Modelos Bayesianos

Ernesto F. L. Amaral
Magna M. Inácio

09 de dezembro de 2010
Tópicos Especiais em Teoria e Análise Política:
Problema de Desenho e Análise Empírica (DCP 859B4)

Objetivos

– Apresentar conceitos básicos de estatística Bayesiana.

– Explicar diferenças entre abordagens tradicionais e Bayesianas para análise de dados.

Uma Breve História

– Estatística Bayesiana é assim chamada por ter sido elaborada por Thomas Bayes (1763).

– Laplace (1812) aplicou esta abordagem a problemas práticos.

– A abordagem frequentista foi desenvolvida entre 1850 e 1950 por Fisher, Pearson, Neyman e outros, tornando-se o método dominante.

– Métodos Bayesianos estão sendo mais utilizados, devido às suas vantagens e desenvolvimento de computadores.

– Há diferenças nos conceitos e métodos das abordagens frequentistas e Bayesianas.

Comparação de Termos Frequentistas e Bayesianos

	Frequentista	Bayesiano
Estimação	Estimativas pontuais Intervalos de confiança	Distribuições a priori e posteriores
Inferência	Intervalos de confiança Testes de significância	Distribuições a priori e posteriores

Análise Bayesiana ≠ Análise Frequentista

– Na **perspectiva Bayesiana**, a probabilidade representa uma abordagem subjetiva do desconhecido.

  – Qualquer quantidade para a qual o valor verdadeiro é incerto, incluindo os parâmetros do modelo, pode se representada com distribuições de probabilidade.

– Na **perspectiva clássica**, não é aceitável colocar distribuições de probabilidade nos parâmetros, porque estes são considerados como quantidades fixas.

  – Somente dados são aleatórios, então distribuições de probabilidade somente podem representar os dados.

Problemas com Estatística Tradicional

(segundo Bayesianos)

– Produz respostas em termos que são difíceis de interpretar e que são mal compreendidos.

– Produz respostas em termos que não são usualmente de grande interesse:

Pr(dados|H), ao invés de Pr(H|dados)

– Falta de métodos claros para incluir dados (conhecimentos) existentes e para lidar com incerteza.

Responda Verdadeiro ou Falso

(*p* é o valor da probabilidade de teste de significância para H_0)

– $p=0,01$: há 1 em 100 chances que H_0 seja verdadeira.
– $p<0,001$ significa que H_0 é falsa.
– H_0 tem maior probabilidade de ser verdadeira se $p>0,9$ do que se $p<0,1$.
– Se 2 amostras A e B não são diferentes significativamente (H_0 não é rejeitada), A e B são assumidas como iguais.
– Se $p<0,05$ (significante) para um conjunto de dados e $p < 0,001$ (muito significante) em outro conjunto, o segundo conjunto de dados indica que H_0 é falsa mais fortemente.
– Se H_0 não for rejeitada e a magnitude do coeficiente for alta, é mais provável que H_0 seja verdadeira do que se H_0 não for rejeitada e a magnitude do coeficiente for baixa.
– Se o intervalo de confiança de 95% de uma estimativa Y vai de A a B, isso significa que o valor verdadeiro de Y está situado entre A e B com probabilidade 0,95.

