

# SME0809 - Inferência Bayesiana - Distribuição Normal

Grupo 13 - Francisco Miranda - 4402962 - Heitor Carvalho - 11833351

Outubro 2021

```
library(tidyverse)
library(effectsize)
library(invgamma)
library(dados)
#remotes::install_github("cienciadedatos/dados")
set.seed(42)
```

## Caso 1: $\mu$ desconhecido e $\sigma$ conhecido

### Verossimilhança da distribuição

$$\mathcal{L}(y|\theta) = \prod_{i=1}^n p(y_i|\theta) = \prod_{i=1}^n e^{-1/2\sigma^2(y_i-\theta)^2}$$

### A *priori* da distribuição

Nós parametrizamos  $p(\theta)$  de modo que  $\theta \sim \mathcal{N}(y_0, \tau_0^2)$  com média  $y_0$  e variância  $\tau_0^2$

$$p(\theta) \propto e^{(-1/2\tau_0^2)(\theta-y_0)^2}$$

### A priori não informativa de Jeffreys

A distribuição a *priori* de Jeffreys é dada por  $p(\theta) \propto \sqrt{J(n/\sigma^2)} \propto 1$

Sabemos que a Informação de Fisher de  $\theta$  através de  $y = y_1, \dots, y_n$  é definida como:

$$I(\theta) = E\left[-\frac{\partial^2 \log p(y|\theta)}{\partial \theta^2}\right]$$

Segue que:

$$\begin{aligned}
& -E\left[\frac{\partial^2}{\partial\theta^2}(-\log(2\pi\sigma^2)/2 - 1/2\sigma^2(\sum_{i=1}^n(y_i - \theta)^2))\right] \\
&= -E\left[\frac{\partial^2}{\partial\theta^2}(-\log(2\pi\sigma^2)/2 - 1/2\sigma^2(\sum_{i=1}^n(y_i^2 - 2\theta n\bar{y} + n\theta^2))\right] \\
&= -E\left[\frac{\partial}{\partial\theta}(-1/2\sigma^2(-2n\bar{y} + 2n\theta))\right] \\
&= -E[-1/2\sigma^2(2n)] \\
&= n/\sigma^2
\end{aligned}$$

## A *posteriori* da distribuição

A *posteriori* é computada assumindo-se que:

1. Cada observação é independentemente distribuída
2. Cada observação tem a mesma variância

$$p(\theta|y) = p(\theta) p(y|\theta) = p(\theta) \prod_{i=1}^n p(y_i|\theta) = e^{(-1/2\tau_0^2)(\theta-\mu_0)^2} \prod_{i=1}^n e^{(-1/2\sigma^2)(y_i-\theta)^2} = e^{(\frac{-1}{2})(1/\tau_0^2(\theta-\mu_0)^2 + \frac{1}{\sigma^2} \sum_{i=1}^n (y_i-\theta)^2)}$$

Desse modo, a distribuição a posteriori da média  $\theta$  depende apenas da média amostral  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ , sendo assim,  $\bar{y}$  é uma estatística suficiente.

Portanto, para  $n$  observações, a posteriori apresenta a seguinte distribuição:

$$p(\theta|y_1, \dots, y_n) = p(\theta | \bar{y}) = \mathcal{N}(\theta|y_n, \tau_n^2)$$

Sendo,

$$\mu_n = \frac{\tau_0^{-2}\mu_0 + n\sigma^{-2}\bar{y}}{\tau_0^{-2} + \sigma^{-2}}, \quad \text{e} \quad \tau_n^{-2} = \tau_0^{-2} + n\sigma^{-2}$$

Podemos reescrever  $p(\theta|y)$  como:

$$p(\theta|y_n) \propto e^{(-1/2\tau_n^2)(\theta-\mu_n)^2}$$

Logo, para uma distribuição Normal com variância conhecida, a média a *posteriori*  $\mu_n$  pode ser interpretada como a média ponderada da média a *priori* e o valor observado  $y = y_1, \dots, y_n$ , sendo os pesos proporcionais às precisões de cada um.

```

sample <- dados::penguins%>%
  filter(especie == "Pinguim-de-barbicha")%>%
  select(comprimento_bico)%>%
  drop_na()
sample2 <- sample %>% sample_n(2) %>% pull()
sample5 <- sample %>% sample_n(5) %>% pull()
sample15 <- sample %>% sample_n(15) %>% pull()
sample30 <- sample %>% sample_n(30) %>% pull()
sample <- sample %>% pull()

```

```

# gera a priori e a posteriori de uma normal com media desconhecida e sigma conhecido
norm <- function(samp, sigma = 40, mu, tau0 = 10000 ){

  n <- length(samp)
  xbar <- mean(samp)

  ver <- function(x) exp(- n/(2*sigma^2) * (xbar - x)^2)

  mu.post <- (tau0^(-2)*mu + n*sigma^(-2)*xbar)/ (tau0^(-2) + n * sigma^(-2))
  sigma.post <- (tau0^(-2) + n*sigma^(-2))^(1)

  theta <- seq(20, 60, 0.02)

  tibble(theta = theta,
          priori = dnorm(theta,mu,tau0),
          post = dnorm(theta,mu.post,sigma.post),
          ver = ver(theta),
          pred = dnorm(theta,mu.post,tau0+sigma),
          tau1 = sigma.post,
          mu1 = mu.post)

}

```

## Exemplo: Comprimento do bico dos pinguins

Escolheu-se o conjunto de dados *palmerpenquins* em sua versão traduzida. Como atributo de interesse escolhemos o comprimento do bico dos pinguins e nos limitamos a análise de uma espécie - nesse caso o **Pinguim-de-barbicha**. Existem 68 pinguins desta espécie em nosso dataset.

