

## AVALIAÇÃO DE TESTES DIAGNÓSTICOS ATRAVÉS DE SIMULAÇÃO COMPUTACIONAL

### EVALUATION OF DIAGNOSTIC TESTS THROUGH COMPUTATIONAL SIMULATION

**Cleber Giuglioli Carrasco**, Mestre em Estatística, UEG/CET, cleber.carrasco@ueg.br  
**Gustavo Henrique Gomes Muniz**, Graduado em Matemática, UEG/CET, gustavohgmuniz@gmail.com

**Resumo:** O uso de testes diagnósticos requer, além de considerações clínicas, o conhecimento de algumas medidas sobre o seu desempenho, em particular, a sensibilidade e a especificidade. Essas medidas são definidas respectivamente como a probabilidade de um teste dar positivo dado que o indivíduo seja portador de uma determinada doença e a probabilidade de um teste dar negativo dado que o indivíduo não seja portador dessa doença. Este trabalho teve como objetivo estudar os testes diagnósticos através de um ponto de vista estatístico, ou seja, avaliar a influência que medidas como a sensibilidade, especificidade e prevalência tem sobre o desempenho de um teste diagnóstico. Com a utilização da linguagem de programação *Python*, realizou-se um estudo de simulação computacional para verificar o comportamento desses testes ao variar os valores dessas medidas. Concluiu-se que um teste com maior sensibilidade acarreta a ocorrência de mais resultados Verdadeiro Positivo e menos resultados Falso Negativo, enquanto um teste com alta especificidade tem mais resultados Verdadeiro Negativo e menos do tipo Falso Positivo. Observou-se também que uma doença com grande prevalência provoca mais resultados Verdadeiro Negativo e Falso Positivo, enquanto uma doença mais rara gera mais resultados Verdadeiro Positivo e Falso Negativo.

**Palavras-chave:** Especificidade; Linguagem Python; Prevalência; Probabilidade; Sensibilidade.

**Abstract:** The use of diagnostic tests requires, in addition to clinical considerations, the knowledge of some measures of their performance, in particular, sensitivity and specificity. These measures are defined respectively as the probability of a test being positive given that the individual is a carrier of a certain disease and the probability of a test being negative given that the individual is not a carrier of that disease. This work aimed to study diagnostic tests from a statistical point of view, that is, to evaluate the influence that measures such as sensitivity, specificity and prevalence have on the performance of a diagnostic test. Using the Python programming language, a computational simulation study was carried out to verify the behavior of these tests when the values of these measures vary. It was concluded that a test with greater sensitivity results in the occurrence of more True Positive results and fewer False Negative results, while a test with high specificity has more True Negative results and fewer False Positive results. It was also observed that a disease with high prevalence causes more True Negative and False Positive results, while a rarer disease generates more True Positive and False Negative results.

**Keywords:** Specificity; Python language; Prevalence; Probability; Sensitivity.

### INTRODUÇÃO

Na Ciências da Saúde, a Estatística vêm sendo muito utilizada e proporcionando grandes avanços nos últimos anos. Em particular, os métodos estatísticos vêm

sendo aplicados na área da saúde para estimar a prevalência de uma determinada doença, ou seja, a proporção de pessoas com essa doença na população e para avaliar o desempenho e a qualidade de testes diagnósticos. Esses testes são apresentados como

métodos capazes de classificar a presença ou não de uma determinada doença (ou condição) nos indivíduos, com certa probabilidade de acerto. E para avaliar e determinar essa probabilidade é necessário utilizar os métodos estatísticos.

Na literatura médica, os estudos sobre a qualidade de testes diagnósticos se preocupam comumente em avaliar a probabilidade de o teste classificar corretamente o indivíduo como tendo ou não determinada doença, ou seja, em avaliar a probabilidade do resultado do teste ser positivo dado que o indivíduo apresenta a doença, e a probabilidade de o resultado do teste ser negativo dado que o indivíduo não apresenta a doença. Essas probabilidades, são respectivamente conhecidas na nomenclatura médica como sendo a sensibilidade e a especificidade do teste diagnóstico.

Dessa forma, a sensibilidade pode ser considerada como sendo a probabilidade de diagnósticos verdadeiros positivos, isto é, resultados positivos corretos em pessoas portadoras da doença, ou ainda, a probabilidade condicional do resultado do teste ser positivo dado que o indivíduo é portador da doença. E, de forma análoga, a especificidade pode ser considerada a probabilidade de diagnósticos verdadeiros negativos, isto é, resultados negativos corretos em pessoas não portadoras da doença, ou seja, a probabilidade condicional do resultado do teste ser negativo dado que o indivíduo não é portador da doença.

Neste trabalho propõem-se avaliar as medidas de sensibilidade e especificidade de testes diagnósticos através de um estudo de simulação computacional e, ainda analisar o efeito das medidas de sensibilidade e especificidade desses testes na estimação da prevalência de uma doença conforme os valores da sensibilidade e especificidade variarem.

