

8. Fundamentos Teóricos da Inferência por Verossimilhança

Conceitos

- Lei da Verossimilhança
- Princípio da Verossimilhança
- Suporte para Inferência Estatística

Tutorial

Lei da Verossimilhança

Como já foi visto no [tutorial](#) sobre a Função de Verossimilhança, a Lei da Verossimilhança pode ser enunciada da seguinte forma:

Dada uma variável aleatória X , cujo comportamento pode ser explicado por duas hipóteses: H_0 e H_1 .

- A hipótese H_0 afirma que a observação x seria observada com probabilidade $f_0(x)$.
- A hipótese H_1 afirma que a observação x seria observada com probabilidade $f_1(x)$.

A observação x é uma evidência em favor de H_0 **vis-a-vis** (*face-a-face*) H_1 se, e somente se,

$f_0(x) > f_1(x)$.

A **força de evidência** em favor de H_0 vis-a-vis H_1 é dada pela **razão de verossimilhança**:

$$\frac{f_0(x)}{f_1(x)}$$

A Observação Empírica comanda a Lei da Verossimilhança

Hipóteses sobre Valores do Parâmetro de um Modelo

Tomemos o exemplo de um laboratório que realizou o seguinte experimento: um certo produto foi ministrado a n cobaias para se comparar duas hipóteses:

- H_0 a probabilidade do produto causar a morte é p_0

- a probabilidade do produto causar a morte é

Um ponto importante é que a observação do número de cobaias mortas é que irá definir qual hipótese é favorecida e qual hipótese é desfavorecida.

Vejam as probabilidades que a hipótese estabelece para cada uma das observações possíveis (1, 2, ..., 20):

```
pa = dbinom(0:20, 20, prob=0.5)
barplot(pa, width=1, space=0.1, col="blue")
axis(1, 1, label=0:20, at=0:20*1.1+(1.1/2) )
```

No caso da hipótese temos:

```
pb = dbinom(0:20, 20, prob=0.3)
barplot(pb, width=1, space=0.1, col="red")
axis(1, 1, label=0:20, at=0:20*1.1+(1.1/2) )
```

A Razão de Verossimilhança para as observações possíveis pode ser facilmente obtida:

```
raz <- pa/pb
barplot(raz, width=1, space=0.1, col="darkgreen")
axis(1, 1, label=0:20, at=0:20*1.1+(1.1/2) )
```

A escala da Razão de Verossimilhança pode facilmente nos confundir. É melhor olharmos a Razão de Verossimilhança em termos da diferença da Log-Verossimilhança Negativa (com a indicação da Razão de 8, como limite definidor):

```
barplot(-log(raz), width=1, space=0.1, col="darkgreen")
axis(1, 1, label=0:20, at=0:20*1.1+(1.1/2) )
abline( h = c(log(8), -log(8)), col="red", lty=2 )
```

Como a transformação inclui a mudança de sinal, a interpretação é que os valores positivos favorecem a hipótese, enquanto que os valores negativos favorecem a hipótese.

Resultado:

- O número de cobaias mortas no experimento definirá qual das duas hipóteses é mais plausível.
- Algumas observações favorecerão (11 ou mais mortes), outras (6 ou menos). Outras ainda não nos permitirão distinguir qual das duas é mais plausível (entre 7 e 10).

CONCLUSÕES:

1. Definido o modelo de trabalho, **os dados** são a única evidência para definir qual a hipótese é mais plausível.
2. A evidência nem sempre é **conclusiva**.

Hipóteses sobre Modelos Diferentes

Os dados também indicam, através da verossimilhança, qual o modelo mais plausível numa situação

de comparação. O arquivo [caxeta-3parcelas.csv](#) tem os dados de DAP (diâmetro à altura do peito) de árvores em três parcelas instaladas em “caixetais” na região do Vale do Ribeira (São Paulo).

Comparemos a distribuição Weibull e a distribuição Gama, como modelos para a distribuição do DAP em cada uma das parcelas.

