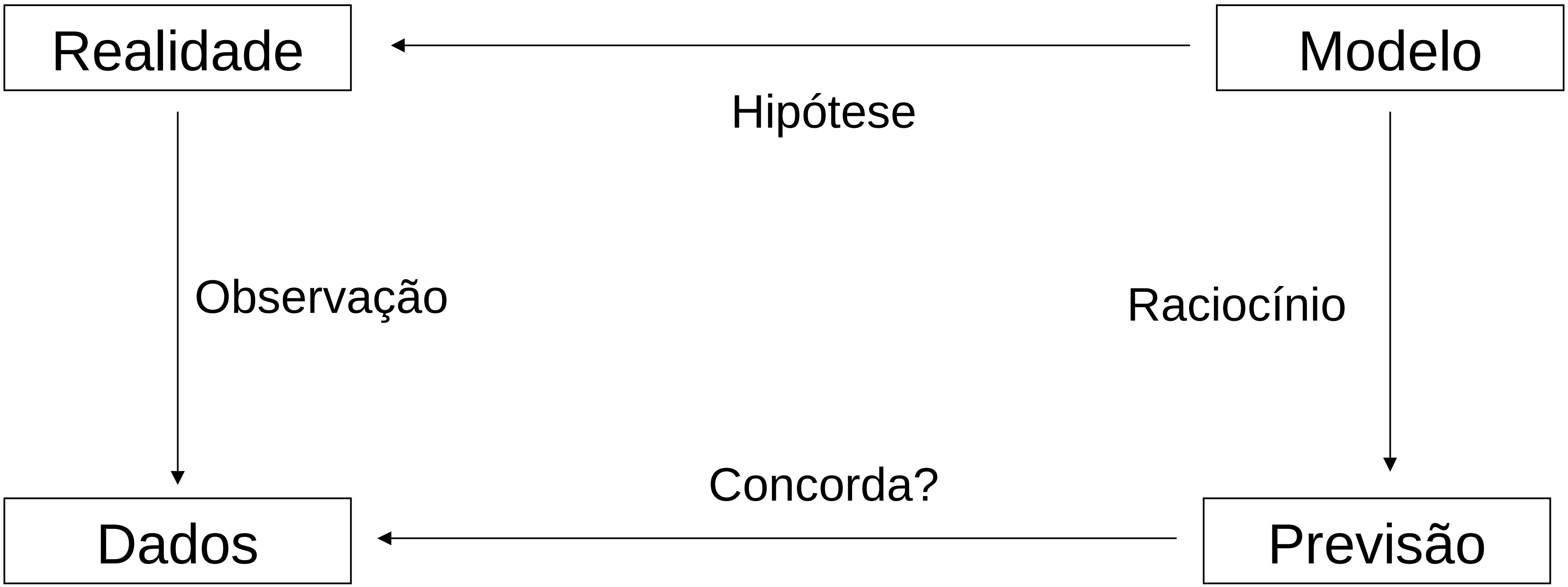
Pensamento crítico 2011/12

Aula 8, 2 e 4-11-11

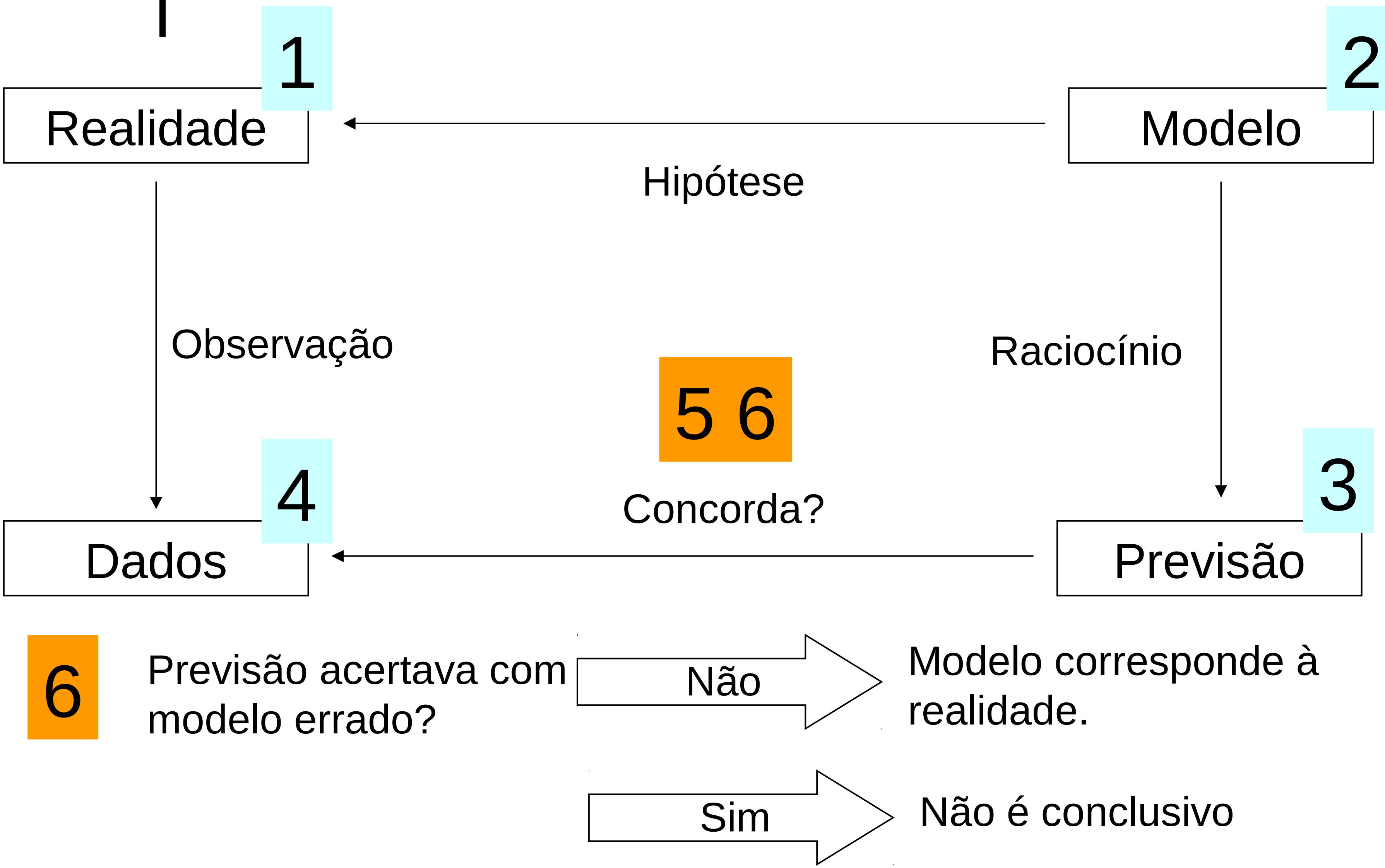
Resumo

- Experiências cruciais
- Ciência Marginal
- Modelos Estatísticos
  - Desenvolvimento de modelos
- Modelos Causais

Modelos Teóricos



Como avaliar



Experiência Crucial

Experiência Crucial

- Dados 2 modelos
  - Experiência tal que os dados possíveis (ou a amostra) só pode ser compatível com um dos modelos.

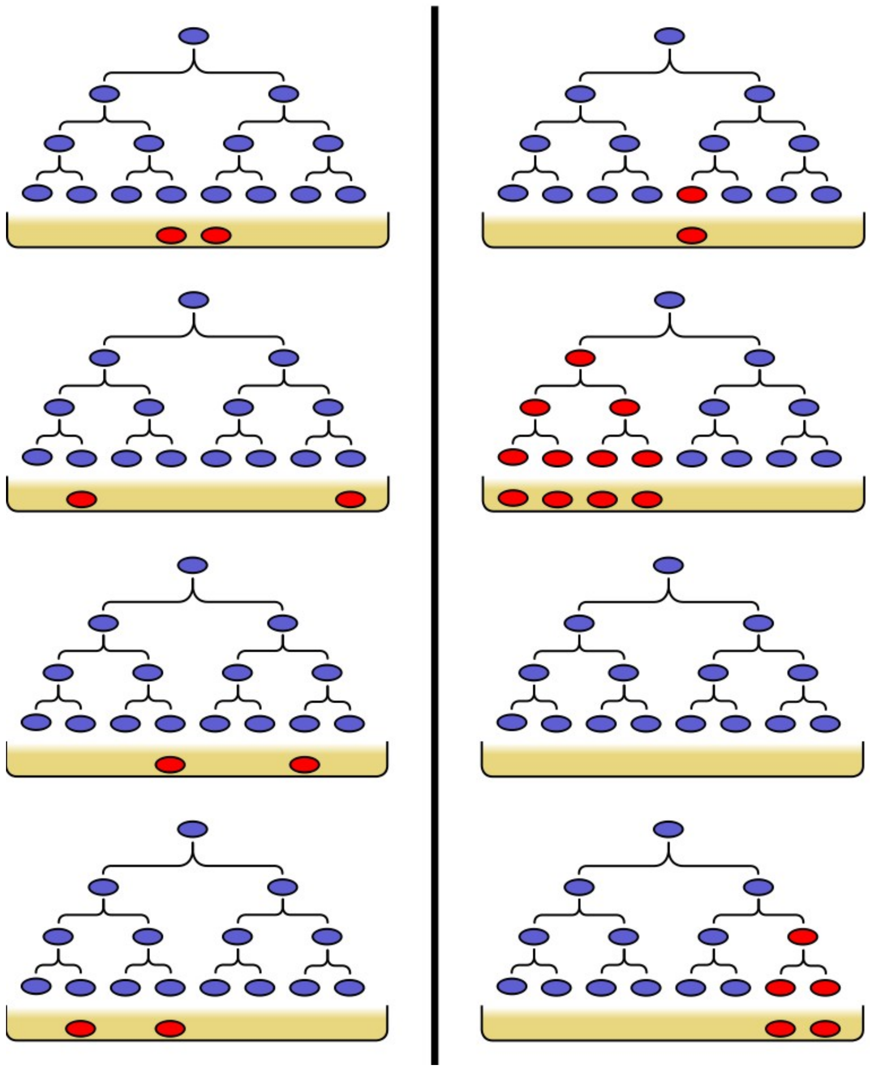
Experiência Crucial

- Mutações
  - Um vírus ao qual uma espécie de bactéria é susceptível.
  - Algumas são resistentes e sobrevivem graças a uma mutação.
  - A mutação surgiu por acaso ou foi induzida pelo vírus?