No estudo de simulação será utilizado um procedimento computacional próprio o qual será implementado no *software Python*. A escolha da utilização do *Python* neste trabalho é que, além de ser gratuito, apresenta código-fonte aberto, podendo ser modificado ou implementado com novos procedimentos desenvolvidos pelo usuário.

## MATERIAIS E MÉTODOS

A utilização de testes diagnósticos requer, além de considerações clínicas, o conhecimento de medidas que caracterizam a sua qualidade. A sensibilidade e a especificidade são medidas usadas para produzir

decisões precisas ao utilizar testes diagnósticos (SOARES; SIQUEIRA, 2012).

A sensibilidade de um teste diagnóstico é a sua capacidade de identificar corretamente indivíduos que apresentam a doença em questão (NUNES *et al*, 2015), ou seja, a probabilidade do teste dar positivo dado que o indivíduo apresenta a doença (VIEIRA, 2010), isto é:

$$\text{sensibilidade} = P(T_+|D_+) \quad (1)$$

A especificidade é a capacidade do teste diagnóstico em identificar corretamente quem não tem a doença investigada (NUNES *et al*, 2015), ou seja, a probabilidade do teste dar negativo dado que o indivíduo não apresenta a doença (VIEIRA, 2010), isto é:

$$\text{especificidade} = P(T_-|D_-) \quad (2)$$

A prevalência de uma doença na população é definida como sendo a proporção de indivíduos portadores dessa doença na população (MARTINES, ACHCAR, LOUZADA-NETO, 2004), isto é:

$$\text{prevalência} = \frac{\text{número de indivíduos com a doença na população}}{\text{número de indivíduos na população}} \quad (3)$$

Para realizar o estudo de simulação inicialmente será gerado uma amostra de tamanho  $n$  através da distribuição de Bernoulli com parâmetro igual a  $p$ . A distribuição de Bernoulli possui apenas dois resultados possíveis (MORETTIN, 2010): **1** (sucesso) ou **0** (fracasso), será considerado o resultado igual a 1 como o indivíduo é portador de uma determinada doença e 0 como o indivíduo não é portador dessa doença. Assim temos:

$$X_i = \begin{cases} 1, & \text{se o indivíduo possui a doença} \\ 0, & \text{se o indivíduo não possui a doença} \end{cases} \quad i = 1, \dots, n \quad (4)$$

onde  $X_i$  segue uma distribuição de Bernoulli com parâmetro igual a  $p$  e ainda,  $P(X_i = 1) = p$  e  $P(X_i = 0) = 1 - p$ . Será fixado o valor do parâmetro  $p$  como sendo igual a prevalência da doença na população (3).

Em seguida, será fixado o valor da sensibilidade e da especificidade do teste em **s** e **e**, respectivamente, para gerar o resultado do teste através de uma outra distribuição de Bernoulli, onde será considerado que o resultado da Bernoulli igual a 1 indica que o teste é positivo e o resultado igual a 0 indica que o teste é negativo.

Adotar-se-á o seguinte procedimento: Se  $X_i = 1$ , gerar uma distribuição de Bernoulli com parâmetro igual a

**s** (sensibilidade), ou seja, cuja probabilidade de gerar o valor 1 (teste positivo) seja igual a **s** e, se  $X_i = 0$ , gerar uma distribuição de Bernoulli com parâmetro igual a  $(1 - e)$ , ou seja, cuja probabilidade de gerar 0 (teste negativo) seja **e** (especificidade). Assim tem-se:

$$\text{Se } X_i = 1, \text{ gerar } Y_i = \begin{cases} 1, \text{ se o teste for positivo} \\ e \\ 0, \text{ se o teste for negativo} \end{cases} \quad (5)$$

onde  $Y_i$  segue uma distribuição de Bernoulli com parâmetro igual a **s** e ainda,  $P(Y_i = 1) = s$  e  $P(Y_i = 0) = 1 - s$ .

$$\text{Se } X_i = 0, \text{ gerar } Z_i = \begin{cases} 1, \text{ se o teste for positivo} \\ e \\ 0, \text{ se o teste for negativo} \end{cases} \quad (6)$$

onde  $Z_i$  segue uma distribuição de Bernoulli com parâmetro igual a  $(1 - e)$  e ainda,  $P(Z_i = 1) = 1 - e$  e  $P(Z_i = 0) = e$ .

Dessa forma, se o indivíduo tem a doença ( $X_i = 1$ ) a probabilidade do teste indicar corretamente positivo será igual a **s**, ou seja, o teste tem sensibilidade igual a **s**, isto é,  $P(T_+|D_+) = s$ . Caso contrário, se o indivíduo não tem a doença ( $X_i = 0$ ) a probabilidade do teste indicar corretamente negativo será igual a **e**, ou seja, o teste tem especificidade igual a **e**, isto é,  $P(T_-|D_-) = e$ .