Primeiramente ler os dados e carregar o pacote “MASS” que possui a função “fitdistr” para ajuste de distribuições por máxima verossimilhança:

```
cax3p = read.csv("caxeta-3parcelas.csv",header=T,as.is=T,sep=";")
library(MASS)
```

O segundo passo é ajustar os modelos (Weibull e Gama) para cada parcela.

```
# Ajuste da Dist. Weibull para as três parcelas
weib1 = fitdistr( cax3p$dap[ cax3p$parcela==1 ] - 47, "weibull",
start=list(shape=1, scale=10) )
weib2 = fitdistr( cax3p$dap[ cax3p$parcela==2 ] - 47, "weibull",
start=list(shape=1, scale=10) )
weib3 = fitdistr( cax3p$dap[ cax3p$parcela==3 ] - 47, "weibull",
start=list(shape=1, scale=10) )

# Ajuste da Dist. Gamma para as três parcelas
gamm1 = fitdistr( cax3p$dap[ cax3p$parcela==1 ] - 47, "gamma",
start=list(shape=1, scale=10) )
gamm2 = fitdistr( cax3p$dap[ cax3p$parcela==2 ] - 47, "gamma",
start=list(shape=1, scale=10) )
gamm3 = fitdistr( cax3p$dap[ cax3p$parcela==3 ] - 47, "gamma",
start=list(shape=1, scale=10) )
```

Comparação dos modelos nas parcelas uma a uma:

```
# Comparação Parcela 1
hist( cax3p$dap[ cax3p$parcela==1 ], prob = TRUE )
curve( dweibull(x, shape=weib1$estimate["shape"] ,
scale=weib1$estimate["scale"]), 40, 250, col="red", add=TRUE)
curve( dgamma(x, shape=gamm1$estimate["shape"] ,
scale=gamm1$estimate["scale"]), 40, 250, col="blue", add=TRUE)
AIC(weib1) - AIC(gamm1)

# Comparação Parcela 2
hist( cax3p$dap[ cax3p$parcela==2 ], prob = TRUE )
curve( dweibull(x, shape=weib2$estimate["shape"] ,
scale=weib2$estimate["scale"]), 40, 400, col="red", add=TRUE)
curve( dgamma(x, shape=gamm2$estimate["shape"] ,
scale=gamm2$estimate["scale"]), 40, 400, col="blue", add=TRUE)
AIC(weib2) - AIC(gamm2)

# Comparação Parcela 3
hist( cax3p$dap[ cax3p$parcela==3 ], prob = TRUE )
curve( dweibull(x, shape=weib3$estimate["shape"] ,
```

```
scale=weib3$estimate["scale"]), 40, 400, col="red", add=TRUE)
curve( dgamma(x, shape=gamm3$estimate["shape"] ,
scale=gamm3$estimate["scale"]), 40, 400, col="blue", add=TRUE)
AIC(weib3) - AIC(gamm3)
```

Questões:

- Qual o modelo mais plausível em cada parcela?
- O modelo mais plausível é sempre o mesmo em todas as parcelas?
- É possível discriminar o modelo mais plausível em todas as parcelas?
- A diferença de plausibilidade entre os modelos segundo o AIC é compatível com as diferenças observadas nos gráficos?

Princípio da Verossimilhança

O **Princípio da Verossimilhança** estabelece que a **Função de Verossimilhança** contém **TODA** evidência contida nos dados a respeito de uma dada hipótese. Esta hipótese pode ser um modelo, ou o valor de parâmetro de um modelo. Assim, a **Razão de Verossimilhança** é uma medida absoluta da força de evidência na comparação de duas hipóteses, independentemente do conjunto de dados utilizado ou das hipóteses comparadas.

