# Capítulo 3

# Contexto e Inferência Bayesiana

### *@ Interpretação de cenários possíveis vs. teoria das medidas*

*@ AIC/BIC*

### Entra Carnap

Feyerabend argumenta (<https://d-van.org/contra-o-método-1-91eb74d200ed>) que boa parte dos avanços científicos significativos aconteceram fora do método. Mais que isso, que usar falsificabilidade e o método hipotético-dedutivo teriam nos feito rejeitar o heliocentrismo e outras ideias chave.

De fato, Popper debateu por muito tempo sobre o status científico da teoria da evolução.

Outro filósofo que contrapôs Popper de maneira brilhante e pouco conhecida foi Rudolf Carnap, do Círculo de Viena. Em “Testability and Meaning” (1936-7), Carnap argumenta que falsificabilidade não difere verificacionismo. Envolve a testagem das sentenças em si, um problema que outros (https://en.wikipedia.org/wiki/Ludwig\_Wittgenstein) também endereçaram.

Diante de resultados inesperados em um experimento, o procedimento padrão do 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.

O cuidado com esses pontos é desejável e desnuda o inevitável calcanhar de Aquiles da falsificabilidade. É impossível obedecer às premissas necessárias para que todo o experimento produza o significado clamado por seus resultados.

### Entra Quine

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 níveis de glicemia são mensuráveis usando números e que estes podem ser comparados com valores em 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.

Por fim, estimar livremente os parâmetros de que falamos é muito mais intuitivo que adequar uma ideia ao racional procedural de testagem de hipóteses.

*Teste de hipóteses*: “Quero comparar A e B. As probabilidades de obter minhas observações supondo igualdade entre A e B são baixas o suficiente supor que A e B são, na verdade, diferentes?”

*Inferência Bayesiana*: “Quero comparar A e B. Como é a distribuição da diferença entre A e B considerando os dados e minhas ideias prévias sobre ela?”

### Bayes

Uma abordagem da matemática aplicada que tem se popularizado é o de inferência Bayesiana. Por princípio, partimos de um ponto diferente: não queremos testar hipóteses (ainda). Temos um modelo modelo teórico e incerteza sobre um parâmetro.

Um parâmetro é um símbolo, uma aproximação (para,"resembling", meter,"measure") para uma ideia. Em geral, usamos parâmetros para representar algo que intuitivamente se comporta como número(e.g: existem elementos que podem ser ordenados por tamanho).

Em uma prova de QI, a idade do indivíduo é uma medida. O tempo total de prova é uma medida, assim como a quantidade de questões acertadas. O valor de QI é uma parâmetro, um número real estimado a partir das medidas citadas.

Inferência Bayesiana nos ajuda a estimar parâmetros. Ela usa a linguagem das probabilidades. Costumamos tratar parâmetros como distribuições. Isso é intuitivo para qualquer pessoa.

Alguém pergunta "que horas você chega no jantar?".

A resposta "20:00" na verdade é uma estimativa pontual, porém sabemos que existe a chance de chegarmos 19:55 ou 20:05. Também sabemos que chegar 19:30 é improvável, o que também vale para 20:30.

Trabalhamos com incertezas o tempo inteiro.

### 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 (μa=0; μb=0.6; σa = σb = 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

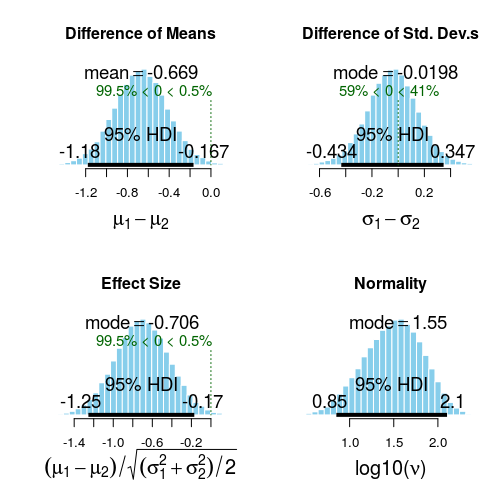
>BESTout <- BESTmcmc(a, b)

### BEST plots

>par(mfrow=c(2,2))

>sapply(c("mean", "sd", "effect", "nu"), function(p) plot(BESTout, which=p))

>layout(1)

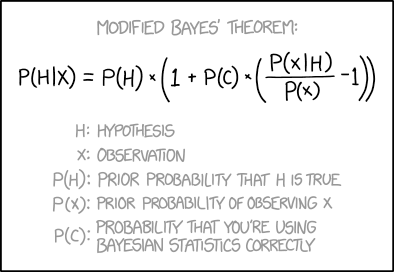


A distribuição no canto superior esquerdo corresponde à nossa estimativa da diferença entre A e B. Com ela, podemos fazer estimativas pontuais (diffμab=-0.669). O intervalo apontado como 95% HDI (High density interval) contém 95% da distribuição.

