ou categorias

Quantitativa:

- Idade : 33 anos

- N° de automóveis : 1 auto

- Deslocamento diário: 13 km

Qualitativa:

Avalie o transporte público da cidade:
() bom (x) razoável () ruim

Avalie de 0 a 5 o tempo de espera por ônibus:









Tipos de dados:

- Dados **primários** são observados diretamente;
- Dados **secundários** são obtidos de outras fontes (publicação, banco de dados e etc);

Questionários:

- Mensurar uma variável;
- Evitar indução nas repostas;
- Elaboração de questionários, tipos, termos e ferramentas;
 - Pré-teste

Variáveis:

- São características que podem ser observadas (medida) em cada elemento da amostra/população;
- Quantitativas são dados quantitativos;
- Qualitativas representam qualidade ou categorias



Pergunta aberta:

- Há quanto tempo você tem CHN?

9 anos

7 anos e 2 meses

Quase 2 anos

Não completou 1 ano





Pergunta com múltiplas variáveis:

- Assinale os meios de transportes que utiliza regularmente:
- (x) a pé
- () bicicleta
-) motocicleta
-) automóvel
- (x) ônibus
-) trem (superfície e metrô)
- () outros:_____





Ferramentas disponíveis:













Tipos de dados:

- Dados **primários** são observados diretamente;
- Dados **secundários** são obtidos de outras fontes (publicação, banco de dados e etc);

Questionários:

- Mensurar uma variável;
- Evitar indução nas repostas;
- Elaboração de questionários, tipos, termos e ferramentas;
- Pré-teste

Variáveis:

- São características que podem ser observadas (medida) em cada elemento da amostra/população;
- Quantitativas são dados quantitativos;
- Qualitativas representam qualidade ou categorias

Codificação dos dados:

- Formato de banco de dados
- Cada linha uma ocorrência do objeto
- Objeto definido por atributos ou vetor de características

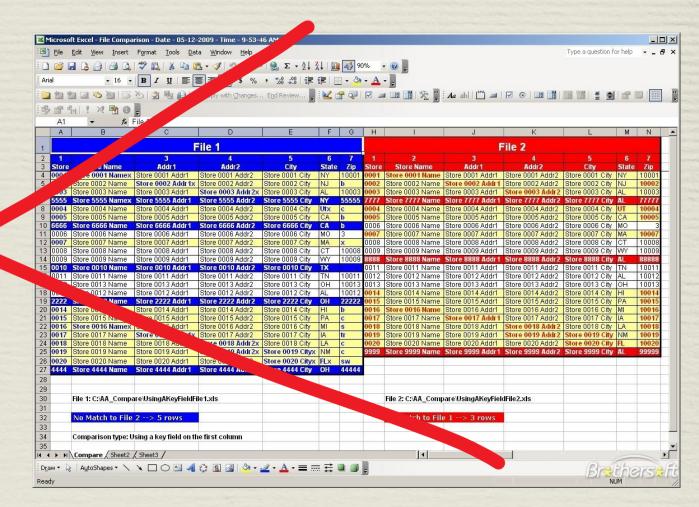
				Passenger- netres	Departures	Passeng	ers carried
	Period	Year	W 14* (billio	International (billion)	World (million)	World (billion)	International (billion)
>		2002	3075	1798	21	1.7	0.6
History	2002-2012		+5.8%	+ b.	+4.0%	+5.6%	+7.5%
I		2012	5401	3350	31	3.0	1.2
	2012-2020		+4.5%	+4.8%	9/	+4.5%	+5.2%
sts		2020	7682	4884	42	4.2	1.7
Forecasts	2020-2030		+4.5%	+4.6%	+3.5%	W.	+4.9%
P.		2030	11872	7672	59	6.4	2.8
	2012-2030		+4.5%	+4.7%	+3.6%	+4.4%	

