

Início dos textos em construção

## Capítulo 5 : Contexto e Inferência Bayesiana

### Probabilidades

“O provável é aquilo que acontece na maioria das vezes”, Aristóteles, Retórica.

Uma abordagem da matemática aplicada que tem se popularizado é o de *Inferência Bayesiana*. Os procedimentos anteriores são usualmente agrupados entre *frequencistas*.

Muitas vezes, a informação obtida é quase idêntica, mas a perspectiva muda de forma considerável.

Por princípio, usamos caminhos diferentes.

Antes, associamos cenários a hipóteses e estimamos parâmetros (probabilidades) para testá-las. Agora, os *parâmetros* têm um papel conceitual mais central.

Um parâmetro é um símbolo, uma aproximação para uma ideia (*para*, “perto”, *metron*, “medida”). Aqui, usamos parâmetros para fenômenos que se comportam como números (e.g: existem elementos que podem ser ordenados por alguma noção de tamanho e operações, como soma e multiplicação).

No capítulo 1, estimamos um parâmetro para a diferença média entre tamanho dos bicos das espécies A e B. No capítulo 2, um parâmetro para a correlação entre expectativa de vida saudável e número de médicos em um país. Mais do que isso, também usamos estatísticas para testar hipóteses envolvendo estes parâmetros.

### Frequencistas e Bayesianos

Abordagens frequencistas situam probabilidades como aproximações para cenários com um número infinito de eventos. Exemplo: se jogarmos uma moeda honesta infinitas vezes, a proporção de *caras* tende a que valor? Notamos que para um grande número de sorteios, a proporção tende a 0,5.

Simulação:

```
> set.seed(2600)
> coin_t <- function(x) {
  sample(size=x,x=c(0,1), prob = c(0.5,0.5), replace = T) %>%
  (function(y) sum(y)/length(y))}
> coin_t(3)
[1] 0.6666667
> coin_t(10)
[1] 0.4
> coin_t(30)
[1] 0.5666667
> coin_t(100)
[1] 0.51
> coin_t(1000)
[1] 0.498
> coin_t(100000)
[1] 0.50098
> coin_t(1000000)
[1] 0.4999367
```

É comum a ideia de populações ou procedimentos hipotéticos infinitos. Geralmente é usada em conjunto com o método hipotético-dedutivo, **associando uma probabilidade aos eventos observados** como método de falseamento.

Calculamos uma probabilidade associada à ocorrência de uma observação. No teste t para duas amostras, definimos a hipótese nula em função das médias dos bicos( $\mu$ ) e outros parâmetros ( $\sigma, df$ ).  $H_0 : \mu_{amostra_1} = \mu_{amostra_2}$ .

O procedimento de imaginar os eventos observados como instâncias de infinitos eventos parecidos é muito eficiente, mas pouco intuitivo.

Prismas bayesianos instrumentalizam probabilidades como noções mais básicas: a *plausibilidade*, *grau de crença*, *expectativa* para uma determinada situação. É uma terminologia pouco precisa e possivelmente compatível com o conceito frequentista. Afinal, a expectativa poderia ser calculada como o valor de convergência para infinitos eventos.

O ponto chave é de que deixamos de guiar os procedimentos objetivando uma probabilidade para os eventos. As probabilidades em si passam a ser entidades centrais.

No caso dos pássaros:

*Inferência Frequentista*: Supondo que a diferença média entre tamanho dos bicos seja 0, qual a probabilidade para minhas observações?

Sendo  $H_0$  definida por  $H_0 : \mu_{amostra_1} = \mu_{amostra_2}$ , queremos saber:

$P(H_0) < 0,05$ ?

*Inferência Bayesiana*: Como posso estimar os valores mais prováveis para a diferença entre  $\mu_{amostra_1}$  e  $\mu_{amostra_2}$ ? Considerando um modelo e os dados, qual a distribuição probabilística de  $\mu_{diff_{1-2}}$

$P(\mu_{diff_{1-2}}) = ?$

Em verdade, a linha divisória é bastante tênue. Veremos a seguir que uma plataforma bayesiana oferece dois recursos poderosos: sensibilidade a informações prévias sobre um fenômeno (*priors*) e estimadores estocásticos, que não dependem de soluções analíticas (*Markov Chain Monte Carlo*).

## Muitos métodos científicos: Feyerabend, Carnap e Quine

No primeiro capítulo, entramos em contato com o método hipotético-dedutivo e a falseabilidade como critério de demarcação científica. Apesar de dominante, esse racional possui vulnerabilidades interessantes. Entenderemos melhor argumentos contrários e propostas alternativas através de três filósofos do século XX. Esse é um momento conveniente, uma vez que tiramos os holofotes das hipóteses.