Experiência Crucial

- Experiência
  - Salvador Luria e Max Delbrück prepararam 20 culturas de bactérias
  - Depois de um tempo, plaquearam em agar com o vírus
    - Se mutação induzida, o número de resistentes seria aproximadamente igual
    - Se mutação aleatória, o número dependeria de quando a mutação tinha surgido e seria muito variável

Experiência Crucial



http://en.wikipedia.org/wiki/Luria–Delbrück_experiment

Experiência Crucial

- Experiência
  - Salvador Luria e Max Delbrück cultivaram 20 caixas de bactérias
  - Inocularam com o vírus
    - O número de resistentes era muito variável entre as caixas
  - Ou seja, a experiência permitiu decidir entre os dois modelos alternativos.

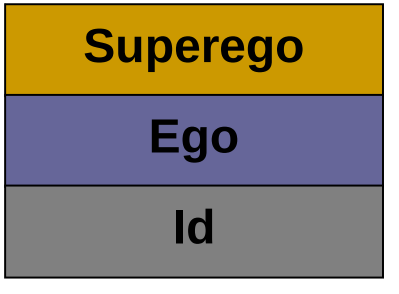
Ciência Marginal

Ciência Marginal

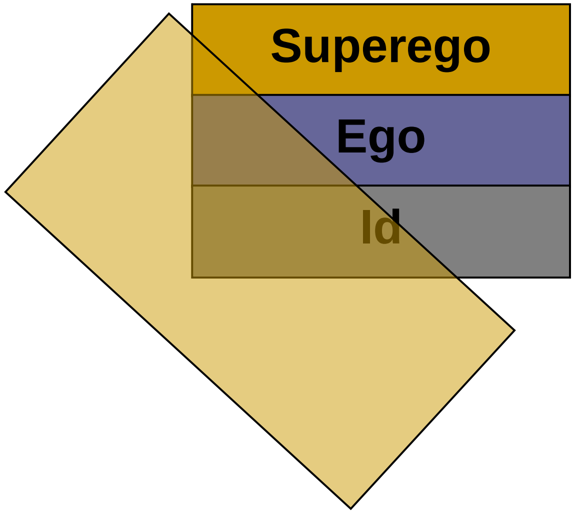
- A investigação é científica se
  - Propõe modelos de algum aspecto da realidade.
  - Os modelos podem ser confrontados com os dados.
  - Os modelos são ajustados de forma progressiva.
- Mas pode ser inconclusiva e ficar à margem.
  - Ou, se persistir em se desviar do confronto, acabar por se tornar regressiva e deixar de ser ciência.

Ciência Marginal

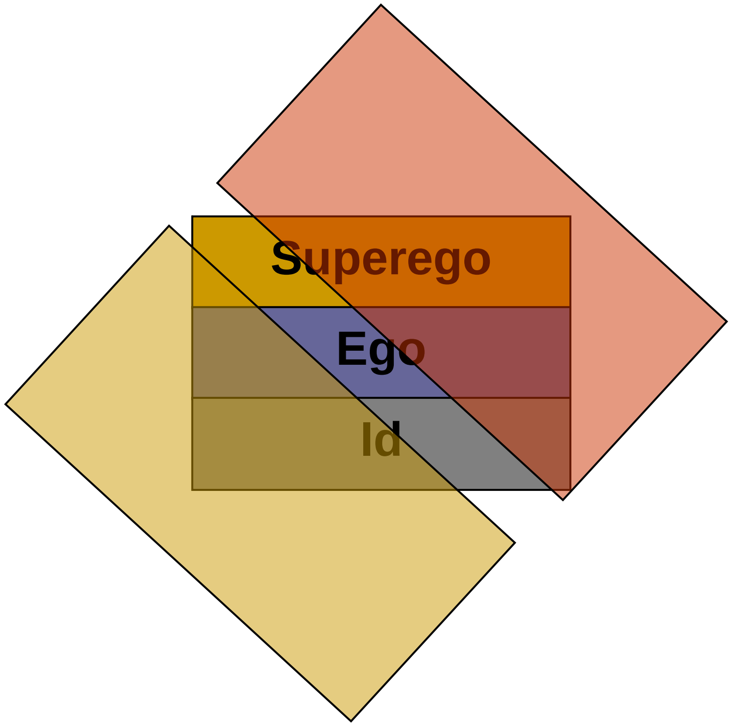
- Exemplo: “Teoria” (modelo) de Freud



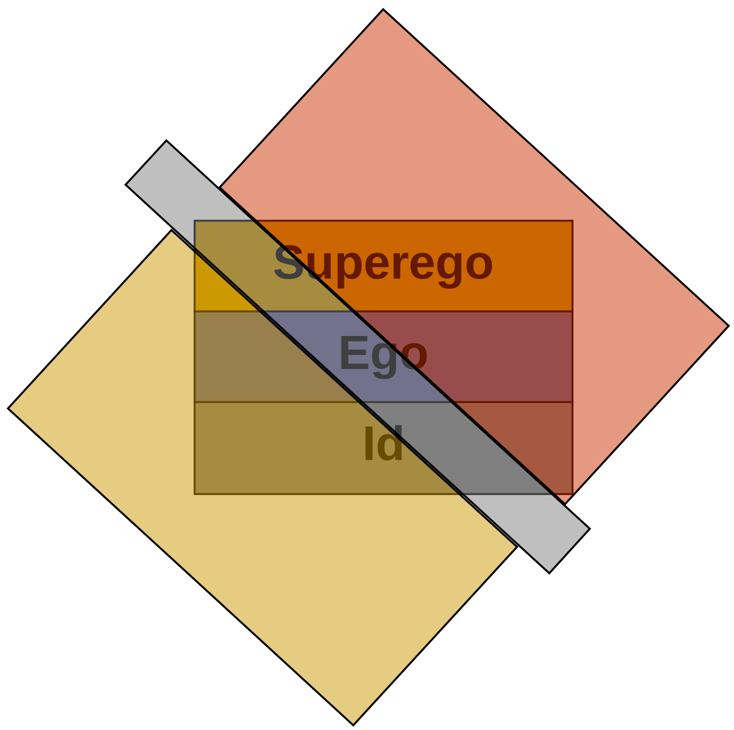
Ciência Marginal



Ciência Marginal



Ciência Marginal



Ciência Marginal

- Hans tinha 4 anos e muito medo de cavalos. Especialmente das palas e arreios (cabeçadas).
- Antes dormia com a mãe, mas agora o pai queria que ele dormisse no quarto dele.
- Passava mais tempo com a ama e foi com esta, no parque, que se assustou com um cavalo.

Ciência Marginal

- Freud propôs
  - Crise de Édipo: o Id de Hans queria a atenção da mãe.
  - O Ego de Hans tinha medo que o pai se zangasse.
  - O superego tinha sido ensinado que se devia amar os pais e por isso forçava a projecção do medo do pai num medo de cavalos.
  - Os arreios e palas eram a barba e óculos do pai.

Ciência Marginal

- Freud disse ao pai de Hans para lhe assegurar que não estava zangado com o miúdo.
- Ao fim de uns tempos o medo de cavalos foi diminuindo

Ciência Marginal

- 1: Realidade
  - O medo do pequeno Hans
- 2: Modelo
  - O conflito entre os desejos do Id, o medo do Ego e os preceitos do Superego levaram a projectar o medo do pai nos cavalos.
  - O problema está nos aspectos sub-conscientes.

Ciência Marginal

- 3: Previsão
  - O modelo diz que tomar consciência destes factores ajuda a resolver o problema.
- 4: Dados
  - Hans ficou melhor com o tratamento.

Ciência Marginal

- 5: Inconsistência?
  - Não.
- 6: Acertava mesmo estando errado?
  - Mesmo sem outro modelo explícito podemos pensar noutras explicações. A criança apanhou um susto e passou-lhe, ou queria atenção da mãe e estava a fazer fita, etc...

Ciência Marginal

- Resumindo:
  - O modelo é testável, em princípio, mas os dados são inconclusivos.
  - Daqui pode progredir se mais dados ajudarem a rejeitar ou confirmar o modelo (rejeitando alternativas)
  - Ou regredir se o modelo for demasiado permissivo, eventualmente deixando de ser ciência.

Modelos Estatísticos

Modelos Estatísticos

- Um modelo estatístico é um caso particular do modelo teórico
  - O aspecto da realidade é uma população, que pode ser de indivíduos, ocorrências, etc
  - Os dados são uma amostra da população
  - O modelo é acerca da distribuição de características na população
  - De onde se infere o que se espera da amostra, com uma margem de erro