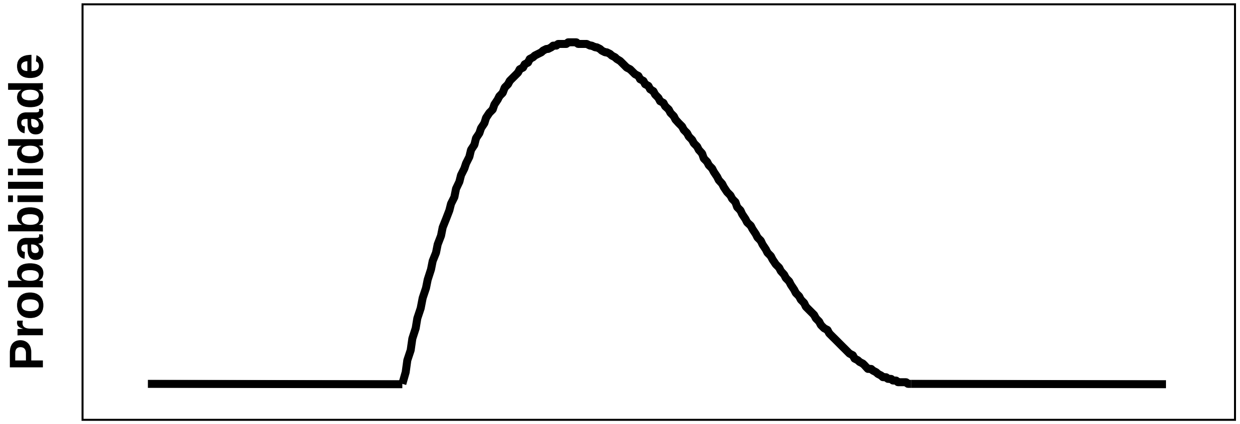
Dois Conceitos Bayesianos Fundamentais

– Coisas que são desconhecidas são representadas por distribuições de probabilidade.

– Coisas que são conhecidas (dados) são usadas para aperfeiçoar nosso conhecimento acerca do problema, a partir do Teorema de Bayes.

O Desconhecido

– Coisas que são desconhecidas são representadas por distribuições de probabilidade.



Quantidade desconhecida de interesse (Quantidade de votos, tamanho da população, parâmetro do modelo)

– A probabilidade da distribuição pode ser contínua (uniforme, normal...) ou discreta (binomial, Poisson).

O Conhecido

– Regra da multiplicação:

$$\Pr(x, y) = \Pr(x)\Pr(y \mid x) = \Pr(y)\Pr(x \mid y)$$

– Teorema de Bayes:

$$\Pr(y \mid x) = \frac{\Pr(x \mid y)\Pr(y)}{\Pr(x)}$$

– Seja y=hipótese sobre o desconhecido e x=dados:

$$p\,(\mathbf{H} \mid \mathbf{dados}) = \frac{p\,(\mathbf{dados} \mid \mathbf{H}) \;\; p\,(\mathbf{H})}{p\,(\mathbf{dados})}$$

– Distribuição posterior é proporcional a:

Posterior $\propto$ Verossimilhança (*Likelihood*) X Priori

Paradigma para Inferência Bayesiana

Posterior $\propto$ Verossimilhança X Priori

– Ou seja:

novo estado do conhecimento $\propto$

informação dos dados novos

X estado atual do conhecimento

– Os dados novos atualizam o estado atual do conhecimento, com base no Teorema de Bayes.

– O resultado é um novo estado do conhecimento, representado pela distribuição posterior.

Objetivos da Estatística Bayesiana

– Representar o desconhecimento a priori sobre os parâmetros do modelo com uma distribuição de probabilidade (**distribuição a priori**).

– Atualizar esse desconhecimento a priori com dados atuais (***likelihood***).

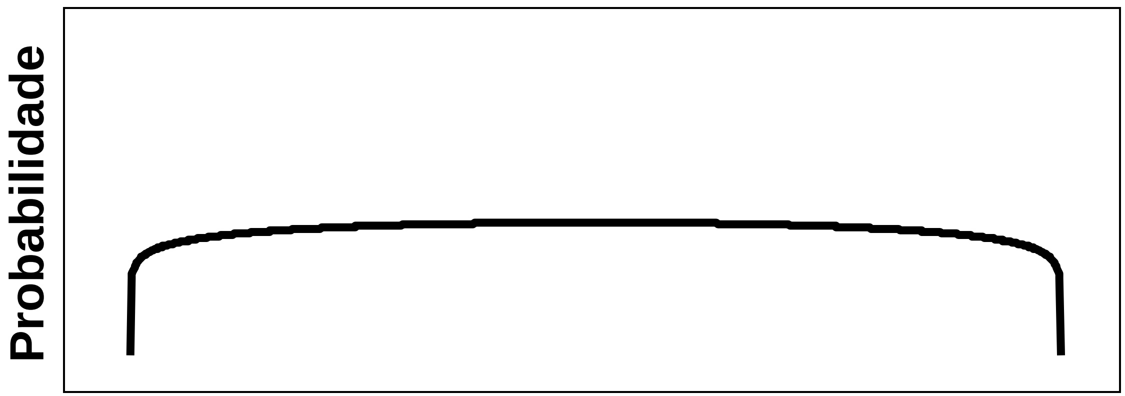
– Produzir uma distribuição de probabilidade para o parâmetro que contenha menos desconhecimento (**distribuição posterior**).