Dois Métodos com a Mesma Evidência

Voltemos ao exemplo da aplicação de um produto em cobaias para verificar a taxa de mortalidade. Nesse caso temos dois laboratórios que testaram o produto:

- **Laboratório 1:** Aplicou o produto em 20 cobaias das quais 6 morreram.
- **Laboratório 2:** Foi aplicando o produto em várias cobaias, com a determinação que quando a sexta morte ocorresse o experimento terminaria. A vigésima cobaia a receber o produto foi a sexta a morrer.

A questão principal agora é saber qual o valor mais plausível para o parâmetro p , que indica a probabilidade de morte das cobaias.

Vejamos as curvas de verossimilhança para nos dois laboratórios:

```
p = seq(0.01, 0.99, by=0.01)
lik.binom = dbinom(6, 20, p)      # Lab 1: dist. Binomial
lik.nbinom = dnbinom(14, 6, p)    # Lab 2: dist. Binomial Negativa
plot(p, lik.binom, type="l", ylab="Verossimilhança", col="blue")
lines(p, lik.nbinom, col="red")
```

Aparentemente as curvas não são as mesmas. Mas devemos nos lembrar que o **Princípio da Verossimilhança** generaliza a **Razão de Verossimilhança**. Portanto devemos apresentar a curva de **Verossimilhança Relativa**¹⁾, que obtemos dividindo cada valor de verossimilhança pelo máximo:

```
lik.binom = lik.binom / max(lik.binom)
lik.nbinom = lik.nbinom / max(lik.nbinom)
```

```
plot(p, lik.binom, type="l", ylab="Verossimilhança Relativa", col="blue",
lwd=8)
lines(p, lik.nbinom, col="red", lwd=2)
```

CONCLUSÕES:

1. As curvas de verossimilhança (relativa/normalizada) são idênticas nos dois laboratórios, mostrando que a observação de 6 cobaias mortas em 20 representa **exatamente** a mesma evidência empírica a respeito do valor mais plausível do parâmetro θ , independentemente de como os dados foram gerados.
2. Portanto, **o espaço amostral é irrelevante**, uma vez definidos os modelos concorrentes e observado o dado.
3. Curvas de **Verossimilhança Relativa** (ou *Verossimilhança Normalizada*) representam a força de evidência a favor das MLE (estimativas de máxima verossimilhança) contra todos os demais valores possíveis, independentemente dos dados utilizados ou do modelo considerado.
4. Consequentemente, **Intervalos de Verossimilhança** (para razão de 8, por exemplo) tem exatamente a mesma interpretação, independentemente dos dados utilizados ou do modelo considerado.

Força de Evidência e Tamanho de Amostra

Consideremos o mesmo exemplo das cobaias, mas comparemos o primeiro laboratório com outros dois laboratórios que possuem mais recursos para o experimento:

- Laboratório 1: Aplicou o produto em 20 cobaias das quais 6 morreram.
- Laboratório 2: Aplicou o produto em 200 cobaias das quais 60 morreram.
- Laboratório 3: Aplicou o produto em 2000 cobaias das quais 600 morreram.

Vejamos as curvas de verossimilhança desses 3 laboratórios:

```
p = seq(0.01, 0.99, by=0.01)
lik.binom1 = dbinom(6, 20, p)           # Lab 1: dist. Binomial
lik.binom2 = dbinom(60, 200, p)        # Lab 2: dist. Binomial
lik.binom3 = dbinom(600, 2000, p)      # Lab 3: dist. Binomial
plot(p, lik.binom1, type="l", ylab="Verossimilhança", col="blue")
lines(p, lik.binom2, col="red")
lines(p, lik.binom3, col="green")
```

Vejamos as curvas de verossimilhança **RELATIVA** desses 3 laboratórios:

```
lik.binom1 = lik.binom1/ max(lik.binom1)
lik.binom2 = lik.binom2/ max(lik.binom2)
lik.binom3 = lik.binom3/ max(lik.binom3)
plot(p, lik.binom1, type="l", ylab="Verossimilhança Relativa", col="blue")
lines(p, lik.binom2, col="red")
lines(p, lik.binom3, col="green")
```

