

MCM5880 - Bioestatística I - Exercícios de Avaliação

Fernanda Fortti

19 de outubro de 2019

```
library(formatR)
library(kableExtra)
library(knitr)
library(pander)
library(tidyverse)
library(VennDiagram)

knitr::opts_chunk$set(base::options(scipen = 999,
                                     digits = 3))

knitr::opts_chunk$set(
  tidy.opts = base::list(blank = FALSE,
                        width.cutoff = 60),
  tidy = TRUE,
  comment = "Resposta:"
)
```

Informações gerais:

- A data máxima para entrega dos exercícios é 19/10/2019.
- A entrega dos 20 exercícios deveria ser feita exclusivamente pelo e-mail mcm5880@gmail.com.
- Os exercícios podem ser respondidos à mão e escaneados, ou digitados, em arquivo único.
- Os exercícios podem ser feitos individualmente ou em dupla. No email de entrega, incluir o nome completo do(a)(s) aluno(a)(s).
- Sugere-se fortemente que os exercícios de treinamento e avaliação sejam feitos no decorrer do curso e não ao seu término.

Exercícios

```
respostas <- base::data.frame(base::matrix(base::vector(mode = "numeric",
  length = 51), nrow = 51, ncol = 1))
base::rownames(respostas) <- base::c("1a", "1b", "1c", "1d",
  "1e", "1f", "1g", "2a", "2b", "2c", "2d", "2e", "2f", "3",
  "4a", "4b", "4c", "4d", "4e", "4f", "5", "6", "7", "8a",
  "8b", "8c", "8d", "8e", "9", "10", "11", "12", "13", "14",
  "15a", "15b", "15c", "15d", "15e", "16", "17", "18a", "18b",
  "18c", "18d", "19a", "19b", "19c", "20a", "20b", "20c")
base::colnames(respostas) <- "resposta"
```

1. Você tem as seguintes informações com relação a uma cidade:

População: 100.000.

Hipertensos: 25%.

Diabéticos: 15%.

Deprimidos: 12%.

Hipertensos e diabéticos: 4%.

Hipertensos e deprimidos: 5%.

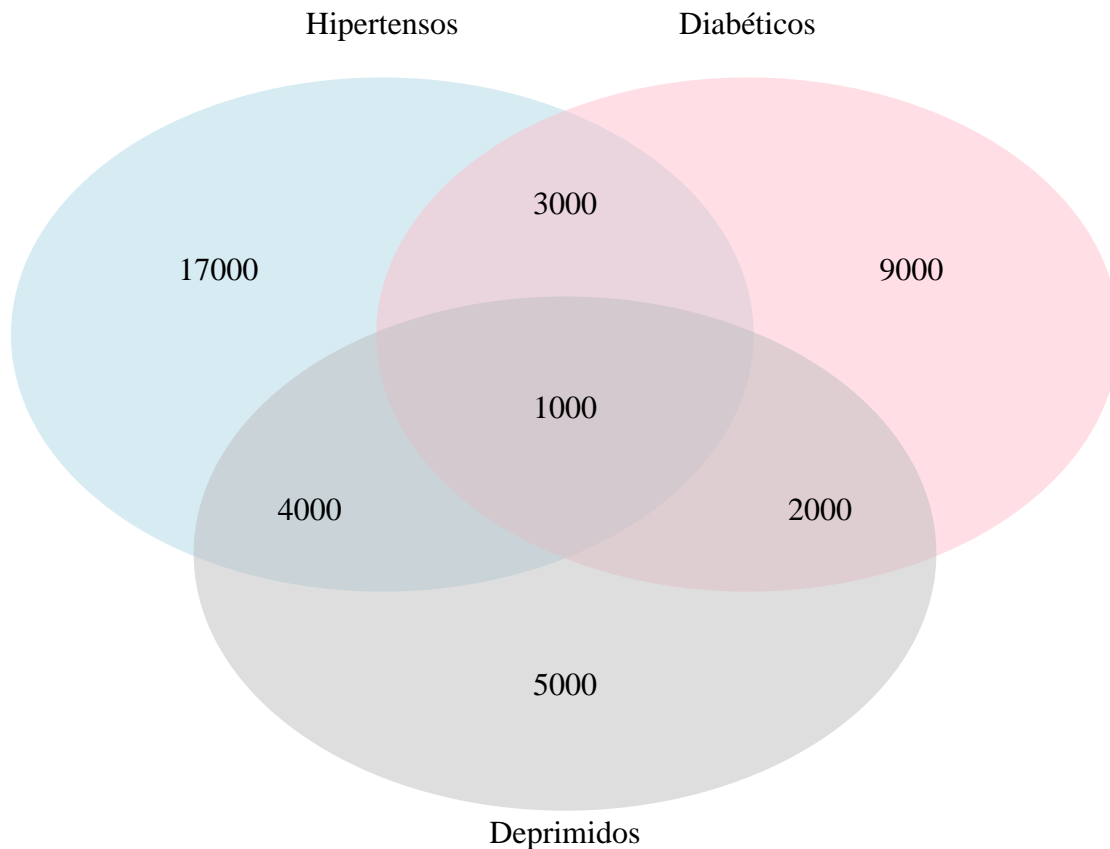
Diabéticos e deprimidos: 3%.
Hipertensos, diabéticos e deprimidos: 1%.

```
n_total <- 100000
exercicio_1a <- base::as.data.frame(x = base::matrix(ncol = 1,
  nrow = 7), row.names = base::c("hipertensos", "diabéticos",
  "deprimidos", "hipertensos e diabéticos", "hipertensos e deprimidos",
  "diabéticos e deprimidos", "diabéticos e hipertensos e deprimidos"))
exercicio_1a[1, 1] <- n_total * 0.25
exercicio_1a[2, 1] <- n_total * 0.15
exercicio_1a[3, 1] <- n_total * 0.12
exercicio_1a[4, 1] <- n_total * 0.04
exercicio_1a[5, 1] <- n_total * 0.05
exercicio_1a[6, 1] <- n_total * 0.03
exercicio_1a[7, 1] <- n_total * 0.01
knitr::kable(x = exercicio_1a, format = "latex", caption = "Condições crônicas.",
  col.names = base::c("n"), row.names = TRUE, digits = 3, align = "c") %>%
  kableExtra::kable_styling(kable_input = ., position = "center",
    latex_options = "hold_position")
```

