



# **Metodologias de Classificação Baseadas em Testes Compostos**

Mestrado em Ciências de Dados

André Filipe Almeida Catarino

Leiria, setembro de 2024





# **Metodologias de Classificação Baseadas em Testes Compostos**

Mestrado em Ciências de Dados

André Filipe Almeida Catarino

Dissertação realizada sob a orientação do Professor Rui Santos e do Professor Miguel Felgueiras, Professores da Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Leiria.

Leiria, setembro de 2024



# **Originalidade e Direitos de Autor**

A presente dissertação é original, elaborada unicamente para este fim, tendo sido devidamente citados todos os autores cujos estudos e publicações contribuíram para a elaborar.

Reproduções parciais deste documento serão autorizadas na condição de que seja mencionado o Autor e feita referência ao ciclo de estudos no âmbito do qual a mesma foi realizada, a saber, Curso de Mestrado em Ciências de Dados, no ano letivo 2023/2024, da Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Leiria, Portugal, e, bem assim, à data das provas públicas que visaram a avaliação destes trabalhos.



# Dedicatória

In life's vast weave, joy and trials blend,  
Family's warmth, and friends who mend.  
Through seasons' shift, they are our light,  
Guiding home, through the darkest night.

*“When do you think people die? When they’re shot through the heart by the bullet of a pistol?  
No. When they’re ravaged by an incurable disease? No. When they drink soup made from a  
poisonous mushroom!?! No! It’s when... they are forgotten.”*

Dr. Hiriluk

À minha irmã, Inês

Aos meus pais, Paulo e Carla

Aos meus avós, Carlos Alberto, Maria Rosa e Justina

Ao meu falecido avô, Mário



# Agradecimentos

Primeiramente, quero agradecer ao professor Rui Santos e ao professor Miguel Felgueiras, pela enorme ajuda, paciência, disponibilidade e constante boa disposição ao longo da realização desta dissertação.

Agradeço aos meus pais, que sem dúvida foram a base para o meu sucesso até hoje, pelos valores que me transmitiram, pelo carinho e apoio que me deram sempre que surgiram dúvidas e obstáculos ao longo da minha vida.

Um obrigado à minha irmã, por ser bondosa e muito amiga. Por ter cuidado de mim numa fase onde tempo era o que mais faltava e mesmo assim dispensava o pouco tempo que tinha comigo. Um muito obrigado!

Especial obrigado aos meus avós, por me inspirarem a ser uma melhor pessoa e por me apoiarem em tudo. Por serem os pilares da minha vida, vou sempre ficar grato!

Quero também agradecer aos meus colegas de mestrado que me ajudaram a chegar até aqui!

As pessoas mais importantes na vida, são aquelas que nela surgem e permanecem e, aos que permanecem na minha, apenas quero dizer, “salud, mi familia”.



# Resumo

A crescente procura por eficiência em testes diagnósticos, especialmente em contextos de baixa prevalência de doenças em populações de elevada dimensão, destaca a necessidade de metodologias que otimizem recursos sem comprometer a precisão do diagnóstico. Este estudo tem como objetivo a comparação de diferentes metodologias de classificação baseadas em testes compostos (metodologia de Dorfman, metodologia hierárquica com 4 etapas e metodologia não hierárquica com e sem *master pool*), que permitem a redução significativa do número de testes necessários ao diagnóstico de toda a população através do agrupamento de amostras. O foco deste estudo está principalmente na eficiência/ custo relativo (quantidade de testes necessários para identificar cada indivíduo) e a probabilidade de existirem erros de classificação (medida pela sensibilidade e especificidade de cada metodologia).

Recorrendo a simulações computacionais e análises estatísticas em linguagem R, esta pesquisa avalia a sensibilidade, especificidade, eficiência e custo-benefício de forma a comparar as diferentes metodologias. Nas simulações foram considerados apenas testes qualitativos, não sendo, por este motivo, considerado o efeito de diluição, ou seja, considerou-se que a fiabilidade do teste composto é igual à do teste individual. Os resultados demonstram que as metodologias de classificação baseadas em testes compostos oferecem uma alternativa viável e eficaz, reduzindo custos e tempo de resposta sem sacrificar de forma significativa a precisão, apesar da escolha da metodologia a utilizar ter de ser feita com base no objetivo que se pretende alcançar e no contexto populacional em estudo.

Deste modo, este estudo contribui para a literatura existente ao fornecer uma análise detalhada sobre a implementação prática e as vantagens e desvantagens das metodologias de classificação baseadas em testes compostos, propondo melhorias para sua aplicação em programas de saúde pública global.

**Palavras-chave:** testes compostos, simulação, eficiência, sensibilidade, especificidade, metodologias de classificação.



# Abstract

The growing demand for efficiency in diagnostic testing, especially in low disease prevalence contexts, highlights the need for methodologies that optimize resources without compromising diagnostic accuracy. This study aims to compare different classification methodologies based on pooled testing (Dorfman's method, a 4-stage hierarchical method, and non-hierarchical methods with and without a master pool), which significantly reduce the number of tests needed by grouping samples. The focus of this study is primarily on efficiency/cost-effectiveness (the number of tests required to identify each individual) and the reliability of classification (measured by the sensitivity and specificity of each methodology).