*inter all and domest

Source: International Civil Aviation Organization

data as of May 2013

4	А	В	С	D		F	G
1			Fluxo de	Caix~ se	mestre		
2	Fluxo de Caixa	Jan	Fev	Mar	Abr	Mai	Jun
3	Saldo Inicial	R\$ 0,00	R\$ 50°	R\$ 800,00	R\$ 1.320,00	R\$ 2.120,00	R\$ 3.220,00
4	Receitas	R\$ 2.500,00	J0,00	R\$ 2.500,00	R\$ 2.800,00	R\$ 3.200,00	R\$ 2.800,00
5	Despesas	R\$ 2.000,00	R\$ 2.000,00	R\$ 1.980,00	R\$ 2.000,00	R\$ 2.100,00	R\$ 2.100,00
5	Lucro/Prejuizo	R\$ 500,00	R\$ 300,00	R\$ 520,00	R\$ 800,00	R\$ 1.100,00	R\$ 700,00
7	Acumulado	R\$ 500,00	R\$ 800,00	R\$ 1.320,00	R\$ 2.120,00	R\$ 2.120,00	R\$ 2.820,00
8							
9	Mês	Fev		Lucro/Prejuizo	R\$ 300		



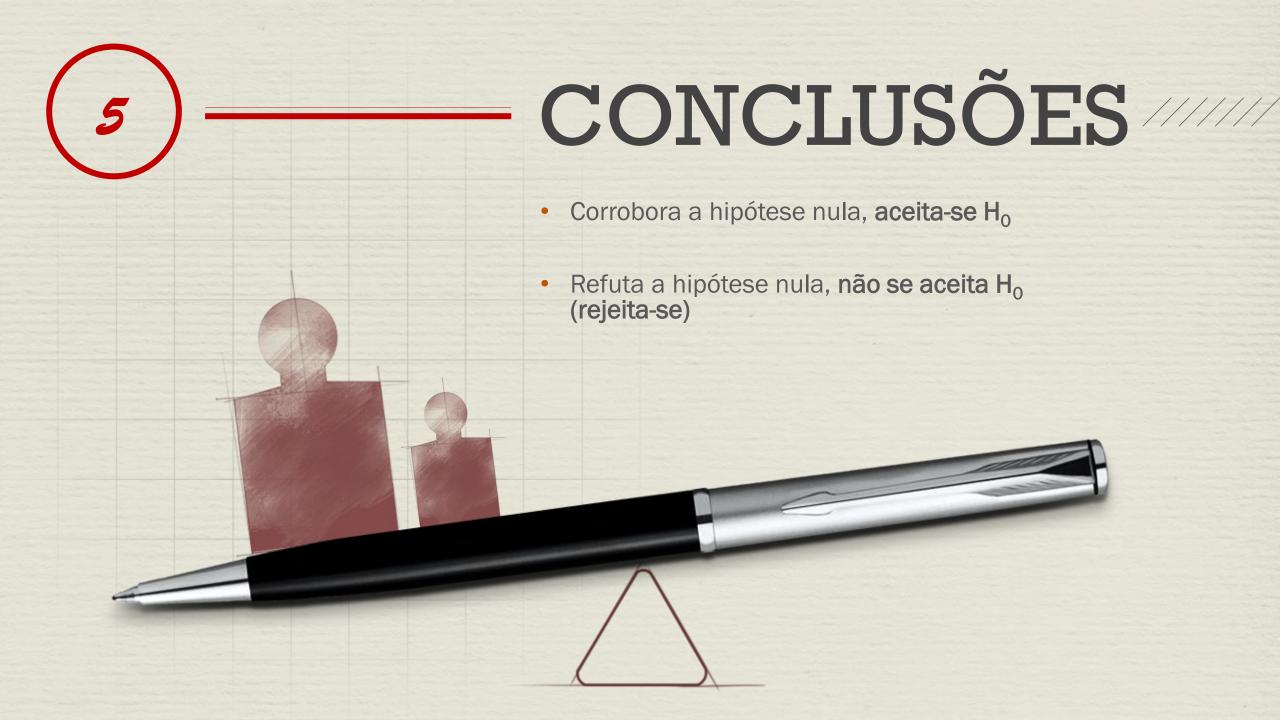
Α	В	С	D	E	F	G	Н	ı	J	К	L	М	N	0	Р	Q	R	S	Т	U	V	W	х	Υ		
	10		0 4:	D -"		ID .					r: b	AT ENG	F:	. ~	0.40:		5 / / 1	h :1.					Pe	so por E	<u>ix</u>	
Tipo ro				Região		ID I	N. Pesage	Hora	Marca		Eixo novo C					rguid			Compi	1	2	3	4	5	<u> </u>	
DUPLA		SP-330 - km53	S		01/03/05	1		12:57:41	MB	1935	213	5	5		pesado		360	tanque			11.680	8.840		10.760		
DUPLA		SP-330 - km53	5		01/03/05	2		12:57:53	Scania	1115	253	5	5		pesado	1	296	baú		4.030		11.220	9.360			
DUPLA		SP-330 - km53	S		01/03/05	3		12:58:03	VW	8-140	2C	2	2		leve		135	baú		2.680	2.420					
DUPLA		SP-330 - km53	S		01/03/05	4		12:58:09	MB	1316	2C	2	2	2			156	basculante		3.120	3.620	4.000	0.040			
DUPLA		SP-330 - km53	S		01/03/05	5		12:58:24	Ford	4030 Cargo	253	5	5		pesado	1	291	contêiner		3.850	4.490	4.880	3.210			