Façamos um “zoom” na curva de log-verossimilhança negativa relativa na vizinhança da MLE:

```
nlik.binom1 = -log(lik.binom1)
```

```
nlik.binom2 = -log(lik.binom2)
nlik.binom3 = -log(lik.binom3)
plot(p, nlik.binom1, type="l", ylab="Log-Veros. Neg. Relativa", col="blue",
xlim=c(0.2,0.4))
lines(p, nlik.binom2, col="red")
lines(p, nlik.binom3, col="green")
```

Questões:

- A **curva de verossimilhança** é sensível ao tamanho da amostra? Como?
- A **curva de verossimilhança RELATIVA** é sensível ao tamanho da amostra? Como?
- A **força de evidência** em favor do MLE aumenta com o tamanho da amostra? Por que?
- Qual o impacto do tamanho da amostra sobre o **intervalo de verossimilhança**? Por que?

Suporte para Inferência Estatística

A consequência **imediate** da combinação da Lei e do Princípio da Verossimilhança é que a função de verossimilhança, ou mais especificamente a *função de log-verossimilhança negativa* é o **suporte necessário e suficiente** à construção da inferência estatística:

- Por **suporte** entende-se a base **teórica** e **empírica** para se construir e implementar a inferência estatística.
- Como suporte **necessário** entende-se que qualquer inferência não baseada nesse suporte não é apropriada.
- como suporte **suficiente** entende-se que nada mais é necessário à inferência estatística além desse suporte.

Suporte para Inferência sobre Parâmetros

Partindo de um modelo assumido como apropriado, qualquer inferência sobre os parâmetros do modelo, ou *funções desses parâmetros*²⁾ pode ser realizada tendo a função de log-verossimilhança negativa como suporte.

Voltemos ao exemplo da distribuição de DAP no caxetal (parcela 2):

```
hist( cax3p$dap[ cax3p$parcela==2 ], prob = TRUE )
curve(dweibull(x, shape=weib2$estimate["shape"],
scale=weib2$estimate["scale"]), 40, 400, col="red", add=TRUE)
```

Inferência sobre os Parâmetros

Vejamos a superfície de log-verossimilhança negativa relativa para inferência sobre os parâmetros:

- Criando a função vetorizada:

```
lweibull = function(forma, escala) -sum(dweibull( cax3p$dap[
cax3p$parcela==2 ]-47, shape=forma, scale=escala, log=TRUE))
```

```
vlweibull = Vectorize( lweibull, c("forma","escala") )
```

- Definido a amplitude de variação dos parâmetros:

```
forma = seq(0.5, 2.5, by=0.05)
escala = seq( 50, 100, by=0.5 )
```

- Calculando a superfície de log-veros. neg. relativa:

```
sup.weibull = outer( forma, escala, vlweibull )
sup.weibull = sup.weibull - min(sup.weibull)
```

- Construindo o gráfico de contorno da superfície:

```
contour(forma, escala, sup.weibull, xlab="Forma", ylab="Escala",
col="purple", levels=c(5,10,20,40,80,120))
```

- Marcando a posição das MLE com linhas tracejadas:

```
abline(v=weib2$estimate[1], col="red", lty=2)
abline(h=weib2$estimate[2], col="red", lty=2)
```

- Marcando a região referente à razão de verossimilhança de 8:

```
contour(forma, escala, sup.weibull, levels=log(8), labels="log(8)", add=T,
col="darkgreen" )
```

Inferência sobre a Média

Na distribuição Weibull a média (valor esperado) é definido em função dos parâmetros da seguinte forma:



. onde:

- é o parâmetro de escala;
- é o parâmetro da forma; e
- é a função gama.