Sabe-se que para a distribuição do comprimento do bico dos pinguins tem-se que  $\sigma = 3,34$ .

Quatro amigos, Cleiton, Eduarda, Larissa e Robertinho resolvem tentar estimar a média do comprimento do bico dos pinguins de barbicha. Para tanto, cada integrante do grupo resolve dar um palpite em relação a média e variância da distribuição.

Desse modo, Cleiton, que nunca foi a um zoológico e nunca viu um pinguim pessoalmente acredita que a distribuição seja próxima a  $\mathcal{N}(80,12)$ , Eduarda, que adora pinguins porém nunca viu um pessoalmente, acredita que  $\mathcal{N}(30,8)$ , Larissa que visita zoológicos com frequência acredita que  $\mathcal{N}(55,5)$  e Robertinho - biólogo que trabalha com pinguins acredita que  $\mathcal{N}(47,3)$ .

Adotaremos três amostras de bico de pinguins, uma com  $n = 5$ , outra com  $n = 15$ , sorteadas independentemente, e a amostra com todos os pinguins, onde  $n = 68$ .

Para a amostra com  $n = 5$  teremos:

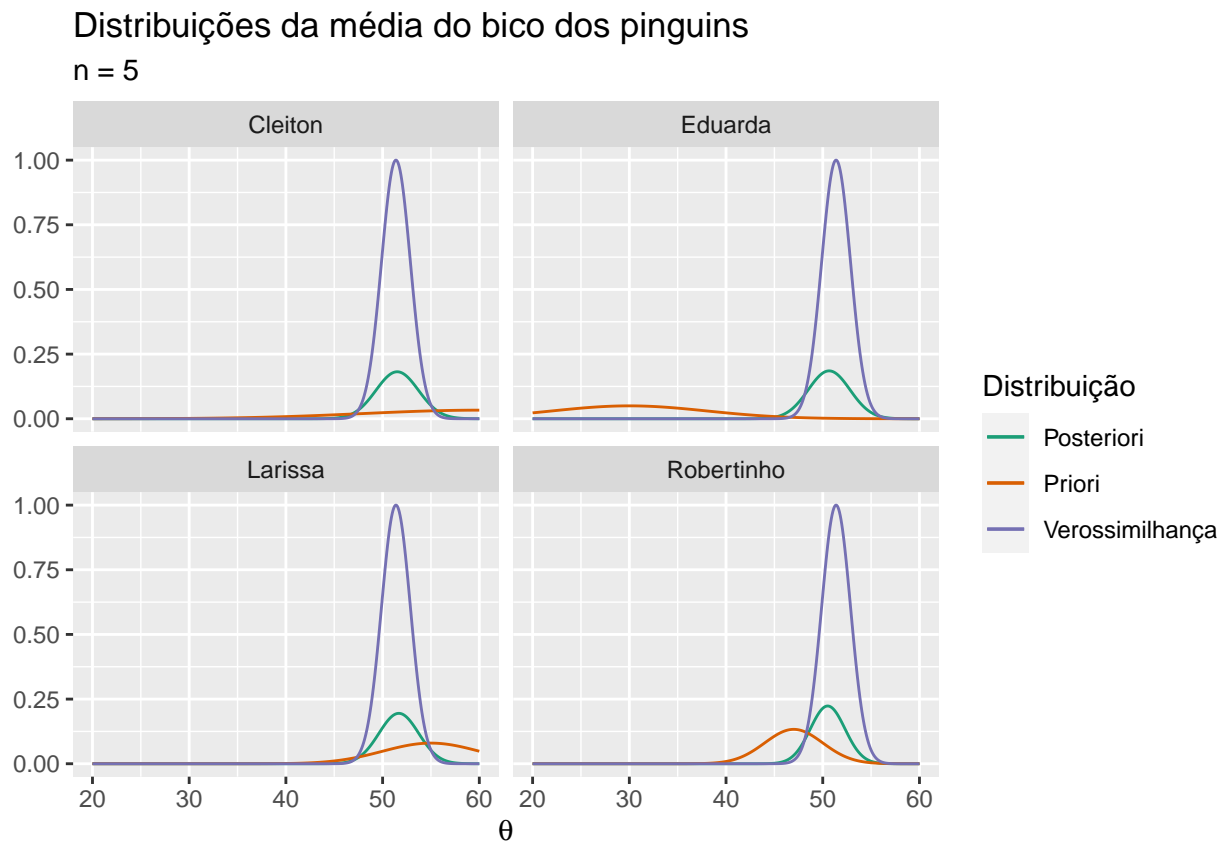
```

a <- norm(samp = sample5, sigma = sd(sample), tau0 = 12, mu = 60) %>% mutate(Priori = "Cleiton")
b <- norm(samp = sample5, sigma = sd(sample), tau0 = 8, mu = 30) %>% mutate(Priori = "Eduarda")
c <- norm(samp = sample5, sigma = sd(sample), tau0 = 5, mu = 55) %>% mutate(Priori = "Larissa")
d <- norm(samp = sample5, sigma = sd(sample), tau0 = 3, mu = 47) %>% mutate(Priori = "Robertinho")

rbind(a,b,c,d) %>%
  ggplot(aes(x = theta)) +
  geom_line(aes(y = post, color = "Posteriori")) +

