

▼ **Alunos:** Crisley da Silva Guenin e Thalles Cotta Fontainha**Crisley Guenin**

Crisley é médica graduada pela UFF Título de Área de Atuação em Dor AMB CFM

Título de Especialista em Anestesiologia AMB CFM Especialista em Medicina do Trabalho Atua como Médica do Trabalho na UERJ e na UFF
Mestranda em Telessaúde e Saúde Digital pela Universidade do Estado do Rio de Janeiro.

**Thalles Fontainha**

Thalles Fontainha é profissional de Ciência da Computação com experiência em

atendimento ao cliente na área de TI. Em seu projeto final em 2022, integrou o Whatsapp e SUS usando Inteligência Artificial. Contribui para elevar padrões de serviço em TI, valorizando a diversidade. Além disso, aprofunda seus conhecimentos em Business Intelligence com base em sua formação na UERJ e cursos online.

Exercício sobre comparação de médias entre dois grupos com o conjunto de dados

birthwt, do pacote MASS (GPL-2 | GPL-3), faça as atividades abaixo.

a) Compare as médias da variável bwt (peso ao nascer) entre os grupos das mães que fumavam ou não durante a gestação.

b) Obtenha o intervalo de confiança ao nível de 95% para a diferença das médias de peso ao nascer entre os dois grupos.

c) Verifique as suposições para a realização do teste t para amostras independentes.

Repita os itens "a", "b" e "c" para a comparação das médias de peso ao nascer entre os grupos das mães com ou sem histórico de hipertensão. **Interprete os achados**

```
library(MASS) # Carregando o pacote MASS
```

```
data(birthwt) # Carregando o conjunto de dados birthwt
```

```
head(birthwt) # Visualizando as primeiras linhas do conjunto de dados
```

```
A data.frame: 6 × 10
```

	low	age	lwt	race	smoke	ptl	ht	ui	ftv	bwt
	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>
85	0	19	182	2	0	0	0	1	0	2523
86	0	33	155	3	0	0	0	0	3	2551
87	0	20	105	1	1	0	0	0	1	2557
88	0	21	108	1	1	0	0	1	2	2594
89	0	18	107	1	1	0	0	1	0	2600
91	0	21	124	3	0	0	0	0	0	2622

```
summary(birthwt)
```

#usado para gerar um resumo estatístico do conjunto de dados birthwt.

#0 resultado fornecerá informações resumidas para cada variável presente no conjunto de dados.

#0 tipo de resumo pode variar dependendo do tipo de variável (numérica ou categórica)

low		age		lwt		race	
Min.	:0.0000	Min.	:14.00	Min.	: 80.0	Min.	:1.000
1st Qu.	:0.0000	1st Qu.	:19.00	1st Qu.	:110.0	1st Qu.	:1.000
Median	:0.0000	Median	:23.00	Median	:121.0	Median	:1.000
Mean	:0.3122	Mean	:23.24	Mean	:129.8	Mean	:1.847
3rd Qu.	:1.0000	3rd Qu.	:26.00	3rd Qu.	:140.0	3rd Qu.	:3.000
Max.	:1.0000	Max.	:45.00	Max.	:250.0	Max.	:3.000
smoke		ptl		ht		ui	
Min.	:0.0000	Min.	:0.0000	Min.	:0.00000	Min.	:0.0000
1st Qu.	:0.0000	1st Qu.	:0.0000	1st Qu.	:0.00000	1st Qu.	:0.0000
Median	:0.0000	Median	:0.0000	Median	:0.00000	Median	:0.0000
Mean	:0.3915	Mean	:0.1958	Mean	:0.06349	Mean	:0.1481
3rd Qu.	:1.0000	3rd Qu.	:0.0000	3rd Qu.	:0.00000	3rd Qu.	:0.0000