Table 1: Condições crônicas.

	n
hipertensos	25000
diabéticos	15000
deprimidos	12000
hipertensos e diabéticos	4000
hipertensos e deprimidos	5000
diabéticos e deprimidos	3000
diabéticos e hipertensos e deprimidos	1000

```
VennDiagram::draw.triple.venn(area1 = exercicio_1a[1, 1], area2 = exercicio_1a[2,
  1], area3 = exercicio_1a[3, 1], n12 = exercicio_1a[4, 1],
  n13 = exercicio_1a[5, 1], n23 = exercicio_1a[6, 1], n123 = exercicio_1a[7,
  1], category = base::c("Hipertensos", "Diabéticos", "Deprimidos"),
  lty = base::c("blank", "blank", "blank"), fill = base::c("light blue",
  "pink", "gray"), alpha = base::c(0.5, 0.5, 0.5), cat.pos = base::c(0,
  0, 180), scaled = TRUE)
```



(a) Determine o número de pessoas com apenas uma das doenças.

```
respostas[1, 1] <- 17000 + 9000 + 5000
base::cat(base::noquote(respostas[1, 1]), sep = "")
```

Resposta: 31000

(b) Quantas pessoas têm ao menos duas dessas doenças?

```
respostas[2, 1] <- 4000 + 3000 + 2000 + 1000
base::cat(base::noquote(respostas[2, 1]), sep = "")
```

Resposta: 10000

(c) Quantos diabéticos são hipertensos ou deprimidos?

```
respostas[3, 1] <- 3000 + 1000 + 2000
base::cat(base::noquote(respostas[3, 1]), sep = "")
```

Resposta: 6000

(d) Quantas pessoas não têm nenhuma das doenças?

```
respostas[4, 1] <- 100000 - (17000 + 4000 + 1000 + 3000 + 9000 +
  2000 + 5000)
base::cat(base::noquote(respostas[4, 1]), sep = "")
```

Resposta: 59000

(e) Qual a probabilidade, nessa cidade, de um indivíduo hipertenso ser diabético?

```
respostas[5, 1] <- base::round(((3000 + 1000)/(17000 + 4000 +
  3000 + 1000)), digits = 3)
base::cat(base::noquote(respostas[5, 1]), sep = "")
```

Resposta: 0.16

(f) Qual a probabilidade, nessa cidade, de um indivíduo diabético e deprimido ser hipertenso?

```
respostas[6, 1] <- base::round(((1000)/(1000 + 2000)), digits = 3)
base::cat(base::noquote(respostas[6, 1]), sep = "")
```

Resposta: 0.333

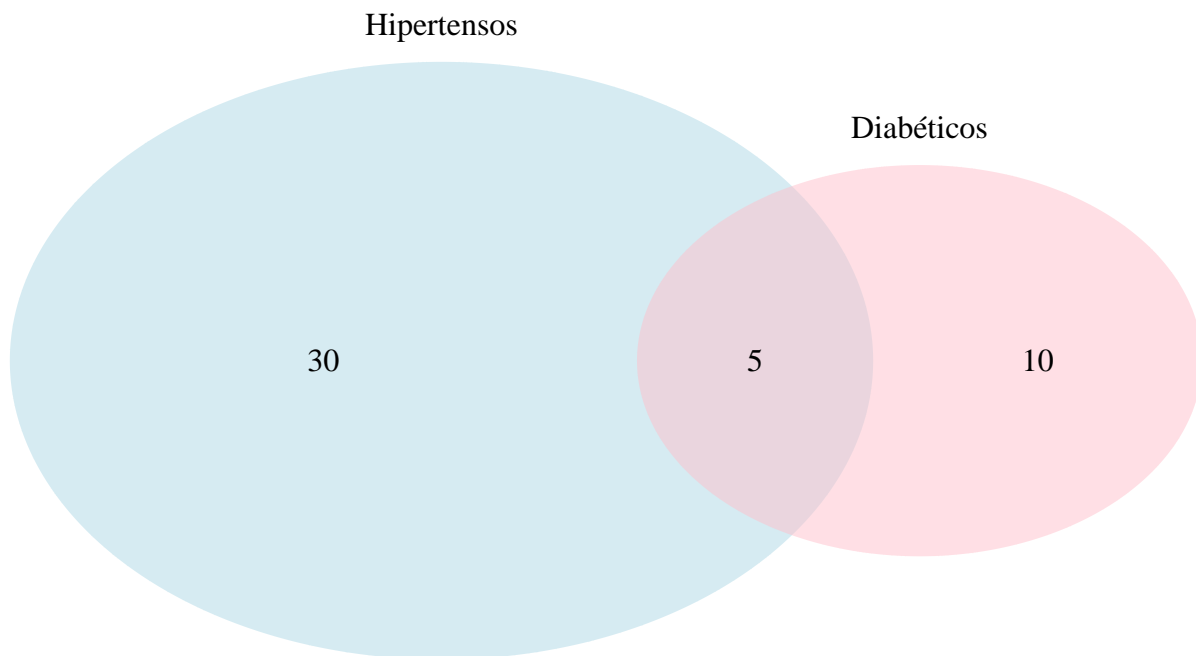
(g) Qual a probabilidade, nessa cidade, de um indivíduo não hipertenso ser diabético e deprimido?

```
respostas[7, 1] <- base::round(((2000)/(100000 - 25000)), digits = 3)
base::cat(base::noquote(respostas[7, 1]), sep = "")
```

Resposta: 0.027

2. Em determinada população, 30% das pessoas são apenas hipertensas, 10% são apenas diabéticas e 5% são hipertensas e diabéticas. Calcule a probabilidade, nessa população,

```
VennDiagram::draw.pairwise.venn(area1 = 35, area2 = 15, cross.area = 5,
  category = base::c("Hipertensos", "Diabéticos"), lty = base::c("blank",
    "blank"), fill = base::c("light blue", "pink"), alpha = base::c(0.5,
    0.5), cat.pos = base::c(0, 0), scaled = TRUE)
```



(a) De uma pessoa ser diabética

```
respostas[8, 1] <- base::round(((5 + 10)/100), digits = 3)
base::cat(base::noquote(respostas[8, 1]), sep = "")
```

