Fundamentos da Inferênica pelo Axioma da Verossimilhança

João Batista e Paulo Inácio Prado

2018

BIE5781 Modelagem Estatística em Ecologia e Recursos Naturais

Sumário

Sumário

Inferência Clássica

Filosofia da Inferência Clássica: Dados

Os dados:

- Provem de uma *população*
- Foram observados segundo um *procedimento*
  - ⇒ levantamento/ observação orientada
  - ⇒ experimento / observação controlada

Conceito de "observação":

- O mesmo procedimento pode ser *repetido* (hipoteticamente)
- *Amostrando* sempre a mesma população
- Os dados observados são:
  - ⇒ *uma* amostra
  - ⇒ de um *grande número* (***infinito***) de amostras possíveis

A inferência quantitativa começa com:

- *Estatísticas* calculadas nos dados observados:
  - $\Rightarrow$ médias, medianas, variâncias, razões, somas,
  - $\Rightarrow$ representam *parâmetros* da população que gerou os dados
- Sobre as estatísticas conhecemos:
  - $\Rightarrow$ um *único* valor observado
  - $\Rightarrow$ mas elas são *variáveis aleatórias*
  - $\Rightarrow$ com sua *distribuição amostral*

Filosofia da Inferência Clássica: Métodos

Realizar a inferência implica em:

- *conhecer* a distribuição amostral da estatística
- *utilizar* essa distribuição amostral
- para *avaliar* o comportamento da estatística através de:
  - $\Rightarrow$ inferência pontual (*estimativa da estatística*)
  - $\Rightarrow$ inferência intervalar (*intervalo de confiança*)
  - $\Rightarrow$ teste de hipóteses (*a respeito dos parâmetros*)

Filosofia da Inferência Clássica: Fundamentos

Pontos Importantes

- Os parâmetros se referem à população (origem os dados)
- As estatísticas se referem
  - ⇒ aos dados observados
  - ⇒ mas inferem sobre a população
- O comportamento das estatísticas dependem:
  - ⇒ da população original
  - ⇒ dos dados observados (amostra)
  - ⇒ mas também do *procedimento* de observação

Filosofia da Inferência Clássica: Fundamentos

Resultado: a *distribuição amostral* (*inferência*) depende

$\Rightarrow$ da *população*

$\Rightarrow$ da *amostra*

$\Rightarrow$ do *procedimento*

O Procedimento define:

$\Rightarrow$ o *espaço amostral* da distribuição amostral

$\Rightarrow$ tudo o que *pode* ser observado

$\Rightarrow$ mas que *não necessariamente* será observado

Exemplo de Inferência Clássica: Dados

Os Dados: Levantamento de Floresta Tropical

	N. Árvores	N. Parcelas	Árvores por Parcela
Encosta	2981	18	165.6
Platô grande	4339	30	144.6
Platô pequeno	3143	22	142.9

Comparação de Duas Áreas: Encosta $\times$ Platô Pequeno

Topografia	Diâmetros: DAP (cm)		
	Valores	Média	Desv. Pad.
Encosta	30 28 28 25 24 26 26 23 23 25 25 23 26 27 26 23 27 26	25.6	1.9745
Platô pequeno	25 24 24 24 22 22 23 22 24 25 31 27 24 25 22 24 24 23 25 25 28 28	24.6	2.2395

Exemplo de Inferência Clássica: Teste t

Teste t: comparação de duas médias

- Hipóteses:

$$H_0 : \mu_1 = \mu_2 \qquad H_\alpha : \mu_1 \neq \mu_2$$

- Distribuição Amostal — pelo Teorema Central do Limite
  - a média amostral tem distribuição *Gaussiana*
  - sob H_0: diferenças entre médias amostrais *de uma mesma população* seguem a estatística t
- Estatística:

$$t = \frac{\overline{x}_1 - \overline{x}_2}{s_\Delta}$$

Exemplo de Inferência Clássica: Teste t (cont.)