### Paul Feyerabend (1924 - 1994)

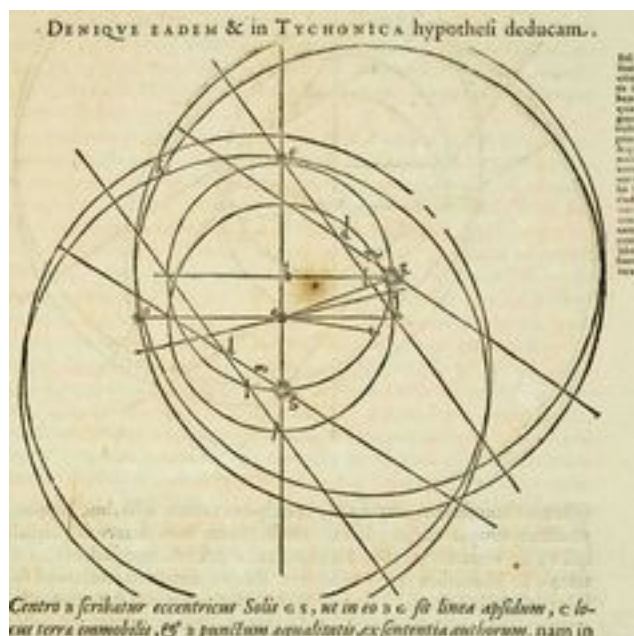
Conhecido pela personalidade ímpar e por ideias radicais, Paul Feyerabend, em *Contra o Método* (1975), argumenta que boa parte dos avanços significativos aconteceram fora do método científico.

Crenças pessoais e detalhes biográficos são responsáveis por mudanças em nosso conhecimento. Mais que isso, usar falsificabilidade (como vimos no *Capítulo I* sobre K. Popper) e o método hipotético-dedutivo teriam nos feito rejeitar o heliocentrismo e outras ideias chave. Na verdade, o sistema geocêntrico (Terra no centro do sistema) de Ptolomeu era mais acurado (!) que o de Copérnico (Sol ao centro) usando um mesmo número de parâmetros para cálculos das órbitas. O sistema copernicano estava mais próximo da realidade como entendida hoje, porém o estágio intermediário de concepção teórica era ‘pior’<sup>1</sup>.

Além de menos acurado, era mais complexo em alguns aspectos, incluindo mais epiciclos. A Revolução Copernicana somente consolidou a mudança de paradigma com contribuições subsequentes de Tycho Brahe, Kepler, Galileu e Newton, cerca de 1 século depois.

---

<sup>1</sup>Stanley E. Babb, “Accuracy of Planetary Theories, Particularly for Mars”, *Isis*, Sep. 1977, pp. 426



Diante das incongruências entre um método e as inevitáveis imprevisibilidades da empreitada humana em conhecer o Universo, Feyerabend propõe o *anarquismo epistêmico* sob o mote “*Anything goes*” (‘Vale tudo’). Isto é, quaisquer recursos são válidos na tentativa e atacar um problema.

É tentador pensar que, dada a profundidade do trabalho, a defesa de uma postura tão contundente é obviamente uma aplicação dos preceitos defendidos no livro como necessários para disseminar uma ideia. Outros filósofos nos ajudam a conceber uma ciência não pautada num método hipotético-dedutivo de maneira mais construtiva.

## Rudolph Carnap (1891 - 1970)

Carnap, do Círculo de Viena, também contrapôs Popper. Em “Testability and Meaning” (1936-7), argumenta que falsificabilidade não difere de verificacionismo. Envolve a testagem de cada assertiva em si, um problema que outros também endereçaram.

Diante de resultados inesperados em um experimento, o procedimento automático para um cientista envolve checar a integridade das condições desenhadas. Verificar a composição da amostra, os métodos de coleta, mecanismos de perda, critérios de exclusão e inclusão, premissas da análise. Isso não é desonestidade intelectual: são fatores menores reais e facilmente abordáveis que podem ter invalidado a teoria de base. O mesmo se dá para técnicas de análise e conceptualização de construtos.

O cuidado com esses pontos é desejável e desnuda o inevitável calcanhar de Aquiles da falsificabilidade. É impossível refutar uma hipótese/assertiva de maneira isolada. Cada procedimento experimental ou lógico envolve a interdependência entre os símbolos usado.