Resposta: 0.15

(b) De uma pessoa ser hipertensa

```
respostas[9, 1] <- base::round(((30 + 5)/100), digits = 3)
base::cat(base::noquote(respostas[9, 1]), sep = "")
```

Resposta: 0.35

(c) De uma pessoa hipertensa ser diabética

```
respostas[10, 1] <- base::round(((5)/(30 + 5)), digits = 3)
base::cat(base::noquote(respostas[10, 1]), sep = "")
```

Resposta: 0.143

(d) De uma pessoa diabética ser hipertensa

```
respostas[11, 1] <- base::round(((5)/(5 + 10)), digits = 3)
base::cat(base::noquote(respostas[11, 1]), sep = "")
```

Resposta: 0.333

(e) De uma pessoa não hipertensa ser diabética

```
respostas[12, 1] <- base::round(((10)/((100) - (30 + 5))), digits = 3)
base::cat(base::noquote(respostas[12, 1]), sep = "")
```

Resposta: 0.154

(f) De uma pessoa não hipertensa não ser diabética

```
respostas[13, 1] <- base::round(((100 - (30 + 5) - 10)/((100) -
  (30 + 5))), digits = 3)
base::cat(base::noquote(respostas[13, 1]), sep = "")
```

Resposta: 0.846

3. Um paciente procura o consultório médico trazendo um exame positivo para determinada doença, que acomete 5% da população geral. O teste dá positivo em 90% dos doentes e em 20% dos não doentes. Qual a probabilidade desse paciente ter a doença?

```
prevalencia <- 0.05
odds_pre <- 0.05/(1 - 0.05)
sensibilidade <- 0.9
especificidade <- 1 - 0.2
razao_verossimilhanca_teste_positivo <- (sensibilidade)/(1 -
  especificidade)
odds_pos_teste_positivo <- odds_pre * razao_verossimilhanca_teste_positivo
prob_pos_teste_positivo <- (odds_pos_teste_positivo)/(1 + (odds_pos_teste_positivo))
respostas[14, 1] <- base::round(prob_pos_teste_positivo, digits = 3)
base::cat(base::noquote(respostas[14, 1]), sep = "")
```

Resposta: 0.191

4. Um estudo de teste diagnóstico observou o seguinte resultado quando comparado ao padrão-ouro:

Table 2: Teste diagnóstico.

	doentes	não doentes	total
teste positivo	240	70	310
teste negativo	60	630	690
total	300	700	1000

Baseado nesses resultados, calcule:

(a) Sensibilidade

```
respostas[15, 1] <- base::round((240/300), digits = 3)
base::cat(base::noquote(respostas[15, 1]), sep = "")
```

Resposta: 0.8

(b) Especificidade

```
respostas[16, 1] <- base::round((630/700), digits = 3)
base::cat(base::noquote(respostas[16, 1]), sep = "")
```

Resposta: 0.9

(c) Valor preditivo positivo

```
respostas[17, 1] <- base::round((240/310), digits = 3)
base::cat(base::noquote(respostas[17, 1]), sep = "")
```

Resposta: 0.774

(d) Valor preditivo negativo

```
respostas[18, 1] <- base::round((630/690), digits = 3)
base::cat(base::noquote(respostas[18, 1]), sep = "")
```

Resposta: 0.913

(e) Razão de verossimilhança para o resultado positivo

```
respostas[19, 1] <- base::round((respostas[15, 1]/(1 - respostas[16,
1])), digits = 3)
base::cat(base::noquote(respostas[19, 1]), sep = "")
```

Resposta: 8

(f) Razão de verossimilhança para o resultado negativo

```
respostas[20, 1] <- base::round(((1 - respostas[15, 1])/respostas[16,
1]), digits = 3)
base::cat(base::noquote(respostas[20, 1]), sep = "")
```

Resposta: 0.222

5. Dadas as probabilidades pré-teste e razões de verossimilhança indicados no quadro abaixo, calcule as probabilidades pós-teste:

Table 3: Teste diagnóstico.

probabilidade pré-teste	razão de verossimilhança
0.20	0.4
0.40	6.0
0.05	9.5

```

exercicio_5 <- base::data.frame(probabilidade_pre_teste = base::c(0.2,
  0.4, 0.05), razao_de_verossimilhanca = base::c(0.4, 6, 9.5))
exercicio_5$odds_pre_teste <- base::c(base::c((exercicio_5[1,
  1])/(1 - exercicio_5[1, 1])), base::c((exercicio_5[2, 1])/(1 -
  exercicio_5[2, 1])), base::c((exercicio_5[3, 1])/(1 - exercicio_5[3,
  1])))
exercicio_5$odds_pos_teste <- base::c(base::c((exercicio_5[1,
  2]) * (exercicio_5[1, 3])), base::c((exercicio_5[2, 2]) *
  (exercicio_5[2, 3])), base::c((exercicio_5[3, 2]) * (exercicio_5[3,
  3])))
exercicio_5$probabilidade_pos_teste <- base::round(base::c(base::c((exercicio_5[1,
  4])/(1 + exercicio_5[1, 4])), base::c((exercicio_5[2, 4])/(1 +
  exercicio_5[2, 4])), base::c((exercicio_5[3, 4])/(1 + exercicio_5[3,
  4]))), digits = 3)
respostas[21, 1] <- base::as.character("Probabilidade pós teste 1 = 0.091;
  probabilidade pós teste 2 = 0.800;
  probabilidade pós teste 3 = 0.333.")
knitr::kable(x = exercicio_5, format = "latex", caption = "Resposta:",
  col.names = base::c("probabilidade pré-teste", "razão de verossimilhança",
    "odds pré-teste", "odds pós-teste", "probabilidade pós-teste"),
  row.names = FALSE, digits = 3, align = "c") %>% kableExtra::kable_styling(kable_input = .,
  position = "center", latex_options = base::c("hold_position",
    "scale_down"))

```

Table 4: Resposta:

probabilidade pré-teste	razão de verossimilhança	odds pré-teste	odds pós-teste	probabilidade pós-teste
0.20	0.4	0.250	0.1	0.091
0.40	6.0	0.667	4.0	0.800
0.05	9.5	0.053	0.5	0.333