DUPLA		SP-330 - km53	S		01/03/05	6		12:58:28	MB	1618	3C	3	3		médio		184	aberta		5.040	9.320	8.930				
DUPLA		SP-330 - km53	S	_	01/03/05	7		12:58:35	MB	1420	3C	3	3		médio		204	sider		5.530	5.920	4.300	7.455	7.000		-
DUPLA		SP-330 - km53	S		01/03/05	8		12:58:55	Ford	4331 Cargo	2S3	5	5		pesado		303	baú		4.950	8.890	7.150	7.150	7.600		
DUPLA		SP-330 - km53	S		01/03/05	9		12:58:59	Ford	1415 Cargo	2C	2	2	2			158	aberta		3.760	5.360					
DUPLA		SP-330 - km53	S		01/03/05	10		12:59:13	MB	L 1620	3C	3	3	3			204	baú		3.510		4.540				
DUPLA		SP-330 - km53	S	_	01/03/05	11		12:59:20	Ford	1317 Cargo	3C	3	3	3			162	baú		2.900	4.120	2.800				
DUPLA		SP-330 - km53	S		01/03/05	12		12:59:29	MB	912	2C	2	2	2			122	baú		2.340	2.710					
DUPLA		SP-330 - km53	S		01/03/05	13		12:59:35	VW	23-220	3C	3	3	3			218	baú		4.840	7.360	5.930				
DUPLA		SP-330 - km53	S		01/03/05	14		12:59:41	MB	1720	3C	3	3	3			204	aberta		5.560	9.950	7.720				
DUPLA	AutoB	SP-330 - km53	S	Capital	01/03/05	15		12:59:47	MB	1418R	2C	2	2	2	leve		184	aberta		3.560	3.480					
DUPLA	AutoB	SP-330 - km53	S	Capital	01/03/05	16	65.068	12:59:51	MB	709	2C	2	2	2	leve		90	aberta	5	2.080	1.650				T.	
DUPLA	AutoB	SP-330 - km53	S	Capital	01/03/05	17	65.069	12:59:55	Ford	814 Cargo	2C	2	2	2	leve		141	aberta	7	2.610	2.680					
DUPLA	AutoB	SP-330 - km53	S	Capital	01/03/05	18	65.070	13:00:02	Volvo	NL12	2S3	5	5	3	pesado	2	337,8	aberta	12	4.930	4.750	6.860				
DUPLA	AutoB	SP-330 - km53	S	Capital	01/03/05	19	65.071	13:00:10	Volvo	NL10	253	5	5	5	pesado		340	tanque	16	4.920	9.980	9.180	9.370	9.810		