## Willard van Orman Quine (1908 – 2000)

Uma escola filosófica parte do problema acima. A tese de Duhem-Quine postula que é impossível testar qualquer hipótese científica, uma vez que sempre há premissas aceitas como verdade.

Em ‘*Os dois dogmas do empiricismo*’, Quine considera as proposições e as relações lógicas entre elas apenas um sistema, que só pode ser estudado em conjunto. Os exercícios ilustrados no volume anterior testa a adequação dos dados à família de distribuições *t*. Também assume que tamanhos dos bicos são mensuráveis usando números e que estes podem ser comparados com valores de outras amostras.

A princípio, essas declarações parecem triviais. Entretanto, considerando os fatores humanos da ciência, a mudança de lentes é significativa. Abandonando o esquema de testagem de hipóteses como eixo, o *valor p* deixa de ter papel central na narrativa. Integra um conjunto de informações maior sobre os parâmetros examinados.

Discutivelmente, abordar um problema dessa maneira é historicamente mais frutífero. As contribuições mais contundentes são advindas de cientistas dedicados a estudar um contexto ou problema como um todo. É raro, talvez inédito, que um grupo operando de forma sistemática com o método hipotético-dedutivo tenha obtido avanços consistentes.

Estimar livremente os parâmetros de que falamos é muito mais intuitivo que adequar uma ideia aos procedimentos de testagem de hipóteses.

## Bayesian estimation

Para a abordagem anterior, ao fazer um test *t*, calculamos a estatística *t* correspondente às diferenças encontradas e então a probabilidade de obter valores iguais ou mais extremos.

Agora, faremos algo mais simples e intuitivo. Vamos estimar um parâmetro: a diferença entre os grupos. Na verdade, valores prováveis dela. Todas as inferências subsequentes serão derivadas da distribuição produzida por nosso procedimento.

Novamente, usaremos 30 observações retiradas de amostras de distribuição normal ( $\mu_a = 0; \mu_b = 0.6; \sigma_a = \sigma_b = 1$ ) normais. Usando a library BEST, é possível usar inferência bayesiana para responder nossa pergunta (“Como é a distribuição da diferença entre A e B (...)?”). Aplicamos o estimado sobre as amostras A e B e, em seguida, plotamos as distribuições.

```
>library(ggthemes)
>library(rstan)
>library(reshape2)
>library(BEST)
>library(ggplot2)
>options(mc.cores = parallel::detectCores() - 1)
>a <- rnorm(n = 30, sd = 1, mean = 0)
```

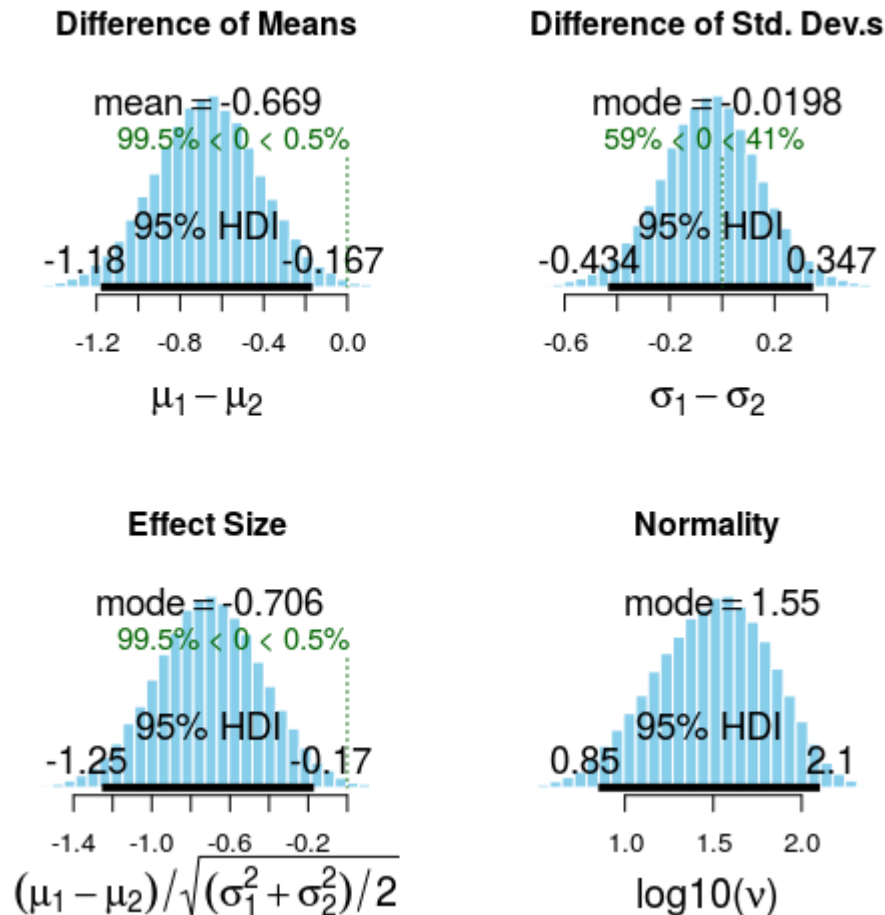
```

>b <- rnorm(n = 30, sd = 1, mean = 0.6)

# BEST
>BESTout <- BESTmcmc(a, b)

### BEST plots
>par(mfrow=c(2,2))
>sapply(c("mean", "sd", "effect", "nu"), function(p) plot(BESTout, which=p))
>layout(1)