Teste t: variâncias equais

$$
\begin{aligned}
s_\Delta &= \left[\frac{(n_1-1)s_1^2+(n2-1)s_2^2}{n_1+n_2+2}\right]\sqrt{\frac{1}{n_1}+\frac{1}{n2}} \\
\Rightarrow \text{g.l.} &= n_1+n_2-2
\end{aligned}
$$

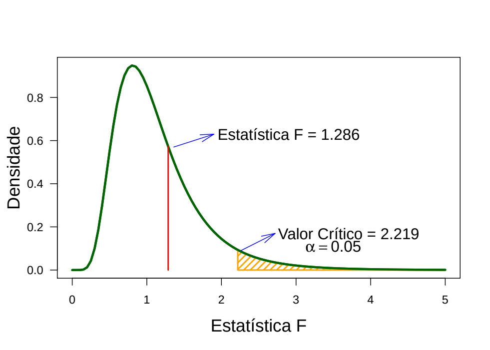
Teste t: variâncias diferentes

$$
\begin{aligned}
s_\Delta &= \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n2}} \\
\Rightarrow \text{g.l.} &= \frac{\left(s_1^2/n_1+s_2^2/n_2\right)^2}{(s_1^2/n_1)^2/(n_1-1)+(s_2^2/n_2)/(n_2-1)}
\end{aligned}
$$

Exemplo de Inferência Clássica: Teste F para as variâncias

Testando a Igualdade de Variâncias

$$F = \frac{\widehat{s}_2^2}{\widehat{s}_1^2} = \frac{2.2395^2}{1.9745^2} = 1.286$$



Filosofia da Inferência Clássica: Exemplo Simples

Teste t de comparação das médias (variâncias iguais)

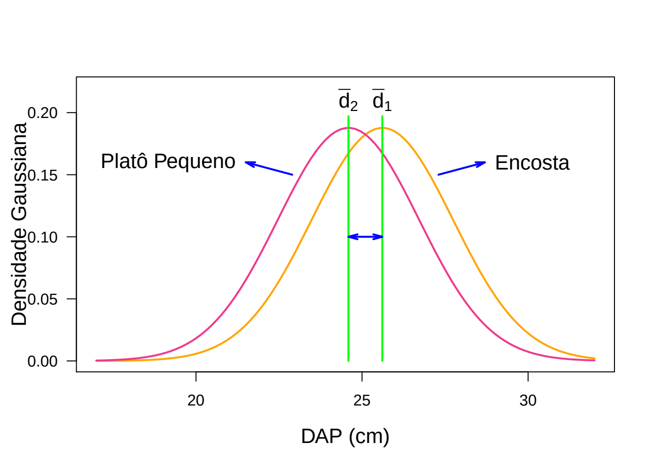
```
> t.test(dap1, dap2, var.equa=TRUE)

        Two Sample t-test

data:  dap1 and dap2
t = 1.5106, df = 38, p-value = 0.1392
alternative hypothesis:
  true difference in means is not equal to 0
95 percent confidence interval:
 -0.3470205  2.3874246
sample estimates:
mean of x mean of y
 25.61111  24.59091
```

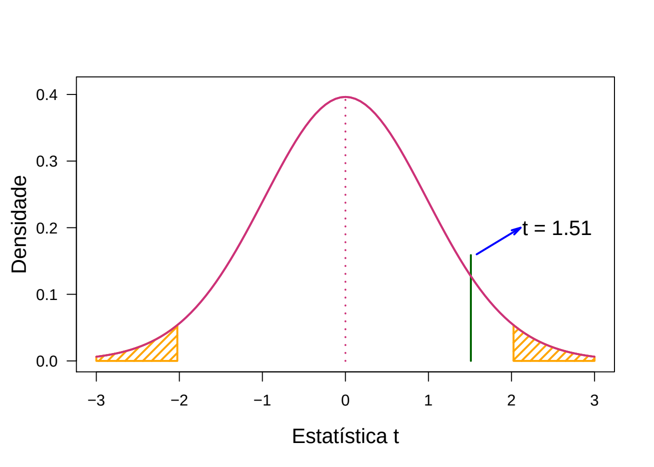
Exemplo de Inferência Clássica: Comparando as médias

Teste t: explicação gráfica



Exemplo de Inferência Clássica: Comparando as médias

Teste t: explicação gráfica (cont.)



Resultado do teste t

Não se rejeita H_0 $(\alpha = 0.05)$

Conclusão:

- Os DAP médios *não* são diferentes
- As florestas de encosta e de pequenos platôs são iguais
- *Nenhum deles!!*

Exemplo de Inferência Clássica: Conclusão (cont.)