### Por trás das cortinas

Obviamente, vamos entender como é possível estimar essas distribuições. A flexibilidade e o poder dos modelos bayesianos permite lidar com uma série de problemas dificilmente tratáveis sob outra perspectiva. 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.



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

Então, seguindo o exercício anterior, precisamos especificar que consideramos as duas amostras vindo de distribuições *t* com médias μ1 e μ2 e desvios-padrão idênticos, σa= σb.

@ Teorema de Bayes: posterior = Likelihood function x prior / evidence

@ Posterior para correlação

@ Posterior para diferença entre médias

@ Posterior para correlação

Arquivo *.stan* especificando o modelo:

@Reestruturar lendo artigo do BEST

data {

int<lower=0> N;

vector[N] y\_1;

vector[N] y\_2;

}

parameters {

real mu\_1;

real mu\_2;

real sigma;

real<lower=0, upper=1> df;

real<lower=0, upper=1000> lambda;

}

model {

//priors

mu\_1 ~ normal(0, 1);

mu\_2 ~ normal(0, 1);

sigma ~ normal(0, 1);

df ~ uniform(0,1);

lambda ~ exponential(df);

//likelihood

for (n in 1:N){

y\_1[n] ~ student\_t(lambda, mu\_1, sigma);

y\_2[n] ~ student\_t(lambda, mu\_2, sigma);

}

}

generated quantities{

real mudiff;

real cohenD;

mudiff = mu\_1 - mu\_2;

cohenD = mudiff/sigma;

}

Com o R:

>sample\_data <- list(y\_1=a,y\_2=b,N=length(a))

fit <- rstan::stan(file="bayesian-philos.stan",

data=sample\_data,

iter=3000, warmup=100, chains = 6)

>print(fit, digits=3,

pars=c('mu\_1', 'mu\_2', 'mudiff', 'cohenD','lambda','lp\_\_'))

Output:

Inference for Stan model: bayesian-philos.

6 chains, each with iter=3000; warmup=100; thin=1;

post-warmup draws per chain=2900, total post-warmup draws=17400.

mean se\_mean sd 2.5% 25% 50% 75% 97.5% n\_eff Rhat

mu\_1 -0.002 0.001 0.178 -0.349 -0.123 -0.001 0.117 0.348 19759 1.000

mu\_2 0.648 0.001 0.179 0.293 0.528 0.649 0.769 0.997 19356 1.000

mudiff -0.650 0.002 0.249 -1.142 -0.820 -0.649 -0.482 -0.160 19201 1.000

cohenD -0.731 0.002 0.289 -1.309 -0.924 -0.728 -0.533 -0.176 18839 1.000

lambda 40.806 1.960 90.819 3.425 7.843 14.599 34.039 269.731 2148 1.002

lp\_\_ -53.380 0.027 1.735 -57.738 -54.255 -53.049 -52.100 -51.061 4255 1.001

Samples were drawn using NUTS(diag\_e) at Fri Nov 23 15:51:26 2018.

For each parameter, n\_eff is a crude measure of effective sample size,

and Rhat is the potential scale reduction factor on split chains (at

convergence, Rhat=1).