Agora é só contar quantas vezes o teste acertou ou não a condição do indivíduo apresentar ou não a doença, ou seja, se  $X_i = 1$  e  $Y_i = 1$  (teste deu positivo para o indivíduo que apresenta a doença) façamos  $\alpha_i = 1$ , caso contrário  $\alpha_i = 0$ ; se  $X_i = 1$  e  $Y_i = 0$  (teste deu negativo para o indivíduo que apresenta a doença) façamos  $\beta_i = 1$ , caso contrário  $\beta_i = 0$ ; se  $X_i = 0$  e  $Z_i = 1$  (teste deu positivo para o indivíduo que não apresenta a doença) façamos  $\delta_i = 1$ , caso contrário  $\delta_i = 0$  e se  $X_i = 0$  e  $Y_i = 0$  (teste deu negativo para o indivíduo que não apresenta a doença) façamos  $\lambda_i = 1$ , caso contrário  $\lambda_i = 0$ . Assim, pode-se calcular o número de acertos e erros cometidos pelo teste da seguinte forma:

$$\alpha = \sum_{i=1}^n \alpha_i, \quad i = 1, 2, 3, \dots, n \quad (7)$$

$$\beta = \sum_{i=1}^n \beta_i, \quad i = 1, 2, 3, \dots, n \quad (8)$$

$$\delta = \sum_{i=1}^n \delta_i, \quad i = 1, 2, 3, \dots, n \quad (9)$$

$$\lambda = \sum_{i=1}^n \lambda_i, \quad i = 1, 2, 3, \dots, n \quad (10)$$

onde:  $\alpha$  é o número de acertos do teste quando o indivíduo apresenta a doença;  $\beta$  é o número de erros do teste quando o indivíduo apresenta a doença;  $\delta$  é o número de erros do teste quando o indivíduo não apresenta a doença;  $\lambda$  é o número de acertos do teste quando o indivíduo não apresenta a doença.

Dessa forma, tem-se quatro possíveis resultados ao aplicar o teste em um indivíduo. Os possíveis resultados e seus valores são apresentados na Tabela 1, observe que tem dois resultados certos e dois resultados errados ao aplicar o teste.

**Tabela 1:** Relação entre o resultado de um teste e a presença da doença.

| Resultado do Teste | Indivíduo Apresenta a Doença |                       | Total             |
|--------------------|------------------------------|-----------------------|-------------------|
|                    | Sim                          | Não                   |                   |
| Positivo           | $\alpha$<br>(acerto)         | $\delta$<br>(erro)    | $\alpha + \delta$ |
| Negativo           | $\beta$<br>(erro)            | $\lambda$<br>(acerto) | $\beta + \lambda$ |
| Total              | $\alpha + \beta$             | $\delta + \lambda$    | $n$               |

Fonte: Elaborado pelo autor, 2025.

Logo pode-se calcular a sensibilidade e a especificidade desse estudo de simulação da seguinte maneira:

$$\text{sensibilidade} = P(T_+|D_+) = \frac{\alpha}{\alpha + \beta} \quad (11)$$

$$\text{especificidade} = P(T_-|D_-) = \frac{\lambda}{\delta + \lambda} \quad (12)$$

e, comparar os valores nominais (fixados) da sensibilidade e especificidade com os calculados através do estudo de simulação.

Por último pode-se estimar a prevalência da população, através de:

$$\text{prevalência} = \frac{\alpha + \delta}{n} \quad (13)$$

Neste trabalho será construído uma rotina na linguagem *Python* para os cálculos necessários para o estudo de simulação, o qual será implementado no próprio *Python* (PYTHON, 2023). A escolha da utilização dessa linguagem computacional neste trabalho é que, além de ser gratuito, apresenta código-fonte aberto, podendo ser modificado ou implementado com novos procedimentos desenvolvidos pelo usuário. Nesse

processo de simulação serão utilizados valores distintos para a sensibilidade (**s**) e especificidade (**e**), para verificar seu efeito na estimação da prevalência e no estudo de simulação.

## RESULTADOS

Para realizar os procedimentos de simulação, gerou-se populações com  $N = 1.000.000$  de indivíduos e, para cada uma delas, retirou-se uma amostra aleatória simples de tamanho  $n = 1.000$ . Para averiguar a influência que diferentes prevalências podem ter nos resultados de um teste diagnóstico, fixou-se os valores das prevalências em 5%, 10%, 20%, 30%, 50%, 70%, 80%, 90% e 95%, abrangendo várias situações, tais como: doenças raras, com baixa prevalência, como situações em que a doença pode estar amplamente presente na população, ou seja, com alta prevalência.

Fixou-se também diferentes valores para a sensibilidade e para a especificidade dos testes simulados. Os valores escolhidos para esses parâmetros foram: 70%, 80%, 90%, 95% e 99%. Essa decisão foi tomada de acordo com valores de sensibilidade e especificidade de testes diagnósticos encontrados na literatura de Soares e Siqueira (2012), Carminati (2016) e Chiara, Sichieri e Martins (2003).

Em resumo, são 225 situações distintas, dada pela multiplicação de 5 valores para sensibilidade, 5 valores para especificidade e 9 valores para prevalência. Em cada caso e para cada combinação de valores dos parâmetros, o processo consiste em gerar uma população e retirar uma amostra, que ao final, totalizam:  $1.000.000 \times 1.000 \times 9 \times 5 \times 5 = 225.000.000.000$  de simulações para analisar a influência da prevalência, sensibilidade e especificidade nos testes.