Assim podemos construir uma superfície para inferência sobre a Média:

- Cálculo da superfície dos valores da média:

```
mean.weibull = function(c, b) (b*gamma( (c+1)/c )+47)/10
sup.mean = outer(forma, escala, mean.weibull)
```

- Gráfico da superfície da média, com a posição das MLE dos parâmetros:

```
contour(forma, escala, sup.mean, col="darkgreen", xlab="Forma",
ylab="Escala", levels=seq(8,18,by=1))
abline(v=weib2$estimate[1], col="red", lty=2)
```

```
abline(h=weib2$estimate[2], col="red", lty=2)
```

- Região de razão de verossimilhança (8) e linha da média amostral:




```
contour(forma, escala, sup.weibull, levels=log(8), labels="log(8)", add=T,
col="orange" )
media.estimada = (weib2$estimate[2] * gamma(1 + (1/weib2$estimate[1])) +
47)/10
contour(forma, escala, sup.mean, levels=media.estimada, col="red", add=T)
contour(forma, escala, sup.mean, levels=mean(dap)/10, col="blue", add=T)
```

Inferência sobre Quantis da Distribuição

Na distribuição Weibull os quantis podem ser determinados a partir da função inversa da função de distribuição:



, onde:

-  é a probabilidade que se deseja o quantil, por exemplo 0.95 (95%);
-  e  parâmetros de escala e forma, respectivamente.

Para construir a superfície para inferência sobre o quantil 95% basta seguir os mesmos passos da construção da superfície sobre a média:

```
dap95.weibull = function(c, b) (b*( log(1/(1-0.95)) )^(1/c) + 47)/10
sup.dap95 = outer(forma, escala, dap95.weibull)
contour(forma, escala, sup.dap95, col="darkseagreen",
levels=c(seq(15,25,by=1),seq(30,60,by=10)))
abline(v=weib2$estimate[1], col="red", lty=2)
abline(h=weib2$estimate[2], col="red", lty=2)
contour(forma, escala, sup.weibull, levels=log(8), labels="log(8)", add=T,
col="orange" )
dap95.est = dap95.weibull( weib2$estimate[1], weib2$estimate[2])
contour(forma, escala, sup.dap95, levels=dap95.est,
labels=round(dap95.est,1), add=T, col="red" )
dap95.amostr = quantile(cax3p$dap[ cax3p$parcela==2 ]/10, 0.95, type=6)
contour(forma, escala, sup.dap95, levels=dap95.amostr,
labels=round(dap95.amostr,1), add=T, col="blue" )
```

Questões

- O que representa a superfície de log-verossimilhança negativa relativa dos parâmetros?
- O que se pode inferir (estatisticamente) a partir dela?
- O que representa a superfície de valores da média a partir dos parâmetros?
- O que se pode inferir (estatisticamente) dessa superfície?
- O que representa a superfície de valores do quantil 95% a partir dos parâmetros?
- O que se pode inferir (estatisticamente) dessa superfície?

Suporte para Inferência sobre Modelos

A inferência sobre modelos consiste na comparação dos modelos dois-a-dois através da razão de verossimilhança.

É comum se utilizar o “*Akaike Information Criterion*” (AIC) como forma de comparação de modelos, sendo que além de considerar a verossimilhança do modelo ele penaliza o número de parâmetros no modelo.

Elementos que tornam essa abordagem mais simples para inferência sobre modelos quando comparada à abordagem “*frequentista*” ou “*bayesiana*” são:

- Não há restrições a respeito do número de modelos ou como eles são formulados (com ou sem inspeção dos dados).
- A log-verossimilhança é *aditiva*, portanto, os modelos podem ser comparados para a amostra como um todo ou para sub-conjuntos dessa.
- Para comparação entre modelos é irrelevante quais (ou quantas) variáveis foram utilizadas como variáveis preditoras/explicativas ou como elas são incorporadas ao modelo.