```

```
geom_line(aes(y = priori, color = "Priori")) +
geom_line(aes(y = ver, color = "Verossimilhança")) +
scale_colour_brewer(name = "Distribuição", type = "qual", palette = "Dark2")+
scale_x_continuous(name = expression(theta), limits = c(20, 60))+
theme(axis.title.y=element_blank()) +
labs(title = "Distribuições da média do bico dos pinguins",
      subtitle = "n = 5") +
facet_wrap(~Priori)
```



```
tibble(
  Prioris = c("Cleiton", "Eduarda", "Larissa", "Robertinho"),
  Media.pri = c(80, 30, 55, 47),
  Media.pos = c(50.6, 43.8, 51.8, 47.5),
  SD.pri = c(12, 8, 5, 3),
  SD.pos = c(22.2, 17.2, 12.1, 6.51),
  IC.025 = c(qnorm(0.025, mean = 50.6, sd = 22.2),
             qnorm(0.025, mean = 43.8, sd = 17.2),
             qnorm(0.025, mean = 51.8, sd = 12.1),
             qnorm(0.025, mean = 47.5, sd = 6.51)),
  IC.975 = c(qnorm(0.975, mean = 50.6, sd = 22.2),
             qnorm(0.975, mean = 43.8, sd = 17.2),
             qnorm(0.975, mean = 51.8, sd = 12.1),
             qnorm(0.975, mean = 47.5, sd = 6.51)))%>%
knitr::kable(digits = 2, caption = "Resumo aposteriori dos quatro amigos (n=5)")
```

Table 1: Resumo aposteriori dos quatro amigos (n=5)

Prioris	Media.pri	Media.pos	SD.pri	SD.pos	IC.025	IC.975
Cleiton	80	50.6	12	22.20	7.09	94.11
Eduarda	30	43.8	8	17.20	10.09	77.51
Larissa	55	51.8	5	12.10	28.08	75.52
Robertinho	47	47.5	3	6.51	34.74	60.26

Procedendo de forma análoga, para  $n = 15$ :

### Distribuições da média do bico dos pinguins

n = 15

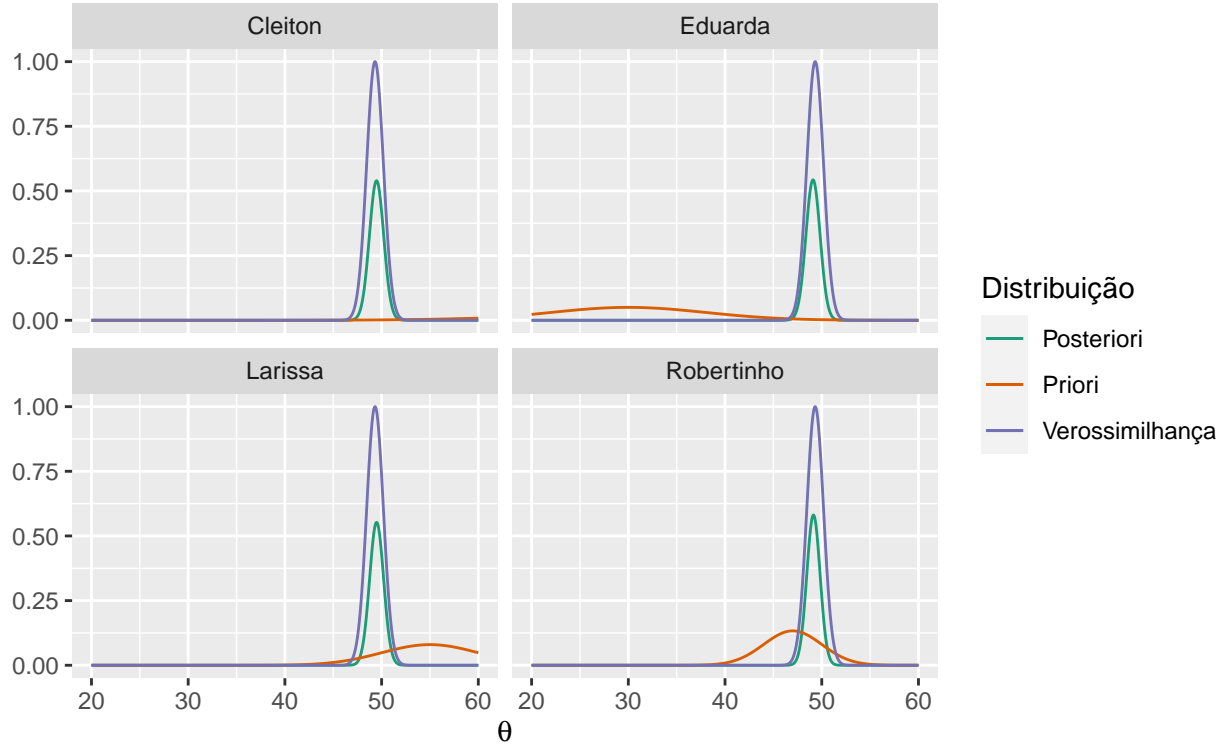


Table 2: Resumo aposteriori dos quatro amigos (n=15)