A Tabela 2 apresenta os resultados obtidos através das combinações geradas pela sensibilidade e especificidade, ambas fixadas em 70% e a prevalência variando de 5% a 95%.

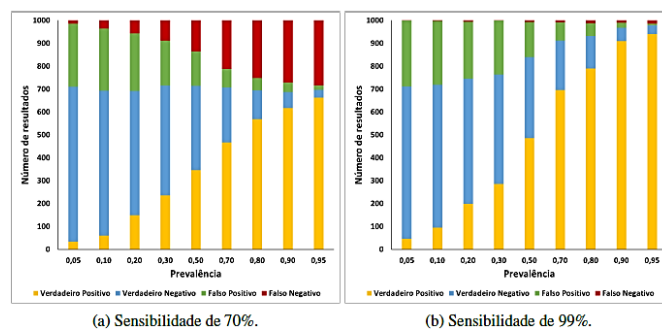
**Tabela 2:** Resultado da simulação para sensibilidade e especificidade de 70%.

| Prevalência | Verdadeiro  | Verdadeiro  | Falso       | Falso       |
|-------------|-------------|-------------|-------------|-------------|
|             | Positivo    | Negativo    | Positivo    | Negativo    |
| 5%          | 33 (3,3%)   | 677 (67,7%) | 275 (27,5%) | 15 (1,5%)   |
| 10%         | 60 (6,0%)   | 632 (63,2%) | 272 (27,2%) | 36 (3,6%)   |
| 20%         | 148 (14,8%) | 543 (54,3%) | 252 (25,2%) | 57 (5,7%)   |
| 30%         | 235 (23,5%) | 480 (48,0%) | 196 (19,6%) | 89 (8,9%)   |
| 50%         | 346 (34,6%) | 367 (36,7%) | 150 (15,0%) | 137 (13,7%) |
| 70%         | 467 (46,7%) | 240 (24,0%) | 80 (8,0%)   | 213 (21,3%) |
| 80%         | 568 (56,8%) | 126 (12,6%) | 54 (5,4%)   | 252 (25,2%) |
| 90%         | 616 (61,6%) | 71 (7,1%)   | 41 (4,1%)   | 272 (27,2%) |
| 95%         | 662 (66,2%) | 34 (3,4%)   | 19 (1,9%)   | 285 (28,5%) |

Fonte: Elaborado pelo autor, 2025.

A Figura 1 apresenta os Resultados da simulação com sensibilidade de 70% e 99% e especificidade fixada em 70%. Na Figura 1(a) tem-se os resultados obtidos das simulações utilizando a sensibilidade fixada em 70%, enquanto na Figura 1(b) encontra-se os resultados obtidos com a sensibilidade fixada em 99%. Em ambas as figuras, a especificidade foi fixada em 70%.

**Figura 1:** Resultado da simulação com sensibilidade de 70% e 99% e especificidade de 70%.



Fonte: Elaborado pelo autor, 2025.

Na Tabela 3 tem-se os resultados obtidos a partir da simulação com os parâmetros de sensibilidade e especificidade fixados em 99% e 70% respectivamente.

**Tabela 3:** Resultado da simulação para sensibilidade 99% e especificidade de 70%.

| Prevalência | Verdadeiro  | Verdadeiro  | Falso       | Falso     |
|-------------|-------------|-------------|-------------|-----------|
|             | Positivo    | Negativo    | Positivo    | Negativo  |
| 5%          | 45 (4,5%)   | 666 (66,6%) | 288 (28,8%) | 1 (0,1%)  |
| 10%         | 95 (9,5%)   | 623 (62,3%) | 279 (27,9%) | 3 (0,3%)  |
| 20%         | 199 (19,9%) | 546 (54,6%) | 249 (24,9%) | 6 (0,6%)  |
| 30%         | 286 (28,6%) | 477 (47,7%) | 235 (23,5%) | 2 (0,2%)  |
| 50%         | 486 (48,6%) | 353 (35,3%) | 154 (15,4%) | 7 (0,7%)  |
| 70%         | 695 (69,5%) | 216 (21,6%) | 80 (8,0%)   | 9 (0,9%)  |
| 80%         | 789 (78,9%) | 142 (14,2%) | 56 (5,6%)   | 13 (1,3%) |
| 90%         | 909 (90,9%) | 58 (5,8%)   | 22 (2,2%)   | 11 (1,1%) |
| 95%         | 939 (93,9%) | 38 (3,8%)   | 8 (0,8%)    | 15 (1,5%) |

Fonte: Elaborado pelo autor, 2025.

A Tabela 4 apresenta os resultados obtidos a partir da simulação com os parâmetros de sensibilidade e especificidade fixados em 70% e 99% respectivamente.