Distribuição a Priori

– Função a priori Pr(H) fornece a probabilidade dos diferentes valores possíveis da quantidade de interesse **antes** dos dados serem considerados.

– Representa o estado do conhecimento anterior aos dados.

– A distribuição a priori pode ser ampla, plana, uniforme se possuímos poucos dados (priori não informativa), ou pode se concentrada com um ápice se possuímos mais informação (priori informativa).



Quantidade desconhecida de interesse (Quantidade de votos, tamanho da população, parâmetro do modelo)

Verossimilhança (*Likelihood*)

– Função de verossimilhança Pr(dados|H) fornece a probabilidade de obter o dado, considerando diferentes valores possíveis da quantidade desconhecida de interesse (hipótese H).

– Verossimilhança: (1) é calculada usando um modelo estatístico que representa o processo que produziu os dados; e (2) conecta os parâmetros do modelo aos dados.

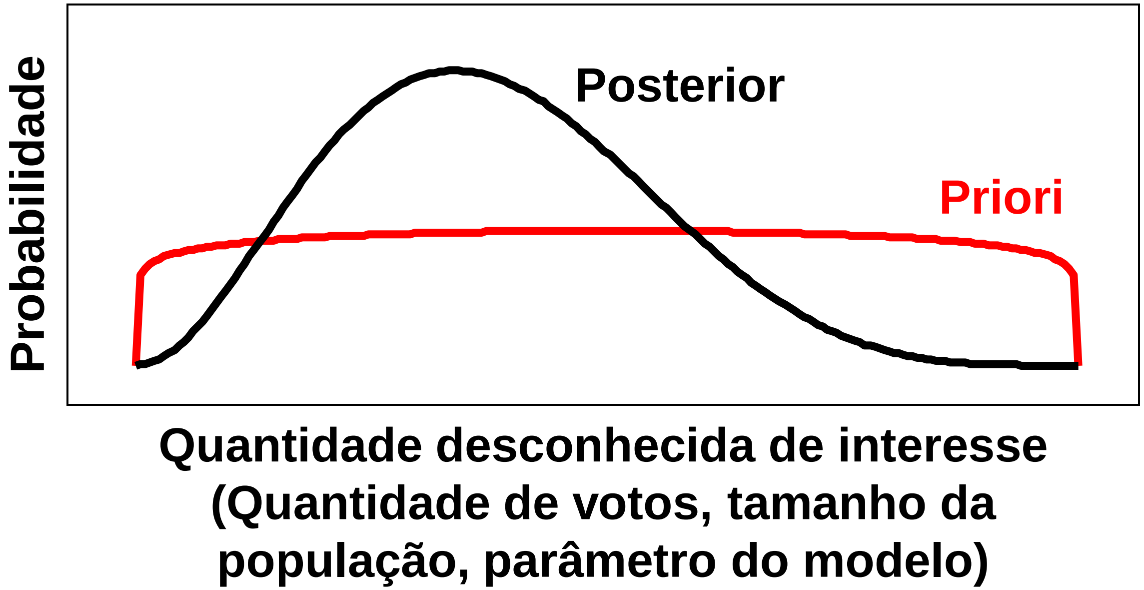
– Também são utilizadas em análises frequentistas.

Distribuição Posterior

– Função posterior Pr(H|dados) fornece a probabilidade dos diferentes valores possíveis da quantidade de interesse **depois** de considerar os dados, representando o estado do conhecimento posterior aos dados.