Comparando Modelos nos Dados Sub-divididos ou Agregados

Voltemos ao exemplo das 3 parcelas em caxetais. Foram ajustados as distribuições Weibull e gama **para cada uma das parcelas**. Assim, podemos usar o AIC para uma comparação **parcela-a-parcela**

```
library(bbmle)
AICtab(weib1, gamm1, base=TRUE, logLik=TRUE)
AICtab(weib2, gamm2, base=TRUE, logLik=TRUE)
AICtab(weib3, gamm3, base=TRUE, logLik=TRUE)
```

Podemos pensar no conjunto dos três ajustes como um só modelo para as três parcelas, com seis parâmetros. Como a log-verossimilhança é aditiva, o AIC para este modelo combinado é a soma do AICs dos modelos componentes:

```
AIC(weib1) + AIC(weib2) + AIC(weib3)
AIC(gamm1) + AIC(gamm2) + AIC(gamm3)
```

Um modelo mais parcimonioso para as três parcelas seria ajustar um só modelo para os **dados agregados**.

```
weib.agr = fitdistr( cax3p$dap - 47, "weibull", start=list(shape=1,
scale=10) )
gamm.agr = fitdistr( cax3p$dap - 47, "gamma", start=list(shape=1, scale=10)
)
AICtab(weib.agr, gamm.agr, base=TRUE, logLik=TRUE)
AIC(weib1) + AIC(weib2) + AIC(weib3)
AIC(gamm1) + AIC(gamm2) + AIC(gamm3)
```

Questões:

- Quais as diferenças na comparação dos modelos nos níveis:
 1. parcela-a-parcela,
 2. combinado,
 3. agregado?
- A fundamentação teórica muda ao se realizar comparações nos diferentes níveis?
- Com os resultados obtidos é possível testar se a melhor abordagem de modelagem é ter um modelo para cada parcela ou ter um modelo para os dados agregados? Como?

Inferência por Verossimilhança e Inferência Frequentista

Inferência de Intervalo

Na abordagem *frequentista* a inferência de intervalo é realizada através do **Intervalo de Confiança**.

O intervalo de confiança apela para o conceito de probabilidade **a longo prazo** que implica na repetição ilimitada do procedimento utilizado para gerar os dados, como se os dados fossem uma amostra de uma população **infinita** de observações possíveis. Dessa forma a construção do intervalo de confiança segue os seguintes passos:

1. definir o **parâmetro de interesse**,
2. encontrar uma **estatística** que pode ser um estimador do parâmetro ou uma transformação do estimador;
3. definir a **distribuição amostral** dessa estatística, isto é, qual seria a distribuição da estatística se fosse possível repetir a amostra infinitas vezes;
4. construir um intervalo para a estatística com base nessa distribuição amostral;
5. converter esse intervalo de volta à escala do parâmetro de interesse.

Exemplo de Árvores Doentes em Floresta Plantada

Considere que numa plantação de *Eucalyptus grandis* foram amostradas aleatoriamente 100 árvores, sendo que dessas 37 se mostraram infectadas por uma certa doença. O objetivo é estimar a taxa de ocorrência da doença com estimativa de intervalo de confiança.

Pela distribuição binomial a MLE da taxa de ocorrência é:

.

e o erro padrão dessa estimativa é:

.

Utilizando a *aproximação normal* para grandes amostras () sabemos que a estatística

tem distribuição amostral igual à distribuição Normal padronizada (média zero e desvio padrão um).

Assim, um intervalo com probabilidade 95% para essa estatística é:



Assim o intervalo de confiança de 95% para estimativa da taxa de ocorrência de doença é:



Intervalo de Verossimilhança

O intervalo de verossimilhança (para razão 8, por exemplo) é obtido inspecionando a vizinhança da MLE na curva de verossimilhança:

```
p = seq(0.20, 0.50, length=100)
lik = dbinom(37, 100, p)
lik = lik / max(lik)
plot(p, lik, type="l", col="red")
abline(h=1/8, lty=2, col="blue")
abline(v=37/100, lty=9, col="red")
```

Segundo Exemplo de Árvores Doentes

Suponha agora que a amostra aleatória de árvores de 100 árvores foi obtida, mas nenhuma das árvores se mostrou doente. Isso indica que a ocorrência de doença é rara e, conseqüentemente, a taxa de infestação é muito pequena, próxima de zero.