Interpretação Apropriada do Resultado:

Não há evidência *suficiente* para se rejeitar H_0 no nível de probabilidade $\alpha = 0,05$.

Conclusão Apropriada:

Não se pode afirmar que dos DAP médios diferem estatisticamente.

Qual a confiabilidade da conclusão da inferência?

$\Rightarrow$ o nível de probabilidade de 5%?

$\Rightarrow$ o coeficiente de confiança de 95%?

$\Rightarrow$ o valor-p de 14% (valor-p=0.1392)

Exemplo de Inferência Clássica: Os Problemas

O *nível de probailidade* (α) é

$\Rightarrow$ margem aceitável de Erro Tipo I

$\Rightarrow$ erro de *se rejeitar* uma H_0 verdadeira

$\Rightarrow$ uma margem de erro de *longo prazo*

Nesse teste específico:

- O α não nos diz *nada*
- Não se rejeitou H_0 $\Rightarrow$ Erro Tipo II (*não se rejeitar* H_0 falsa)
  - $\Rightarrow$ Erro Tipo II: não é possível se calcular
- Segurança com que se conclui que não há diferença
  - $\Rightarrow$ *Completamente desconhecida!*

Inferência pelo Axioma da Verossimilhança

Inferência pelo Axioma da Verossimilhança

Dois Componentes:

- *Lei da Verossimilhança*
- *Princípio da Verossimilhança*

Lei da Verossimilhança

É de aceitação geral entre teóricos e práticos

Princípio da Verossimilhança

É o ponto de discórdia

Axioma da Verossimilhança: Lei

Situação:

- $X = x$ é uma observação de uma dada variável aleatória.
- Hipótese $A \Rightarrow p_A(x)$.
- Hipótese $B \Rightarrow p_B(x)$.

Resultado:

- A observação $X = x$ é evidência
- a favor de A *vis-à-vis* B se:

$$p_A(x) > p_B(x) \Longrightarrow R_L = \frac{p_A(x)}{p_B(x)} > 1$$

- A *Razão de Verossimilhança* R_L mede a *força* dessa evidência.

Lei de Verossimilhança

Duas Hipóteses: A e B:

- a evidência estatística é sempre **relativa**:
- A vis-à-vis B.
- teste sempre *condicionado* ao que foi observado (dados)

A Função de Verossimilhança:

- define a *plausibilidade* de cada hipótese
- plausibilidade é *relativa* (A vis-à-vis B)

Princípio de Verossimilhança: Situações

A Razão de Verossimilhança é Razão de Plausibilidade

- define a *força de evidência* em favor de uma das hipóteses
- *sempre* condicionadas aos dados (o que foi observado)

Primeira Situação

- Primeiro conjunto de dados: observações $X = x$,
- Duas hipóteses simples relativas a ela: $f_1(x)$ e $f_2(x)$.

Segunda Situação

- Segundo conjunto de dados: observações $Y = y$,
- Duas hipóteses simples relativas a ela: $g_1(y)$ e $g_2(y)$.

Princípio de Verossimilhança: Universalidade

Razão de Verossimilhança é universal

- Se a razão de verossimilhança for igual:

$$\frac{f_1(x)}{f_2(x)} = \frac{g_1(y)}{g_2(y)}$$

- então: a observação $X = x$ relativa a f_1 vis-à-vis f_2
- e a observação $Y = y$ relativa a g_1 vis-à-vis g_2
- são *EQUIVALENTES* em termos de evidência estatística.
- A razão de verossimilhança
  - tem *interpretação universal*
  - contem *toda informação* nos dados a respeito das hipóteses

Os Dados

- Provem de um *cenário estocástico*
- São a única *evidência empírica* que temos a respeito do fenômeno em estudo
- Foram observados segundo um *procedimento*
  - $\Rightarrow$ que tem *fundamentos teóricos*
  - $\Rightarrow$ segue certas *considerações práticas*

Filosofia da Inferência por Verossimilhança (cont.)

O Cenário Estocástico

- Gera os dados segundo um *Modelo Operante*
- Inferência baseada num *Modelo de Aproximação*
- *Melhor aproximação* ao cenário estocástico?
- É o modelo que *melhor explica* os dados

Filosofia da Inferência por Verossimilhança (cont.)