>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,

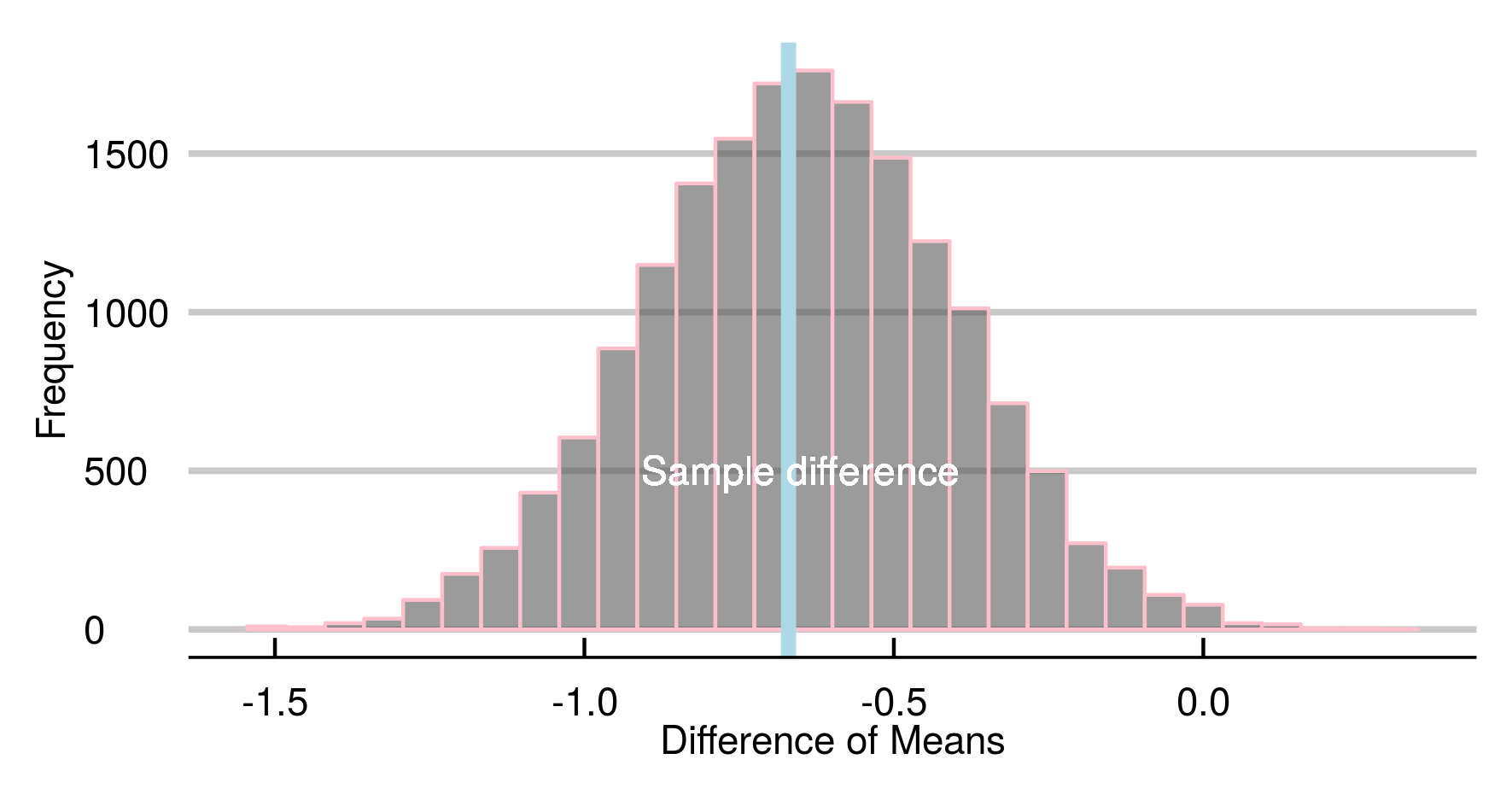
color="light blue",size=2)+ # line for observed difference

xlab("Difference of Means")+ylab("Frequency")+

geom\_text(label="Sample difference",

color="white",x=mean(muDiff),y=500)+

theme\_economist\_white(gray\_bg = F)



### Correlação

@Reproduzindo análise de dataset WHO/Worldbank

Arquivo *.stan* especificando o modelo:

data {

int<lower=1> N;

vector[2] x[N];

}

parameters {

vector[2] mu;

real<lower=0> sigma[2];

real<lower=1> nu;

real<lower=-1, upper=1> rho;

}

transformed parameters {

// Covariance matrix

cov\_matrix[2] cov = [[ sigma[1] ^ 2 , sigma[1] \* sigma[2] \* rho],

[sigma[1] \* sigma[2] \* rho, sigma[2] ^ 2 ]];

}

model {

// Likelihood

// Bivariate Student's t-distribution instead of normal for robustness

x ~ multi\_student\_t(nu, mu, cov);