6. Um paciente procura o consultório médico com resultado de exame positivo, querendo saber a probabilidade dele ter a doença. A sensibilidade do exame é de 90%, a especificidade é 90% e a prevalência da doença, em indivíduos de mesma idade e sexo, é 5%.

```

prevalencia <- 0.05
odds_pre <- 0.05/(1 - 0.05)
sensibilidade <- 0.9
especificidade <- 0.9
razao_verossimilhanca_teste_positivo <- (sensibilidade)/(1 -
  especificidade)
odds_pos_teste_positivo <- odds_pre * razao_verossimilhanca_teste_positivo
prob_pos_teste_positivo <- (odds_pos_teste_positivo)/(1 + (odds_pos_teste_positivo))
VPP <- prob_pos_teste_positivo

```



```
respostas[22, 1] <- base::round(prob_pos_teste_positivo, digits = 3)
base::cat(base::noquote(respostas[22, 1]), sep = "")
```

Resposta: 0.321

7. Um paciente procura o consultório médico com resultado de exame negativo, querendo saber a probabilidade dele ter a doença. A sensibilidade do exame é de 90%, a especificidade é 90% e a prevalência da doença, em indivíduos de mesma idade e sexo, é 5%.

```
prevalencia <- 0.05
odds_pre <- 0.05/(1 - 0.05)
sensibilidade <- 0.9
especificidade <- 0.9
razao_verossimilhanca_teste_negativo <- (1 - sensibilidade)/(especificidade)
odds_pos_teste_negativo <- odds_pre * razao_verossimilhanca_teste_negativo
prob_pos_teste_negativo <- (odds_pos_teste_negativo)/(1 + (odds_pos_teste_negativo))
VPN <- 1 - prob_pos_teste_negativo
respostas[23, 1] <- base::round(prob_pos_teste_negativo, digits = 3)
base::cat(base::noquote(respostas[23, 1]), sep = "")
```

Resposta: 0.006

8. Dada a seguinte amostra ($n = 20$),

```
n_total <- 100000
exercicio_8 <- base::as.data.frame(x = base::matrix(ncol = 1,
  nrow = 20))
exercicio_8[1, 1] <- 4
exercicio_8[2, 1] <- 3
exercicio_8[3, 1] <- 7
exercicio_8[4, 1] <- 6
exercicio_8[5, 1] <- 0
exercicio_8[6, 1] <- 3
exercicio_8[7, 1] <- 3
exercicio_8[8, 1] <- 4
exercicio_8[9, 1] <- 5
exercicio_8[10, 1] <- 8
exercicio_8[11, 1] <- 0
exercicio_8[12, 1] <- 6
exercicio_8[13, 1] <- 3
exercicio_8[14, 1] <- 2
exercicio_8[15, 1] <- 1
exercicio_8[16, 1] <- 2
exercicio_8[17, 1] <- 5
exercicio_8[18, 1] <- 3
exercicio_8[19, 1] <- 8
exercicio_8[20, 1] <- 7
knitr::kable(x = exercicio_8, format = "latex", caption = "Amostra.",
  col.names = "observações", row.names = FALSE, digits = 3,
  align = "c") %>% kableExtra::kable_styling(kable_input = .,
  position = "center", latex_options = "hold_position")
```

Table 5: Amostra.

observações
4
3
7
6
0
3
3
4
5
8
0
6
3
2
1
2
5
3
8
7

Calcule:

(a) média

```
exercicio_8 <- base::c(4, 3, 7, 6, 0, 3, 3, 4, 5, 8, 0, 6, 3,
  2, 1, 2, 5, 3, 8, 7)
respostas[24, 1] <- base::round((base::sum(exercicio_8)/base::length(exercicio_8)),
  digits = 3)
base::cat(base::noquote(respostas[24, 1]), sep = "")
```

Resposta: 4

```
# ou, utilizar a função pronta: mean(exercicio_8)
```

(b) mediana

```
exercicio_8 <- base::sort(exercicio_8)
# sabendo que length(exercicio_8) = 20
respostas[25, 1] <- base::round(((exercicio_8[10] + exercicio_8[11])/2),
  digits = 3)
base::cat(base::noquote(respostas[25, 1]), sep = "")
```

Resposta: 3.5

```
# ou, utilizar a função pronta: median(exercicio_8)
```

(c) moda(s)

```
getmode <- function(x) {
  uniq_x <- base::unique(x)
  uniq_x[base::which.max(base::tabulate(base::match(x, uniq_x)))]
}
x <- exercicio_8
respostas[26, 1] <- base::round(getmode(x), digits = 3)
base::cat(base::noquote(respostas[26, 1]), sep = "")
```

Resposta: 3

(d) variância

```
variancia <- (((exercicio_8[1] - base::mean(exercicio_8))^2) +
  ((exercicio_8[2] - base::mean(exercicio_8))^2) + ((exercicio_8[3] -
  base::mean(exercicio_8))^2) + ((exercicio_8[4] - base::mean(exercicio_8))^2) +
  ((exercicio_8[5] - base::mean(exercicio_8))^2) + ((exercicio_8[6] -
  base::mean(exercicio_8))^2) + ((exercicio_8[7] - base::mean(exercicio_8))^2) +
  ((exercicio_8[8] - base::mean(exercicio_8))^2) + ((exercicio_8[9] -
  base::mean(exercicio_8))^2) + ((exercicio_8[10] - base::mean(exercicio_8))^2) +
  ((exercicio_8[11] - base::mean(exercicio_8))^2) + ((exercicio_8[12] -
  base::mean(exercicio_8))^2) + ((exercicio_8[13] - base::mean(exercicio_8))^2) +
  ((exercicio_8[14] - base::mean(exercicio_8))^2) + ((exercicio_8[15] -
  base::mean(exercicio_8))^2) + ((exercicio_8[16] - base::mean(exercicio_8))^2) +
  ((exercicio_8[17] - base::mean(exercicio_8))^2) + ((exercicio_8[18] -
  base::mean(exercicio_8))^2) + ((exercicio_8[19] - base::mean(exercicio_8))^2) +
  ((exercicio_8[20] - base::mean(exercicio_8))^2))/(base::length(exercicio_8) -
  1)
respostas[27, 1] <- base::round(variancia, digits = 3)
base::cat(base::noquote(respostas[27, 1]), sep = "")
```