Estimativa da taxa:

Erro padrão da estimativa:

Como utilizar a aproximação normal nesse caso? Não é possível obter um intervalo de confiança de 95% por essa abordagem. Uma nova abordagem, com outra estatística e outra distribuição amostral, deverá ser concebida.

O que muda no intervalo de verossimilhança? Absolutamente nada:

```
p = seq(0.0, 0.05, length=100)
lik = dbinom(0, 100, p)
lik = lik / max(lik)
plot(p, lik, type="l", col="red")
abline(h=1/8, lty=2, col="blue")
abline(v=0, lty=9, col="red")
```

Teste de Hipótese

A forma de teste de hipótese de uso mais geral na estatística frequentista é o o **teste de significância**.

Essa abordagem consiste em enunciar duas hipóteses:

- Hipótese nula: que estabelece um valor específico para o parâmetro sendo testado.
- Hipótese alternativa: que deve ser *complementar* à hipótese nula.

O teste de significância segue os seguintes passos:

- Define-se uma estatística e se deduz a distribuição amostral dessa estatística **sob a hipótese nula**, isto é, assumindo a hipótese nula como verdadeira.
- Com esta distribuição calcula-se, então, o **valor-p** que é a probabilidade de se observar o valor observado da estatística **ou um valor mais extremo** sob a hipótese nula.
- Compara-se o valor-p com o **nível de significância** previamente definido.
- Se o valor-p for menor que o nível de significância, rejeita-se a hipótese nula em favor da hipótese alternativa.

Exemplo dos Dois Laboratórios

Voltemos ao exemplo da aplicação de um produto em cobaias para verificar a taxa de mortalidade com os dois laboratórios:

- Laboratório 1: Aplicou o produto em 20 cobaias das quais 6 morreram.
- Laboratório 2: Foi aplicando o produto em várias cobaias, com a determinação que quando a sexta morte ocorresse o experimento terminaria. A vigésima cobaia a receber o produto foi a sexta a morrer.

A questão agora é testar as seguintes hipóteses:

- Hipótese Nula: .
- Hipótese Alternativa: .

Laboratório A: o modelo deste experimento é uma distribuição binomial. A probabilidade de obter seis **ou menos** mortes em 20 tentativas sob a hipótese de que é dada pela probabilidade acumulada da binomial:

```
pbinom(q=6, size=20, prob=0.5)
```

Laboratório B: o modelo do experimento é uma distribuição binomial negativa. A probabilidade de obter seis mortes em 20 **ou mais** tentativas é:

```
1 - pnbinom(q=14, size=6, prob=0.5)
```

Conclusão: Tomando-se o **nível de significância de 5% (0.05)**, o laboratório A não rejeitaria a hipótese nula, mas o laboratório B a rejeitaria.

Mesmo nível de significância e mesmos dados, mas conclusões diferentes.

Na inferência por verossimilhança, como vimos, os dois laboratório tem exatamente a mesma curva de verossimilhança e, portanto, a evidência estatística também:

```
p = seq(0.01, 0.99, by=0.01)
lik.binom = dbinom(6, 20, p)      # Lab 1: dist. Binomial
lik.binom = lik.binom / max(lik.binom)
lik.nbinom = dnbinom(14, 6, p)    # Lab 2: dist. Binomial Negativa
lik.nbinom = lik.nbinom / max(lik.nbinom)
plot(p, lik.binom, type="l", ylab="Verossimilhança Relativa", col="blue",
     lwd=8)
lines(p, lik.nbinom, col="red", lwd=2)
```

Entretanto, não seria apropriado estabelecer as hipóteses na forma

- hipótese A: contra
- hipótese B: .

Pois a hipótese A indica um ponto na curva enquanto que a hipótese B indica uma região.