DUPLA	AutoB	SP-330 - km53	S	Capital	01/03/05	20	65.072	13:00:18	Volvo	NL12	2S3	5	5	4	pesado	1	360	basculante	15	4.800	6.080	4.360	2.720			
DUPLA	AutoB	SP-330 - km53	S	Capital	01/03/05	21	65.073	13:00:24	Volvo	NL12	253	5	5	5	pesado		337,8	basculante	15	5.100	12.140	9.490	9.250	9.830		
DUPLA	AutoB	SP-330 - km53	S	Capital	01/03/05	22	65.074	13:00:32	Scania	112HW	253	5	5	5	pesado		337,8	tanque	16	5.000	10.180	9.060	13.700	6.160		
DUPLA	AutoB	SP-330 - km53	S	Capital	01/03/05	23	65.075	13:00:38	Volvo	FH12	3S3	6+	6		pesado		369	contêiner	17	4.300	13.260	1.420	3.670	13.690	10.950	
DUPLA	AutoB	SP-330 - km53	S	Capital	01/03/05	24	65.076	13:00:45	MB	1935	2S3	5	5	4	pesado	1	360	aberta	16	4.590	5.180	4.360	2.980			
DUPLA	AutoB	SP-330 - km53	S	Capital	01/03/05	25	65.077	13:00:57	MB	1519	2S2	4	4	4	médio		215	baú	19	3.700	5.700	5.410	4.650			
DUPLA		SP-330 - km53	S		01/03/05	26		13:01:08	MB	1929	253	5	5		pesado	1	310	sider		4.700		5.740				
DUPLA		SP-330 - km53	S	_	01/03/05	27		13:01:13	VW	8-120	2C	2	2	2			120	baú		2.580	1.920					
DUPLA		SP-330 - km53	S		01/03/05	28		13:01:20	MB	914C	2C	2	2	2			136	baú		2.690	3.510					
DUPLA		SP-330 - km53	S	_	01/03/05	29		13:01:25	Ford	1215 Cargo	2C	2	2	2			158	baú		3.280	4.820					
DUPLA		SP-330 - km53	S		01/03/05	30		13:01:32	VW	18-310	253	5	5		pesado	1	303	tanque		4.040		4.640	2.980			
DUPLA		SP-330 - km53	S	_	01/03/05	31		13:01:38	VW	18-310	253	5	5		pesado	2	303	contêiner			3.940					