Resposta: 6

```
# ou, utilizar a função pronta: var(exercicio_8)
```

(e) desvio padrão

```
respostas[28, 1] <- base::round(base::sqrt(variancia), digits = 3)
base::cat(base::noquote(respostas[28, 1]), sep = "")
```

Resposta: 2.449

```
# ou, utilizar a função pronta: sd(exercicio_8)
```

9. Determine um valor z tal que a probabilidade do valor de uma observação estar entre $[-z; +z]$ é de 99%

Observação: No caso deste exercício, será considerada a distribuição normal padrão, que apresenta média igual 0 e desvio padrão igual a 1.

```
respostas[29, 1] <- base::round(stats::qnorm(p = 0.99, mean = 0,
  sd = 1), digits = 3)
base::cat(base::noquote(respostas[29, 1]), sep = "")
```

Resposta: 2.326

*# esta função converte valores de probabilidades (p;
semelhantes a quantis de distribuição) em valores de z de
uma distribuição normal com determinada média (mean) e
determinado desvio padrão (sd)*

10. Os valores de pressão arterial sistólica de uma população têm distribuição normal $N(120, 625)$. Encontre a probabilidade de observar níveis entre 140 e 160 mmHg.

Observação: No caso deste exercício, será considerada a distribuição normal padrão, que apresenta média igual 0 e desvio padrão igual a 1.

```
X1 <- 140
X2 <- 160
mu <- 120
sigma <- base::sqrt(625)
z1 <- (X1 - mu)/sigma
p1 <- stats::pnorm(q = z1, mean = 0, sd = 1)
z2 <- (X2 - mu)/sigma
p2 <- stats::pnorm(q = z2, mean = 0, sd = 1)
p <- (p2 - p1)
respostas[30, 1] <- base::round(p, digits = 3)
base::cat(base::noquote(respostas[30, 1]), sep = "")
```

Resposta: 0.157

11. Construa o IC 95% para a prevalência de hipertensão obtida a partir de uma amostra de 19600 indivíduos que verificou que 4900 deles tinham a doença.

Observação: No caso deste exercício, será considerada a distribuição normal padrão, que apresenta média igual 0 e desvio padrão igual a 1.

```
z <- stats::qnorm(p = 0.95, mean = 0, sd = 1)
amostra_total <- 19600
amostra_doentes <- 4900
prevalencia <- (amostra_doentes/amostra_total)
limite_inferior <- base::round(prevalencia - z * (base::sqrt((1/(4 *
  amostra_total)))), digits = 3)
limite_superior <- base::round(prevalencia + z * (base::sqrt((1/(4 *
  amostra_total)))), digits = 3)
respostas[31, 1] <- base::as.character("IC=[0.244;0.256]")
base::cat(base::noquote(respostas[31, 1]), sep = "")
```

Resposta: IC=[0.244;0.256]

12. Construa o IC 99% para o índice de falsos-positivos de um determinado exame baseado no fato de que uma amostra de 40000 exames positivos havia 800 indivíduos sem a doença.

Observação: No caso deste exercício, será considerada a distribuição normal padrão, que apresenta média igual 0 e desvio padrão igual a 1.

```

z <- stats::qnorm(p = 0.99, mean = 0, sd = 1)
amostra_total_positivos <- 40000
amostra_falsos_positivos <- 800
prop_falsos_positivos <- (amostra_falsos_positivos/amostra_total_positivos)
limite_inferior <- base::round(prop_falsos_positivos - z * (base::sqrt((1/(4 *
  amostra_total_positivos)))), digits = 3)
limite_superior <- base::round(prop_falsos_positivos + z * (base::sqrt((1/(4 *
  amostra_total_positivos)))), digits = 3)
respostas[32, 1] <- base::as.character("IC=[0.014;0.026]")
base::cat(base::noquote(respostas[32, 1]), sep = "")

```

Resposta: IC=[0.014;0.026]

13. Construa o IC 95% para a média de glicemia de uma amostra de $n = 100$ que resultou em média 100 mg/dl e variância 100 (mg/dl)²

```

t <- stats::qt(p = 0.95, df = (amostra_total - 1))
amostra_total <- 100
amostra_media <- 100
S <- base::sqrt(100)
limite_inferior <- base::round(amostra_media - t * (S/base::sqrt(amostra_total)),
  digits = 3)
limite_superior <- base::round(amostra_media + t * (S/base::sqrt(amostra_total)),
  digits = 3)
respostas[33, 1] <- base::as.character("IC=[98.355;101.645]")
base::cat(base::noquote(respostas[33, 1]), sep = "")

```

Resposta: IC=[98.355;101.645]

14. Construa o IC 99% para a média de valores de natremia (presença de sódio no sangue) de uma amostra de $n = 144$ que resultou em média 140 mEq/L e variância 9 (mEq/L)²

```

t <- stats::qt(p = 0.99, df = (amostra_total - 1))
amostra_total <- 144
amostra_media <- 140
S <- base::sqrt(9)
limite_inferior <- base::round(amostra_media - t * (S/base::sqrt(amostra_total)),
  digits = 3)
limite_superior <- base::round(amostra_media + t * (S/base::sqrt(amostra_total)),
  digits = 3)
respostas[34, 1] <- base::as.character("IC=[139.409;140.591]")
base::cat(base::noquote(respostas[34, 1]), sep = "")

```

Resposta: IC=[139.409;140.591]

15. Explique os seguintes conceitos:

(a) Poder do estudo de 80%

```

respostas[35, 1] <- base::as.character("Significa que, se a associação
  de interesse existe na população, há 80% de probabilidade que
  isso seja evidenciado no estudo.")
base::cat(base::noquote(respostas[35, 1]))

```

Resposta: Significa que, se a associação
Resposta: de interesse existe na população, há 80% de probabilidade que
Resposta: isso seja evidenciado no estudo.