```



A distribuição no canto superior esquerdo corresponde à nossa estimativa da diferença entre A e B. Com ela, podemos fazer estimativas pontuais ( $diff_{\mu_a \mu_b} = -0.669$ ). O intervalo apontado como 95% HDI (High density interval) contém 95% da distribuição.

### Por trás das cortinas

Obviamente, vamos entender a arte permitindo estimar essas distribuições. A flexibilidade e o poder dos modelos bayesianos permite lidar com uma série de problemas dificilmente tratáveis de outra forma. Entretanto, é fácil cair em armadilhas ou esbarrar em dificuldades durante o processo.

Nesse framework, lidamos com distribuições. É extremamente importante entender os componentes envolvidos para não cometer erros importantes.

MODIFIED BAYES' THEOREM:

$$P(H|X) = P(H) \times \left( 1 + P(C) \times \left( \frac{P(X|H)}{P(X)} - 1 \right) \right)$$

H: HYPOTHESIS  
 X: OBSERVATION  
 P(H): PRIOR PROBABILITY THAT H IS TRUE  
 P(X): PRIOR PROBABILITY OF OBSERVING X  
 P(C): PROBABILITY THAT YOU'RE USING  
 BAYESIAN STATISTICS CORRECTLY

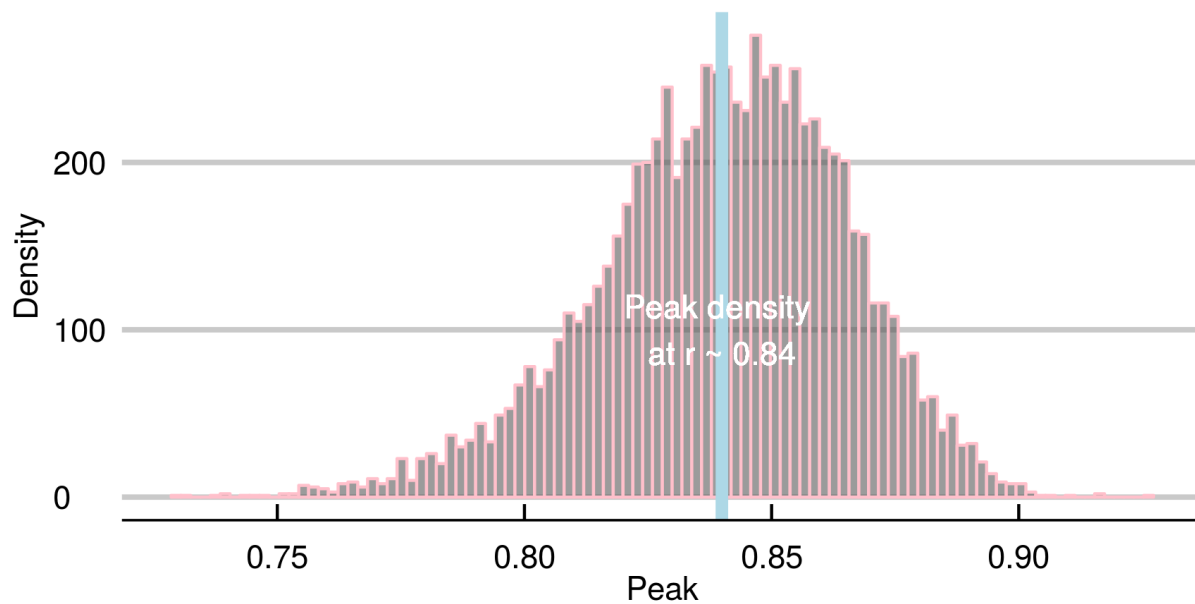
Figure 1: <https://xkcd.com/2059/>

Então, seguindo o exercício anterior, precisamos especificar que consideramos as duas amostras vindo de distribuições t com médias  $\mu_1$  e  $\mu_2$  e desvios-padrão idênticos,  $\sigma_a = \sigma_b$ .