Procedimento de Observação e Inferência

- *Uma vez* que os dados foram observados
- O procedimento se torna *irrelevante* para a inferência
- O procedimento de observação
  - $\Rightarrow$ é *planejado segundo* o cenário estocástico em estudo
  - $\Rightarrow$ logo, foi *arbitrariamente* definido
  - $\Rightarrow$ por isso, *não é informativo* sobre o cenário

Sequência para Inferência por Verossimilhança

Cenário Estocástico, Hipóteses e Dados

1. *Abstração* do cenário estocástico

   $\Rightarrow$ a partir das hipóteses sobre o fenômeno em estudo
2. *Definição* do procedimento de observação

   $\Rightarrow$ método de coleta de dados
3. *Descrição* dos dados obtidos

   $\Rightarrow$ comparação e seleção dos modelos candidatos
4. *Explanação* da relação cenário-hipóteses-dados

   $\Rightarrow$ inferência baseada no modelo selecionado

Sequência para Inferência por Verossimilhança (cont.)

Passos para Inferência Baseada em Modelos

1. *Especificação* dos modelos de aproximação candidatos
2. *Estimação* dos modelos de aproximação
3. *Comparação* dos modelos de aproximação
4. *Seleção* do modelo mais apropriado
5. *Inferência* baseada no modelo selecionado

Sequência para Inferência por Verossimilhança (cont.)

Especificação dos modelos candidatos

(A) Dúvidas quanto ao componente estocástico que gera a variabilidade/diversidade observada

- Famílias de distribuição estocásticas *distintas*
- Primeira etapa: comparação e seleção de *famílias de aproximação*

(B) Dúvida *apenas* quanto ao componente determinínico:

- família operante é definida pela *especificação do modelo*
- comparar *formas funcionais* para o valor esperado/variância
- comparar *estimativas* para os parâmetros

⇒ Frequentemente *(A) e (B)*

- são *indissociáveis* no processo de inferência
- mas,. . . pensar como *questões distintas* ajuda na inferência

Exemplo de Inferência por Verossimilhança: comparando distribuições

Exemplo simples de Inferência por Verossimilhança

Cenário Estocástico e Hipóteses

- Floresta tropical no Maranhão
- Condições variáveis de topografia:
  - ⇒ encosta,
  - ⇒ pequenos platôs e
  - ⇒ grandes platôs
- As florestas em condições ditintas de topografia têm *estruturas* diferentes?

Exemplo simples de Inferência por Verossimilhança (cont.)

Observações: parcelas de $5000\ m^2$ em faixa ($10 \times 500\ m$)

	N. Árvores	N. Parcelas	Árvores por Parcela
Encosta	2981	18	165.6
Platô grande	4339	30	144.6
Platô pequeno	3143	22	142.9

Detalhamento das Hipóteses:

- A estrutura da floresta pode ser entendida como a *distribuição de tamanho* das árvores:
    - distribuição de tamanho $\Rightarrow$ *distribuição de diâmetros*
    - topografia diferentes $\Rightarrow$ *diferentes* distribuições de diâmetros

Exemplo de Inferência por Veros.: Dados

Cenário, Hipóstes e Dados

- Comparação de Duas Áreas: Encosta $\times$ Platô Pequeno
  - Dados:

Topografia	*Parcela*	*Árvores*	*Diâmetros (DAP - cm)*
Encosta	502	149	$14.9, 71.0, \ldots, 21.3$
18 parcelas	503	185	$22.0, 17.8, \ldots, 22.9$
2981 árvores	$\cdots$	$\cdots$	$\cdots$
	522	136	$24.8, 30.2, \ldots, 16.2$
Platô pequeno	102	179	$48.0, 14.3, \ldots, 24.2$
22 parcelas	103	166	$23.8, 23.2, \ldots, 17.2$
3143 árvores	$\cdots$	$\cdots$	$\cdots$
	414	170	$20.7, 21.6, \ldots, 29.9$

Exemplo de Inferência por Veros.: Modelos 1

Abordagem

- Ignorar que os DAP estão *agregados* nas parcelas
- *Modelo Único:* Uma mesma distribuição para duas topografias
- *Modelo por Topografia:* Duas distribuições distintas