**Tabela 4:** Resultado da simulação para sensibilidade 70% e especificidade de 99%.

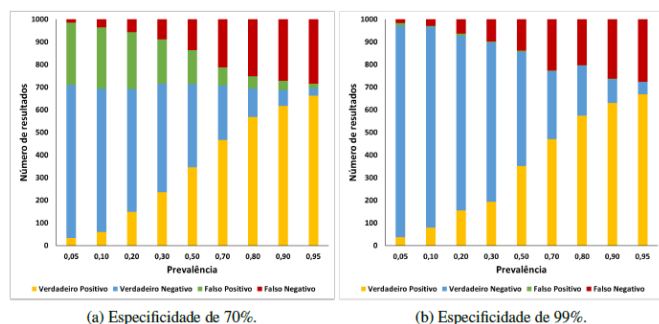
| Prevalência | Verdadeiro Positivo | Verdadeiro Negativo | Falso Positivo | Falso Negativo |
|-------------|---------------------|---------------------|----------------|----------------|
| 5%          | 37 (3,7%)           | 935 (93,5%)         | 11 (1,1%)      | 17 (1,7%)      |
| 10%         | 79 (7,9%)           | 888 (88,8%)         | 4 (0,4%)       | 29 (2,9%)      |
| 20%         | 155 (15,5%)         | 772 (77,2%)         | 9 (0,9%)       | 64 (6,4%)      |
| 30%         | 193 (19,3%)         | 704 (70,4%)         | 4 (0,4%)       | 99 (9,9%)      |
| 50%         | 352 (35,2%)         | 504 (50,4%)         | 5 (0,5%)       | 139 (13,9%)    |
| 70%         | 470 (47,0%)         | 300 (30,0%)         | 3 (0,3%)       | 227 (22,7%)    |
| 80%         | 574 (57,4%)         | 221 (22,1%)         | 1 (0,1%)       | 204 (20,4%)    |
| 90%         | 630 (63,0%)         | 105 (10,5%)         | 2 (0,2%)       | 263 (26,3%)    |
| 95%         | 668 (66,8%)         | 56 (5,6%)           | 0 (0,0%)       | 276 (27,6%)    |

Fonte: Elaborado pelo autor, 2025.



A Figura 2 ilustra os resultados obtidos das simulações utilizando a sensibilidade fixada em 70% e especificidade de 70% e 99%.

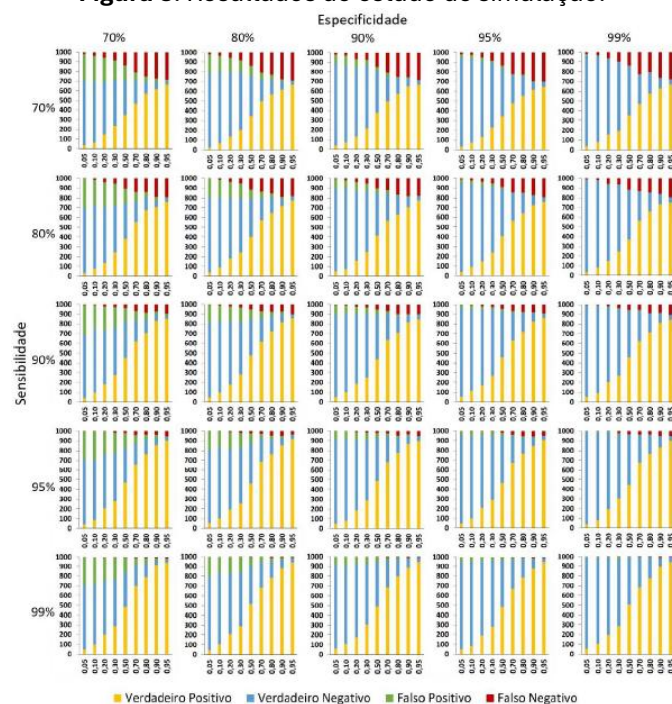
**Figura 2:** Resultado da simulação com sensibilidade de 70% e especificidade de 70% e 99%.



Fonte: Elaborado pelo autor, 2025.

A Figura 3 apresenta todos os resultados do estudo de simulação computacional. Nela, pode-se analisar o comportamento dos resultados das 225 combinações distintas em relação a sensibilidade, especificidade e a prevalência.

**Figura 3:** Resultados do estudo de simulação.



Fonte: Elaborado pelo autor, 2025.

A Tabela 5 apresenta os resultados obtidos a partir da simulação com os parâmetros de sensibilidade e especificidade fixados em 99%.