Computational simulations and statistical analyses in R are used to assess sensitivity, specificity, efficiency, and cost-effectiveness, with the goal of comparing hierarchical and non-hierarchical pooled testing methods. Since only qualitative tests were conducted, the dilution effect will not be considered, meaning equal reliability in pooled and individual tests. The results show that classification methodologies based on pooled testing offers a viable and effective alternative, reducing costs and response time without significantly sacrificing accuracy, although the choice of the methodology must be based on the proposed goals and on the scenario in which it will be applied.

Hence, this study contributes to existing literature by providing a detailed analysis of the practical implementation and advantages and disadvantages of classification methodologies based on pooled testing, proposing improvements for its application in global public health programs.

**Keywords:** compound tests, simulation, efficiency, sensitivity, specificity, classification methodologies.



# Índice

|  |              |
|--|--------------|
| <b>Originalidade e Direitos de Autor .....</b>                       | <b>iii</b>   |
| <b>Dedicatória .....</b>   | <b>v</b>     |
| <b>Agradecimentos .....</b>  | <b>vii</b>   |
| <b>Resumo .....</b>  | <b>ix</b>    |
| <b>Abstract .....</b>  | <b>xi</b>    |
| <b>Lista de Figuras .....</b>  | <b>xix</b>   |
| <b>Lista de tabelas .....</b>  | <b>xxiii</b> |
| <b>Lista de siglas e acrónimos.....</b>                              | <b>xxv</b>   |
| <b>1. Introdução .....</b>   | <b>1</b>     |
| <b>2. Noções e metodologias .....</b>                                | <b>3</b>     |
| <b>2.1. Testes de classificação .....</b>                            | <b>3</b>     |
| <b>2.2. Testes compostos.....</b>                                    | <b>4</b>     |
| <b>2.3. Diluição em testes compostos.....</b>                        | <b>6</b>     |
| <b>2.4. Testes qualitativos versus testes quantitativos.....</b>     | <b>7</b>     |
| <b>2.5. Eficiência e custo relativo.....</b>                         | <b>9</b>     |
| <b>2.6. Sensibilidade e especificidade.....</b>                      | <b>11</b>    |
| <b>2.7. Acurácia.....</b>  | <b>13</b>    |
| <b>2.8. Valor preditivo negativo e valor preditivo positivo.....</b> | <b>14</b>    |
| <b>2.10. Linguagem R e Simulação .....</b>                           | <b>17</b>    |
| <b>2.11. Fiabilidade em testes compostos.....</b>                    | <b>19</b>    |
| 2.11.1. Grupo doente versus grupo saudável.....                      | 19           |
| 2.11.2. Sensibilidade e especificidade em testes compostos .....     | 20           |

|  |           |
|--|-----------|
| <b>3. Metodologias de classificação baseadas em testes compostos .....</b> | <b>22</b> |
| <b>3.1. Metodologia de Dorfman .....</b>                                   | <b>22</b> |
| <b>3.2. Metodologia hierárquica.....</b>                                   | <b>23</b> |
| <b>3.3. Metodologias não hierárquicas .....</b>                            | <b>27</b> |
| <b>3.4. Fiabilidade de uma metodologia de classificação .....</b>          | <b>32</b> |
| 3.4.1. Sensibilidade, especificidade e acurácia de uma metodologia.....    | 33        |
| 3.4.2. VPP e VPN da metodologia .....                                      | 34        |
| <b>4. Análise de resultados.....</b>                                       | <b>35</b> |
| <b>4.1. Configuração das simulações .....</b>                              | <b>35</b> |
| <b>4.2. Sensibilidade igual à especificidade .....</b>                     | <b>37</b> |
| 4.2.1. Média total de testes e eficiência.....                             | 37        |
| 4.2.2. Acurácia da metodologia.....  | 41        |
| 4.2.3. Sensibilidade da metodologia.....                                   | 45        |
| 4.2.4. Especificidade da metodologia.....                                  | 46        |
| <b>4.3. Sensibilidade e especificidade com valores distintos.....</b>      | <b>47</b> |
| 4.3.1. Média total de testes .....   | 47        |
| 4.3.2. Acurácia da metodologia.....  | 51        |
| 4.3.3. Sensibilidade da metodologia.....                                   | 54        |
| 4.3.4. Especificidade da metodologia.....                                  | 56        |
| <b>4.4. Metodologia não hierárquica .....</b>                              | <b>57</b> |
| 4.4.1. Média de total de testes .....                                      | 58        |
| 4.4.2. Acurácia da metodologia.....  | 62        |
| 4.4.3. Sensibilidade da metodologia.....                                   | 63        |
| 4.4.4. Especificidade da metodologia.....                                  | 64        |
| <b>5. Discussão de resultados.....</b>                                     | <b>67</b> |
| <b>6. Conclusão .....</b>  | <b>74</b> |
| <b>Bibliografia .....</b>  | <b>76</b> |
| <b>Anexos .....</b>  | <b>80</b> |