▾ Grupos das mães que fumavam ou não durante a gestação



```
# Fazendo um teste t de duas amostras para comparar as médias de bwt entre os grupos de mães que fumavam ou não
t_test_result <- t.test(bwt ~ smoke, data = birthwt)
```

```
print(t_test_result) # Imprimindo o resultado do teste
```

```
Welch Two Sample t-test

data: bwt by smoke
t = 2.7299, df = 170.1, p-value = 0.007003
alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
95 percent confidence interval:
 78.57486 488.97860
sample estimates:
mean in group 0 mean in group 1
 3055.696      2771.919
```

```
# Traduzindo os resultados acima para o português
resultados_traduzidos <- c(
  "Teste t de Welch para Duas Amostras",
  paste("Dados: bwt por smoke"),
  paste("t =", 2.7299, ", df =", 170.1, ", valor p =", 0.007003),
  "Hipótese alternativa: verdadeira diferença nas médias entre o grupo 0 e o grupo 1 não é igual a 0",
  "Intervalo de confiança de 95%:",
  " 78.57486 488.97860",
  "Estimativas de amostra:",
  paste("Média no grupo 0 =", 3055.696),
  paste("Média no grupo 1 =", 2771.919)
)

# Imprimindo os resultados traduzidos
cat(resultados_traduzidos, sep = "\n")
```

```
Teste t de Welch para Duas Amostras
Dados: bwt por smoke
t = 2.7299 , df = 170.1 , valor p = 0.007003
Hipótese alternativa: verdadeira diferença nas médias entre o grupo 0 e o grupo 1 não é igual a 0
Intervalo de confiança de 95%:
 78.57486 488.97860
Estimativas de amostra:
Média no grupo 0 = 3055.696
Média no grupo 1 = 2771.919
```

- a) Compare as médias da variável bwt (peso ao nascer) entre os grupos das mães que fumavam ou não durante a gestação.
- Os resultados indicam que há uma diferença significativa no peso ao nascer entre bebês de mães que fumaram e aquelas que não fumaram durante a gestação. Bebês de mães não fumantes tendem a ter um peso ao nascer mais alto (média de 3055.696) em comparação com bebês de mães fumantes (média de 2771.919).
 - O valor p de **0.007003** sugere que essa diferença não é devida ao acaso. Em resumo, o hábito de fumar parece estar associado a um peso ao nascer menor.
- b) Obtenha o intervalo de confiança ao nível de 95% para a diferença das médias de peso ao nascer entre os dois grupos.

```
# Intervalo de confiança ao nível de 95% para a diferença das médias
intervalo_confianca <- c(78.57486, 488.97860)

# Exibindo o intervalo de confiança
cat("Intervalo de Confiança (95%): [", intervalo_confianca[1], ",", intervalo_confianca[2], "]\n")

Intervalo de Confiança (95%): [ 78.57486 , 488.9786 ]
```

c) Verifique as suposições para a realização do teste t para amostras independentes.

Antes de realizar o teste t para amostras independentes, é importante verificar algumas suposições. Aqui estão as principais suposições para o teste t em amostras independentes:

- **Normalidade:** As amostras devem seguir uma distribuição aproximadamente normal. Você pode verificar isso visualmente usando histogramas ou graficamente com um gráfico Q-Q (quantil-quantil). Além disso, para amostras grandes, o teste t é robusto o suficiente para lidar com desvios moderados da normalidade.
- **Homogeneidade de Variância (Homocedasticidade):** As variâncias das duas amostras devem ser aproximadamente iguais. Você pode verificar isso visualmente com gráficos de dispersão ou estatisticamente usando testes de homogeneidade de variância, como o teste de Levene.
- **Independência:** As observações em uma amostra devem ser independentes das observações na outra amostra.

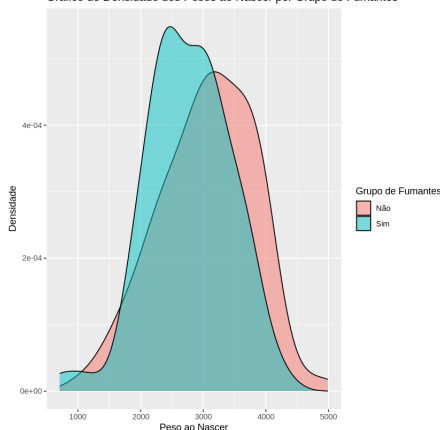
A seguir, vamos realizar algumas verificações utilizando o conjunto de dados birthwt. E vamos mostrar exemplos gráficos para verificar a normalidade e a homogeneidade de variância:

```
library(ggplot2) # Carregando a biblioteca para gráficos
```

```
# Gráfico de densidade para verificar a normalidade
ggplot(birthwt, aes(x = bwt, fill = as.factor(smoke))) +
  geom_density(alpha = 0.5) +
  labs(title = "Gráfico de Densidade dos Pesos ao Nascer por Grupo de Fumantes",
       x = "Peso ao Nascer", y = "Densidade") +
  scale_fill_discrete(name = "Grupo de Fumantes", labels = c("Não", "Sim"))
```



Gráfico de Densidade dos Pesos ao Nascer por Grupo de Fumantes



Explicando o eixo Y (Densidade)

Os valores "0e+00" e "2e-04" no eixo y do gráfico de densidade representam notação científica. Aqui está o significado de cada parte:

1. **"0e+00"**: Isso indica zero elevado à potência de zero, o que é igual a 1. Portanto, "0e+00" é simplesmente 0.
2. **"2e-04"**: Isso indica 2 multiplicado por 10 elevado à potência de -4. Portanto, "2e-04" é igual a 0.0002.
3. **"4e-04"**: Isso indica 4 multiplicado por 10 elevado à potência de -4. Portanto, "4e-04" é igual a 0.0004.