**Tabela 5:** Resultado da simulação para sensibilidade e especificidade de 99%.

| Prevalência | Verdadeiro  | Verdadeiro  | Falso     | Falso     |
|-------------|-------------|-------------|-----------|-----------|
|             | Positivo    | Negativo    | Positivo  | Negativo  |
| 5%          | 51 (5,1%)   | 938 (93,8%) | 11 (1,1%) | 0 (0,0%)  |
| 10%         | 102 (10,2%) | 884 (88,4%) | 12 (1,2%) | 2 (0,2%)  |
| 20%         | 195 (19,5%) | 794 (79,4%) | 8 (0,8%)  | 3 (0,3%)  |
| 30%         | 287 (28,7%) | 697 (69,7%) | 12 (1,2%) | 4 (0,4%)  |
| 50%         | 511 (51,1%) | 479 (47,9%) | 7 (0,7%)  | 3 (0,3%)  |
| 70%         | 683 (68,3%) | 306 (30,6%) | 3 (0,3%)  | 8 (0,8%)  |
| 80%         | 778 (77,8%) | 210 (21,0%) | 3 (0,3%)  | 9 (0,9%)  |
| 90%         | 896 (89,6%) | 90 (9,0%)   | 1 (0,1%)  | 13 (1,3%) |
| 95%         | 939 (93,9%) | 52 (5,2%)   | 0 (0,0%)  | 9 (0,9%)  |

Fonte: Elaborado pelo autor, 2025.

## DISCUSSÃO

Analisando os resultados do estudo de simulação, observa-se na Tabela 2 claramente a influência da prevalência de uma doença nos resultados obtidos. Quando é simulada uma população com prevalência de 5% (doença rara), os resultados mais comuns são Verdadeiro Negativo, com uma frequência de 677 resultados (67,7%), seguidos pelos Falso Positivo, com uma frequência de 275 resultados (27,5%). Já os resultados que aparecem com menos frequência, são os Verdadeiros Positivos e Falsos Negativos, com 33 e 15 ocorrências (3,3% e 1,5%), respectivamente.

A proporção entre Verdadeiro Positivo e Verdadeiro Negativo é de aproximadamente 1:20, enquanto a proporção encontrada entre Falso Positivo e Falso Negativo, é de aproximadamente 18:1. Esse resultado era esperado, uma vez que, por se tratar de uma condição pouco comum (doença rara), existem menos pessoas doentes e mais pessoas saudáveis que realizarão o teste, pois ela está menos presente na população. Ou seja, quando o teste acerta o resultado, é mais provável que ele classifique corretamente o indivíduo como saudável, aumentando assim a quantidade de Verdadeiros Negativos. Ao mesmo tempo, quando o teste erra o diagnóstico, é mais provável que ele tenha classificado erroneamente um indivíduo saudável como doente, aumentando então a quantidade de Falsos Positivos.

A medida em que a prevalência vai aumentando, observa-se na Tabela 2 que o cenário vai se invertendo. Ao mesmo tempo em que a quantidade de Verdadeiro Negativo e Falso Positivo diminui, o número de Verdadeiro Positivo e Falso Negativo aumentam. A proporção entre Verdadeiro Positivo e Verdadeiro Negativo para a prevalência de 50% é de aproximadamente 1:1. O mesmo ocorre na proporção entre Falso Positivo e Falso Negativo, que também é de aproximadamente 1:1. A mudança fica ainda mais evidente quando são comparados os resultados obtidos com a prevalência de 5% e 95%. O que é observado na segunda situação, é o oposto do que é

visto na primeira. Na condição mais comum, com prevalência de 95%, os resultados mais obtidos são Verdadeiro Positivo, com 662 representantes (66,2%), seguidos pelos Falso Negativos, com 285 indivíduos (28,5%). Já os resultados que eram mais comuns na primeira situação, Verdadeiro Negativo e Falso Positivo, passam a se tornar mais raros no segundo caso, com apenas 34 e 19 indivíduos (3,4% e 1,9%), respectivamente. A proporção agora observada entre Verdadeiros Positivos e Verdadeiros Negativos e entre Falsos Positivos e Falsos Negativos é de aproximadamente 19:1 e 1:5, respectivamente.

Esse fato faz sentido, uma vez que, por se tratar de uma condição extremamente comum, existem mais pessoas doentes e menos pessoas saudáveis que foram submetidas a testagem, resultando no oposto do ocorrido na primeira situação. Desse modo, quando o teste acerta o resultado, é mais provável que ele classifique corretamente o indivíduo como doente, aumentando assim a quantidade de Verdadeiro Positivo. Ao mesmo tempo, quando o teste erra o diagnóstico, é mais provável que ele tenha classificado de forma errada um indivíduo doente como saudável, aumentando então a quantidade de Falsos Negativos.

Na Figura 1(a), observa-se que à medida que a prevalência aumenta, diminuem os casos Verdadeiro Negativo (azul) e Falso Positivo (verde) e crescem os casos Verdadeiro Positivo (amarelo) e Falso Negativo (vermelho). Analisando as Figura 1 (a) e 1 (b), fica fácil a visualização do comportamento descrito anteriormente. Ao aumentarmos a sensibilidade de um teste, a quantidade de Verdadeiro Positivo (amarelo) aumenta, enquanto o número de Falso Negativo (vermelho) diminui.

A partir da comparação entre os resultados da Tabela 2 e da Tabela 3 pode-se perceber a influência da sensibilidade sobre os resultados de um teste diagnóstico. Para a prevalência de 50% os resultados Verdadeiro Positivo passaram de 346 (34,6%) na Tabela 2, com sensibilidade de 70%, para 486 (48,6%) na Tabela 3, onde a sensibilidade é de 99%, um aumento de 40,5%. Não por coincidência, ocorreram 130 resultados do tipo Falso Negativo a menos, saindo de um total de 137 (13,7%) resultados desse tipo, para apenas 7 (0,7%). A proporção entre Falso Negativo para Verdadeiro Positivo passa de 1:3 na Tabela 2 para 1:69 na Tabela 3.