(b) Nível de significância de 5%

```
respostas[36, 1] <- base::as.character("0 nível de significância é um  
valor de probabilidade arbitrado no início do estudo, a ser  
comparado com o valor de probabilidade encontrado no estudo  
(p-valor), com o intuito de decidir se a hipótese nula será  
ou não rejeitada. Quando o nível de significância é 5%, a  
hipótese nula só será rejeitada caso o valor de probabilidade  
encontrado no estudo tenha valor inferior a 5%.")  
base::cat(base::noquote(respostas[36, 1]))
```

Resposta: 0 nível de significância é um
Resposta: valor de probabilidade arbitrado no início do estudo, a ser
Resposta: comparado com o valor de probabilidade encontrado no estudo
Resposta: (p-valor), com o intuito de decidir se a hipótese nula será
Resposta: ou não rejeitada. Quando o nível de significância é 5%, a
Resposta: hipótese nula só será rejeitada caso o valor de probabilidade
Resposta: encontrado no estudo tenha valor inferior a 5%.

(c) Erro beta é de 10%

```
respostas[37, 1] <- base::as.character("Probabilidade de 10% da  
hipótese nula ser aceita quando, na verdade, ela é falsa.  
Esta probabilidade está associada ao poder do estudo (1 - poder).")  
base::cat(base::noquote(respostas[37, 1]))
```

Resposta: Probabilidade de 10% da
Resposta: hipótese nula ser aceita quando, na verdade, ela é falsa.
Resposta: Esta probabilidade está associada ao poder do estudo (1 - poder).

(d) Erro alpha é de 5%

```
respostas[38, 1] <- base::as.character("Probabilidade de 5% da  
hipótese nula ser rejeitada quando, na verdade, ela é  
verdadeira. Esta probabilidade está associada ao nível de  
significância do estudo.")  
base::cat(base::noquote(respostas[38, 1]))
```

Resposta: Probabilidade de 5% da
Resposta: hipótese nula ser rejeitada quando, na verdade, ela é
Resposta: verdadeira. Esta probabilidade está associada ao nível de
Resposta: significância do estudo.

(e) p-valor de 0.012

```
respostas[39, 1] <- base::as.character("Caso a hipóteses nula
  seja verdadeira na população, a probabilidade do resultado
  encontrado na amostra do estudo ocorrer é de 1.2%. Ou seja,
  a probabilidade do resultado observado na amostra ter ocorrido
  devido ao acaso na seleção dessa amostra é de 1.2%.")
base::cat(base::noquote(respostas[39, 1]))
```

Resposta: Caso a hipóteses nula

Resposta: seja verdadeira na população, a probabilidade do resultado

Resposta: encontrado na amostra do estudo ocorrer é de 1.2%. Ou seja,

Resposta: a probabilidade do resultado observado na amostra ter ocorrido

Resposta: devido ao acaso na seleção dessa amostra é de 1.2%.

16. Um pesquisador, querendo concluir se existe associação entre obesidade e transtorno do pânico (TP), observou os seguintes resultados em uma amostra de 300 indivíduos:

Table 6: Obesidade e transtorno do pânico (TP).

	obeso	não obeso	total
com TP	20	15	35
sem TP	130	135	265
total	150	150	300

Considerando um nível de significância de 5%, é possível dizer que há associação entre essas condições clínicas?

```
alpha <- 0.05
m <- 2
n <- 2
df <- (m - 1) * (n - 1)
qui_quadrado_valor_critico <- stats::qchisq(p = 0.95, df)
observado <- base::data.frame()
observado[1, 1] <- 20
observado[1, 2] <- 15
observado[1, 3] <- 35
observado[2, 1] <- 130
observado[2, 2] <- 135
observado[2, 3] <- 265
observado[3, 1] <- 150
observado[3, 2] <- 150
observado[3, 3] <- 300
base::row.names(observado) <- base::c("com TP", "sem TP", "total")
esperado <- base::data.frame()
esperado[1, 1] <- (150 * 35)/300
esperado[1, 2] <- (150 * 35)/300
esperado[1, 3] <- 35
esperado[2, 1] <- (150 * 265)/300
esperado[2, 2] <- (150 * 265)/300
esperado[2, 3] <- 265
esperado[3, 1] <- 150
esperado[3, 2] <- 150
esperado[3, 3] <- 300
```

```
base::row.names(esperado) <- base::c("com TP", "sem TP", "total")
knitr::kable(x = observado, format = "latex", caption = "Observado.",
  col.names = base::c("obeso", "não obeso", "total"), row.names = TRUE,
  digits = 3, align = "c") %>% kableExtra::kable_styling(kable_input = .,
  position = "center", latex_options = "hold_position")
knitr::kable(x = esperado, format = "latex", caption = "Esperado.",
  col.names = base::c("obeso", "não obeso", "total"), row.names = TRUE,
  digits = 3, align = "c") %>% kableExtra::kable_styling(kable_input = .,
  position = "center", latex_options = "hold_position")
```

Table 7: Observado.

	obeso	não obeso	total
com TP	20	15	35
sem TP	130	135	265
total	150	150	300

Table 8: Esperado.

	obeso	não obeso	total
com TP	17.5	17.5	35
sem TP	132.5	132.5	265
total	150.0	150.0	300

```
qui_quadrado <- ((observado[1, 1] - esperado[1, 1])^2/esperado[1,
  1]) + ((observado[1, 2] - esperado[1, 2])^2/esperado[1, 2]) +
  ((observado[2, 1] - esperado[2, 1])^2/esperado[2, 1]) + ((observado[2,
  2] - esperado[2, 2])^2/esperado[2, 2])
respostas[40, 1] <- base::ifelse(qui_quadrado > qui_quadrado_valor_critico,
  "Sim. É possível dizer que há associação entre as condições clínicas",
  "Não. Não é possível dizer que há associação entre as condições clínicas.")
base::cat(base::noquote("Qui quadrado = 0.809; qui-quadrado valor crítico = 3.84."))
```

Resposta: Qui quadrado = 0.809; qui-quadrado valor crítico = 3.84.

```
base::cat(base::noquote(respostas[40, 1]), sep = "")
```

Resposta: Não. Não é possível dizer que há associação entre as condições clínicas.