Modelos Estatísticos

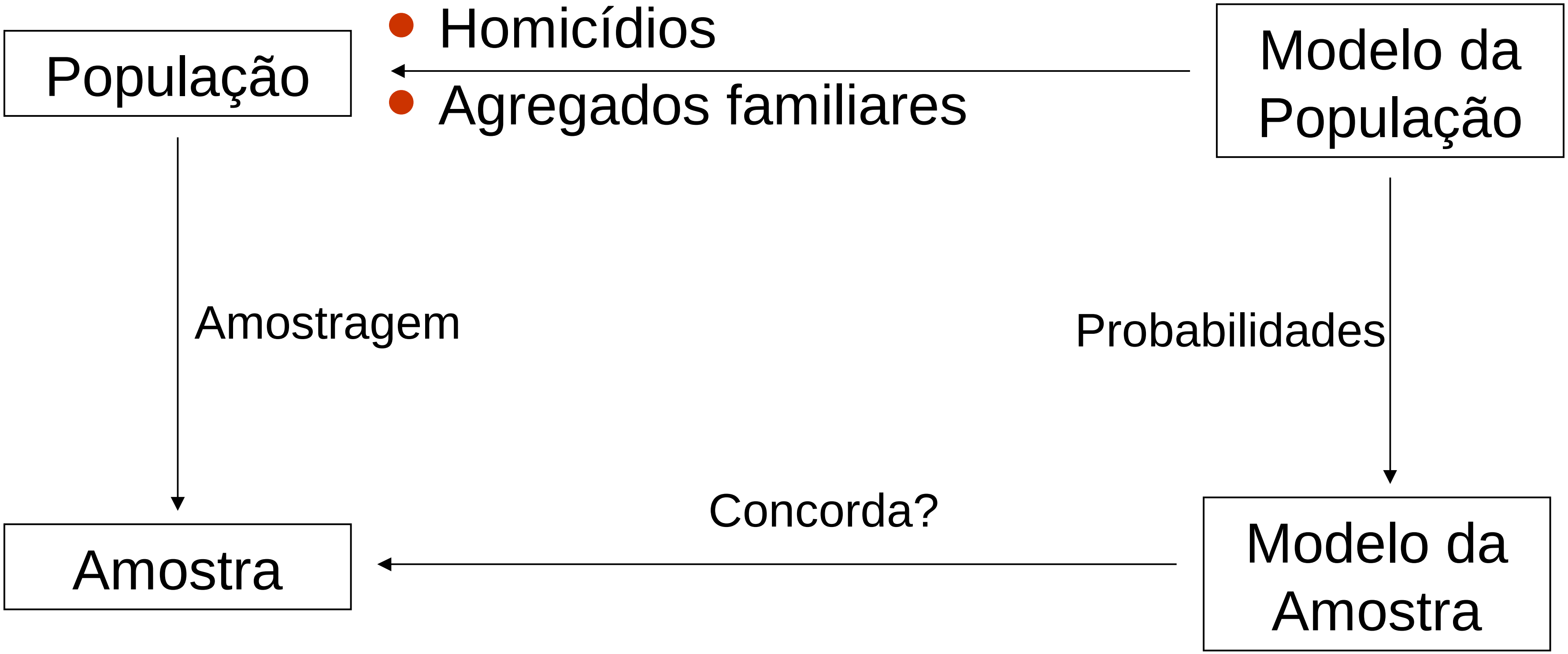
- Estudar uma população por amostragem.
  - Relação entre escolaridade e rendimento.
  - Colesterol é factor de risco para ataques cardíacos?

Modelos Estatísticos

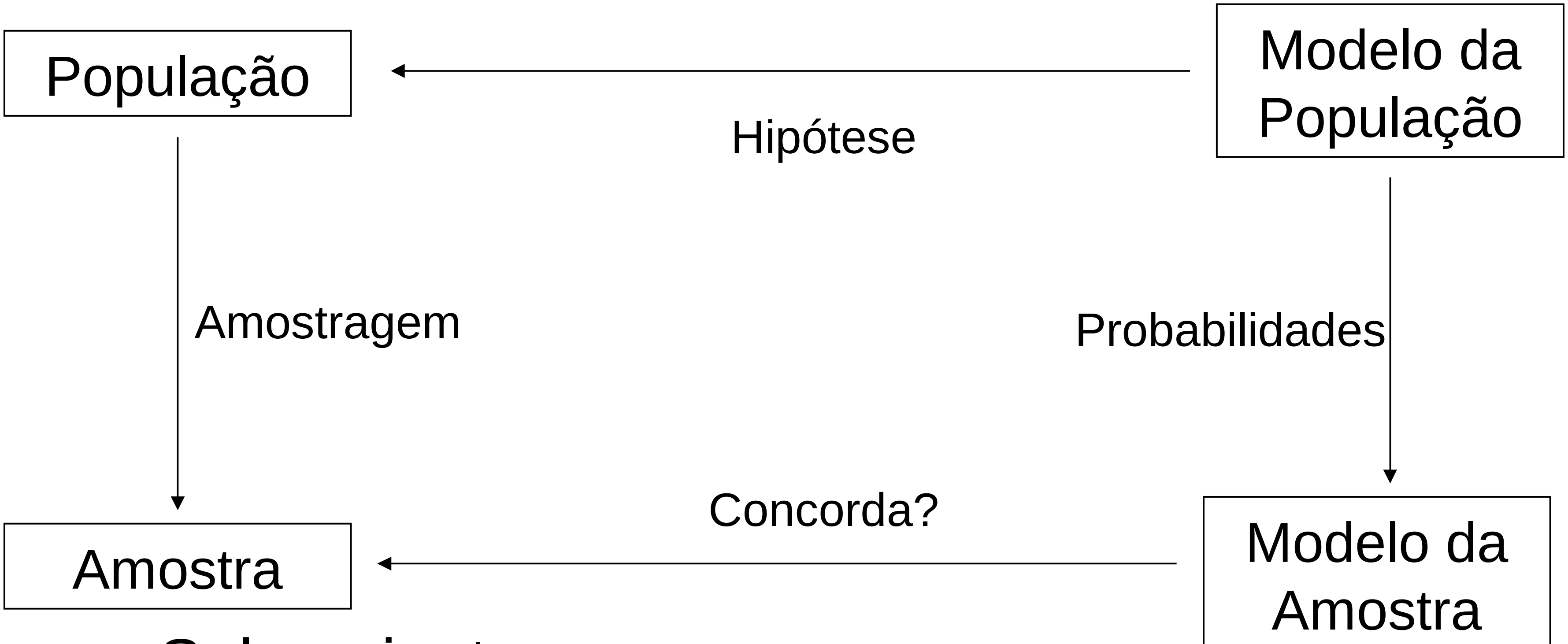
- Estudar uma população por amostragem.
  - Relação entre escolaridade e rendimento.
  - Colesterol é factor de risco para ataques cardíacos?
- Probabilidades.
  - A hipótese do modelo corresponder à população é avaliada pela probabilidade da amostra dado o modelo.
    - Atenção: não é a probabilidade do modelo!