Famílias Candidatas

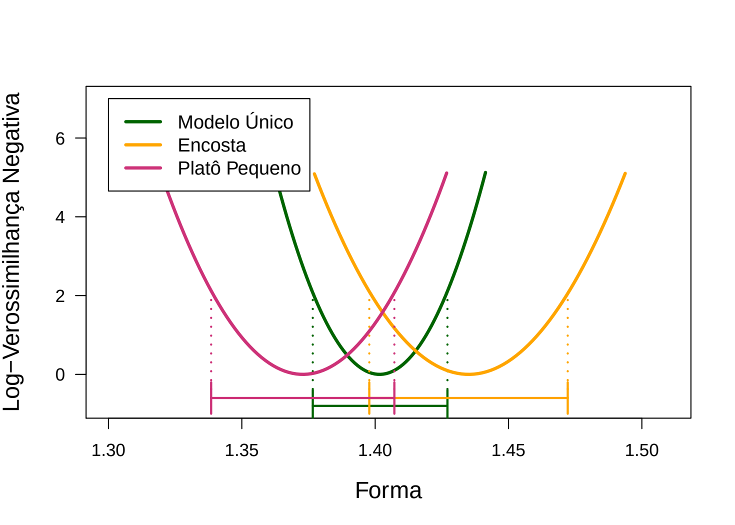
- Para simplificar: uma única família *Weibull*
- Justificativa: modelo de uso muito frequente na literatura

Exemplo de Inferência por Veros.: Resultados

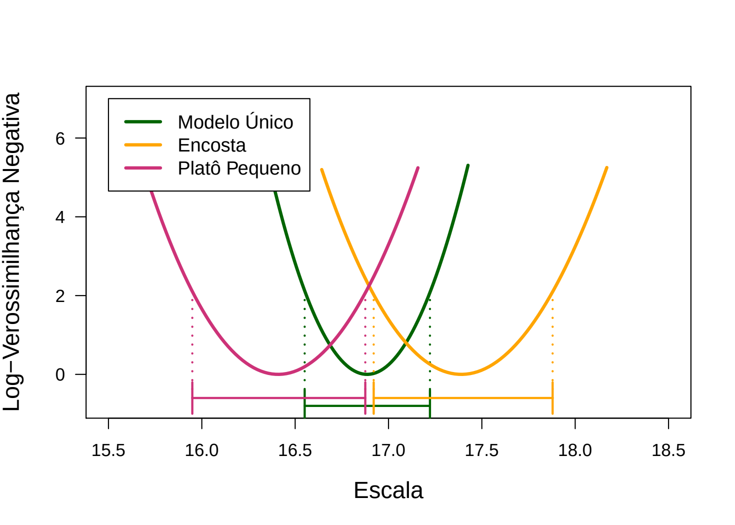
Modelo	*Parâm.*	*Estimativas*		*Log-Veros.*	*AIC*
		Escala	*Forma*		
Único	2	1.4017	16.8854	-22199.24	44402.48
		(0.0123)	(0.1637)		
por Topografia	4	—	—	-22193.49	44394.99
Encosta	2	1.4351	17.3904	-10851.22	21706.45
		(0.0183)	(0.2360)		
Peq. Platô	2	1.3729	16.4089	-11342.27	22688.54
		(0.01676)	(0.2268)		

- O Modelo por Topografia é melhor que o Modelo Único
- Razão de verossimilhança: $\exp(5.7464) \approx 313$ vezes.
- Diferença do AIC: 7.49

Exemplo de Inferência por Veros.: Weibull — Forma



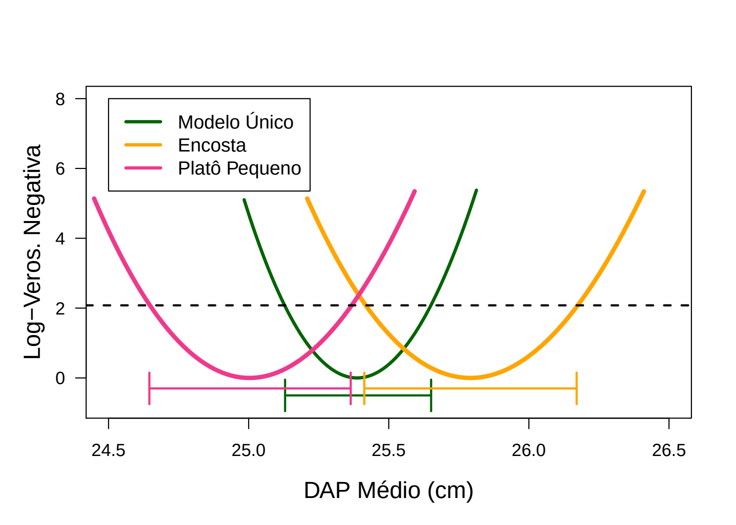
Exemplo de Inferência por Veros.: Weibull — Escala