Em relação aos Verdadeiros Negativos e Falsos Positivos, a quantidade passou de 150 (15%) para 154 (15,4%) Falsos Positivos e a quantidade de Verdadeiro Negativo diminuiu de 367 (36,7%) para 353 (35,3%), mantendo a proporção aproximada de 1:2 em ambas as tabelas.

Esse resultado também era esperado, uma vez que a sensibilidade, é definida por  $P(T+|D+)$  e pode ser entendida como a probabilidade de classificar indivíduos Verdadeiro Positivo. Ao aumentar-se ela, evidentemente deve-se encontrar mais resultados desse tipo. A diminuição de resultados Falso Negativo se deve a esse ser o complementar do Verdadeiro Positivo.

Outra coisa que se pode notar nas Tabelas 2 e 3 é o impacto que a prevalência tem nos resultados de um teste extremamente sensível. Quando a prevalência simulada é baixa (5%), representando uma doença rara, não faz tanta diferença o teste ser tão sensível, pois a magnitude da quantidade de erro, tanto Falso Positivo quanto Falso Negativo, continuam a mesma, passando de  $275 + 15 = 290$  (29%) na Tabela 2 para  $288 + 1 = 289$  (28,9%) na Tabela 3. Entretanto, quando a prevalência da doença é maior, 95% por exemplo, quanto mais sensível o teste, melhor. Isso fica claro ao observar que nessa situação,  $19 + 285 = 304$  ou 30,4% dos indivíduos foram classificados erroneamente na Tabela 2 enquanto na Tabela 3 apenas  $8 + 15 = 23$  erros foram cometidos, isto é, 2,3%.

Uma análise análoga pode ser realizada para analisar a influência da especificidade sobre os resultados de um teste diagnóstico. A partir da comparação entre a Tabela 2 e a Tabela 4 e com prevalência fixada em 50%, pode-se perceber mudanças evidentes da primeira para a segunda. Ao comparar a quantidade de resultados do tipo Verdadeiro Positivo e Falso Negativo, essa mudança não é tão significativa. A quantidade de Verdadeiros Positivos aumentou de 346 (34,6%) para 352 (35,2%) casos. Também foi registrada uma diferença de apenas 2 indivíduos do tipo Falso Negativo de uma tabela para a outra, com 137 (13,7%) casos na primeira 139 (13,9%) na segunda. Em ambas as tabelas, a proporção entre Falsos Negativos e Verdadeiros Positivos se manteve próxima de 1:3.

Entretanto, a diferença fica discrepante nos outros tipos de resultados. A quantidade de indivíduos classificados como Verdadeiro Negativo aumentou de 367 (36,7%) para 504 (50,4%), um aumento de 37,3%. Ao mesmo tempo, houve uma diminuição de 145 resultados do tipo Falso Positivo, saindo de um total de 150 (15%) resultados desse tipo, para apenas 5 (0,5%). A proporção entre Falso Positivo e Verdadeiro Negativo passou de 1:2 na Tabela 2 para 1:101 na Tabela 4.

De modo semelhante ao que aconteceu com a sensibilidade, o resultado obtido com a especificidade é lógico, uma vez que essa é definida como  $P(T-|D-)$ , ou seja, e probabilidade de gerar resultados Verdadeiro Negativo. É de se esperar que esse resultado, apareça de forma mais frequentemente que os demais, da mesma

forma que se espera menos ocorrências do seu complementar, o Falso Positivo.

Pode-se perceber também, através da análise da Tabela 4, a influência que a prevalência tem em testes diagnósticos mais específicos. Quando a prevalência da doença é baixa, faz toda diferença utilizar um teste mais específico, visto que esse diminuiu em 262 os erros cometidos quando a prevalência foi fixada em 5%, saindo de 290 (29%) classificações erradas na Tabela 2 para 28 (2,8%) erros na Tabela 4. Entretanto, essa medida não é tão relevante quando busca-se detectar doenças mais comuns, com maiores prevalências. Como visto nas tabelas, com o valor da prevalência fixada em 95%, a primeira cometeu 304 (30,4%) erros, enquanto a segunda cometeu 276 (27,6%).

Observa-se na Figura 2 que ao aumentar a especificidade de um teste, a quantidade de Verdadeiro Negativo (azul) aumenta, enquanto o número de Falso Positivo (verde) diminui.

É claro que ao alterar a sensibilidade e especificidade, aumentando ambas ao mesmo tempo, obtém-se um teste mais preciso, que comete menos erros. Esse fato fica evidente ao analisarmos a Tabela 5. Percebe-se que a quantidade de erros, tanto do tipo Falso Positivo, quanto do tipo Falso Negativo, diminui quando comparados com a Tabela 2. O teste menos sensível e específico obteve 287 (28,7%) resultados errados com a prevalência fixada em 50%. Já o teste mais preciso obteve apenas 10 (1%) resultados errados, uma queda de 96,5%.