Modelos Estatísticos

- Adultos <50 anos
- Homicídios
- Agregados familiares



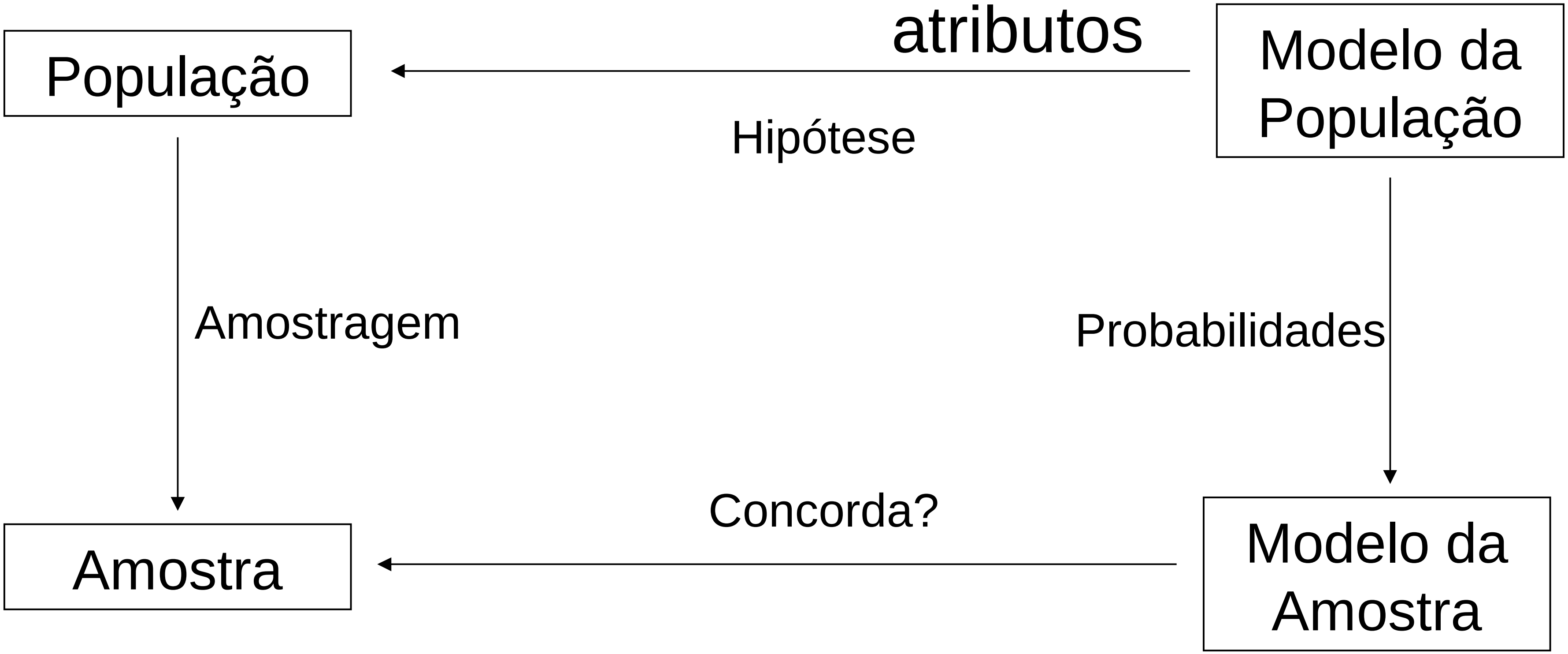
Modelos Estatísticos



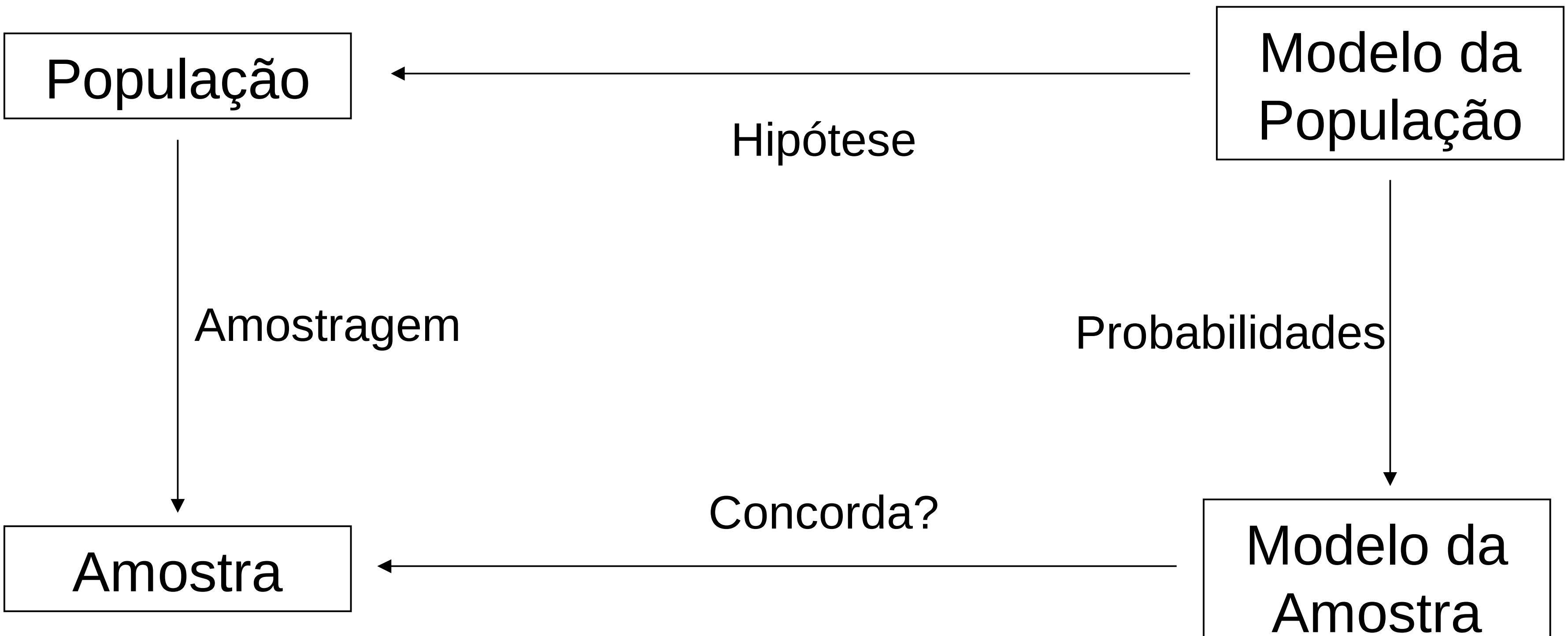
- Subconjunto
  - Representativo (?)

Modelos Estatísticos

- Distribuição de atributos



Modelos Estatísticos



- Características esperadas
  - Probabilidade

Modelos Estatísticos

- 50% vermelhos
- 50% brancos

População

Modelo da População

Hipótese

- Saco com berlindes vermelhos e brancos.

Amostragem

Probabilidades

Concorda?

Amostra

Modelo da Amostra

- 5 Berlindes
  - 3 vermelhos
  - 2 brancos

- ?

Probabilidades

- Variável aleatória
  - Atributo que pode tomar vários valores
  - E.g. Cor
- Probabilidade
  - Proporção do valor (ou combinação de valores) na população.
  - $P(V)=0.5$

Probabilidades

- Modelo 1
  - Cor:
    - B 0.70
    - V 0.30
  - Tamanho
    - L 0.50
    - S 0.50

Probabilidades

- Conjunção
  - Se as variáveis são independentes a probabilidade de V1=X e V2=Y é P(X)*P(Y).
- Exemplo
  - P(B e L)=0.7*0.5=0.35

Probabilidades

- Disjunção
  - Se os valores são mutuamente exclusivos, a probabilidade de ser um ou outro é a soma das probabilidades
- Exemplo
  - P(B ou V)=0.7+0.3 = 1
  - P(2 ou 3 num dado de 6)=1/6+1/6=1/3

Probabilidades

- Modelo A

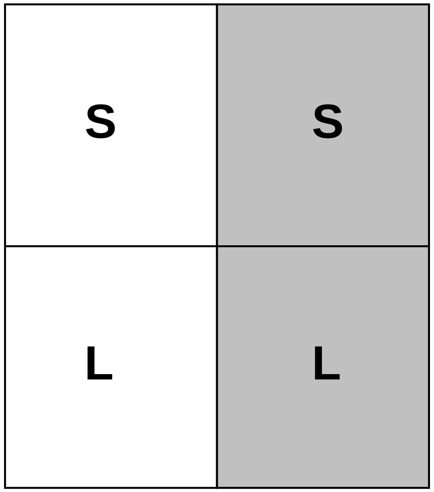
Cor:		Tamanho:
B	0.50	0.25 S, 0.25 L
V	0.50	0.25 S, 0.25 L

- Modelo B

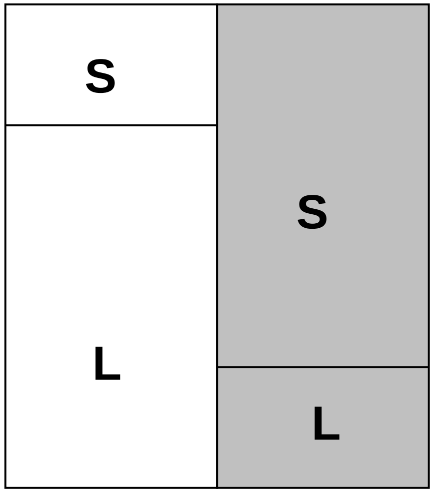
Cor:		Tamanho:
B	0.50	0.10 S, 0.40 L
V	0.50	0.40 S, 0.10 L

Probabilidades

- Correlação
- Modelo A



Modelo B



Probabilidades

- Conjunção
  - Se as variáveis não são independentes a probabilidade de V1=X e V2=Y é P(X)*P(Y/X).
  - P(Y/X) é P(Y) sabendo que V1=X (probabilidade condicionada)
- Exemplo (modelo B)
  - P(B e L)=0.5*P(L/B)=0.5*0.8=0.4

Probabilidades

- Disjunção
  - Se os valores não são mutuamente exclusivos, a probabilidade de ser um ou outro é a soma das probabilidades menos a probabilidade de ser ambos
- Exemplo
  - $P(B \text{ ou } L)=0.5+0.5-P(B \text{ e } L)=1.0-0.4$

Modelos Estatísticos

- 50% vermelhos
- 50% brancos

População

- Saco com berlindes vermelhos e brancos.

Modelo da População

Hipótese:

Bom ajuste?

Amostra

- 5 Berlindes
  - 3 vermelhos
  - 2 brancos

- ?

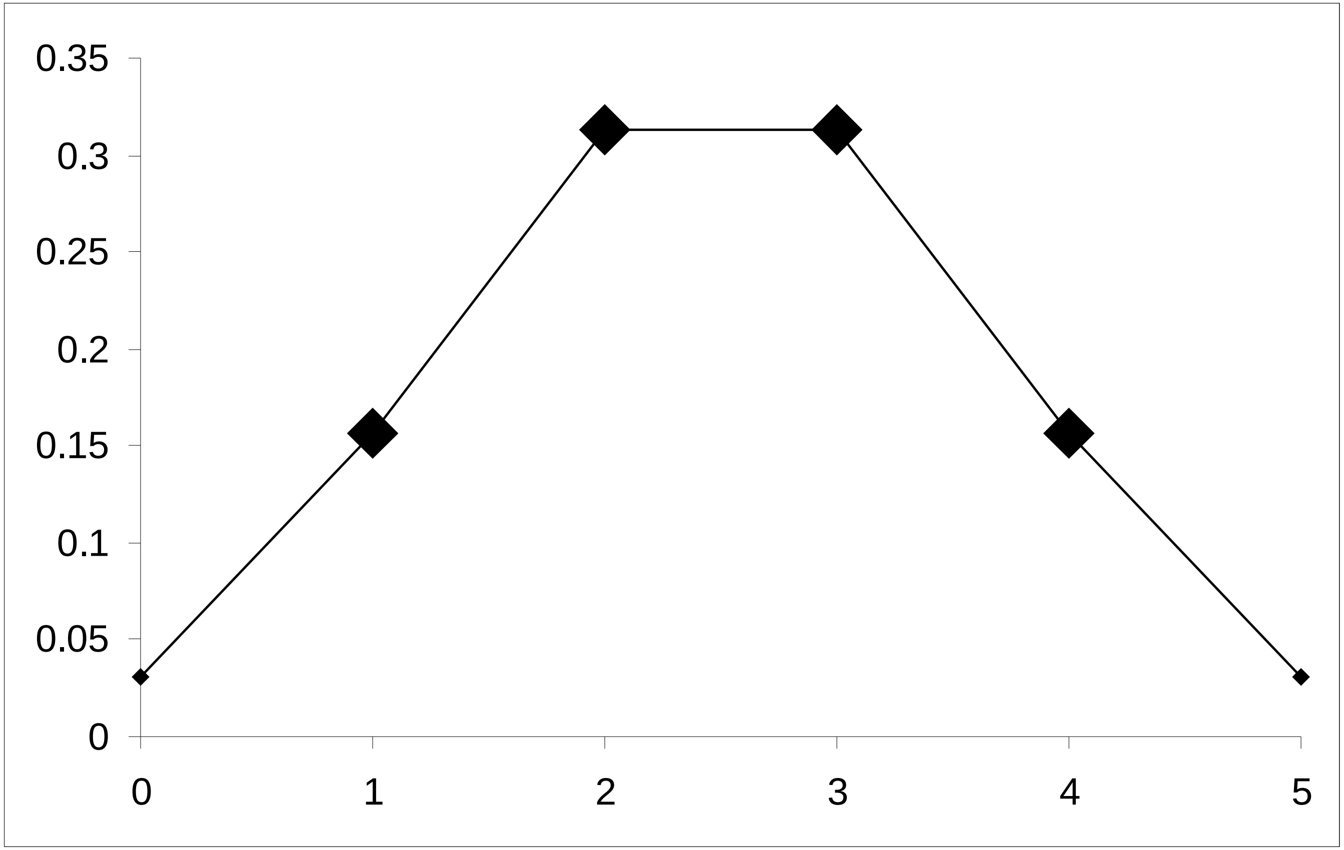
Modelo da Amostra

Probabilidades

- P(0 V) = 1/2*1/2*1/2*1/2*1/2=1/32
- P(1 V) = 5*1/32 = 0,156
- P(2 V) = 5 * 4 / 2 * 1/32 = 0,312
- P(3 V) = P (2 B) = P (2 V) = 0,312
- ...