Confronto: modelos × parâmetros

- Modelo por Topografia melhor que Modelo Único:
    - *ainda que* os intervalos de verossimilhança tenha *sobreposição*
- Sobreposição de intervalos de veros. *não implica* em
    - igualdade de estimativas
    - igualdade de parâmetros
- A *performace* de um modelo depende
    - da sua estrtura
    - das suas MLEs

Mas, e a diferença entre as médias?



Problema da Inferência Clássica: Influência do Procedimento

Dois Experimentos: Inferências Clássicas Distintas

Objetivo

Testar a eficácia de uma droga $\Rightarrow$ probabilidade de cura.

Controle

A probabilidade de morte na ausência de tratamento é de 50%.

Hipóteses

- Parâmetro — Probabilidade de morte: p
- Teste: Efeito $\times$ Ausência de Efeito
- $H_0 : p \geq 0,5$
- $H_\alpha : p < 0,5$

Dois Experimentos: Laboratório Rico

Três de Doze Morreram

- Fez o teste em 12 cobaias e contabilizou as mortes
- Dado: de 12 cobaias (n=12), 3 morreram (x=3)
- Modelo: Distribuição *Binomial*

Experimento Binomial

- Estatística: x = número de mortes em 12 cobaias
- $H_\alpha \to x$ pequeno

$$\begin{aligned} P(X \leq 2 | n = 12; p = 0,5) &= 0.0193 \\ P(X \leq 3 | n = 12; p = 0,5) &= 0.0730 \end{aligned}$$

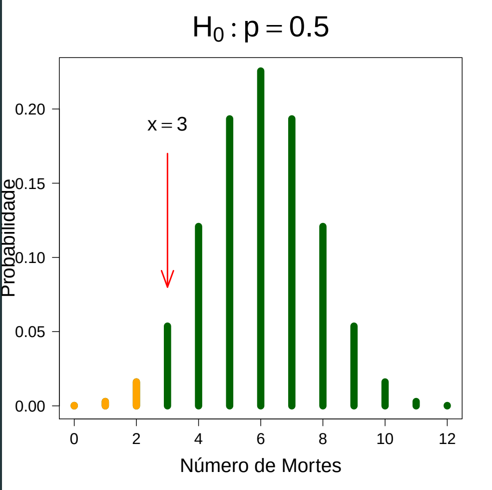
- Região de rejeição para $\alpha = 0,05$: $\{0, 1, 2\}$

Dois Experimentos: Laboratório Rico (cont.)

Conclusão no Lab. Rico

- Estatística observada: $x = 3 \Rightarrow$ fora da região de rejeição
- Sob H_α, algo mais extremo é $x > 3$.
- p-valor $= P(X \geq 3 | n = 12; p = 0,5) = 0.0730$
- Decisão: *não rejeita* H_0
- Conclusão: *ausência* de efeito da droga

Dois Experimentos: Laboratório Rico (cont.)



Dois Experimentos: Laboratório Pobre

Três de Doze Morreram

- Testou uma cobaia por vez, até contabilizar 3 mortes
- Dado: a terceira morte (n=3) ocorreu após 9 curas (x=9)
- Modelo: distribuição *Binomial Negativa*

Experimento Binomial Negativo

- Estatística: $x =$ número de curas antes de 3 mortes
- $H_\alpha \to x$ grande

$$
\begin{aligned}
P(X \geq 7 | n = 3; p = 0,5) &= 0.0547 \\
P(X \geq 8 | n = 3; p = 0,5) &= 0.0327
\end{aligned}
$$

- Região de rejeição para $\alpha = 0,05$: $[8, +\infty)$

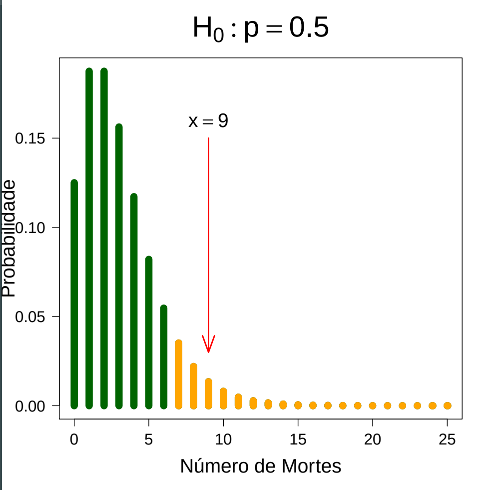
Dois Experimentos: Laboratório Pobre (cont.)