|   |           |
|---|-----------|
| <b>Anexo 1: Total de testes para <math>p = 0.025</math> utilizando a metodologia hierárquica e de Dorfman quando <math>SI = EI</math> .....</b>                             | <b>80</b> |
| <b>Anexo 2: Total de testes para <math>p = 0.01</math> utilizando a metodologia hierárquica e de Dorfman quando <math>SI = EI</math> .....</b>                              | <b>81</b> |
| <b>Anexo 3: Total de testes para <math>p = 0.005</math> utilizando a metodologia hierárquica e de Dorfman quando <math>SI = EI</math> .....</b>                             | <b>82</b> |
| <b>Anexo 4: Acurácia da metodologia hierárquica e de Dorfman para <math>p = 0.025</math> quando <math>SI = EI</math>.....</b>   | <b>83</b> |
| <b>Anexo 5: Acurácia da metodologia hierárquica e de Dorfman para <math>p = 0.01</math> quando <math>SI = EI</math>.....</b>  | <b>84</b> |
| <b>Anexo 6: Acurácia da metodologia hierárquica e de Dorfman para <math>p = 0.005</math> quando <math>SI = EI</math>.....</b>   | <b>85</b> |
| <b>Anexo 7: Sensibilidade da metodologia hierárquica e de Dorfman para <math>p = 0.025</math> quando <math>SI = EI</math>.....</b>  | <b>86</b> |
| <b>Anexo 8: Sensibilidade da metodologia hierárquica e de Dorfman para <math>p = 0.01</math> quando <math>SI = EI</math>.....</b>   | <b>87</b> |
| <b>Anexo 9: Sensibilidade da metodologia hierárquica e de Dorfman para <math>p = 0.005</math> quando <math>SI = EI</math>.....</b>  | <b>88</b> |
| <b>Anexo 10: Especificidade da metodologia hierárquica e de Dorfman para <math>p = 0.025</math> quando <math>SI = EI</math>.....</b>  | <b>89</b> |
| <b>Anexo 11: Especificidade da metodologia hierárquica e de Dorfman para <math>p = 0.01</math> quando <math>SI = EI</math>.....</b>   | <b>90</b> |
| <b>Anexo 12: Especificidade da metodologia hierárquica e de Dorfman para <math>p = 0.005</math> quando <math>SI = EI</math>.....</b>  | <b>91</b> |
| <b>Anexo 13: Total de testes para <math>p = 0.025</math> utilizando a metodologia hierárquica e de Dorfman quando <math>SI</math> e <math>EI</math> são distintos .....</b> | <b>92</b> |

|  |            |
|--|------------|
| <b>Anexo 14: Total de testes para <math>p = 0.01</math> utilizando a metodologia hierárquica e de Dorfman quando SI e EI são distintos.....</b>        | <b>93</b>  |
| <b>Anexo 15: Total de testes para <math>p = 0.005</math> utilizando a metodologia hierárquica e de Dorfman quando SI e EI são distintos.....</b>       | <b>94</b>  |
| <b>Anexo 16: Acurácia da metodologia hierárquica e de Dorfman para <math>p = 0.025</math> quando SI e EI são distintos .....</b>                       | <b>95</b>  |
| <b>Anexo 17: Acurácia da metodologia hierárquica e de Dorfman para <math>p = 0.01</math> quando SI e EI são distintos .....</b>                        | <b>96</b>  |
| <b>Anexo 18: Acurácia da metodologia hierárquica e de Dorfman para <math>p = 0.005</math> quando SI e EI são distintos .....</b>                       | <b>97</b>  |
| <b>Anexo 19: Sensibilidade da metodologia hierárquica e de Dorfman para <math>p = 0.025</math> quando SI e EI são distintos .....</b>                  | <b>98</b>  |
| <b>Anexo 20: Sensibilidade da metodologia hierárquica e de Dorfman para <math>p = 0.01</math> quando SI e EI são distintos .....</b>                   | <b>99</b>  |
| <b>Anexo 21: Sensibilidade da metodologia hierárquica e de Dorfman para <math>p = 0.005</math> quando SI e EI são distintos .....</b>                  | <b>100</b> |
| <b>Anexo 22: Especificidade da metodologia hierárquica e de Dorfman para <math>p = 0.025</math> quando SI e EI são distintos .....</b>                 | <b>101</b> |
| <b>Anexo 23: Especificidade da metodologia hierárquica e de Dorfman para <math>p = 0.01</math> quando SI e EI são distintos .....</b>                  | <b>102</b> |
| <b>Anexo 24: Especificidade da metodologia hierárquica e de Dorfman para <math>p = 0.005</math> quando SI e EI são distintos .....</b>                 | <b>103</b> |
| <b>Anexo 25: Total de testes para <math>p = 0.025</math> utilizando a metodologia não hierárquica com e sem <i>master pool</i> quando SI = EI.....</b> | <b>104</b> |
| <b>Anexo 26: Total de testes para <math>p = 0.01</math> utilizando a metodologia não hierárquica com e sem <i>master pool</i> quando SI = EI.....</b>  | <b>105</b> |