// Noninformative priors on all parameters

sigma ~ normal(0,100);

mu ~ normal(0, 100);

nu ~ gamma(2, 0.1);

}

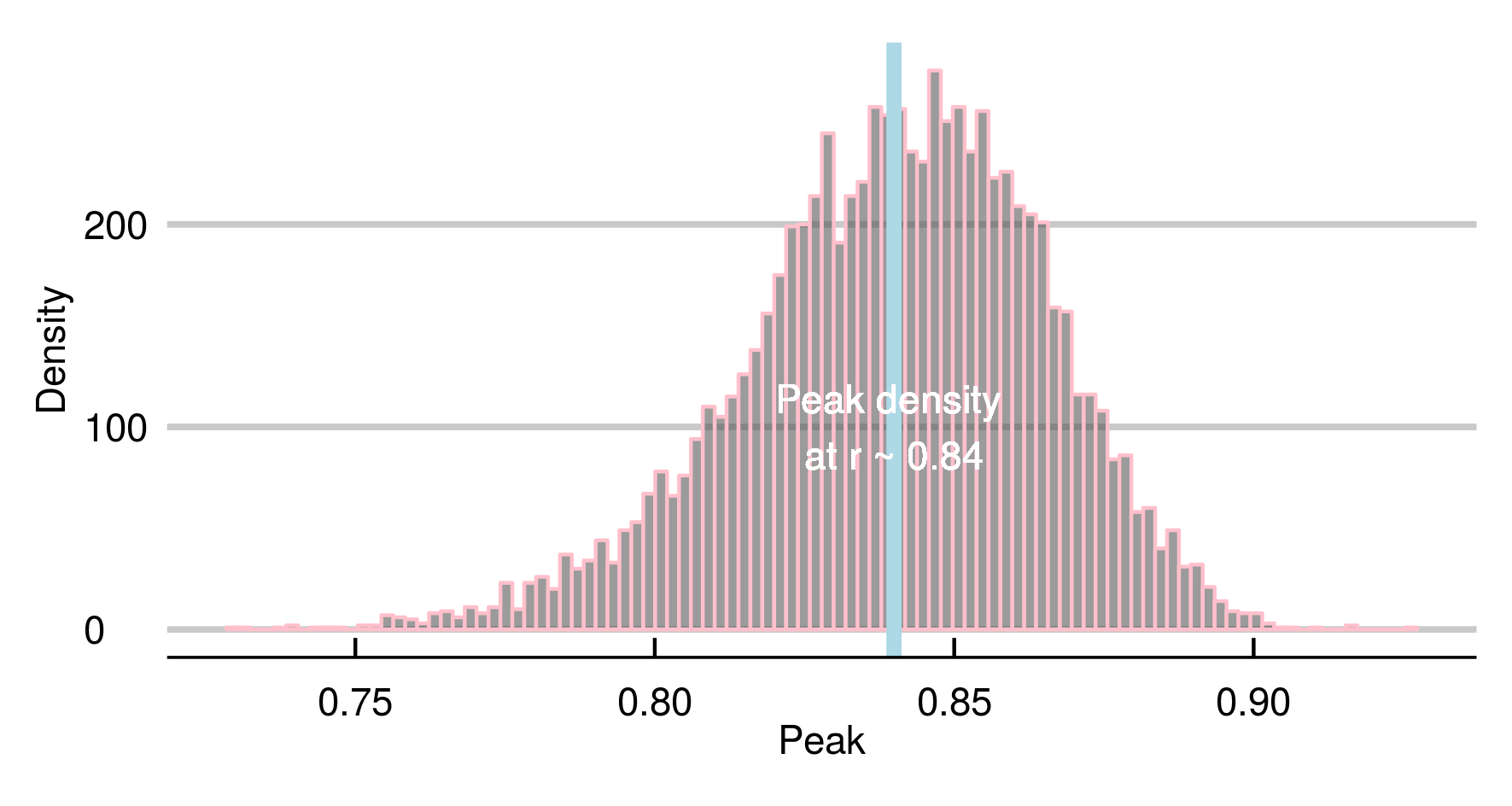
generated quantities {

// Random samples from the estimated bivariate t-distribution (for assessment of fit)

vector[2] x\_rand;

x\_rand = multi\_student\_t\_rng(nu, mu, cov);

}



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)