Prioris	Media.pri	Media.pos	SD.pri	SD.pos	IC.025	IC.975
Cleiton	80	49.6	20	1.24	47.17	52.03
Eduarda	30	49.1	8	0.74	47.66	50.54
Larissa	55	49.5	5	0.72	48.08	50.92
Robertinho	47	49.1	3	0.69	47.75	50.45

Para a amostra completa:

## Distribuições da média do bico dos pinguins

n = 68

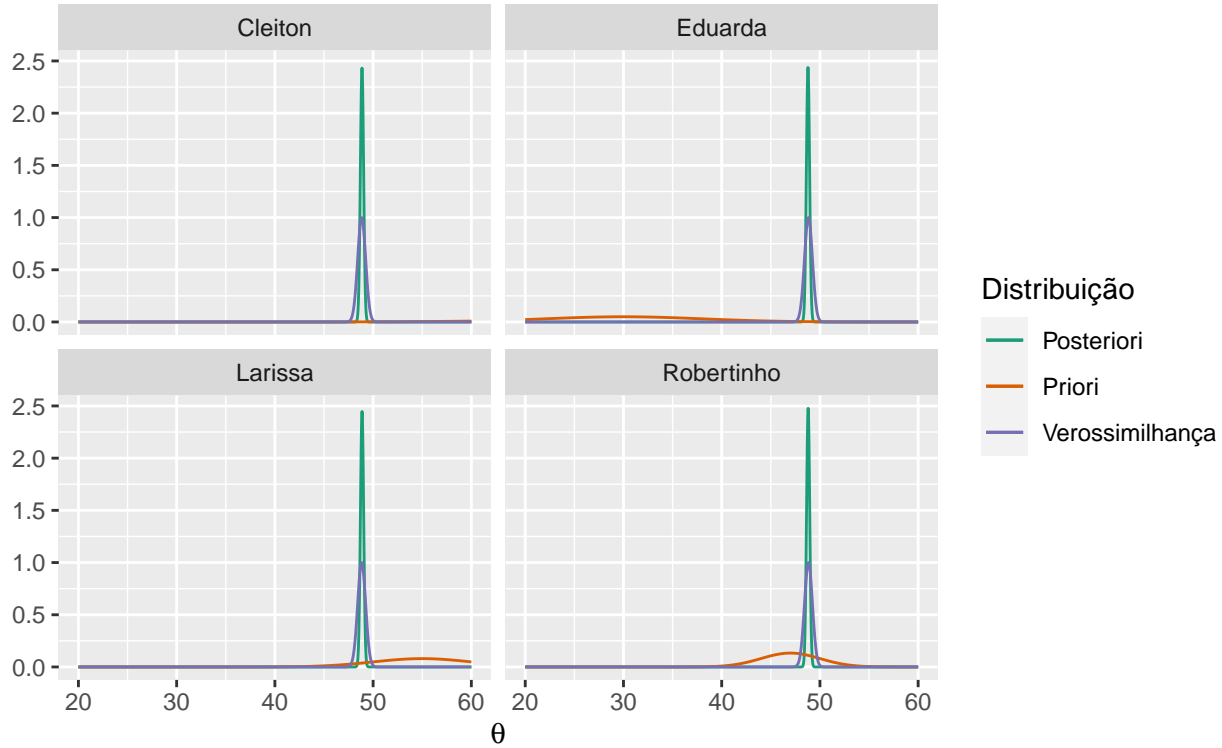


Table 3: Resumo apriori dos quatro amigos para n = 68

Prioris	Media.pri	SD.pri	IC.025	IC.975
Cleiton	80	20	40.80	94.11
Eduarda	30	8	-3.71	77.51
Larissa	55	5	31.28	75.52
Robertinho	47	3	34.24	60.26

## Caso 2: $\mu$ conhecido e $\sigma$ desconhecido

### Distribuições a priori

Seja  $Y_i$  uma amostra aleatória simples de uma distribuição  $Y \sim N(\theta, \sigma^2)$ , com  $\theta$  conhecido.

Primeiramente, vamos encontrar a função de verossimilhança de  $\sigma^2$ .

$$\mathcal{L}(y|\sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma}} e^{-(y_i - \theta)^2 / 2\sigma^2} \propto (\sigma^2)^{-\frac{n}{2}} e^{-\left(\frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \theta)^2\right)}$$

### Priori não informativa

Definimos a log-verossimilhança em nosso caso como sendo:

$$\log(\mathcal{L}(y|\sigma^2)) \propto -\frac{n}{2} \log(\sigma^{-2}) - \sigma^{-2} \sum_{i=1}^n (y_i - \theta)^2$$

A distribuição a *priori* de Jeffreys é dada por  $\pi(\sigma^2) \propto \sqrt{J(\sigma^2)}$ .

$$\begin{aligned} J(\sigma^2) &\propto E \left( -\frac{\partial^2}{\partial \theta^2} \log(L(\theta)) \right) = E \left( -\frac{\partial^2}{\partial \theta^2} \left( -\frac{n}{2} \log(\sigma^{-2}) - (\sigma^2)^{-1} \sum_{i=1}^n (y_i - \theta)^2 \right) \right) \\ &= E \left( -\frac{\partial}{\partial \theta} \left( -\frac{n}{2} (\sigma^2)^{-1} + (\sigma^2)^{-2} \sum_{i=1}^n (y_i - \theta)^2 \right) \right) = E \left( -\frac{n}{2\sigma^2} + 2(\sigma^2)^{-3} \sum_{i=1}^n (y_i - \theta)^2 \right) \\ &= -\frac{n}{2\sigma^2} + 2\sigma^{-4} \sum_{i=1}^n (E(y_i) - \theta)^2 = -\frac{n}{\sigma^2} + 2\sigma^{-4} \sum_{i=1}^n (\theta - \theta)^2 = -\frac{n}{\sigma^2} \propto \sigma^{-2} \end{aligned}$$