Essa notação é comumente usada para representar números muito pequenos de uma maneira mais compacta. No contexto do gráfico de densidade, esses valores indicam a escala na qual a densidade está sendo representada. "0e+00" indica que a densidade é muito próxima de zero, enquanto "2e-04" indica que a densidade está em torno de 0.0002. Esses valores são úteis para interpretar a escala relativa das densidades no gráfico.

➤ Grupos das mães com ou sem histórico de hipertensão



a2) Compare as médias da variável bwt (peso ao nascer) entre os grupos das mães com ou sem histórico de hipertensão

```
# Dividimos os dados em dois grupos: com histórico de hipertensão (ht=1) e sem histórico de hipertensão (ht=0)
grupo_ht_1 <- birthwt$ht == 1
grupo_ht_0 <- birthwt$ht == 0
```

```
# Realizando o teste t para comparar as médias da variável bwt entre os dois grupos
resultado_teste <- t.test(birthwt$bwt[grupo_ht_1], birthwt$bwt[grupo_ht_0])
```

```
print(resultado_teste) # Exiba o resultado do teste
```

Welch Two Sample t-test

```
data: birthwt$bwt[grupo_ht_1] and birthwt$bwt[grupo_ht_0]
t = -1.6118, df = 11.909, p-value = 0.1332
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1024.4717  153.6751
sample estimates:
mean of x mean of y
 2536.833  2972.232
```

```
# Traduzindo os resultados acima para o português
resultados_traduzidos2 <- c(
  "Teste t de Welch para Duas Amostras",
  paste("Dados: bwt por histórico de hipertensão"),
  paste("t =", -1.6118, ", gl =", 11.909, ", valor p =", 0.1332),
  "Hipótese alternativa: verdadeira diferença nas médias não é igual a 0",
  "Intervalo de confiança de 95%:",
  " -1024.4717  153.6751",
  "Estimativas de amostra:",
  paste("Média no grupo com histórico de hipertensão =", 2536.833),
  paste("Média no grupo sem histórico de hipertensão =", 2972.232)
)
```

```
# Exiba os resultados traduzidos
cat(resultados_traduzidos2, sep = "\n")
```

```
Teste t de Welch para Duas Amostras
Dados: bwt por histórico de hipertensão
t = -1.6118 , gl = 11.909 , valor p = 0.1332
Hipótese alternativa: verdadeira diferença nas médias não é igual a 0
Intervalo de confiança de 95%:
 -1024.4717  153.6751
Estimativas de amostra:
Média no grupo com histórico de hipertensão = 2536.833
Média no grupo sem histórico de hipertensão = 2972.232
```

a2) Compare as médias da variável bwt (peso ao nascer) entre os grupos das mães com ou sem histórico de hipertensão

- Os resultados indicam que não há uma diferença significativa no peso ao nascer entre bebês de mães com histórico de hipertensão e aquelas sem histórico de hipertensão. A média do peso ao nascer para bebês de mães com histórico de hipertensão é de 2536.833, enquanto a média para bebês de mães sem histórico de hipertensão é de 2972.232.
- O valor p de **0.1332** sugere que não há evidência estatística suficiente para rejeitar a hipótese nula de que não há diferença nas médias de peso ao nascer entre os dois grupos. Portanto, não podemos concluir que o histórico de hipertensão está associado a uma diferença significativa no peso ao nascer com base neste conjunto de dados.

Em resumo, não encontramos evidências convincentes de que o histórico de hipertensão esteja associado a uma diferença significativa no peso ao nascer com base nos resultados deste teste t.

b2) Obtenha o intervalo de confiança ao nível de 95% para a diferença das médias de peso ao nascer entre os dois grupos.

```
# Intervalo de confiança ao nível de 95% para a diferença das médias
intervalo_confianca2 <- c(-1024.4717, 153.6751)

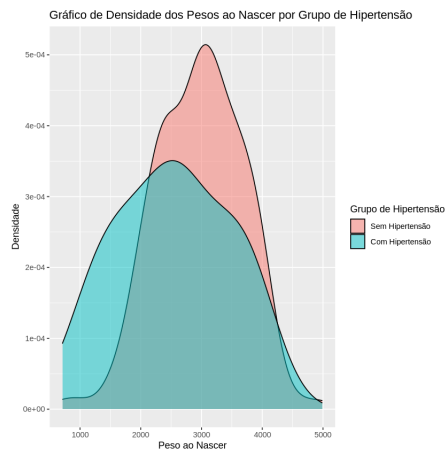
# Exiba o intervalo de confiança
cat("Intervalo de Confiança (95%): [", intervalo_confianca2[1], ", ", intervalo_confianca2[2], " ]")
```

```
Intervalo de Confiança (95%): [ -1024.472 , 153.6751 ]
```