17.Duas medicações estão sendo testadas para tratamento de mieloma múltiplo em 250 pacientes voluntários. Os resultados estão expressos na tabela a seguir:

Table 9: Medicações para tratamento de mieloma múltiplo.

	eficaz	não eficaz	total
droga A	80	70	150
droga B	60	40	100
total	140	110	250

Considerando um nível de significância de 5%, é possível dizer que uma medicação é superior à outra? Se sim, qual delas?

```
alpha <- 0.05
m <- 2
n <- 2
df <- (m - 1) * (n - 1)
qui_quadrado_valor_critico <- stats::qchisq(p = 0.95, df)
observado <- base::data.frame()
observado[1, 1] <- 80
observado[1, 2] <- 70
observado[1, 3] <- 150
observado[2, 1] <- 60
observado[2, 2] <- 40
observado[2, 3] <- 100
observado[3, 1] <- 140
observado[3, 2] <- 110
observado[3, 3] <- 250
base::rownames(observado) <- base::c("droga A", "droga B", "total")
esperado <- base::data.frame()
esperado[1, 1] <- (150 * 140)/250
esperado[1, 2] <- (150 * 110)/250
esperado[1, 3] <- 150
esperado[2, 1] <- (100 * 140)/250
esperado[2, 2] <- (100 * 110)/250
esperado[2, 3] <- 100
esperado[3, 1] <- 140
esperado[3, 2] <- 110
esperado[3, 3] <- 250
base::rownames(esperado) <- base::c("droga A", "droga B", "total")
knitr::kable(x = observado, format = "latex", caption = "Observado.",
  col.names = base::c("eficaz", "não eficaz", "total"), row.names = TRUE,
  digits = 3, align = "c") %>% kableExtra::kable_styling(kable_input = .,
  position = "center", latex_options = "hold_position")
knitr::kable(x = esperado, format = "latex", caption = "Esperado.",
  col.names = base::c("eficaz", "não eficaz", "total"), row.names = TRUE,
  digits = 3, align = "c") %>% kableExtra::kable_styling(kable_input = .,
  position = "center", latex_options = "hold_position")
```

Table 10: Observado.

	eficaz	não eficaz	total
droga A	80	70	150
droga B	60	40	100
total	140	110	250

```
qui_quadrado <- ((observado[1, 1] - esperado[1, 1])^2/esperado[1,
  1]) + ((observado[1, 2] - esperado[1, 2])^2/esperado[1, 2]) +
  ((observado[2, 1] - esperado[2, 1])^2/esperado[2, 1]) + ((observado[2,
  2] - esperado[2, 2])^2/esperado[2, 2])
respostas[41, 1] <- base::ifelse(qui_quadrado > qui_quadrado_valor_critico,
  "Sim. É possível dizer que uma medicação é superior a outra",
```

Table 11: Esperado.

	eficaz	não eficaz	total
droga A	84	66	150
droga B	56	44	100
total	140	110	250

```
"Não. Não é possível dizer que uma medicação é superior à outra.")
base::cat(base::noquote("Qui quadrado = 1.08; qui-quadrado valor crítico = 3.84."))
```

Resposta: Qui quadrado = 1.08; qui-quadrado valor crítico = 3.84.

```
base::cat(base::noquote(respostas[41, 1]), sep = "")
```

Resposta: Não. Não é possível dizer que uma medicação é superior à outra.

18.Considere um estudo hipotético de caso-controle (50 casos e 50 controles) para estudar a associação entre abuso/dependência de bebida alcoólica (variável dicotômica, sim/não) e infarto do miocárdio (variável dicotômica, sim/não). Há no banco de dados variáveis sociodemográficas (sexo, idade, escolaridade, raça) história de fatores de risco cardiovasculares prévios (hipertensão, tabagismo, sedentarismo, diabetes, dislipidemia) e dados laboratoriais (colesterol total, LDL-colesterol, HDL-colesterol, triglicérides, glicemia de jejum).

(a) Cite um teste estatístico que permite avaliar a associação de interesse (abuso/dependência de bebida alcoólica e infarto do miocárdio) sem ajuste para outras variáveis

```
respostas[42, 1] <- base::as.character("Teste qui-quadrado.")
base::cat(base::noquote(respostas[42, 1]))
```

Resposta: Teste qui-quadrado.

(b) Cite um teste estatístico que permite avaliar a associação de interesse (abuso/dependência de bebida alcoólica e infarto do miocárdio) com ajuste para outras variáveis

```
respostas[43, 1] <- base::as.character("Regressão logística.")
base::cat(base::noquote(respostas[43, 1]))
```

Resposta: Regressão logística.

(c) Cite um exemplo de associação entre duas variáveis do banco de dados que possa ser avaliada utilizando-se o teste t

```
respostas[44, 1] <- base::as.character("Associação entre raça e glicemia de jejum.")
base::cat(base::noquote(respostas[44, 1]))
```

Resposta: Associação entre raça e glicemia de jejum.

(d) Cite um exemplo de associação entre duas variáveis do banco de dados que possa ser avaliada utilizando-se regressão linear simples

```
respostas[45, 1] <- base::as.character("Associação entre colesterol total e glicemia de jejum.")
base::cat(base::noquote(respostas[45, 1]))
```

Resposta: Associação entre colesterol total e glicemia de jejum.

19. Em um estudo de coorte de 3000 fumantes e 5000 não fumantes foram detectados os resultados apresentados na tabela abaixo:

Table 12: Doença arterial coronariana (DAC) e tabagismo.

	DAC presente	DAC ausente	total
fumantes	84	2916	3000
não fumantes	87	4913	5000
total	171	7829	8000

(a) Qual a incidência de doença coronariana entre fumantes e não fumantes?

```
incidencia_geral <- 171/8000
incidencia_fumantes <- 84/3000
incidencia_nao_fumantes <- 87/5000
respostas[46, 1] <- base::as.character("Incidência geral = 0.0214;
    incidência em fumantes = 0.028;
    incidência em não fumantes = 0.0174.")
base::cat(base::noquote(respostas[46, 1]))
```

Resposta: Incidência geral = 0.0214;
 Resposta: incidência em fumantes = 0.028;
 Resposta: incidência em não fumantes = 0.0174.

(b) Quais as medidas de associação mais apropriadas para este tipo de estudo?

```
respostas[47, 1] <- base::as.character("Risco relativo (RR),
    odds ratio (OR) e
    risco atribuível (RA).")
base::cat(base::noquote(respostas[47, 1]))
```

Resposta: Risco relativo (RR),
 Resposta: odds ratio (OR) e
 Resposta: risco atribuível (RA).