Assim,  $\pi(\sigma) \propto \sqrt{\sigma^{-2}} = \sigma^{-1}$ . Seu parâmetro  $\Phi$  de escala que faz com que  $\theta$  mude somente em locação pode ser obtido através do cálculo de

$$\phi \propto \int \pi(\sigma^2) d\sigma^2 = \int \frac{1}{\sigma^2} d\sigma^2 = \log |\sigma^2| + k \propto \log \sigma^2$$

$\phi$  é uma distribuição imprópria, pois  $\int_0^{+\infty} \log(\sigma^2) d\sigma^2$  é divergente. Assim, a *priori* não favorece nenhuma escala em detrimento de outra.

### Conjugadas Naturais

O suporte de nosso parâmetro de interesse  $\sigma > 0$  permite-nos adotar três distribuições de probabilidade estudadas durante o curso: **Gama**, **Gama-Inversa** e **Qui-Quadrado**. Note que as duas primeiras estão relacionadas via uma transformação simples e a última é um caso particular delas. Dessa forma, as três distribuições servem como conjugada natural da Normal, em nosso caso optou-se por utilizar a distribuição *Gama Inversa*.

$$p_\theta(\theta|\alpha, \beta) = \frac{\beta^\alpha \theta^{-(\alpha+1)} e^{-\beta/\theta}}{\Gamma(\alpha)} \propto \theta^{-(\alpha+1)} e^{-\beta/\theta}, \quad \alpha > 0, \beta > 0, \theta > 0$$

Fazendo  $\theta = \sigma^2$ , temos uma *priori* da forma:

$$\sigma^{-(\alpha+1)} e^{-\beta/\sigma^2} \Rightarrow \pi(\sigma) \sim \text{Gama-Inv}(\alpha, \beta)$$

Se quisermos torná-la não informativa, basta utilizarmos  $\alpha \rightarrow 0, \beta \rightarrow 0$ .

### Distribuição a *posteriori*

$$\begin{aligned} \pi(\sigma|y) &\propto \mathcal{L}(y|\sigma^2) \pi(\sigma) = (\sigma^2)^{-\frac{n}{2}} e^{-\left(\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \theta)^2\right)} \sigma^{-(\alpha-1)} e^{-\beta/\sigma^2} \\ &= (\sigma^2)^{-(\alpha+\frac{n}{2}+1)} e^{-\frac{1}{\sigma^2} \left(\beta + \frac{1}{2} \sum_{i=1}^n (y_i - \theta)^2\right)} \end{aligned}$$

Dessa forma,

$$\pi(\sigma^2|y) \sim \text{Gama-Inv}\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2} \sum_{i=1}^n (y_i - \theta)^2\right)$$

## Exemplo: Comprimento do bico dos pinguins

Cleiton, Eduarda, Larissa e Robertinho estão estudando sobre os Pinguins-de-barbicha. Sabe-se que o comprimento do bico deles tem distribuição Normal com média 48.833 e desvio padrão desconhecido. Os quatro amigos decidem estimar este desvio padrão, cada um define sua *priori* da seguinte forma:

- **Cleiton** nunca viu um pinguim-de-barbicha na vida, nem em fotografia. Dessa forma, ele decide adotar uma *priori* não informativa  $\text{Gama-Inv}(\alpha = 0.01, \beta = 0.01)$
- **Eduarda** sabe tudo sobre pinguins, mas nunca viu um pessoalmente. Ela opta por uma  $\text{Gama-Inv}(\alpha = 1, \beta = 1.5)$
- **Larissa** adora ir ao zoológico visitar aos pinguins. Ela decide adotar uma  $\text{Gama-Inv}(\alpha = 0.5, \beta = 3)$
- **Robertinho** é um biólogo com muita experiência, que consulta suas anotações sobre pinguins e decide adotar uma *priori*  $\text{Gama-Inv}(\alpha = 35, \beta = 186)$

Assim como no caso anterior, utilizaremos três amostras de bico de pinguins, uma com  $n = 5$ , outra com  $n = 30$ , sorteadas independentemente, e o conjunto de dados completo, com  $n = 68$ .

```
# gera a priori e a posteriori de uma normal com media conhecida e sigma desconhecido
SigmaNorm <- function(samp, theta = 48.833, alpha = 0.001, beta = 0.001){

  n <- length(samp)
  s <- sum(((samp - theta)/2)^2)

  l_sigma2 <- function(sigma2) sigma2^(-(n/2)) * exp(- 1/sigma2 * s)

  a.post <- alpha + n/2
  b.post <- beta + s

  sigma2 <- seq(0.02, 40, 0.02)

  tibble(sigma2 = sigma2,
    priori = (dinvgamma(sigma2,alpha,beta)),
    post = (dinvgamma(sigma2,a.post,b.post)),
    ver = normalize(l_sigma2(sigma2))/3,
    alpha1 = a.post,
    beta1 = b.post,
    alpha0 = alpha,
    beta0 = beta)
}
```