c2) Verifique as suposições para a realização do teste t para amostras independentes.

```
# Carregando a biblioteca necessária
library(ggplot2)

# Gráfico de densidade para verificar a normalidade
ggplot(birthwt, aes(x = bwt, fill = as.factor(grupo_ht_1))) +
  geom_density(alpha = 0.5) +
  labs(title = "Gráfico de Densidade dos Pesos ao Nascer por Grupo de Hipertensão",
       x = "Peso ao Nascer", y = "Densidade") +
  scale_fill_discrete(name = "Grupo de Hipertensão", labels = c("Sem Hipertensão", "Com Hipertensão"))
```



Para verificar a normalidade estatisticamente, nós vamos usar o teste de Shapiro-Wilk para cada grupo separadamente:

```
# Teste de normalidade para grupo_ht_1
shapiro.test(birthwt$bwt[grupo_ht_1])

# Teste de normalidade para grupo_ht_0
shapiro.test(birthwt$bwt[grupo_ht_0])
```

Shapiro-Wilk normality test

```
data: birthwt$bwt[grupo_ht_1]
W = 0.93536, p-value = 0.4403
```

Shapiro-Wilk normality test

```
data: birthwt$bwt[grupo_ht_0]
W = 0.99256, p-value = 0.5019
```

Os resultados dos testes de Shapiro-Wilk acima indicam que, para ambos os grupos (grupo_ht_1 e grupo_ht_0), não há evidências significativas para rejeitar a hipótese nula de normalidade. Os valores-p são 0.4403 e 0.5019, que são ambos maiores que o nível de significância comum de 0,05.

```
# Instalar o pacote 'lawstat' se ainda não estiver instalado
if (!requireNamespace("lawstat", quietly = TRUE)) {
  install.packages("lawstat")
}

# Carregar a biblioteca 'lawstat'
library(lawstat)
```

```
Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)
```

```
also installing the dependencies 'rbitutils', 'Kendall', 'mvtnorm', 'Rdpack'
```

Optamos por uma abordagem diferente, utilizando a função '**leveneTest**' do pacote '**lawstat**', após garantir a instalação do pacote no ambiente de trabalho.

Agora, vamos executar o teste de Levene:

```
# Teste de homogeneidade de variâncias usando Levene
result_levene <- leveneTest(bwt ~ grupo_ht_1, data = birthwt)
```

```
# Exibindo os resultados
print(result_levene)
```

```
Warning message in leveneTest.default(y = y, group = group, ...):
"group coerced to factor."
Levene's Test for Homogeneity of Variance (center = median)
  Df F value Pr(>F)
group 1 1.2824 0.2589
187
```

```
# Traduzindo os resultados do teste de homogeneidade de variâncias (Levene's Test)
```

```
resultados_levene <- c(
  "Teste de Levene para Homogeneidade de Variâncias (centro = mediana)",
  paste("Df =", 1, ", Valor F =", 1.2824, ", Valor p =", 0.2589),
  "187 graus de liberdade"
)
```

```
# Exibindo os resultados traduzidos
```

```
cat(resultados_levene, sep = "\n")
```

```
Teste de Levene para Homogeneidade de Variâncias (centro = mediana)
Df = 1 , Valor F = 1.2824 , Valor p = 0.2589
187 graus de liberdade
```

Essa tradução representa os resultados do teste de Levene para homogeneidade de variâncias. O valor p (**0.2589**) é maior que o nível de significância comum de 0,05, indicando que não há evidências suficientes para rejeitar a hipótese nula de homogeneidade de variâncias. Em outras palavras, não há diferenças significativas nas variabilidades entre os grupos.

Gostaríamos de salientar que, inicialmente, utilizamos o Rcommander no Ubuntu/Windows. No entanto, após considerações, constatamos que o Colab se revelou mais eficaz para conduzirmos as análises desta atividade, dada a composição da nossa equipe, composta por um Cientista da Computação e uma Médica. Optamos por realizar reuniões online para discutir a atividade, e apresento abaixo um print contendo apenas os comandos iniciais. Isso evidencia que o gráfico analisado no Colab seria o mesmo obtido pelo RCommander.