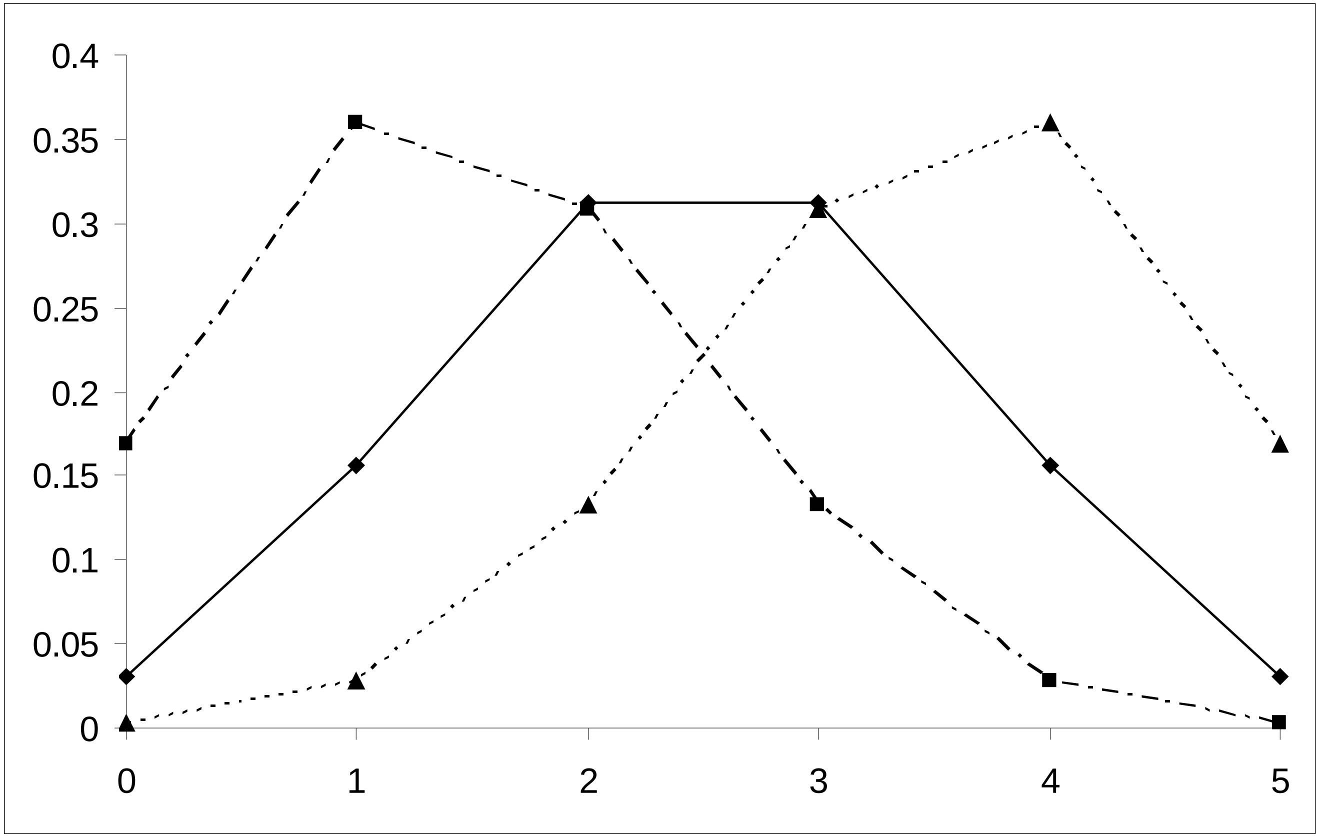
Modelos Estatísticos

- 5 berlindes:



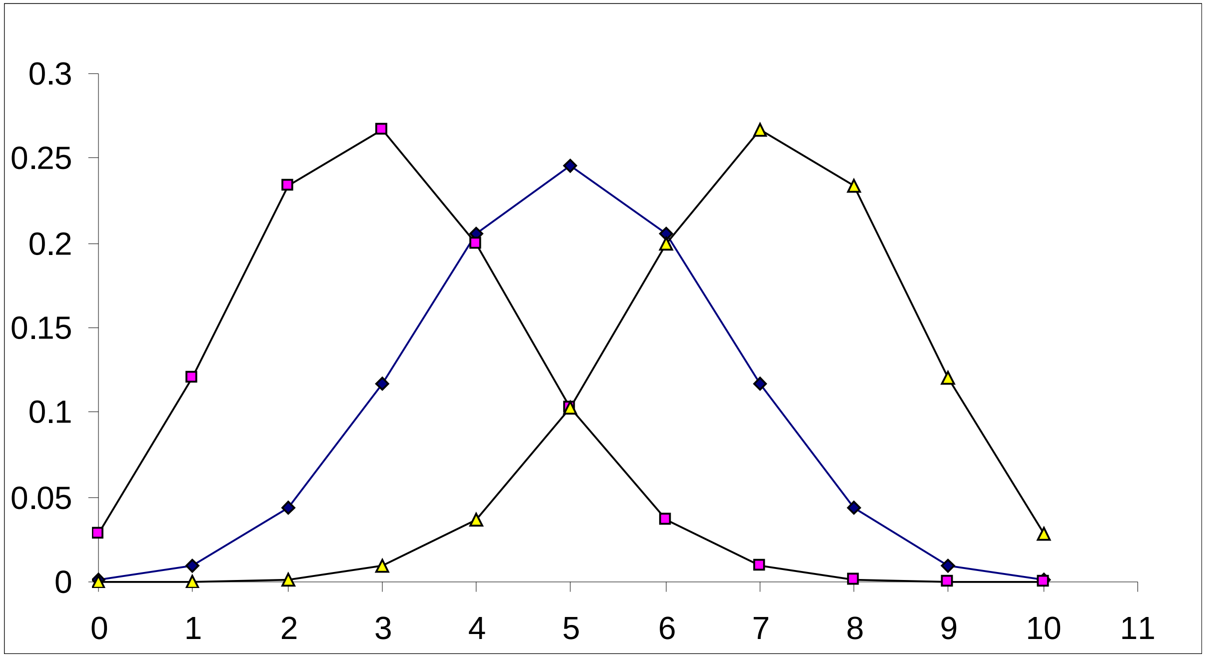
Modelos Estatísticos

- 5 berlindes:



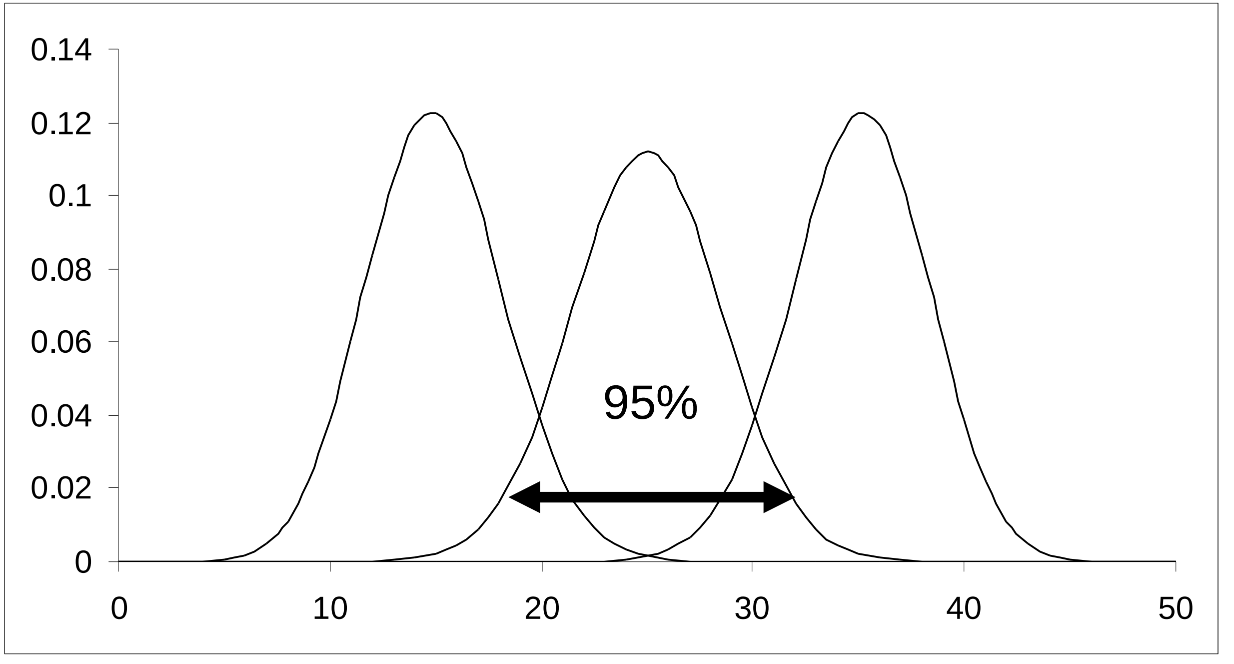
Modelos Estatísticos

- 10 berlindes:



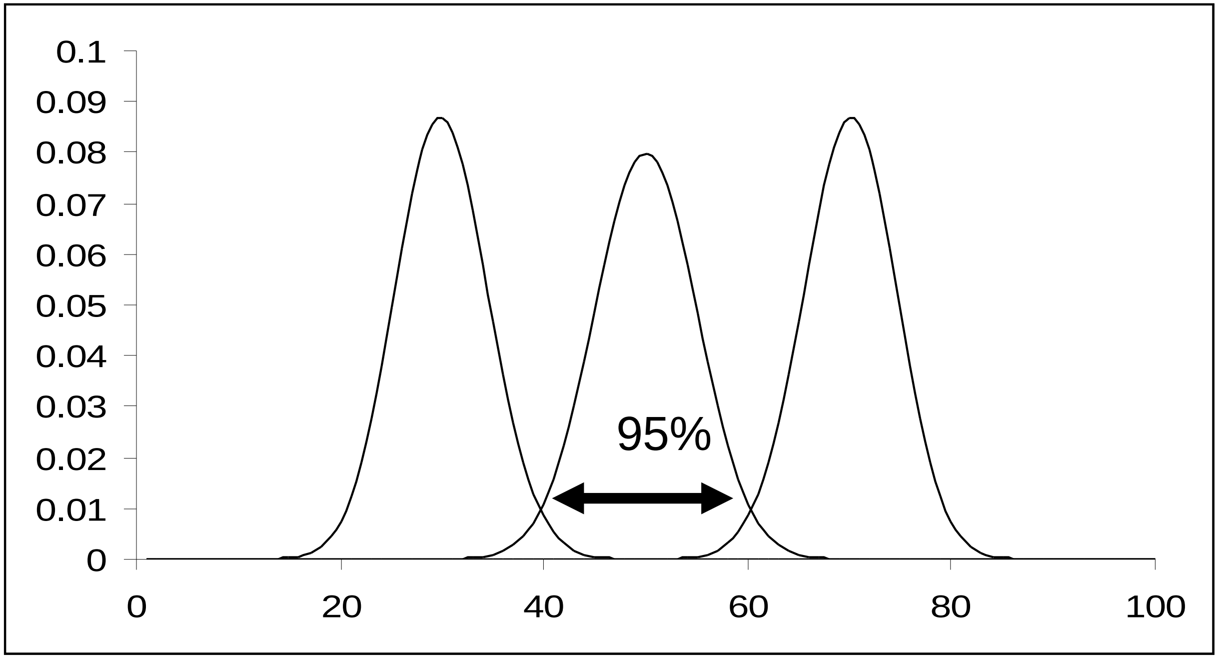
Modelos Estatísticos

- 50 berlindes:



Modelos Estatísticos

- 100 berlindes:



Amostra

- Tamanho da amostra
  - Amostras maiores permitem discriminar melhor entre distribuições (modelos) alternativos.

Modelos Estatísticos

- Quanto maior a amostra menor a margem de erro para um valor de confiança.
- Quanto maior a confiança, maior a margem de erro:
  - 60%: 1 desvio padrão
  - **95%: 2 desvios padrão**

Modelos Estatísticos

- Confiança
  - Probabilidade da estatística da amostra dentro daquele intervalo **se o modelo corresponder à população.**
  - Se a amostra sai fora de um intervalo de confiança de 95%, é pouco plausível que o modelo esteja correcto.
  - Nota: não é a probabilidade do modelo estar correcto.
    - A correspondência entre modelo e população não é uma variável aleatória.

Modelos Estatísticos

- 50% vermelhos
- 50% brancos

Modelo da População

Hipótese

População

- Saco com berlindes vermelhos e brancos.

Amostragem

Probabilidades

Modelo da Amostra

Concorda?

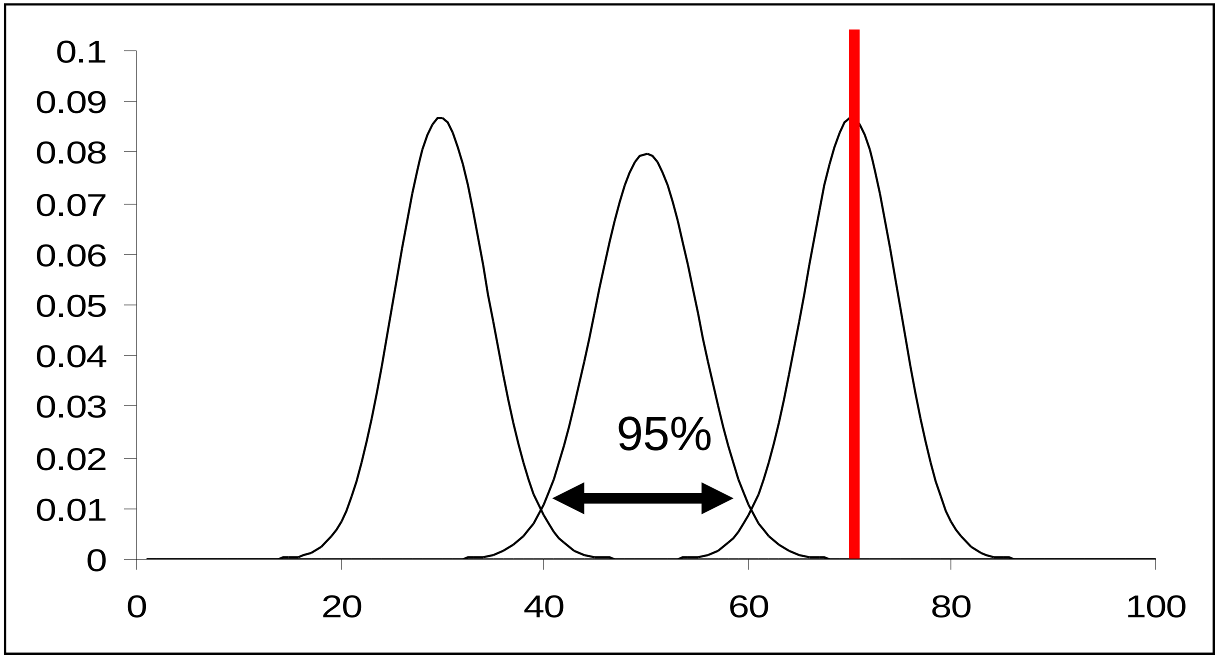
Amostra

- 100 Berlindes
  - 70 vermelhos
  - 30 brancos

- 50±10

Modelos Estatísticos

- 100 berlindes:



Modelos Estatísticos

- 50% vermelhos
- 50% brancos

Modelo da População

Hipótese

População

- Saco com berlindes vermelhos e brancos.

Amostragem

Probabilidades

Concorda?

Amostra

- 100 Berlindes
  - 70 vermelhos
  - 30 brancos

Modelo da Amostra

- 50±10
- Modelo rejeitado

Modelos Estatísticos

- 70% vermelhos
- 30% brancos

Modelo da População

Hipótese

População

- Saco com berlindes vermelhos e brancos.

Amostragem

Probabilidades

Modelo da Amostra

Concorda?

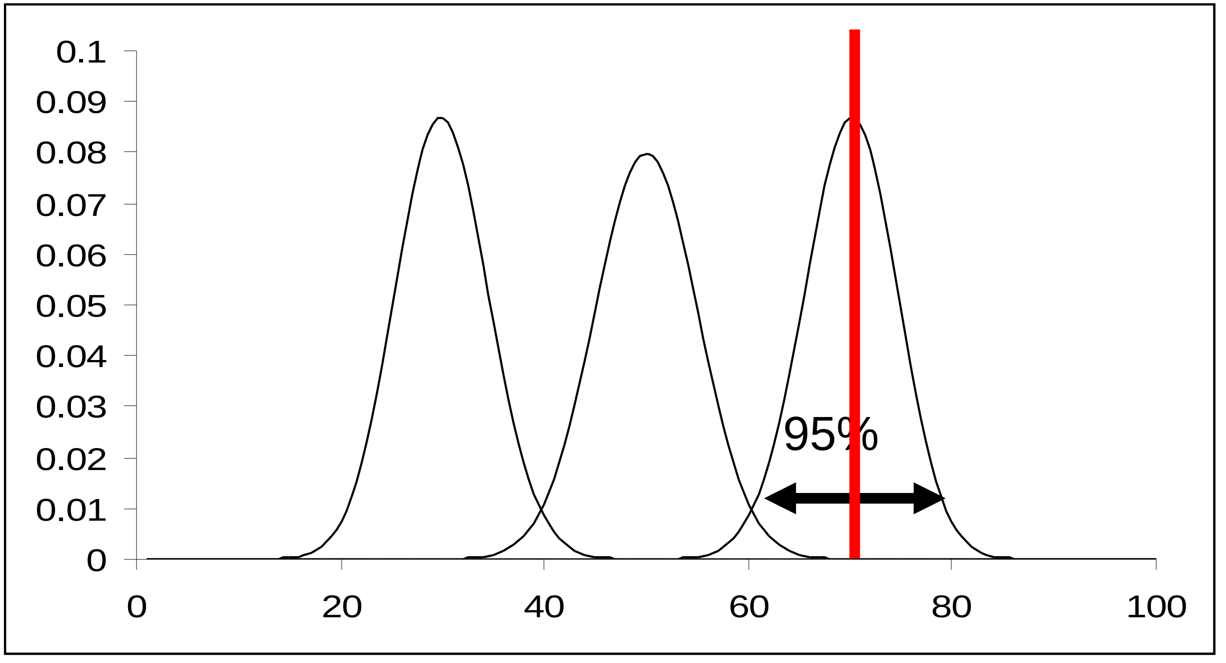
Amostra

- 100 Berlindes
  - 70 vermelhos
  - 30 brancos

- 70±10
- 30 ±10
- Não rejeitado
  - Acertaria mesmo errado?

Modelos Estatísticos

- Modelos que não rejeitaríamos.



Modelos Estatísticos

- Dada a distribuição, amostras vão cair 95% das vezes dentro da margem de erro.
- Portanto, dada uma amostra:
  - Podemos **estimar** o valor na população a partir do valor da amostra.
    - Esse valor será o da amostra +- a **margem de erro**.
  - O intervalo corresponde ao conjunto de modelos que não podemos rejeitar.
- 60-80% V; 40-20% B

Modelos Estatísticos

- Em rigor, o cálculo é mais complexo
  - Depende das funções de distribuição
  - Mas aqui usamos margens de erro aproximadas

Amostra	Erro
25	± .25
100	± .10
500	± .05
2000	± .02
10000	± .01

Desenvolvimento

- Quando os dados são inconsistentes com uma versão do modelo, ou não o suportam devidamente.
- Mas uma alteração ao modelo resolve o problema
- 1,2,3,4,5 -> inconsistente
- 1,2,3,4,5,6.
  - 50% B e V -> 70% V 30% B
- (Mas é preciso continuar o processo)

Modelos Estatísticos

- Comparação de distribuições
  - Duas amostras correspondem a duas distribuições diferentes se não há um modelo aceitável consistente com ambas.
  - Se há um modelo consistente com ambas não podemos dizer que venham de distribuições diferentes.