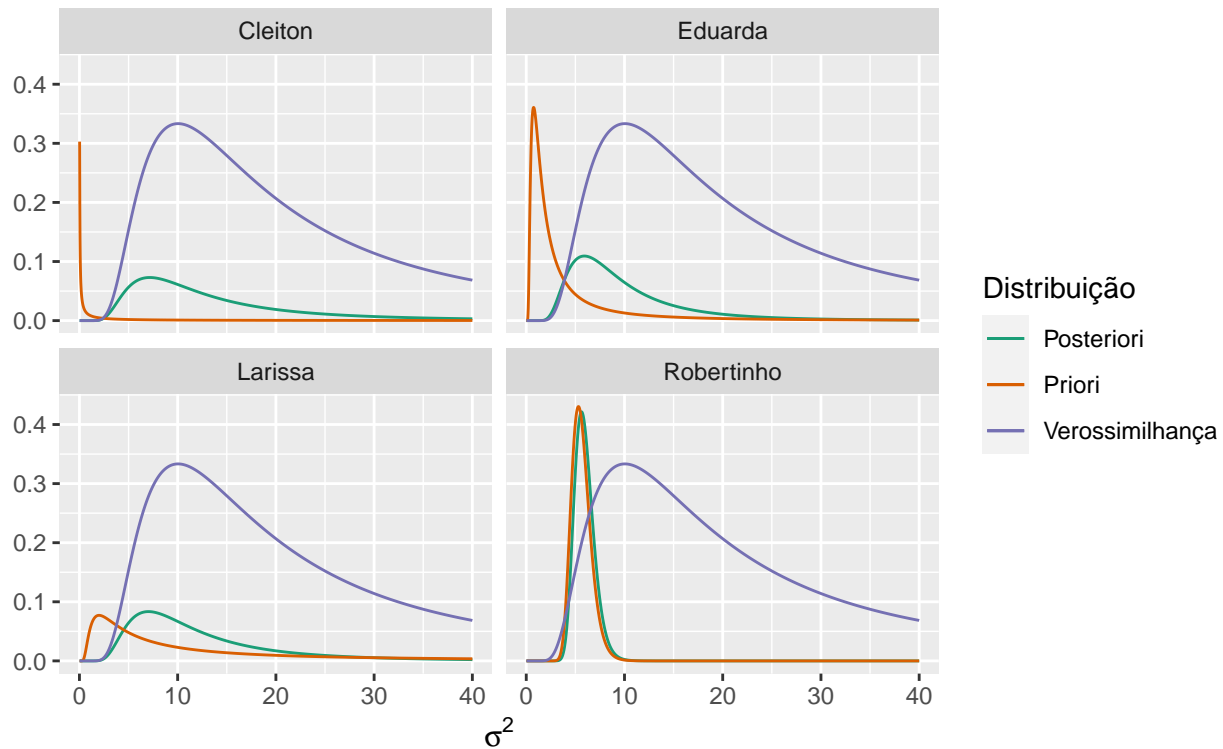
```
a <- SigmaNorm(sample5, alpha = 0.01, beta = 0.01) %>% mutate(Priori = "Cleiton")
b <- SigmaNorm(sample5,alpha = 1, beta = 1.5) %>% mutate(Priori = "Eduarda")
c <- SigmaNorm(sample5,alpha = 0.5, beta = 3) %>% mutate(Priori = "Larissa")
d <- SigmaNorm(sample5,alpha = 34, beta = 186) %>% mutate(Priori = "Robertinho")

rbind(a,b,c,d) %>%
  ggplot(aes(x = sigma2)) +
  geom_line(aes(y = post, color = "Posteriori")) +
  geom_line(aes(y = priori, color = "Priori")) +
  geom_line(aes(y = ver, color = "Verossimilhança")) +
  scale_colour_brewer(name = "Distribuição", type = "qual", palette = "Dark2")+
  scale_x_continuous(name = expression(sigma^2), limits = c(0, 40))+
```



```
theme(axis.title.y=element_blank()) +
labs(
  title = "Distribuições normalizadas da variância do comprimento do bico dos pinguins",
  subtitle = "n=5" +
facet_wrap(~Priori)
```

## Distribuições normalizadas da variância do comprimento do bico dos pinguins n=5



Neste caso, com o  $n$  pequeno, é interessante notar que uma verossimilhança difusa a respeito de  $\sigma^2$ , se traduz em uma *posteriori* difusa para os colegas, exceto para Robertinho, que não aprendeu muito aqui.

```
tabDesc <- function(alpha, beta){
  if(alpha < 1)
    med <- NA
  else med <- beta/(alpha-1)
  if(alpha < 2)
    v <- NA
  else v <- beta^2/((alpha-1)^2*(alpha-2))

  tibble( alpha = alpha,
    beta = beta,
    media = med,
    var = v,
    moda = beta/(alpha+1),
    IC2.5 = qinvgamma(0.025, alpha, beta),
    IC97.5 = qinvgamma(0.975, alpha, beta))
}
```

```
cbind( Priori = c("Cleiton", "Eduarda", "Larissa", "Robertinho"),
      rbind(tabDesc(a$alpha0[1], a$beta0[1]),
            tabDesc(b$alpha0[1], b$beta0[1]),
            tabDesc(c$alpha0[1], c$beta0[1]),
            tabDesc(d$alpha0[1], d$beta0[1])
      )) %>%
knitr::kable(digits = 2, caption = "Resumo a priori (n = 5)")
```