(c) Calcule as medidas de associação que você citou na questão anterior e as interprete.

```
risco_relativo <- (incidencia_fumantes)/(incidencia_nao_fumantes)
odds_fumantes <- ((84/3000)/(2916/3000))
odds_nao_fumantes <- ((87/5000)/(4913/5000))
odds_ratio <- odds_fumantes/odds_nao_fumantes
risco_atribuivel <- incidencia_fumantes - incidencia_nao_fumantes
respostas[48, 1] <- base::as.character("Risco relativo (RR) = 1.61;
    odds ratio (OR) = 1.63;
    risco atribuível (RA) = 0.0106.")
base::cat(base::noquote(respostas[48, 1]))
```

Resposta: Risco relativo (RR) = 1.61;
 Resposta: odds ratio (OR) = 1.63;
 Resposta: risco atribuível (RA) = 0.0106.

20. Um estudo de associação entre câncer de pulmão e tabagismo, com 200 casos e 400 controles mostrou os resultados mostrados na tabela abaixo:

```
exercicio_20 <- base::as.data.frame(x = base::matrix(ncol = 3,
  nrow = 3), row.names = base::c("fumantes", "não fumantes",
  "total"))
exercicio_20[1, 1] <- 112
exercicio_20[1, 2] <- 176
exercicio_20[1, 3] <- 288
exercicio_20[2, 1] <- 88
exercicio_20[2, 2] <- 224
exercicio_20[2, 3] <- 312
exercicio_20[3, 1] <- 200
exercicio_20[3, 2] <- 400
exercicio_20[3, 3] <- 600
knitr::kable(x = exercicio_20, format = "latex", caption = "Câncer de pulmão e tabagismo.",
  col.names = base::c("casos", "controle", "total"), row.names = TRUE,
  digits = 3, align = "c") %>% kableExtra::kable_styling(kable_input = .,
  position = "center", latex_options = "hold_position")
```

Table 13: Câncer de pulmão e tabagismo.

	casos	controle	total
fumantes	112	176	288
não fumantes	88	224	312
total	200	400	600

(a) Por que você não pode calcular o risco relativo em estudos de caso controle?

```
respostas[49, 1] <- base::as.character("Em estudos de caso controle,
  não é possível calcular a incidência da doença, pois começamos
  o estudo com pessoas doentes e não doentes. Uma vez que o risco
  relativo é uma medida de associação definida pela razão das
  incidências de determinado evento em diferentes grupos, não é
  possível calcular o risco relativo em estudos de caso controle.")
base::cat(base::noquote(respostas[49, 1]))
```

Resposta: Em estudos de caso controle,
 Resposta: não é possível calcular a incidência da doença, pois começamos
 Resposta: o estudo com pessoas doentes e não doentes. Uma vez que o risco
 Resposta: relativo é uma medida de associação definida pela razão das
 Resposta: incidências de determinado evento em diferentes grupos, não é
 Resposta: possível calcular o risco relativo em estudos de caso controle.

(b) Este tipo de estudo é o ideal para testar a associação entre tabagismo e câncer pulmão? Justifique.

```

respostas[50, 1] <- base::as.character("Sim. Em câncer o período de
latência (período de tempo compreendido entre a data da primeira
exposição e a data do diagnóstico) é bastante longo. O período de
latência entre o início da exposição à fumaça de tabaco e a ocorrência
do câncer de pulmão é de cerca de três décadas (WEISS, 1997 apud
WUNSCH FILHO et al, 2010). Dessa forma, realizar um estudo de coorte
seria extremamente custoso e demorado; enquanto que, realizar um
estudo de caso controle, reduz o investimento financeiro e o tempo
necessário para que se possa tirar conclusões sobre a existência de
associação entre tabagismo e câncer de pulmão. Além disso, de forma
geral (no Brasil e no mundo), a exposição ao tabaco não é uma exposição
rara a ponto de tornar um estudo de caso-controle ineficiente. Portanto,
apesar de em um estudo de caso controle não ser possível calcular a
medida de associação que permite investigar se o grupo de fumantes
apresenta ou não maior risco de desenvolver câncer de pulmão, em
comparação ao grupo de não fumantes (risco relativo), mas o cálculo
de odds ratio é um medida que pode ser utilizada nesse tipo de estudo
para investigar a existência de associação entre a exposição (tabagismo)
e o desfecho (câncer de pulmão).")
base::cat(base::noquote(respostas[50, 1]))

```

Resposta: Sim. Em câncer o período de

Resposta: latência (período de tempo compreendido entre a data da primeira

Resposta: exposição e a data do diagnóstico) é bastante longo. O período de

Resposta: latência entre o início da exposição à fumaça de tabaco e a ocorrência

Resposta: do câncer de pulmão é de cerca de três décadas (WEISS, 1997 apud

Resposta: WUNSCH FILHO et al, 2010). Dessa forma, realizar um estudo de coorte

Resposta: seria extremamente custoso e demorado; enquanto que, realizar um

Resposta: estudo de caso controle, reduz o investimento financeiro e o tempo

Resposta: necessário para que se possa tirar conclusões sobre a existência de

Resposta: associação entre tabagismo e câncer de pulmão. Além disso, de forma

Resposta: geral (no Brasil e no mundo), a exposição ao tabaco não é uma exposição

Resposta: rara a ponto de tornar um estudo de caso-controle ineficiente. Portanto,

Resposta: apesar de em um estudo de caso controle não ser possível calcular a

Resposta: medida de associação que permite investigar se o grupo de fumantes

Resposta: apresenta ou não maior risco de desenvolver câncer de pulmão, em

Resposta: comparação ao grupo de não fumantes (risco relativo), mas o cálculo

Resposta: de odds ratio é um medida que pode ser utilizada nesse tipo de estudo

Resposta: para investigar a existência de associação entre a exposição (tabagismo)

Resposta: e o desfecho (câncer de pulmão).

(c) Calcule o odds ratio (razão de chances) para câncer de pulmão em relação ao hábito de fumar.

```

odds_fumantes <- ((112/288)/(176/288))
odds_nao_fumantes <- ((88/312)/(224/312))
odds_ratio <- base::round(odds_fumantes/odds_nao_fumantes, digits = 3)
respostas[51, 1] <- odds_ratio
base::cat(base::noquote(respostas[51, 1]))

```

Resposta: 1.62

```
utils::write.csv(respostas, file = "MCM5880_respostas_exercicios_avaliacao_FernandaFortti.csv")
```