Analisando o comportamento dos resultados das 225 combinações distintas em relação a sensibilidade, especificidade e a prevalência, conforme apresentado na Figura 3, observa-se que para valores maiores de sensibilidade, os resultados Falso Negativo (vermelho) diminuem e Verdadeiro Positivo (amarelo) aumentam. Enquanto à medida que se aumenta a especificidade, os resultados Falso Positivo (verde) diminuem e um maior número de Verdadeiro Negativo (azul) é observado. Em relação a prevalência, para valores baixos, a tendência é ter a ocorrência maior de resultados Verdadeiro Negativo (azul), enquanto verifica-se uma maior frequência de Falso Positivo (verde) em relação a Falso Negativo (vermelho). A medida que a prevalência aumenta, esses resultados vão se invertendo, ocorrendo uma maior frequência de resultados Verdadeiro Positivo (amarelo) e menor ocorrência de erros do tipo Falso Positivo (verde) em relação a erros do tipo Falso Negativo (vermelho).

Também se observa na Figura 3 que a medida que ambas, sensibilidade e especificidade, aumentam, os números de Falso Positivo (verde) e Falso Negativo (vermelho) diminuem, ocasionando em mais resultados

certos para quaisquer valores de prevalência. Ou seja, um teste com alta sensibilidade e alta especificidade seria ideal para diminuir a quantidade de diagnósticos errados.

## CONCLUSÕES

Este trabalho comprovou, através de um estudo de simulação, a influência que a prevalência, sensibilidade e especificidade têm sobre o desempenho de um teste diagnóstico. Essa informação nos permite compreender em qual situação é melhor prezar por um teste mais sensível ou específico.

Neste estudo observou-se que a prevalência de uma doença influencia diretamente nos resultados de um teste diagnóstico. Doenças raras, com prevalências baixas, apresentam mais resultados do tipo Verdadeiro Negativo e Falso Positivo. Enquanto doenças mais comuns (prevalências altas), tendem a ter mais resultados do tipo Verdadeiro Positivo e Falso Negativo.

Verificou-se que um teste mais sensível (maior sensibilidade), aumenta a quantidade de resultados do tipo Verdadeiro Positivo e diminui o número de erros do tipo Falso Negativo. Apesar disso, esse teste não tem influência considerável sobre os resultados Verdadeiro Negativo e Falso Positivo. Visto isso, um teste sensível é indicado para identificar doenças com maior prevalência.

Observou-se também que a especificidade de um teste influencia diretamente nos resultados do tipo Verdadeiro Negativo e Falso Positivo, aumentando a frequência em que o primeiro ocorre, e diminuindo a frequência do segundo. No entanto, ela não possui relação com os resultados Verdadeiro Positivo e Falso Negativo. Neste caso, um teste mais específico é apropriado quando a prevalência da doença for baixa.

Por fim, conclui-se que o ideal é um teste diagnóstico com alta sensibilidade e especificidade, desde que possível, pois assim consegue-se diminuir ambos os erros, Falso Positivo e Falso Negativo, ocasionando em testes diagnósticos com maior probabilidade de acerto.

## AGRADECIMENTOS

O presente trabalho foi financiado com recursos do Programa Próprio de Fomento à Pesquisa, Pós-Graduação e Inovação da Universidade Estadual de Goiás/Pró-Laboratórios.

## REFERÊNCIAS

CARMINATI, R. **Estudo da sensibilidade e especificidade de quatro testes Elisa e utilização da**

**técnica de PCR para diagnóstico de linfadenite caseosa em caprinos.** Instituto de Ciências da Saúde, Salvador, 2016.

CHIARA, V.; SICHIERI, R.; MARTINS, P. D. **Sensibilidade e especificidade de classificação de sobrepeso em adolescentes.** Revista de Saúde Pública, Rio de Janeiro, v. 37, p. 226–231, 2003.

MARTINEZ, E. Z; ACHCAR, J. A; LOUZADA NETO. F. **Metodologia Estatística para Testes Diagnósticos e Laboratoriais com Respostas Dicotomizadas.** Revista de Matemática Estatística, São Paulo, V22, n.3, p.21-32, 2004.

MORETTIN, L. G. **Estatística Básica: probabilidade e inferência.** Volume único. São Paulo: Pearson Prentice Hall, 2010.

NUNES; A. A. *et al.* **Testes diagnósticos no contexto da avaliação de tecnologias em saúde: abordagens, métodos e interpretação.** Revista Medicina (Ribeirão Preto), v.48 n.1, p.8-18, 2015.

PYTHON. Disponível em: <<https://www.python.org/>>. Acessado em: 10 abril de 2025.

SOARES, J. F; SIQUEIRA, A. L. **Introdução à Estatística Médica.** 2ª Edição: COOPMED Editora Médica, 2012.

VIEIRA, S. **Bioestatística – Tópicos Avançados.** Editora Elsevier. 3ª Edição. 2010.