|  |            |
|--|------------|
| <b>Anexo 27: Total de testes para <math>p = 0.05</math> utilizando a metodologia não hierárquica com e sem <i>master pool</i> quando <math>SI = EI</math> .....</b>  | <b>106</b> |
| <b>Anexo 28: Acurácia da metodologia não hierárquica com e sem <i>master pool</i> para <math>p = 0.025</math> utilizando quando <math>SI = EI</math>.....</b>  | <b>107</b> |
| <b>Anexo 29: Acurácia da metodologia não hierárquica com e sem <i>master pool</i> para <math>p = 0.025</math> utilizando quando <math>SI = EI</math>.....</b>  | <b>108</b> |
| <b>Anexo 30: Acurácia da metodologia não hierárquica com e sem <i>master pool</i> para <math>p = 0.005</math> utilizando quando <math>SI = EI</math>.....</b>  | <b>109</b> |
| <b>Anexo 31: Sensibilidade da metodologia não hierárquica com e sem <i>master pool</i> para <math>p = 0.025</math> utilizando quando <math>SI = EI</math>.....</b>   | <b>110</b> |
| <b>Anexo 32: Sensibilidade da metodologia não hierárquica com e sem <i>master pool</i> para <math>p = 0.01</math> utilizando quando <math>SI = EI</math>.....</b>  | <b>111</b> |
| <b>Anexo 33: Sensibilidade da metodologia não hierárquica com e sem <i>master pool</i> para <math>p = 0.005</math> utilizando quando <math>SI = EI</math>.....</b>   | <b>112</b> |
| <b>Anexo 34: Especificidade da metodologia não hierárquica com e sem <i>master pool</i> para <math>p = 0.025</math> utilizando quando <math>SI = EI</math>.....</b>  | <b>113</b> |
| <b>Anexo 35: Especificidade da metodologia não hierárquica com e sem <i>master pool</i> para <math>p = 0.01</math> utilizando quando <math>SI = EI</math>.....</b>   | <b>114</b> |
| <b>Anexo 36: Especificidade da metodologia não hierárquica com e sem <i>master pool</i> para <math>p = 0.005</math> utilizando quando <math>SI = EI</math>.....</b>  | <b>115</b> |
| <b>Anexo 37: Programação elaborada através da linguagem R para a realização de testes compostos utilizando a metodologia de Dorfman para uma taxa de prevalência <math>p = 0.9</math> .....</b>                      | <b>116</b> |
| <b>Anexo 38: Programação elaborada através da linguagem R para a realização de testes compostos utilizando a metodologia hierárquica para uma taxa de prevalência <math>p = 1</math> e <math>n = 125</math>.....</b> | <b>118</b> |

**Anexo 39: Programação elaborada através da linguagem R para a realização de testes compostos utilizando a metodologia não hierárquica com e sem *master pool* ..... 120**

# Lista de Figuras

|   |    |
|---|----|
| Figura 1 - Fluxograma do processo de testes compostos.....  | 6  |
| Figura 2 - Fluxograma da metodologia hierárquica em três etapas .....   | 24 |
| Figura 3 - Matriz quadrada com todos os indivíduos saudáveis .....  | 28 |
| Figura 4 - Matriz quadrática com 1 indivíduo doente .....   | 28 |
| Figura 5 - Matriz quadrada com 2 indivíduos doentes na mesma linha .....  | 29 |
| Figura 6 - Matriz quadrada com 2 indivíduos doentes em linhas e colunas diferentes.....   | 29 |
| Figura 7 - Matriz quadrada com 3 indivíduos doentes .....   | 30 |
| Figura 8 - Matriz quadrada com 3 indivíduos doentes em linhas e colunas distintas .....   | 30 |
| Figura 9 - Matriz Quadrática com 3 indivíduos doentes em 3 colunas distintas e 2 linhas .....   | 31 |
| Figura 10 - Gráfico da média total de testes usando a metodologia hierárquica, para uma taxa de prevalência de 2.5% quando SI = EI.....                 | 80 |
| Figura 11 - Gráfico da média total de testes usando a metodologia de Dorfman, para uma taxa de prevalência de 2.5% quando SI = EI.....                  | 80 |
| Figura 12 - Gráfico da média total de testes usando a metodologia hierárquica, para uma taxa de prevalência de 1% quando SI = EI.....                   | 81 |
| Figura 13 - Gráfico da média total de testes usando a metodologia de Dorfman, para uma taxa de prevalência de 1% quando SI = EI.....                    | 81 |
| Figura 14 - Gráfico da média total de testes usando a metodologia hierárquica, para uma taxa de prevalência de 0.5% quando SI = EI.....                 | 82 |
| Figura 15 - Gráfico da média total de testes usando a metodologia de Dorfman, para uma taxa de prevalência de 0.5% quando SI = EI.....                  | 82 |
| Figura 16 - Gráfico da média da acurácia da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 2.5% quando SI = EI .....     | 83 |
| Figura 17 - Gráfico da média da acurácia da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 2.5% quando SI = EI .....      | 83 |
| Figura 18 - Gráfico da média da acurácia da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 1% quando SI = EI .....       | 84 |
| Figura 19 - Gráfico da média da acurácia da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 1% quando SI = EI .....        | 84 |
| Figura 20 - Gráfico da média da acurácia da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 0.5% quando SI = EI .....     | 85 |
| Figura 21 - Gráfico da média da acurácia da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 0.5% quando SI = EI .....      | 85 |
| Figura 22 - Gráfico da média da sensibilidade da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 2.5% quando SI = EI..... | 86 |