– Posterior é combinação da priori (o que sabemos antes) com a verossimilhança (o que os dados nos disseram).

– Diferença entre priori e posterior indica o quanto aprendemos com os dados.



Exemplo: Tabela de Contingência

	Grupo 1	Grupo 2
Tipo A	1	4
Tipo B (não A)	8	6

- Questão principal:
  - A fração do tipo A é diferente entre os dois grupos?

- Pressupostos:
  - Amostras são aleatórias e independentes.
  - A e B são únicos tipos possíveis.

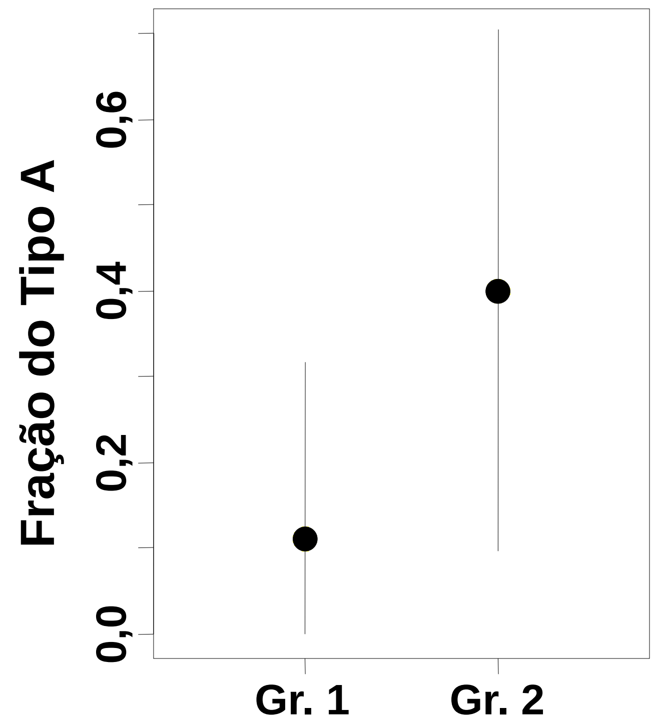
Análise Frequentista

	Grupo 1	Grupo 2
Tipo A	1	4
Tipo B (não A)	8	6

– Teste de Pearson: p=0,36.

– Teste de Fisher: p=0,30.

– Conclusão: hipótese nula de que a fração do tipo A é igual nos dois grupos não é rejeitada.



Análise Bayesiana

	Grupo 1	Grupo 2
Tipo A	1	4
Tipo B (não A)	8	6

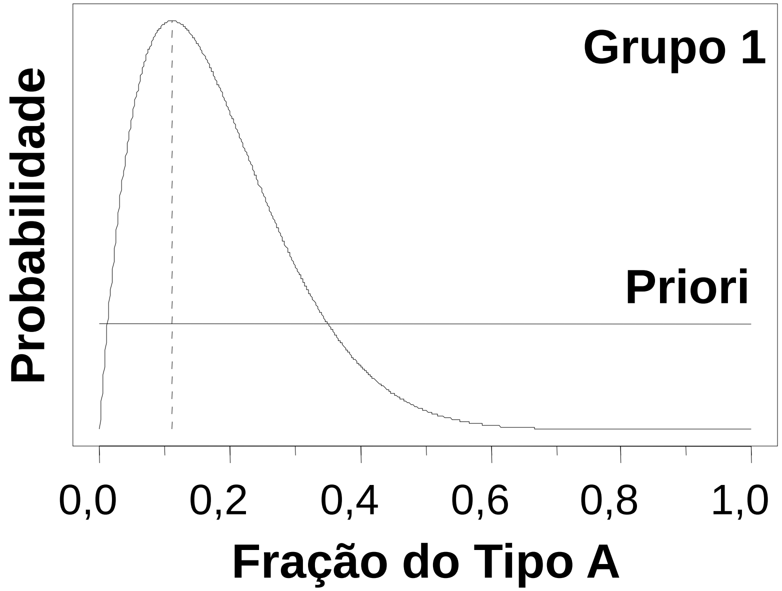
– Distribuição a priori uniforme:

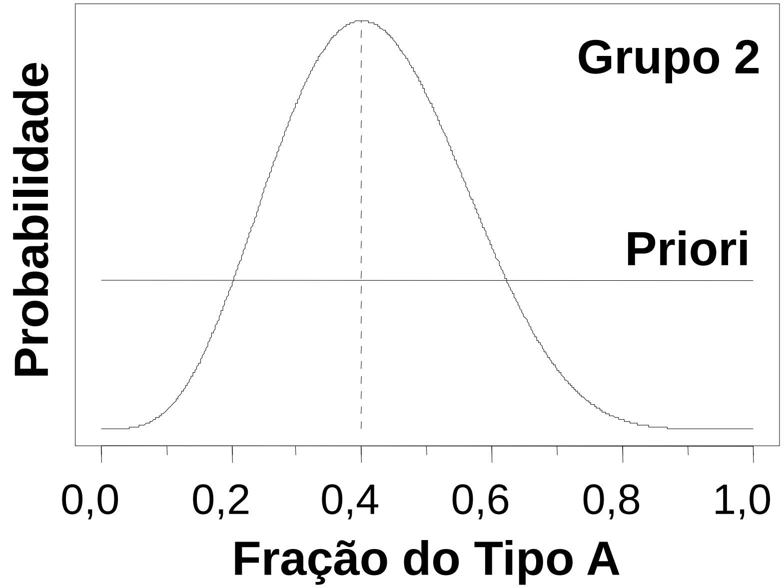
Pr(x)~U(0,1)

– Verossimilhança binomial:

$$\Pr(x \mid n_A, n_B) \propto x^{n_A} (1-x)^{n_B}$$

onde x é a fração do tipo A.

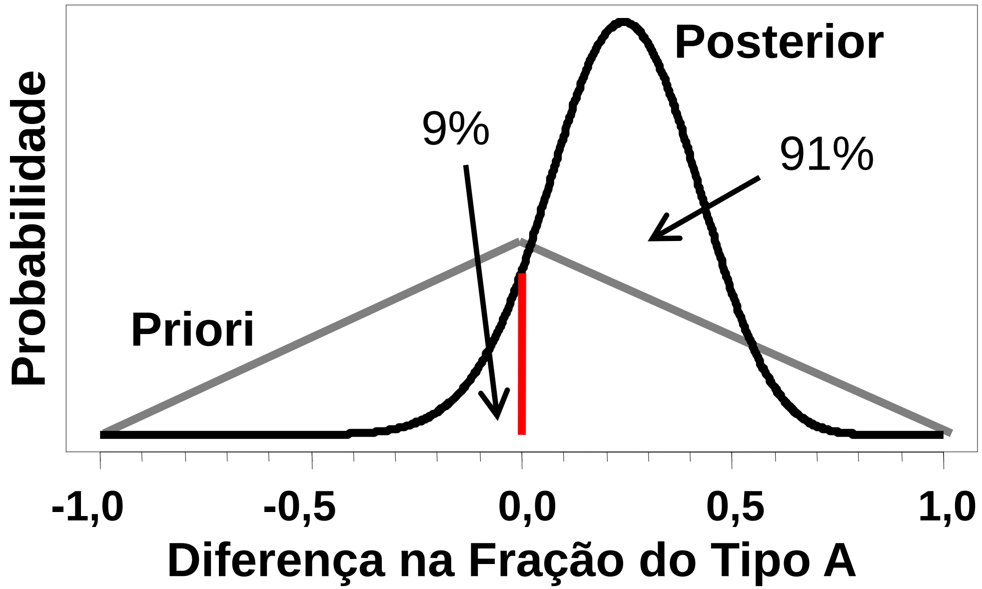




Análise Bayesiana: Mais Clara e Direta

– Como estamos interessados na **diferença** da fração do tipo A entre o grupo 1 e o grupo 2, calculamos a distribuição posterior sobre a diferença.