ANÁLISES E RESULTADOS /////////

- Replicação do experimento para avaliar o erro experimental
- Aleatorização do experimento
- Análise descritiva
- Testes de hipóteses
- Ferramentas computacionais





Probabilidade

Definição

Espaço amostral

Probabilidade condicional

Teorema de Bayes

Probabilidade

"probability is a way of quantifying the uncertainty associated with events chosen from a some universe of events"

Joel Grus

Definição

HIPÓTESES, CONJETURAS, ETC. Distribuição **Modelos** de Probabilísticos frequência **RESULTADOS OU DADOS OBSERVADOS**

Definição

Estudar os processos que envolvem variabilidade, aleatoriedade ou incerteza.

Modelos construídos por suposição dos processos:

Intuitivamente as pessoas procuram tomar decisões em função de fatos que têm maior probabilidade

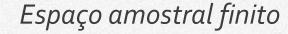
Modelos baseados em dados observados no passado:

A <u>incerteza</u> inerente às decisões que podem ser tomadas sobre determinado problema

Presença de algum experimento aleatório

Espaço Amostral

Conjunto de todos os possíveis resultados do experimento, denotado pela letra grega Ω



Formado por um número limitado de resultados possíveis

Espaço amostral infinito

Formado por um número infinito de resultados, enumeráveis ou reais



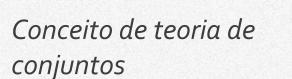
$$\Omega = \{1,2,3,4,5,6\}$$

Eventos

Qualquer subconjunto do espaço amostral.

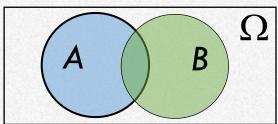
A é um evento se:

$$A\subseteq \Omega$$

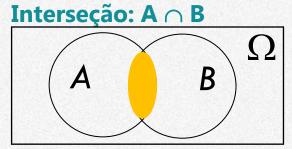


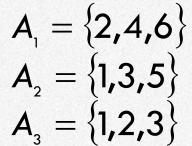
A é um evento se $A \subseteq \Omega$

União: A ∪ B



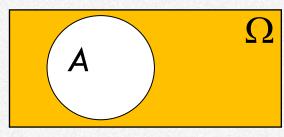
 $A \cap B = 0$





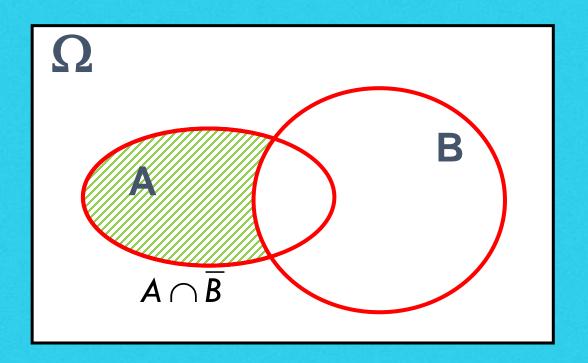
$$A_{3} = \{1,2,3\}$$

Complementar: \overline{A}





EXEMPLO: Defina a área destacada na figura a seguir:



Teoria dos Conjuntos

Axiomas e propriedades

Definição clássica:

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

Seja um experimento aleatório em um espaço amostral, cada evento E_i deve satisfazer aos

axiomas:

$$0 \le P(E_i) \le 1$$

$$P(\Omega)=1$$

$$P(E_1 \cup E_2 \cup ... \cup E_n) = P(E_1) + P(E_2) + ... + P(E_n)$$

Axiomas e propriedades

Propriedades básicas:

$$P(\phi) = 0$$

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

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

Calcular a probabilidade de ocorrência de um evento A, dada a ocorrência de um evento B. Calcular a probabilidade de ocorrência de A condicionada à ocorrência prévia de B

P(A|B) = probabilidade de A dado B



Entrega de Bagagem	Empresa X	Empresa Y	Empresa Z	TOTAL
Normal (N)	500	4.500	1.500	6.500
Defeito (D)	30	270	50	350
Total	530	4.770	1.550	6.850

Qual a probabilidade da bagagem ser entregue com algum defeito?

$$P(D) = \frac{350}{6.850} = 0.051$$

Qual a probabilidade da bagagem ser entregue pela empresa Y?

$$P(Y) = \frac{4.770}{6.850} = 0,696$$

Calcular a probabilidade de ocorrência de um evento A, dada a ocorrência de um evento B. Calcular a probabilidade de ocorrência de A condicionada à ocorrência prévia de B

P(A|B) = probabilidade de A dado B



Entrega de Bagagem	Empresa X	Empresa Y	Empresa Z	TOTAL
Normal (N)	500	4.500	1.500	6.500
Defeito (D)	30	270	50	350
Total	530	4.770	1.550	6.850

Qual a probabilidade da bagagem ser entregue com algum defeito, sabendo-se que a entrega foi realizada pela Empresa Y?