Questões motivadoras para a discussão

1. Na inferência por verossimilhança o espaço amostral é irrelevante, uma vez que não se considera o que poderia se amostrar, mas sim o que foi amostrado. Quais as implicações dessa característica na definição do delineamento amostral? Seria diferente do delineamento amostral quando se usa Inferência Clássica?
2. Na análise de dados usando a Lei da Verossimilhança não se compara uma determinada hipótese a uma hipótese nula, e sim comparam-se hipóteses entre si (podendo haver mais de duas). Podemos dizer esse tipo de análise depende mais fortemente da habilidade do pesquisador para formular hipóteses e criar modelos que explicarão melhor seus dados (já que as hipóteses não estão prontas, elas devem ser formuladas)? Seria a estatística frequentista mais adequada a pesquisadores menos experientes, menos hábeis a formularem modelos mais coerentes com a realidade?
3. O Princípio da Verossimilhança afirma que a função de verossimilhança contém toda informação que um conjunto de dados tem sobre um dado modelo. Quais as vantagens de se realizar a inferência com base apenas na função de verossimilhança? Quais as desvantagens?

Recursos para Estudo

Leituras

Principais

- Royall, R. M. (2007) The likelihood paradigm for statistical evidence. **In:** The nature of scientific evidence (eds. ML Taper and SR Lele), University of Chicago Press, pp 119-152.
- Lewin-Koh N., Taper, M. L. & Lele, S. R. (2004). A brief tour of statistical concepts. **In:** The nature of scientific evidence (eds. ML Taper and SR Lele), University of Chicago Press, pp 3 -16.

Complementares

- Sober, E. 2008. Evidence and Evolution: the logic behind the science. Cambridge, Cambridge University Press. Cap.1.
- [Likelihood](#): Uma palestra de A. W. F. Edwards.
- Berger, J.O. & Wolpert, R.L. 1984. [The likelihood principle](#). Lecture Notes-IMS Monograph Series, Volume 6.

Sobre p-valor e testes de significância

- Ioannidis, John P.A. Why most published research findings are false. [PLoS medicine](#), v. 2, n. 8, p. e124, 2005.
- Forum sobre testes de significância x seleção de modelos: [Ecology 95: 609-667, 2014](#).
- [The problem with p-values is how they're used](#), Comentário de Andrew Gelman que não foi incluído no forum acima por que o periódico cobrou uma taxa do autor, depois de convidá-lo. (<http://andrewgelman.com/2014/05/17/forum-ecology-p-values-model-selection/>).
- Cohen, J. 1994. The Earth Is Round ($p < .05$). *Amer. Psychologist*, 49, 997-1003.
- Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: context, process, and purpose. *The American Statistician*. doi: [10.1080/00031305.2016.1154108](https://doi.org/10.1080/00031305.2016.1154108). Resumo das recomendações da American Statistical Society para uso de p-valores em uma abordagem frequentista. O material suplementar tem 23 artigos com diferentes pontos de vista sobre testes de significância e interpretação do valor p
- Nuzzo, R. (2014). Statistical errors. [Nature](#), 506(7487), 150-152.

Na Internet

- ["Likelihood"](#): Uma palestra de A. W. F. Edwards
- Berger, J.O. & Wolpert, R.L. 1984. [The likelihood principle](#). Lecture Notes-IMS Monograph Series, Volume 6.
- Página do filósofo [Elliot Sober](#), com vários artigos sobre a fundamentação lógica das diferentes abordagens estatísticas.
- Página do filósofo [Malcom Foster](#), especialista em epistemologia da ciências quantitativas, e que tem várias análises críticas do papel de ajuste de modelos nestas.

Resumo da Aula

- slides

1)

Verossimilhança Normalizada, na terminologia de Edwards

2)

propriedade da invariância dos MLES, [tutorial 3](#)

From:

<http://cmq.esalq.usp.br/BIE5781/> - **BIE 5781 Modelagem Estatística para Ecologia e Recursos Naturais**

Permanent link:

<http://cmq.esalq.usp.br/BIE5781/doku.php?id=08-inferencia:08-inferencia>

Last update: **2018/11/24 12:17**