– Embora as distribuições a priori são uniformes na fração do tipo A para cada grupo, a distribuição a priori na diferença entre grupos não é uniforme.



– A probabilidade que a fração do tipo A seja maior no grupo 2 é de 0,91 (acima de zero).

– É 10 vezes mais provável (91% / 9%) que a fração do tipo A seja maior no grupo 2 do que seja menor.

Análise Bayesiana: Distribuição Posterior

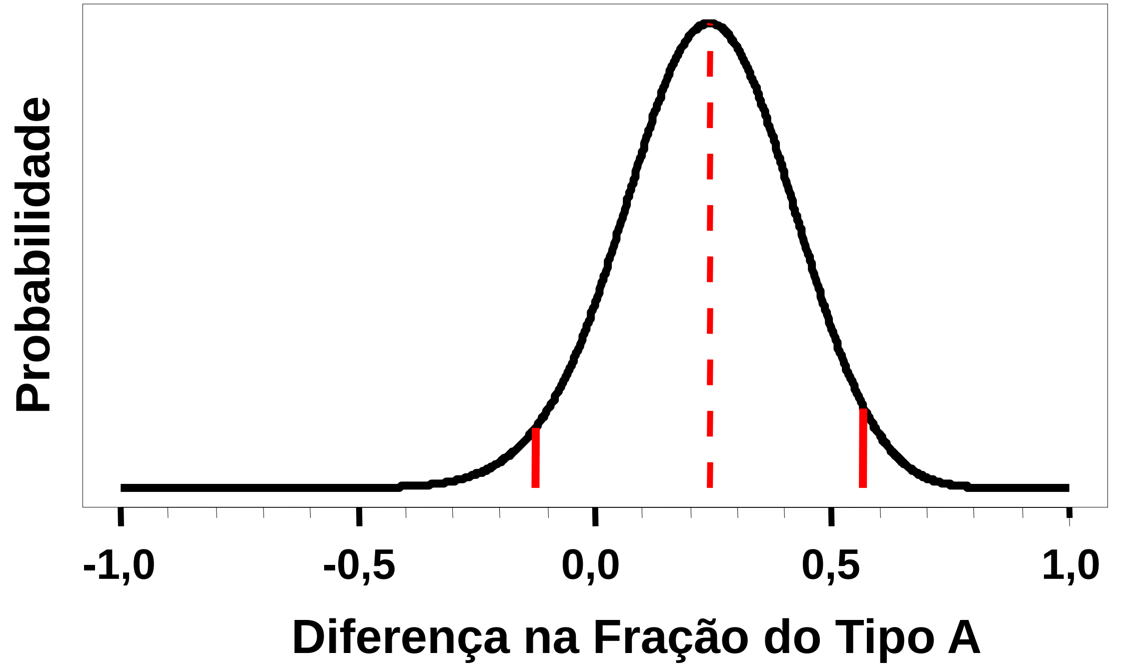
– A distribuição posterior da diferença na fração do tipo A entre grupo 2 e grupo 1 é o novo estado do conhecimento.

– Média = 0,241.

– Mediana = 0,235.

– Moda = 0,241.

– Intervalo de confiança de 95%: -0,13 a 0,57.