$$P(D|Y) = \frac{270}{4.770} = 0.057$$

$$P(D|Y) = \frac{270}{4.770} = \frac{270}{6.850}$$

Calcular a probabilidade de ocorrência de um evento A, dada a ocorrência de um evento B. Calcular a probabilidade de ocorrência de A condicionada à ocorrência prévia de B

P(A|B) = probabilidade de A dado B



Entrega de Bagagem	Empresa X	Empresa Y	Empresa Z	TOTAL
Normal (N)	500	4.500	1.500	6.500
Defeito (D)	30	270	50	350
Total	530	4.770	1.550	6.850

Qual a probabilidade da bagagem ser entregue com algum defeito, sabendo-se que a entrega foi realizada pela Empresa Y?

$$P(D|Y) = \frac{270}{4.770} = 0.057$$

$$P(D|Y) = \frac{P(D \cap Y)}{P(Y)}$$

Calcular a probabilidade de ocorrência de um evento A, dada a ocorrência de um evento B. Calcular a probabilidade de ocorrência de A condicionada à ocorrência prévia de B

P(A|B) = probabilidade de A dado B

Sejam A e B eventos quaisquer, sendo P(B) > o. A probabilidade condicionada é dada por:

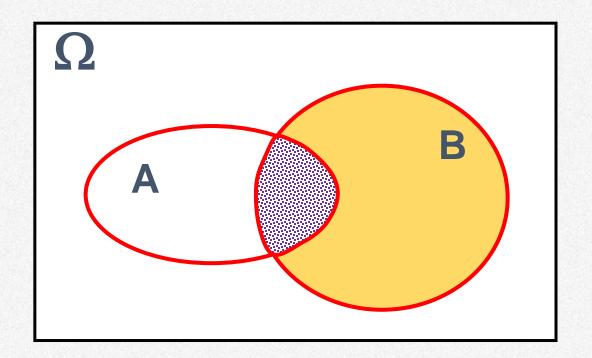
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Ressalta-se que a operação de interseção é comutativa:

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

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

Calcular a probabilidade de ocorrência de um evento A, dada a ocorrência de um evento B.



Regra do produto

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

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

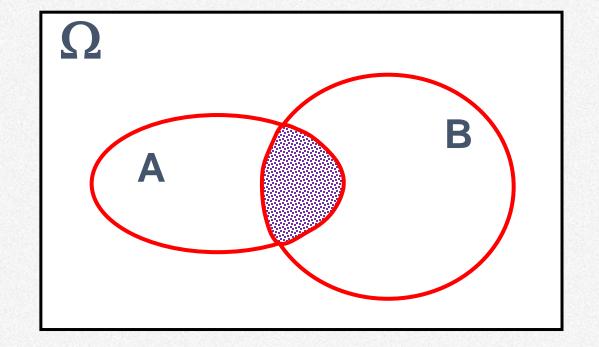
$$P(B \mid A) = \frac{P(B \cap A)}{P(A)}$$

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

Calcular a probabilidade de ocorrência de um evento A, dada a ocorrência de um evento B. Regra do produto

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

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



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

$$P(A \mid B) \cdot P(B) = P(B \mid A) \cdot P(A)$$

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

Teorema da probabilidade total

O teorema da
probabilidade total
permite solucionar
problemas quando não se
sabe o evento a priori.

Seja um evento \mathbf{F} qualquer em um espaço amostral Ω , sujeito a:

$$E_i \cap E_j = \emptyset$$

 $E_1 \cup E_2 \cup ... \cup E_k = \Omega$
 $P(E_i) > 0 \text{ para } i = 1,2,...,k$

$$P(F) = \sum_{i=1}^{k} P(E_i) \cdot P(F \mid E_i)$$

O teorema de Bayes permite obter a probabilidade de que um dos eventos E_i ocorra, sabendo-se que F ocorreu.