Modelos Estatísticos

- Exemplo: não distintas
  - Saco 1, 100, 0.6 V (ME 0.10)
  - Saco 2, 50, 0.8 V (ME 0.15)
  - Modelo 0.7 V é consistente com ambos
    - 0.6±0.1
    - 0.8±0.15
  - Os intervalos intersectam-se

Modelos Estatísticos

- Exemplo: distintas
  - Saco 1, 500, 0.6 V (ME 0.05)
  - Saco 2, 500, 0.8 V (ME 0.05)
  - Nenhum modelo é consistente com ambas
    - 0.6±0.05
    - 0.8±0.05
  - Os intervalos são disjuntos

Modelos Estatísticos

- Correlação
  - O valor de uma variável afecta a distribuição de outra.
  - A correlação é simétrica.

Modelos Estatísticos

- Correlação, exemplo:
  - Amostra de 600 berlindes
    - 500 vermelhos
      - 325 são grandes
    - 100 de outras cores
      - 35 são grandes
  - $P(L/V) = 0.65 \pm 0.05$ $(n=500)$
  - $P(L/\sim V) = 0.35 \pm 0.10$ $(n=100)$
  - Há correlação:
    - as populações V e ~V são distintas.

Avaliar Modelos Estatísticos

- 1: População
  - Identificar a população de interesse. Notar possível discrepância entre esta e a população amostrada.
- 2: Os dados na amostra
  - Identificar a amostra e os dados relevantes nesta.
- 3: O modelo estatístico
  - Explicitar o modelo (Correlação? Distribuição? Diferenças?)

Avaliar Modelos Estatísticos

- 4: Amostragem Aleatória
  - Avaliar o processo de amostragem:
  - A probabilidade de cada elemento ser seleccionado é igual para todos (a probabilidade de seleccionar um valor é igual à fracção na população)
  - Não há correlação entre as escolhas para a amostra

Avaliar Modelos Estatísticos

- 5: Avaliar a hipótese
  - O modelo é adequado aos dados?
  - É consistente?
  - Há evidência de correlação ou de diferença entre distribuições (se for esse o modelo)?
  - Forte ou fraca?
- 6: Resumo da avaliação
  - Notar os factores principais (amostra, significância, etc).

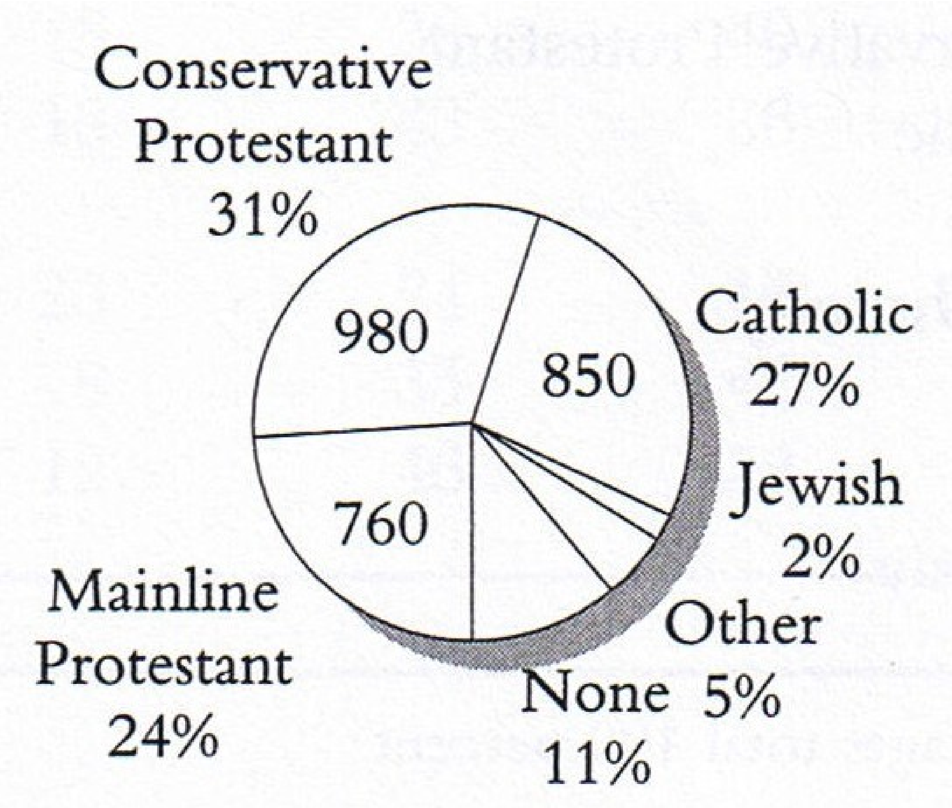
Modelo estatístico

- National Health and Social Life Survey (NHSLS) 1992
  - 3000 adultos (18-49)
    - Entrevistas

Modelo estatístico

- Zero parceiros nos últimos 12 meses
  - Católicos: 12%
  - Protestantes Conservadores: 13%
  - Judeus: 3%
- Há diferenças?

Modelo estatístico



- Com zero parceiros:
  - Católicos 12%
  - Protestantes 13%
  - Judeus 3%

- Amostra:
  - Católicos 1000
  - Protestantes 1000
  - Judeus 50

Modelo estatístico

25	$\pm .25$
100	$\pm .10$
500	$\pm .05$
2000	$\pm .02$
10000	$\pm .01$

- Com zero parceiros:
  - Católicos 12%
  - Protestantes 13%
  - Judeus 3%

- Amostra:
  - Católicos 1000
  - Protestantes 1000
  - Judeus 50

Modelo estatístico

- Zero parceiros nos últimos 12 meses
  - Católicos: 12 ± 3%
  - Protestantes: 13 ± 3%
  - Judeus: 3 ± 20%
- Não há diferença

Modelo estatístico

TABLE 6.4
SEX PARTNERS PAST TWELVE MONTHS

	0	1	2 TO 4	5+
Total	12%	71%	14%	3%
Gender				
Men	10	67	18	5
Women	14	74	10	2

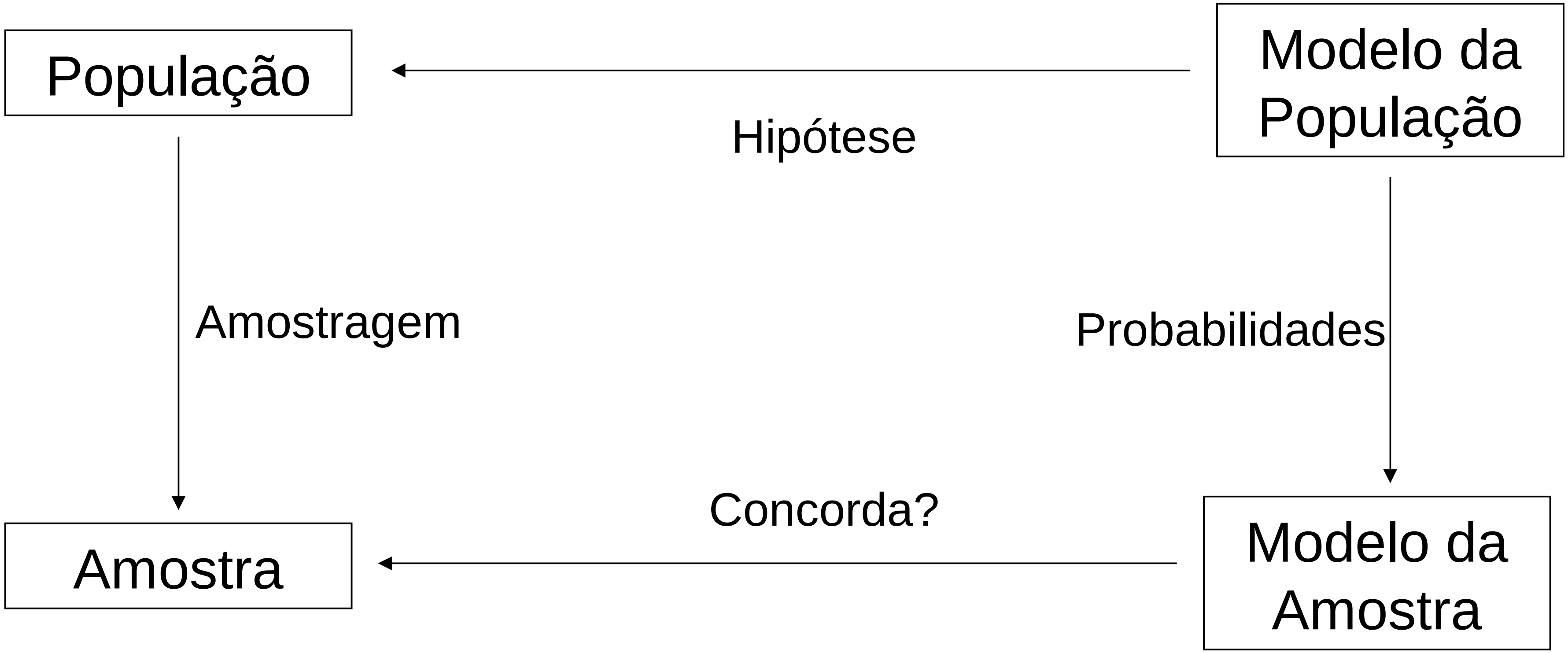
Modelo estatístico

- Mais que 1 parceiro nos últimos 12 meses? (±2%)

- Homens: 23 ± 2%
- Mulheres: 12 ± 2%

- Há diferença

Modelo estatístico



Modelo estatístico

- 1: Qual a população no mundo real
- 2: Estatísticas da amostra
- 3: Modelo estatístico
- 4: Amostra aleatória?
- 5: Avaliar a hipótese
- 6: Resumo

Modelo estatístico

- 1: Qual a população no mundo real
  - Americanos, 18-49 anos

Modelo estatístico

- 1: Qual a população no mundo real
- 2: Estatísticas da amostra
  - mais que 1 parceiro nos últimos 12 meses:
    - Homens: 23%
    - Mulheres: 12%

Modelo estatístico

- 1: Qual a população no mundo real
- 2: Estatísticas da amostra
- 3: Modelo estatístico
  - 2 Variáveis:
    - Sexo (M, F)
    - Mais que 1 parceiro (Sim, Não)

Modelo estatístico

- 1: Qual a população no mundo real
- 2: Estatísticas da amostra
- 3: Modelo estatístico
- 4: Amostra aleatória?
  - Sim
    - Participantes escolhidos aleatoriamente de forma a corresponder à distribuição de idade, educação, estado civil e etnias.