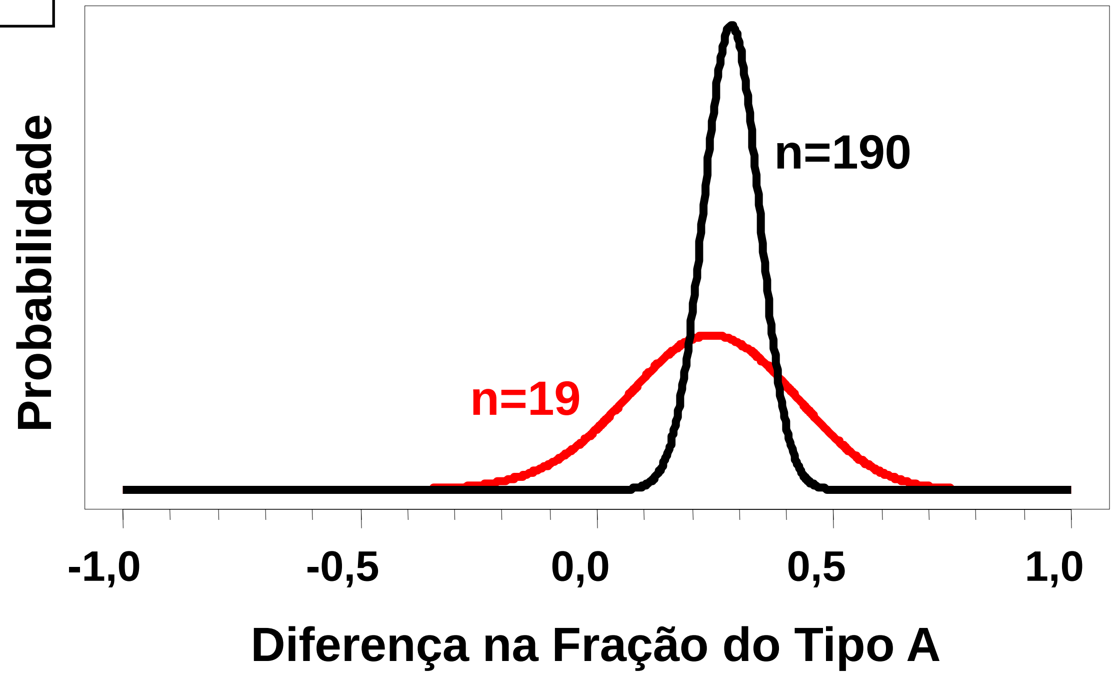


Análise Bayesiana com Amostra Maior

– Suponha que tenhamos dados adicionais, de modo que o tamanho da amostra seja muito maior.

– A distribuição posterior da diferença será mais concentrada.

	Grupo 1	Grupo 2
Tipo A	10	40
Tipo B (não A)	80	60



Problemas com Estatística Bayesiana

(segundo não-Bayesianos)

– Distribuições a priori introduzem julgamento subjetivo na análise de dados.

– Distribuições a priori afetam os resultados, de modo que diferentes pesquisadores podem chegar em respostas distintas com mesmos dados.

– É muito complicado, já que não existem programas computacionais simples como nosso amigo Stata.

Contra Argumentação: Interpretação

– As **estimativas de intervalo Bayesianos** possuem interpretação clara e direta:

  – Simplesmente dizemos que um parâmetro se encontra em um intervalo com determinada probabilidade.

– Os **intervalos de confiança clássicos** indicam a probabilidade de obter uma estimativa de intervalo que contem o parâmetro de interesse, sob a hipótese de realização de várias amostras.

Contra Argumentação: Hipóteses

– A **abordagem Bayesiana** pode naturalmente incorporar resultados de pesquisas anteriores na distribuição a priori.

  – O processo de construção de hipóteses é formalizado com sua incorporação como parte do modelo.

– A **abordagem clássica** não possui meios para usar resultados prévios nas análises atuais, além de especificar hipóteses.

Contra Argumentação: Estatísticas

– Na **abordagem Bayesiana**, são estimadas estatísticas dos parâmetros mais detalhadas, tais como média, mediana, moda e intervalo interquartil.

– Na **abordagem clássica**, simplesmente obtemos a estimativa do parâmetro e do erro padrão.

Contra Argumentação: Subjetividade

– A **abordagem Bayesiana** é acusada de ser subjetiva por incorporar distribuições a priori.

– Porém, **toda estatística é subjetiva**:

– Em um trabalho com estatística frequentista, por exemplo, podemos escolher uma distribuição normal e aplicar um modelo MQO, ou podemos escolher uma distribuição binomial e aplicar um modelo logístico.

– O nível de significância (α) é definido subjetivamente.

– Por fim, a influência de distribuições a priori pode ser avaliada após estimação da distribuição posterior.