|   |    |
|---|----|
| Figura 23 - Gráfico da média da sensibilidade da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 2.5% quando SI = EI.....            | 86 |
| Figura 24 - Gráfico da média da sensibilidade da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 1% quando SI = EI.....             | 87 |
| Figura 25 - Gráfico da média da sensibilidade da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 1% quando SI = EI.....              | 87 |
| Figura 26 - Gráfico da média da sensibilidade da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 0.5% quando SI = EI.....           | 88 |
| Figura 27 - Gráfico da média da sensibilidade da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 0.5% quando SI = EI.....            | 88 |
| Figura 28 - Gráfico da média da especificidade da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 2.5% quando SI = EI.....          | 89 |
| Figura 29 - Gráfico da média da especificidade da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 2.5% quando SI = EI.....           | 89 |
| Figura 30 - Gráfico da média da especificidade da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 1% quando SI = EI.....            | 90 |
| Figura 31 - Gráfico da média da especificidade da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 1 % quando SI = EI.....            | 90 |
| Figura 32 - Gráfico da média da especificidade da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 0.5% quando SI = EI.....          | 91 |
| Figura 33 - Gráfico da média da especificidade da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 0.5 % quando SI = EI.....          | 91 |
| Figura 34 - Gráfico da média total de testes usando a metodologia hierárquica, para uma taxa de prevalência de 2.5% quando SI e EI são distintas .....            | 92 |
| Figura 35 - Gráfico da média total de testes usando a metodologia de Dorfman, para uma taxa de prevalência de 2.5% quando SI e EI são distintas .....             | 92 |
| Figura 36 - Gráfico da média total de testes usando a metodologia hierárquica, para uma taxa de prevalência de 1% quando SI e EI são distintas .....              | 93 |
| Figura 37 - Gráfico da média total de testes usando a metodologia de Dorfman, para uma taxa de prevalência de 1% quando SI e EI são distintas .....               | 93 |
| Figura 38 - Gráfico da média total de testes usando a metodologia hierárquica, para uma taxa de prevalência de 0.5% quando SI e EI são distintas .....            | 94 |
| Figura 39 - Gráfico da média total de testes usando a metodologia de Dorfman, para uma taxa de prevalência de 0.5% quando SI e EI são distintas .....             | 94 |
| Figura 40 - Gráfico da média da acurácia da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 2.5% quando SI e EI são distintas ..... | 95 |
| Figura 41 - Gráfico da média da acurácia da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 2.5% quando SI e EI são distintas .....  | 95 |
| Figura 42 - Gráfico da média da acurácia da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 1% quando SI e EI são distintas .....   | 96 |

|   |     |
|---|-----|
| Figura 43 - Gráfico da média da acurácia da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 1% quando SI e EI são distintas .....          | 96  |
| Figura 44 - Gráfico da média da acurácia da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 0.5% quando SI e EI são distintas .....       | 97  |
| Figura 45 - Gráfico da média da acurácia da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 0.5% quando SI e EI são distintas .....        | 97  |
| Figura 46 - Gráfico da média da sensibilidade da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 2.5% quando SI e EI são distintas .....  | 98  |
| Figura 47 - Gráfico da média da sensibilidade da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 2.5% quando SI e EI são distintas .....   | 98  |
| Figura 48 - Gráfico da média da sensibilidade da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 1% quando SI e EI são distintas .....    | 99  |
| Figura 49 - Gráfico da média da sensibilidade da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 1% quando SI e EI são distintas .....     | 99  |
| Figura 50 - Gráfico da média da sensibilidade da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 0.5% quando SI e EI são distintas .....  | 100 |
| Figura 51 - Gráfico da média da sensibilidade da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 0.5% quando SI e EI são distintas .....   | 100 |
| Figura 52 - Gráfico da média da especificidade da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 2.5% quando SI e EI são distintas ..... | 101 |
| Figura 53 - Gráfico da média da especificidade da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 2.5% quando SI e EI são distintas .....  | 101 |
| Figura 54 - Gráfico da média da especificidade da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 1% quando SI e EI são distintas .....   | 102 |
| Figura 55 - Gráfico da média da especificidade da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 1 % quando SI e EI são distintas .....   | 102 |
| Figura 56 - Gráfico da média da especificidade da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 0.5% quando SI e EI são distintas ..... | 103 |
| Figura 57 - Gráfico da média da especificidade da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 0.5 % quando SI e EI são distintas ..... | 103 |
| Figura 58 - Gráfico da média total de testes usando a metodologia não hierárquica com <i>master pool</i> , para uma taxa de prevalência de 2.5% quando SI = EI .....    | 104 |
| Figura 59 - Gráfico da média total de testes usando a metodologia não hierárquica sem <i>master pool</i> , para uma taxa de prevalência de 2.5% quando SI = EI .....    | 104 |
| Figura 60 - Gráfico da média total de testes usando a metodologia não hierárquica com <i>master pool</i> , para uma taxa de prevalência de 1% quando SI = EI .....      | 105 |
| Figura 61 - Gráfico da média total de testes usando a metodologia não hierárquica sem <i>master pool</i> , para uma taxa de prevalência de 1% quando SI = EI .....      | 105 |
| Figura 62 - Gráfico da média total de testes usando a metodologia não hierárquica com <i>master pool</i> , para uma taxa de prevalência de 0.5% quando SI = EI .....    | 106 |