Table 4: Resumo a priori (n = 5)

Priori	alpha	beta	media	var	moda	IC2.5	IC97.5
Cleiton	0.01	0.01	NA	NA	0.01	0.21	2.838743e+158
Eduarda	1.00	1.50	Inf	NA	0.75	0.41	5.925000e+01
Larissa	0.50	3.00	NA	NA	2.00	1.19	6.109550e+03
Robertinho	34.00	186.00	5.64	0.99	5.31	4.01	7.900000e+00

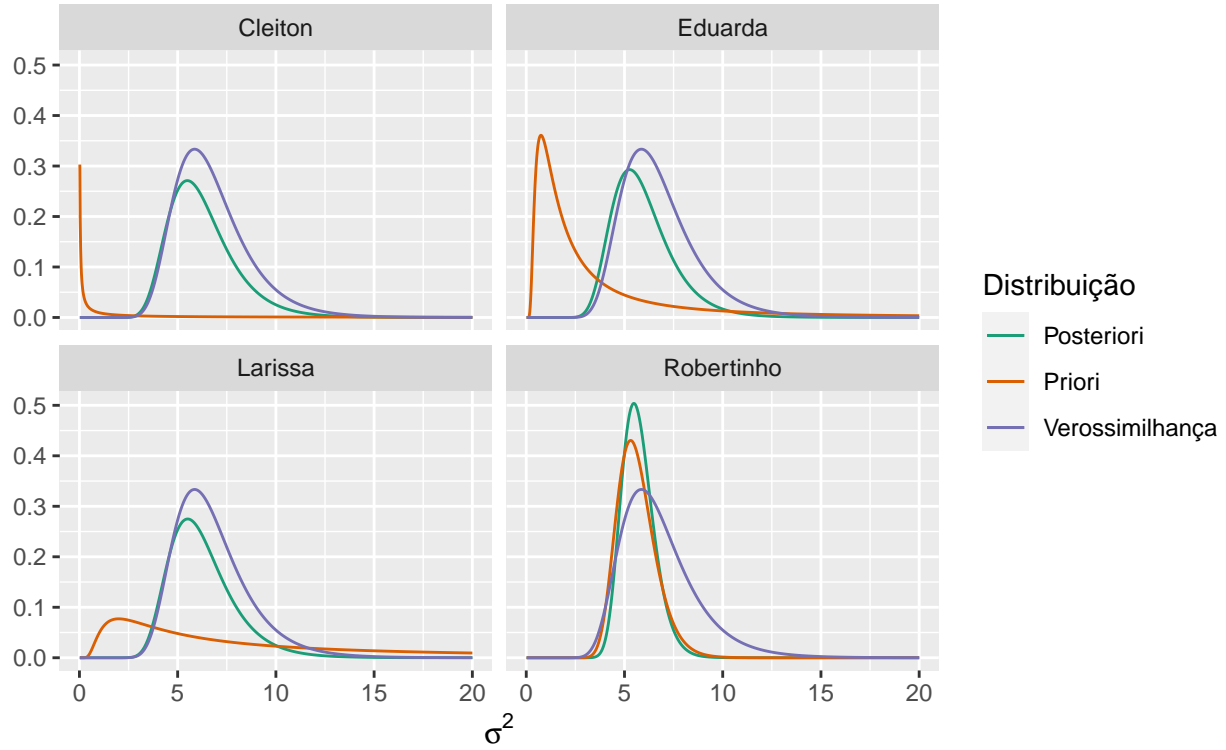
```
cbind( Priori = c("Cleiton", "Eduarda", "Larissa", "Robertinho"),
      rbind(tabDesc(a$alpha1[1], a$beta1[1]),
            tabDesc(b$alpha1[1], b$beta1[1]),
            tabDesc(c$alpha1[1], c$beta1[1]),
            tabDesc(d$alpha1[1], d$beta1[1])
      )) %>%
knitr::kable(digits = 2, caption = "Resumo a posteriori (n = 5)")
```

Table 5: Resumo a posteriori (n = 5)

Priori	alpha	beta	media	var	moda	IC2.5	IC97.5
Cleiton	2.51	25.12	16.63	542.47	7.16	3.90	59.89
Eduarda	3.50	26.61	10.64	75.51	5.91	3.32	31.49
Larissa	3.00	28.11	14.05	197.49	7.03	3.89	45.43
Robertinho	36.50	211.11	5.95	1.03	5.63	4.29	8.24

Procedendo de forma análoga, geramos os gráficos e tabelas para os outros valores de  $n$ .

# Distribuições normalizadas da variância do comprimento do bico dos pinguins n=30



Agora, o  $n$  parece ser adequado. Vemos que todos os participantes obtiveram bastante informação a respeito de  $\sigma$  com essa amostra.