Modelo estatístico

- 1: Qual a população no mundo real
- 2: Estatísticas da amostra
- 3: Modelo estatístico
- 4: Amostra aleatória?
- 5: Avaliar a hipótese
  - Amostra 1500, erro de 2-3%
  - 23 – 12 = 11 > 4 a 6% de margem
    - Diferença significativa

Modelo estatístico

- 6: Resumo
  - Esta amostra sugere que, na população considerada, os homens têm mais probabilidade de ter vários parceiros sexuais no período de seis meses.
    - Ou, pelo menos, de dizer que tiveram...

Modelos Causais

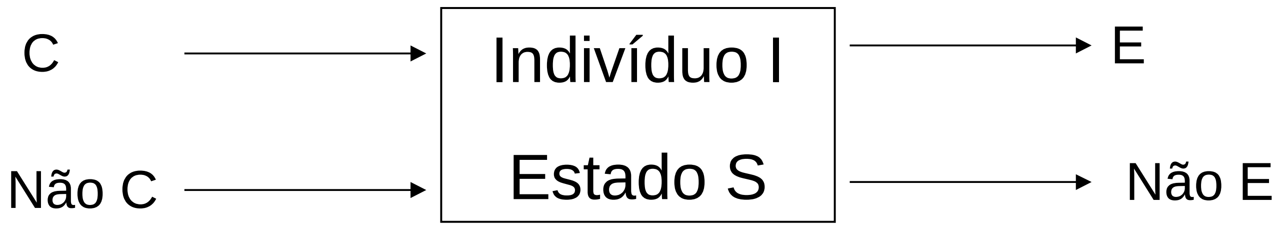
- Correlação e Causalidade
  - Correlação é simétrica:
    - Se há correlação entre A e B há correlação entre B e A.
  - Causalidade é assimétrica:
    - Se A causa B, B não causa A

Modelos Causais

- Correlação e Causalidade
  - Causa:
    - A causa correlaciona-se com o efeito e há um mecanismo pelo qual a causa produz o efeito.
  - Contrafactuais
    - Um modelo causal não nos diz apenas o que acontece, mas o que tem de acontecer (o que aconteceria se).

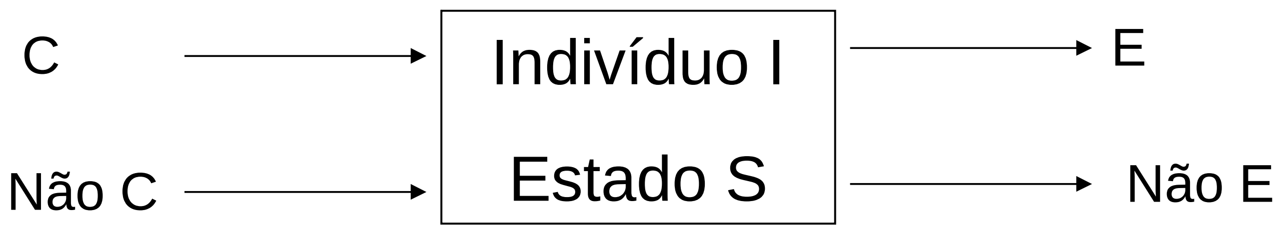
Modelos Causais

- Modelo Determinista
  - C é causa de E para I em S



Modelos Causais

- Modelo Determinista

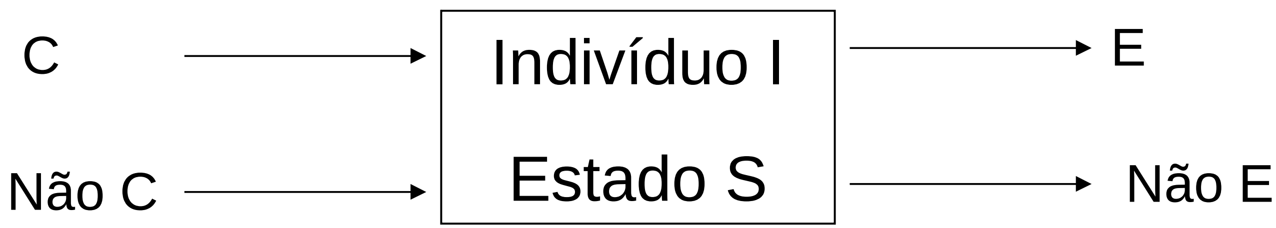


Para o indivíduo I no estado S, C é um factor causal:

- Positivo se C produz E e Não C produz Não E
- Negativo se Não C produz E e C produz Não E

Modelos Causais

- Modelo Probabilístico



Para o I em S, C é um factor causal probabilistico:

- Positivo se P(E|C) > P(E|Não C)
- Negativo se P(E|C) < P(E|Não C)

Modelos Causais

- Um ser humano:
  - Modelo determinista ou probabilístico?

Modelos Causais

- Um ser humano:
  - Modelo determinista ou probabilístico?
  - Porque é que umas pessoas adoecem e outras não?
    - P: Porque calha, não há razão
    - D: Tem de haver alguma diferença
  - Felizmente, importa pouco na prática...

Modelos Causais

Pop.   Real (U)

$P_U(E)$

Modelos Causais

Pop. Hip. Todos C (X)	$P_X(E)$
Pop. Real (U)	$P_U(E)$

Modelos Causais

Pop. Hip. Todos C (X)	$P_X(E)$
Pop. Real (U)	$P_U(E)$
Pop. Hip. Não C (K)	$P_K(E)$

Modelos Causais

Pop. Hip. Todos C (X)	$P_X(E)$
Pop. Real (U)	$P_U(E)$
Pop. Hip. Não C (K)	$P_K(E)$

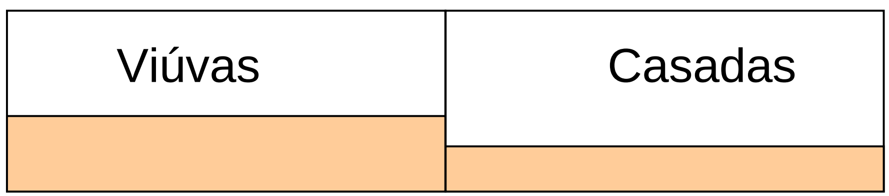
- Modelo População
  - Para a população U, C é um factor causal:
    - positivo se $P_X(E) > P_K(E)$
    - negativo se $P_X(E) < P_K(E)$
    - irrelevante se $P_X(E) = P_K(E)$
  - Para a análise da população não importa se os modelos dos I são deterministas ou probabilísticos.
    - O estado S é diferente conforme o individuo

Modelos Causais

- Ataques cardíacos são mais comuns entre viúvas. Por isso há quem sugira que enviuvar causa ataques cardíacos.
  - Enviuvar é um factor causal positivo?

Modelos Causais

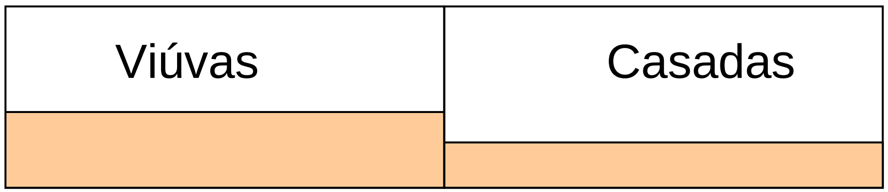
U



Correlação

Modelos Causais

U



Correlação

Modelo Causal (Hipotético)

Para as mesmas mulheres:

X

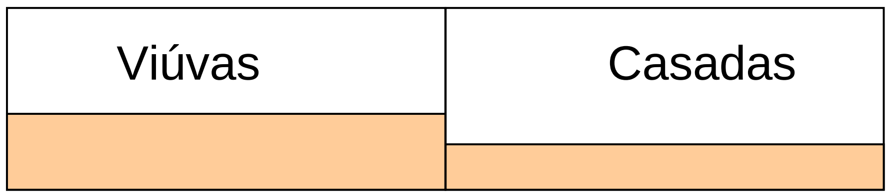
O que aconteceria se todas enviuvassem

K

O que aconteceria se nenhuma enviuvasse

Modelos Causais

U

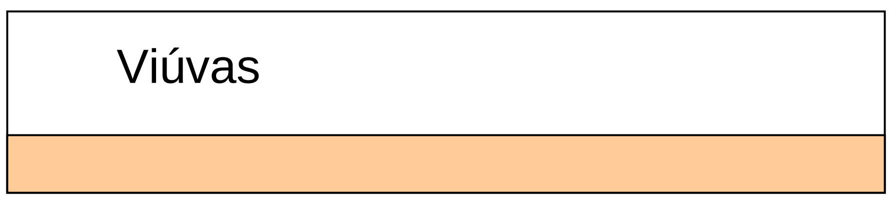


Correlação

Modelo Causal (Hipotético)

Se for irrelevante

X



O que aconteceria se todas enviuvassem

K

O que aconteceria se nenhuma enviuvasse

Próximas

- Teórica 9
  - Análise de modelos causais, exemplos
- Práticas 7, 8, 9
  - Modelos teóricos, estatísticos, causais
  - Turnos P4 e P5: Aula 6 e 7 no dia 8-11
- Ficha 3 (28-11-2011)
  - Um destes
  - Identificar, analisar.

Dúvidas?

- Raciocínio científico
  - Conhecimento
    - crença verdadeira justificada
- Modelos teóricos
  - Realidade, modelo, dados, previsão
  - Consistência, alternativas
- Modelos estatísticos
  - Caso particular, inferir atributos da população pela amostra
- Causalidade
  - A causa geraria um efeito se presente