|  |     |
|--|-----|
| Figura 63 - Gráfico da média total de testes usando a metodologia não hierárquica sem <i>master pool</i> , para uma taxa de prevalência de 0.5% quando SI = EI.....                  | 106 |
| Figura 64 - Gráfico da média da acurácia da metodologia usando a metodologia não hierárquica com <i>master pool</i> , para uma taxa de prevalência de 2.5% quando SI = EI.....       | 107 |
| Figura 65 - Gráfico da média da acurácia da metodologia usando a metodologia não hierárquica sem <i>master pool</i> , para uma taxa de prevalência de 2.5% quando SI = EI.....       | 107 |
| Figura 66 - Gráfico da média da acurácia da metodologia usando a metodologia não hierárquica com <i>master pool</i> , para uma taxa de prevalência de 1% quando SI = EI.....         | 108 |
| Figura 67 - Gráfico da média da acurácia da metodologia usando a metodologia não hierárquica sem <i>master pool</i> , para uma taxa de prevalência de 1% quando SI = EI.....         | 108 |
| Figura 68 - Gráfico da média da acurácia da metodologia usando a metodologia não hierárquica com <i>master pool</i> , para uma taxa de prevalência de 0.5% quando SI = EI.....       | 109 |
| Figura 69 - Gráfico da média da acurácia da metodologia usando a metodologia não hierárquica sem <i>master pool</i> , para uma taxa de prevalência de 0.5% quando SI = EI.....       | 109 |
| Figura 70 - Gráfico da média da sensibilidade da metodologia usando a metodologia não hierárquica com <i>master pool</i> , para uma taxa de prevalência de 2.5% quando SI = EI.....  | 110 |
| Figura 71 - Gráfico da média da sensibilidade da metodologia usando a metodologia não hierárquica sem <i>master pool</i> , para uma taxa de prevalência de 2.5% quando SI = EI.....  | 110 |
| Figura 72 - Gráfico da média da sensibilidade da metodologia usando a metodologia não hierárquica com <i>master pool</i> , para uma taxa de prevalência de 1% quando SI = EI.....    | 111 |
| Figura 73 - Gráfico da média da sensibilidade da metodologia usando a metodologia não hierárquica sem <i>master pool</i> , para uma taxa de prevalência de 1% quando SI = EI.....    | 111 |
| Figura 74 - Gráfico da média da sensibilidade da metodologia usando a metodologia não hierárquica com <i>master pool</i> , para uma taxa de prevalência de 0.5% quando SI = EI.....  | 112 |
| Figura 75 - Gráfico da média da sensibilidade da metodologia usando a metodologia não hierárquica sem <i>master pool</i> , para uma taxa de prevalência de 0.5% quando SI = EI.....  | 112 |
| Figura 76 - Gráfico da média da especificidade da metodologia usando a metodologia não hierárquica com <i>master pool</i> , para uma taxa de prevalência de 2.5% quando SI = EI..... | 113 |
| Figura 77 - Gráfico da média da especificidade da metodologia usando a metodologia não hierárquica sem <i>master pool</i> , para uma taxa de prevalência de 2.5% quando SI = EI..... | 113 |
| Figura 78 - Gráfico da média da especificidade da metodologia usando a metodologia não hierárquica com <i>master pool</i> , para uma taxa de prevalência de 1% quando SI = EI.....   | 114 |
| Figura 79 - Gráfico da média da especificidade da metodologia usando a metodologia não hierárquica sem <i>master pool</i> , para uma taxa de prevalência de 1% quando SI = EI.....   | 114 |
| Figura 80 - Gráfico da média da especificidade da metodologia usando a metodologia não hierárquica com <i>master pool</i> , para uma taxa de prevalência de 0.5% quando SI = EI..... | 115 |
| Figura 81 - Gráfico da média da especificidade da metodologia usando a metodologia não hierárquica sem <i>master pool</i> , para uma taxa de prevalência de 0.5% quando SI = EI..... | 115 |

# Lista de tabelas

|  |    |
|--|----|
| Tabela 1 - Matriz de confusão .....  | 12 |
| Tabela 2 – Dimensão de grupo com melhor média de testes e eficiência usando a metodologia hierárquica numa taxa de prevalência de 2.5% quando SI = EI .....                            | 37 |
| Tabela 3 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia de Dorfman numa taxa de prevalência de 2.5% quando SI = EI .....                             | 38 |
| Tabela 4 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia hierárquica numa taxa de prevalência de 1% quando SI = EI .....                              | 39 |
| Tabela 5 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia de Dorfman numa taxa de prevalência de 1% quando SI = EI .....                               | 39 |
| Tabela 6 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia hierárquica numa taxa de prevalência de 0.5% quando SI = EI .....                            | 40 |
| Tabela 7 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia tradicional de Dorfman numa taxa de prevalência de 0.5% quando SI = EI .....                 | 41 |
| Tabela 8 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia hierárquica numa taxa de prevalência de 2.5% quando SI e EI são distintas .....              | 48 |
| Tabela 9 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia de Dorfman numa taxa de prevalência de 2.5% quando SI e EI são distintas .....               | 48 |
| Tabela 10 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia hierárquica numa taxa de prevalência de 1% quando SI e EI são distintas .....               | 49 |
| Tabela 11 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia de Dorfman numa taxa de prevalência de 1% quando SI e EI são distintas .....                | 50 |
| Tabela 12 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia hierárquica numa taxa de prevalência de 0.5% quando SI e EI são distintas .....             | 50 |
| Tabela 13 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia de Dorfman numa taxa de prevalência de 0.5% quando SI e EI são distintas .....              | 51 |
| Tabela 14 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia não hierárquica com <i>master pool</i> numa taxa de prevalência de 2.5% quando SI = EI..... | 58 |
| Tabela 15 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia não hierárquica sem <i>master pool</i> numa taxa de prevalência de 2.5% quando SI = EI..... | 59 |
| Tabela 16 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia não hierárquica com <i>master pool</i> numa taxa de prevalência de 1% quando SI = EI .....  | 60 |
| Tabela 17 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia não hierárquica sem <i>master pool</i> numa taxa de prevalência de 1% quando SI = EI .....  | 60 |
| Tabela 18 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia não hierárquica com <i>master pool</i> numa taxa de prevalência de 0.5% quando SI = EI..... | 61 |
| Tabela 19 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia não hierárquica sem <i>master pool</i> numa taxa de prevalência de 0.5% quando SI = EI..... | 61 |