$$P(E_i \mid F) = \frac{P(E_i \cap F)}{P(F)}$$

Seja um evento \mathbf{F} qualquer em um espaço amostral Ω , sujeito a:

$$E_i \cap E_j = \emptyset$$

 $E_1 \cup E_2 \cup ... \cup E_k = \Omega$
 $P(E_i) > 0 \text{ para } i = 1,2,...,k$

Bayes Theorem:

- The geometry of changing beliefs
- The quick proof
- The medical test paradox

EXEMPLO:

A: teste acusa emissão em excesso

B: veículo com emissão em excesso

$$P(B) = 0,25$$

$$P(\overline{B}) = 0,75$$

$$P(A | B) = 0,99$$

$$P(A \mid \overline{B}) = 0.17$$

Teorema de Bayes

De acordo com a avaliação de controle de emissão veicular, considere os dados e um determinado teste:

- (1) 25% dos carros emitem poluentes em excesso;
- (2) Dos carros poluentes em excesso, o teste detecta 99% dos veículos infratores
- (3) Dos carros regulares, o teste aponta 17% dos veículos como infratores, ou seja, erro do teste;

Qual a probabilidade de que um carro acusado no teste emita realmente poluentes em excesso?

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

 $P(A) = 0.99 \cdot 0.25 + 0.17 \cdot 0.75 = 0.375$

$$P(B \mid A) = \frac{P(A \mid B) \cdot P(B)}{P(A)} = \frac{0,99 \cdot 0,25}{0,375} = 0,66$$

Inferência Bayesiana simplesmente atualiza suas crenças após considerar novas evidências

$$Posteriori = \frac{priori \cdot likelihood}{\sum priori \cdot likelihood}$$

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

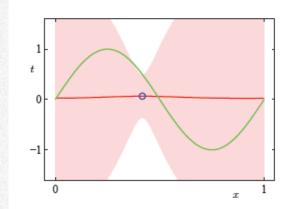
$$P(\theta \mid X) = \frac{P(X \mid \theta) \cdot P(\theta)}{P(X)}$$

Sendo:

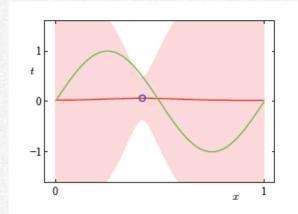
 θ : parâmetros

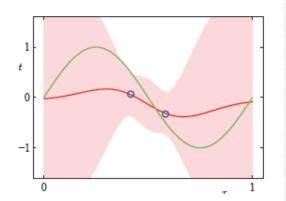
X: dados

$$Posteriori = \frac{priori \cdot likelihood}{\sum priori \cdot likelihood}$$

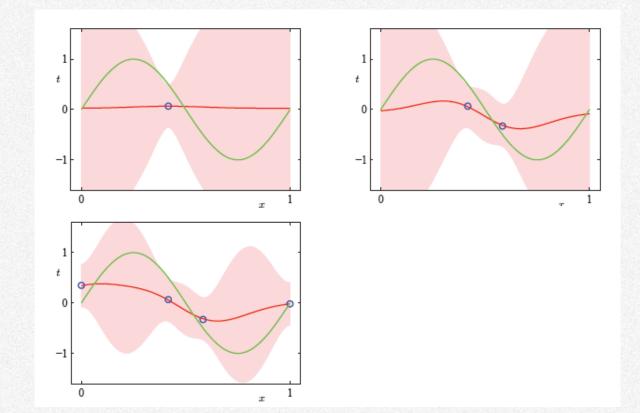


$$Posteriori = \frac{priori \cdot likelihood}{\sum priori \cdot likelihood}$$

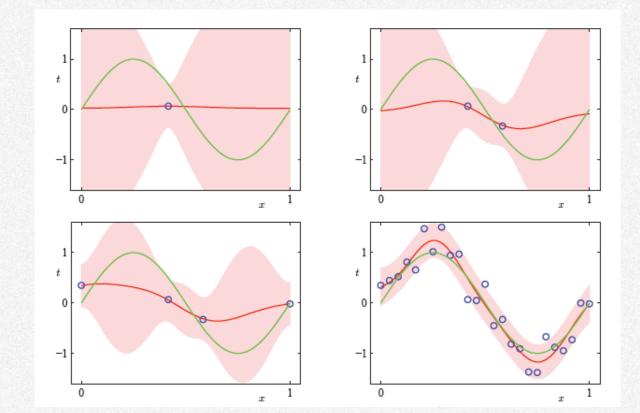




$$Posteriori = \frac{priori \cdot likelihood}{\sum priori \cdot likelihood}$$



$$Posteriori = \frac{priori \cdot likelihood}{\sum priori \cdot likelihood}$$

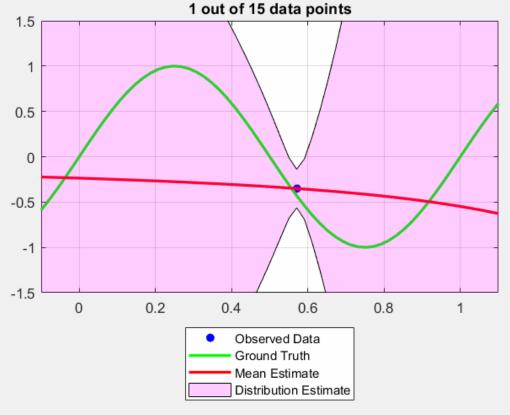


Inferência Bayesiana simplesmente atualiza suas crenças após considerar novas evidências

 $Posteriori = \frac{priori \cdot likelihood}{\sum priori \cdot likelihood}$



- Foundation of Bayesian Inference
- <u>Course 2022</u>
- Github material



EXEMPLO:

Há quatro estradas que ligam a localidade X à localidade Y. Se uma pessoa vai de X a Y pela Estrada I, Estrada II, Estrada III ou Estrada IV, as probabilidades de que ela chegue atrasada devido a congestionamento de trânsito são 30%, 20%, 60% e 35%, respectivamente.

Suponha que uma pessoa escolha uma estrada aleatoriamente e que ela chegue atrasada. Qual é a probabilidade de que ela tenha escolhido a Estrada III?

Teorema de Bayes

EXEMPLO:

Há quatro estradas que ligam a localidade X à localidade Y. Se uma pessoa vai de X a Y pela Estrada I, Estrada II, Estrada III ou Estrada IV, as probabilidades de que ela chegue atrasada devido a congestionamento de trânsito são 30%, 20%, 60% e 35%, respectivamente.

Suponha que uma pessoa escolha uma estrada aleatoriamente e que ela chegue atrasada. Qual é a probabilidade de que ela tenha escolhido a Estrada III?

:	Soma de Quantio	dade	ESTRADAx	▼											
	TRÁFEGOy	V	I		II	Ш	IV	Total Geral			P(I TRÁFEGOy)	P(II TRÁFEGOy)	P(III TRÁFEGOy)	P(IV TRÁFEGOy)	
	congestionado			7.50%	5.00%	15.00%	8.75%	36.25%	conge	estionado	20.7%	13.8%	41.4%	24.1%	100.0%
I	normal			17.50%	20.00%	10.00%	16.25%	63.75%		normal	27.5%	31.4%	15.7%	25.5%	100.0%
	Total Geral			25.00%	25.00%	25.00%	25.00%	100.00%							
				- 1	II	Ш	IV								
P(conges	stionado ESTRA	DAx)		30.0%	20.0%	60.0%	35.0%								
P((normal ESTRA	DAx)		70.0%	80.0%	40.0%	65.0%								
				100.0%	100.0%	100.0%	100.0%								

$$Posteriori = \frac{priori \cdot likelihood}{\sum priori \cdot likelihood}$$