@ Teorema de Bayes: posterior = Likelihood function x prior / evidence @ Posterior para correlação @ Posterior para diferença entre médias @ Posterior para correlação

```
>sample_data <- list(y_1=a,y_2=b,N=length(a))
>fit <- rstan::stan(file="aux/best.stan",
  data=sample_data,
  iter=3000, warmup=100, chains = 6)
```

```
>muDiff <- extract(fit, par='mudiff')$mudiff
>means <- lapply(list(a,b), mean)
>sample_diff <- diff(as.numeric(means)) # observed in data
>ggplot(as.data.frame(muDiff),aes(x=muDiff))+
  geom_histogram(alpha=0.6,color="pink")+
  geom_vline(xintercept=-sample_diff,
    color="light blue",size=2)+ # line for observed difference
  xlab("Difference of Means")+ylab("Frequency")+
  geom_text(label="Sample difference",
    color="white",x=mean(muDiff),y=500)+
  theme_economist_white(gray_bg = F)
```



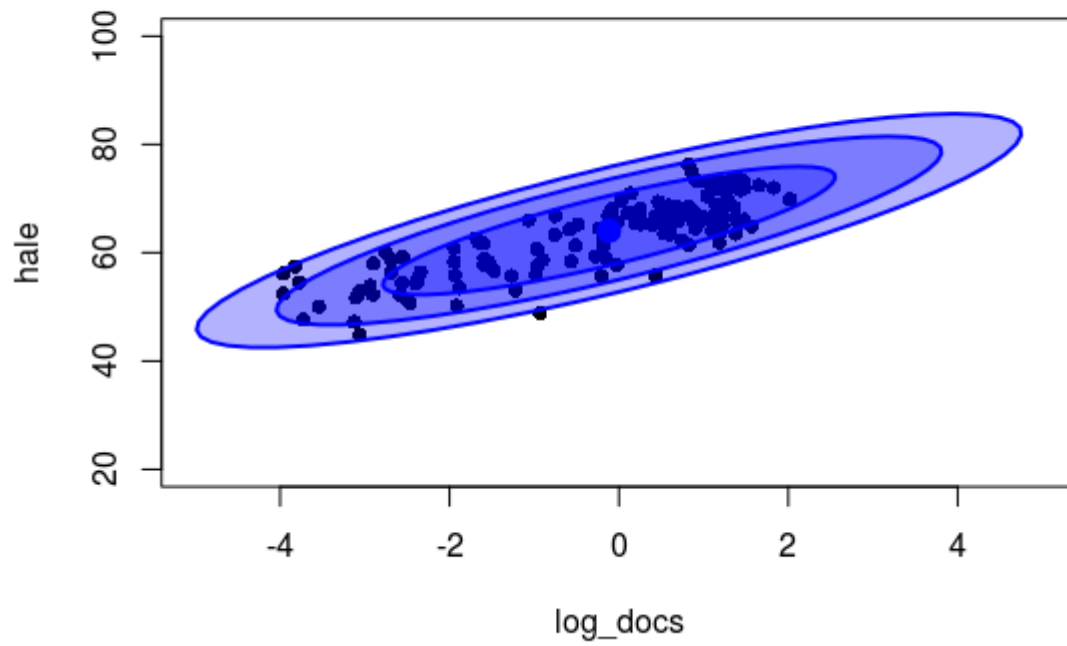
@Correlacao

```
>fit <- rstan::stan(file="aux/corr-docs.stan",
  data=uni_df,
  iter=3000, warmup=100, chains = 6)
```

Ellipse:

```
>x.rand = extract(fit, c("x_rand"))[[1]]
>plot(uni_df[,c("log_docs","hale")],
  xlim=c(-5,5), ylim=c(20, 100), pch=16)
>dataEllipse(x.rand, levels = c(0.75,0.95,0.99),
  fill=T, plot.points = FALSE)
```





- Classificação
  - Regressão logística @ Numero de euler
  - Modelos hierárquicos
- Flexibilidade Bayesiana
  - Usando priors
  - O estimador Markov Chain Monte Carlo