Lab. Pobre: Conclusão

- Estatística observada: $x = 9 \Rightarrow$ *dentro* da região de rejeição
- Sob H_α, algo mais extremo é $x > 9$.
- p-valor $= P(X \geq 9 | n = 3; p = 0,5) = 0.0193$
- Decisão: *se rejeita* H_0
- Conclusão: *presença* de efeito da droga

Dois Experimentos: Laboratório Pobre (cont.)

Conclusão no Lab. Pobre



Dois Experimentos: Lab. Rico × Lab. Pobre

Inferência Clássica: conclusões diferentes

- Lab. Rico não rejeita H_0
- Lab. Pobre rejeita H_0
- Procedimentos diferentes geraram:
  - distribuições amostrais diferentes
  - regiões de rejeição diferentes
  - conclusões diferentes
  - *a partir do mesmo resultado*: três mortes em doze cobaias

Dois Experimentos: Lab. Rico $\times$ Lab. Pobre (cont.)

Experimentos diferentes:

- Os experimentos são diferentes
- As estatísticas são diferentes
- As distribuições amostrais são diferentes
- As regiões de rejeição são diferentes
- *Espaços Amostrais Distintos*
- Poderia ter acontecido resultados diferentes
- Mas, *não aconteceram*

Dois Experimentos: Lab. Rico $\times$ Lab. Pobre (cont.)

Resultados iguais:

- O resultado é o mesmo
- O *procedimento experimental* é relevante?
- A *evidência* deste resultado a respeito destas hipóteses

  muda com o procedimento?
- O *espaço amostral* deve influenciar o resultado?
- Por que o que *poderia acontecer*

  (mas não aconteceu)

  deve influenciar o resultado?

Dois Experimentos com a mesma Evidência

Redefinindo as Hipóteses

- Parâmetro de interesse: p probabilidade de morte
- Valor observado para o parâmetro:

$$\widehat{p}_{\text{MLE}} = \frac{3}{12} = 0.25$$

- Hipóteses:

$$\begin{aligned} H_A : &\qquad p = 0.5 \\ H_B : &\qquad p = \widehat{p}_{\text{MLE}} = 0.25 \end{aligned}$$

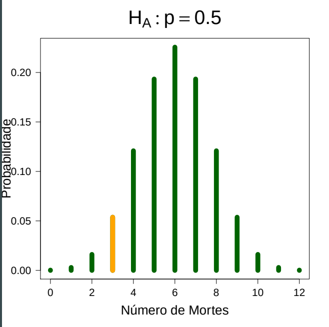
- Modelos:
  - Lab. Rico: Dist. *Binomial*: $n = 12$ e $p = ?$
  - Lab. Probre: Dist. *Binomial Negativa*: $n = 3$ e $p = ?$

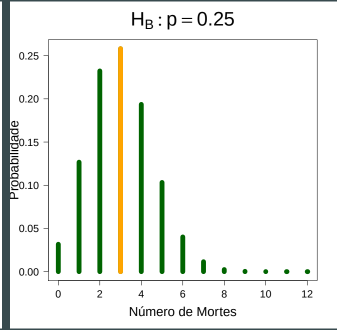
Dois Experimentos com a mesma Evidência (cont.)

Laboratório Rico

$$\frac{P_{\text{Binomial}}(H_B|x=3)}{P_{\text{Binomial}}(H_A|x=3)} = \frac{0.2581}{0.0537} = \mathit{4{,}8}$$

H_B é 4,8 mais plausível que H_A.