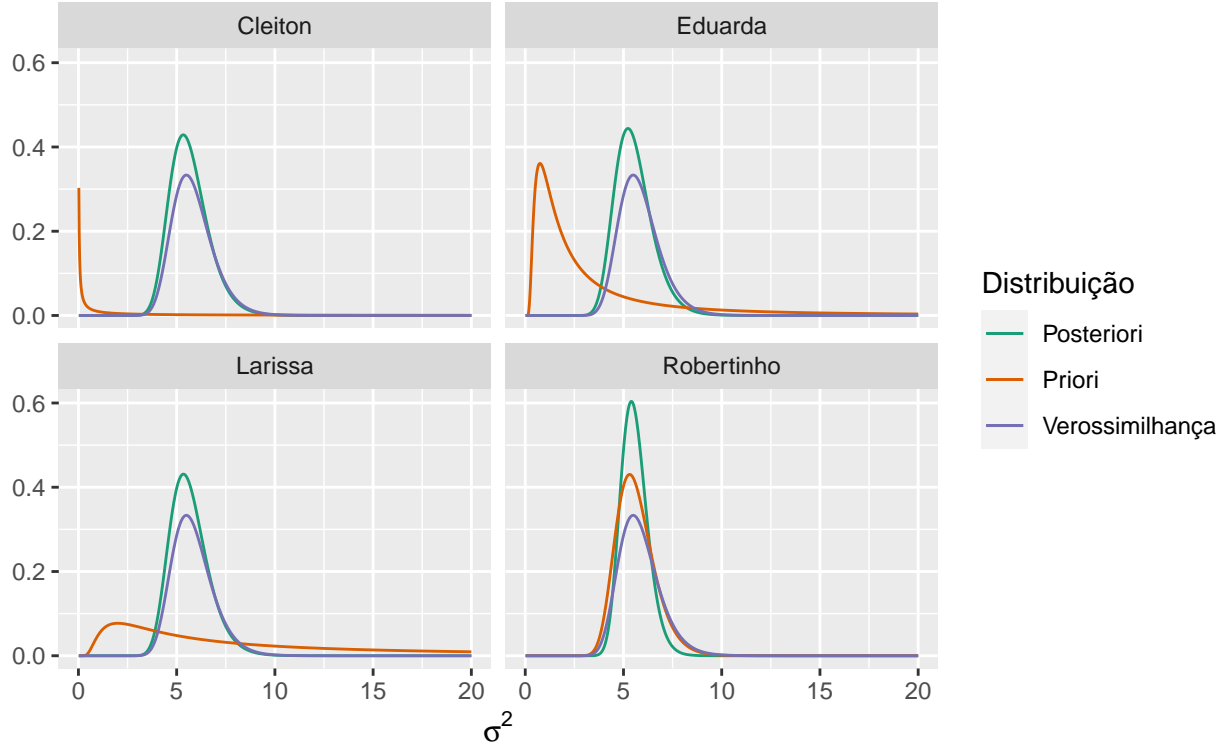
Table 6: Resumo a priori (n = 30)

Priori	alpha	beta	media	var	moda	IC2.5	IC97.5
Cleiton	0.01	0.01	NA	NA	0.01	0.21	2.838743e+158
Eduarda	1.00	1.50	Inf	NA	0.75	0.41	5.925000e+01
Larissa	0.50	3.00	NA	NA	2.00	1.19	6.109550e+03
Robertinho	34.00	186.00	5.64	0.99	5.31	4.01	7.900000e+00

Table 7: Resumo a posteriori (n = 30)

Priori	alpha	beta	media	var	moda	IC2.5	IC97.5
Cleiton	15.01	87.94	6.28	3.03	5.49	3.74	10.46
Eduarda	16.00	89.43	5.96	2.54	5.26	3.61	9.78
Larissa	15.50	90.93	6.27	2.91	5.51	3.77	10.37
Robertinho	49.00	273.93	5.71	0.69	5.48	4.30	7.56

## Distribuições normalizadas da variância do comprimento do bico dos pinguins n=68



Aqui, não há grandes alterações se comparado com o anterior. Aparentemente, o conhecimento que os colegas acumularam a respeito de  $\sigma$  foi parecido em  $n = 30$  e  $n = 68$ .

Table 8: Resumo a priori (n = 68)

Priori	alpha	beta	media	var	moda	IC2.5	IC97.5
Cleiton	0.01	0.01	NA	NA	0.01	0.21	2.838743e+158
Eduarda	1.00	1.50	Inf	NA	0.75	0.41	5.925000e+01
Larissa	0.50	3.00	NA	NA	2.00	1.19	6.109550e+03
Robertinho	34.00	186.00	5.64	0.99	5.31	4.01	7.900000e+00

Table 9: Resumo a posteriori (n = 68)

Priori	alpha	beta	media	var	moda	IC2.5	IC97.5
Cleiton	34.01	186.78	5.66	1.00	5.34	4.03	7.93
Eduarda	35.00	188.27	5.54	0.93	5.23	3.96	7.72
Larissa	34.50	189.77	5.66	0.99	5.35	4.04	7.92
Robertinho	68.00	372.77	5.56	0.47	5.40	4.38	7.06

Neste trabalho, pudemos ver a forma das distribuições *a priori* e *posteriori* da média, com a variância conhecida, e da variância, com a média conhecida de uma distribuição Normal comportaram-se para diferentes valores de  $n$ , utilizando o conjunto de dados *palmerpenguins*. Foi uma aprendizagem interessante tanto do ponto de vista teórico quanto prático, aqui procuramos elencar todo o nosso conhecimento acumulado em SME0809, e utilizá-lo para realizar a análise de um conjunto de dados reais, fazendo algumas suposições, como a normalidade dos dados. Para próximos trabalhos teóricos nesse tópico, poderiam ser calculadas as distribuições preditivas, assim como as posteriores para o caso da *Gama* e da *Qui-Quadrado*.