Tabela 20 - Comparação das metodologias estudadas, numa taxa de prevalência  $p = 0.01$  e  $SI = EI = 0.95$  ..69

Tabela 21 - Comparação das metodologias estudadas, numa taxa de prevalência  $p = 0.005$  e  $SI = EI = 0.99$  71

## Lista de siglas e acrónimos

|            |                                    |
|------------|------------------------------------|
| <b>CR</b>  | Custo relativo                     |
| <b>EI</b>  | Especificidade do teste individual |
| <b>EF</b>  | Eficiência                         |
| <b>EM</b>  | Especificidade da metodologia      |
| <b>EP</b>  | Erro padrão                        |
| <b>FN</b>  | Falso negativo                     |
| <b>FP</b>  | Falso positivo                     |
| <i>n</i>   | Dimensão de grupo                  |
| <b>NC</b>  | Número de colunas                  |
| <b>NL</b>  | Número de linhas                   |
| <i>p</i>   | Taxa de prevalência                |
| <b>SI</b>  | Sensibilidade do teste individual  |
| <b>SM</b>  | Sensibilidade da metodologia       |
| <b>VN</b>  | Verdadeiro negativo                |
| <b>VP</b>  | Verdadeiro positivo                |
| <b>VPN</b> | Valor preditivo negativo           |
| <b>VPP</b> | Valor preditivo positivo           |



# 1. Introdução

A evolução das metodologias de diagnóstico tem sido um pilar fundamental no avanço da medicina e da saúde pública, permitindo não só a identificação precisa de doenças, mas também a implementação de medidas preventivas e terapêuticas eficazes. Dentro deste contexto, os testes compostos surgem como uma abordagem inovadora que promete revolucionar os processos de triagem e diagnóstico em larga escala. Este trabalho dedica-se ao estudo e à avaliação de metodologias de classificação baseadas em testes compostos, destacando a sua relevância e aplicabilidade em cenários de saúde onde a precisão e a economia de recursos são cruciais.

A pertinência deste tema deriva da necessidade crescente de métodos de diagnóstico que não apenas aumentem a eficiência operacional, mas que também sejam economicamente viáveis, especialmente em países e regiões com recursos mais limitados. Os testes compostos, ao reduzirem o número de procedimentos necessários para identificar indivíduos doentes em populações de grande dimensão, oferecem uma solução potencialmente transformadora para o campo da saúde global.

Os objetivos desta dissertação estão centrados na comparação de várias metodologias de testes compostos, tanto hierárquicas quanto não hierárquicas, avaliando a sua acurácia, sensibilidade, especificidade, eficiência e custo-benefício em diferentes taxas de prevalência. O estudo visa identificar quais dessas metodologias oferecem os maiores benefícios em diferentes cenários, proporcionando uma análise sobre a eficácia e a fiabilidade dessas metodologias.

Para alcançar estes objetivos, o estudo emprega uma metodologia que inclui a realização de simulações computacionais e análises estatísticas. Utilizando a linguagem R, um ambiente de software livre para computação estatística e gráficos, este trabalho analisa dados simulados para testar várias configurações de testes compostos, comparando-os com metodologias de teste em termos de eficiência, custo e precisão diagnóstica.

A estrutura desta dissertação está organizada em cinco capítulos principais, além desta introdução e das conclusões finais. O primeiro capítulo apresenta uma revisão da literatura sobre métodos de classificação com destaque em testes compostos, estabelecendo o contexto

teórico para as análises subsequentes. O segundo capítulo descreve as metodologias específicas utilizadas no estudo, detalhando os procedimentos e as técnicas de simulação adotadas. Os terceiro e quarto capítulos apresentam, respetivamente, os resultados das simulações e a discussão destes, interpretando as implicações dos resultados no contexto das hipóteses propostas. Finalmente, o quinto capítulo oferece uma síntese das contribuições do estudo para o campo da saúde pública e da medicina diagnóstica, sugerindo direções para pesquisas futuras.

Esta dissertação pretende contribuir para o campo da epidemiologia e do diagnóstico médico, oferecendo informações sobre como os testes compostos podem ser otimizados para maximizar a sua eficácia e eficiência.

## 2. Noções e metodologias

Neste capítulo, será explorado detalhadamente o conceito de testes compostos. Inicialmente será feita uma revisão das bases teóricas que fundamentam esta metodologia, seguida de uma análise das vantagens e limitações associadas ao seu uso. Serão abordadas também as diferentes estratégias de implementação dos testes compostos, destacando como eles podem ser utilizados para maximizar a eficiência e a acurácia de diagnóstico em diversas situações. Além disso, serão apresentados exemplos e estudos para ilustrar a aplicação real desta metodologia.

### 2.1. Testes de classificação

Os testes de classificação são ferramentas essenciais em diversas áreas, como educação, psicologia, recursos humanos e muitas outras. Estes testes são projetados para categorizar ou agrupar indivíduos com base em certas características ou habilidades específicas. A principal função destes testes é avaliar atributos que não são facilmente mensuráveis de forma direta, utilizando, em vez disso, uma série de perguntas ou tarefas cuja análise permite inferir traços ou capacidades particulares dos sujeitos testados.