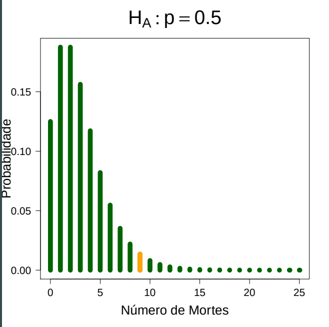


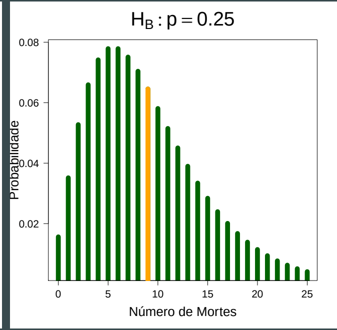
Dois Experimentos com a mesma Evidência (cont.)

Laboratório Pobre

$$\frac{P_{\text{Binomial Negativa}}(H_B|x=9)}{P_{\text{Binomial Negativa}}(H_A|x=9)} = \frac{0.0645}{0.0134} = \textit{4,8}$$

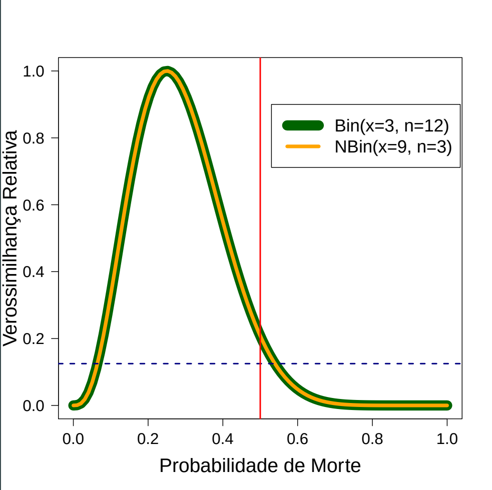
H_B é 4,8 mais plausível que H_A.





Dois Experimentos com a mesma Evidência (cont.)

Os dois experimentos produzem ***evidências equivalentes***



Dois Experimentos: Princípio da Verossimilhança

Por que não há diferenças?

- A função de verossimilhança caracteriza
  - $\Rightarrow$ *toda evidência* contida nos dados a respeito de uma hipótese
  - $\Rightarrow$ curvas idênticas implicam em *evidências equivalentes*
- Modelos diferentes podem
  - $\Rightarrow$ gerar curvas de verossimilhança idênticas
  - $\Rightarrow$ para um mesmo conjunto de dados.
- *Como* os dados foram *gerados*
  - $\Rightarrow$ se torna irrelevante em termos de evidência estatística
  - $\Rightarrow$ não influencia os resultados obtidos

Resumo da Aula

Resumindo: Inferência Clássica

Fundamento: Dados são uma amostra de infinitas possíveis

População: tem parâmetros

Amostra: tem estatísticas

Inferência: distribuição amostral da estatística sob H_0

Espaço Amostral: procedimento define a *"população"*

Resultados obtidos

- Estimativas
- Intervalos de Confiança
- Teste de hipótese

  $\Rightarrow$ decisão: rejeitar $\times$ não rejeitar
- Função ou transformação das estimativas?

Resumindo: Inferência pelo Axioma da Verossimilhança

Fundamento: fenômeno estudado $\Rightarrow$ cenário estocástico

Modelo Operante: modelo por trás do cenário

$\Rightarrow$ não há *"população"*

Amostra: observações do modelo operante

Inferência: em duas etapas

$\Rightarrow$ seleção do modelo de aproximação apropriado

$\Rightarrow$ inferência com base no modelo de aproximação

Resultados obtidos

- Modelo de aproximação ajustado
- Estimativas (MLEs)
- Curvas e Intervalos de Verossimilhança
- Qualquer função ou transformação das MLEs

Resumindo: Clássica $\times$ Verossimilhança

	Inferência	
Característica	*Clássica*	*Verossimilhança*
Fundamento	população	cenário estocástico
Geração dos Dados	amostra	modelo operante
Inferência	por amostragem	por modelagem
Estimação	várias técnicas	máxima verossimilhança
Hipóteses	$H_0 \times H_\alpha$	muitas; comparação 2-a-2
Base de Teste	distribuição amostral	razão de verossimilhança
Pressuposições	geralm. restritivas	geralm. simples
Modelo Estatístico	geralm. implícito	sempre explícito
Procedimento	espaço amostral	irrelevante
Análise Padronizada	fácil	difícil
Amostra Pequena	por desenvolvimento	problemática
Medida de Evidência	não há	razão de verossimilhança
Compara Modelos	somente aninhados	sem restrições

Grato pela Atenção!