Os testes de classificação visam atribuir uma categoria ou classe a cada indivíduo avaliado. Isso é feito a partir da análise das respostas ou desempenhos em diferentes atividades propostas pelo teste. A categorização pode variar conforme o propósito do teste, podendo incluir níveis de habilidade, tipos de personalidade, estilos de aprendizagem ou outros aspectos relevantes. O objetivo central dos testes de classificação é fornecer uma visão estruturada e quantificável das diferenças individuais, facilitando, assim, a tomada de decisões em contextos como seleção de pessoal, diagnóstico clínico, entre outros. Estes testes são elaborados de maneira a garantir validade e confiabilidade, assegurando que os resultados sejam consistentes e representem com precisão o constructo que pretendem medir.

Os testes de classificação têm uma ampla gama de aplicações práticas. Na educação, são utilizados para identificar estilos de aprendizagem e adaptar métodos de ensino às necessidades dos alunos. Em psicologia, ajudam a diagnosticar transtornos e a entender traços de personalidade. No mundo corporativo, são essenciais para processos de recrutamento e seleção, e para desenvolvimento de carreiras. Além disso, com o avanço das

tecnologias digitais, os testes de classificação têm-se tornado cada vez mais acessíveis e sofisticados, permitindo avaliações *online* que podem ser analisadas em tempo real, oferecendo *feedback* imediato e personalizado.

Os testes de classificação representam um recurso valioso para a compreensão e gestão das diferenças individuais em vários contextos. A sua correta aplicação e interpretação são fundamentais para garantir que as decisões tomadas a partir dos seus resultados sejam justas e precisas. Com o desenvolvimento contínuo de novas metodologias e tecnologias, a eficácia e a abrangência desses testes tendem a crescer, consolidando ainda mais a sua importância no futuro.

No campo da saúde, os testes de classificação desempenham um papel crucial no diagnóstico e tratamento de diversas condições médicas. Esses testes são utilizados para categorizar pacientes com base em sintomas, históricos médicos e resultados de exames, facilitando a identificação de doenças e a definição de estratégias de tratamento. Por exemplo, na psiquiatria, testes de classificação ajudam a diagnosticar transtornos mentais como depressão, ansiedade e esquizofrenia, ao agrupar sintomas em categorias específicas que correspondem a critérios diagnósticos estabelecidos. Na medicina, classificações baseadas em fatores de risco e manifestações clínicas são essenciais para a gestão de doenças crônicas, como diabetes e hipertensão, permitindo que os profissionais de saúde personalizem os planos de tratamento de acordo com as necessidades individuais dos pacientes. Com o avanço das tecnologias de saúde digital, como a inteligência artificial, os testes de classificação estão a tornar-se cada vez mais precisos e eficientes, proporcionando uma assistência médica mais eficaz e centrada no paciente.

## **2.2. Testes compostos**

Os testes individuais permitem detetar imediatamente uma determinada doença de um sujeito, como por exemplo, doente ou não doente. Por outro lado, os testes compostos são valiosos por permitirem economizar no número total de testes necessários, sem comprometer significativamente a eficácia e a sensibilidade do diagnóstico. A testagem com recurso a testes compostos visa otimizar os recursos, procurando um equilíbrio entre precisão e eficiência na avaliação clínica.

Nesta abordagem, a amostra observada é previamente agrupada em grupos de dimensão  $n$  e, em seguida, é feito um teste a cada grupo. Caso um grupo obtenha um resultado negativo,

assume-se que nenhum indivíduo do grupo está doente. Caso um grupo obtenha um resultado positivo, assume-se que pelo menos um membro do grupo está doente, e é então tradicionalmente necessário realizar testes individuais subsequentes em cada elemento do grupo para identificar os indivíduos doentes dentro desse grupo.

A implementação de testes compostos traz benefícios significativos em termos de eficiência e economia de recursos, especialmente em situações de triagem em larga escala. Esta metodologia permite reduzir o número de testes necessários ao agrupar amostras, testando-as de forma coletiva inicialmente. Assim, quando um grupo apresentar um resultado negativo, como todos os indivíduos desse grupo são considerados negativos, com um único teste foram classificados  $n$  indivíduos, poupando-se  $n-1$  testes em relação à aplicação de testes individuais. Por outro lado, como um resultado positivo requer testes subsequentes em cada indivíduo do grupo para identificar quais são os indivíduos positivos, serão realizados, nestes casos,  $n+1$  testes para classificar os  $n$  indivíduos do grupo (mais um teste do que na aplicação direta de testes individuais). Em situação de baixa taxa de prevalência, os grupos tendem a ser negativos (nenhum elemento do grupo está doente), pelo que o número de testes necessário para classificar a população será menor do que na aplicação de testes individuais. Assim, esta estratégia otimiza o uso de recursos laboratoriais e diminui o tempo necessário para a triagem de grandes populações, mantendo um nível elevado de precisão diagnóstica.

Além da eficiência, os testes compostos também simplificam a interpretação dos resultados diagnósticos e melhoram a comunicação da performance do teste entre os profissionais de saúde. O uso de testes compostos facilita uma resposta mais rápida em emergências de saúde pública, onde é necessário realizar um grande número de testes num curto período. Esta metodologia é particularmente útil para gerir surtos de doenças ou para campanhas de rastreamento populacional, permitindo uma triagem eficaz sem comprometer os bons diagnósticos. [1, 2]

É possível representar a temática dos testes compostos através do esquema apresentado na Figura 1, onde o sangue de 5 indivíduos ( $n=5$ ) é misturado e, após tornar a mistura homogênea, corresponderá à amostra composta.

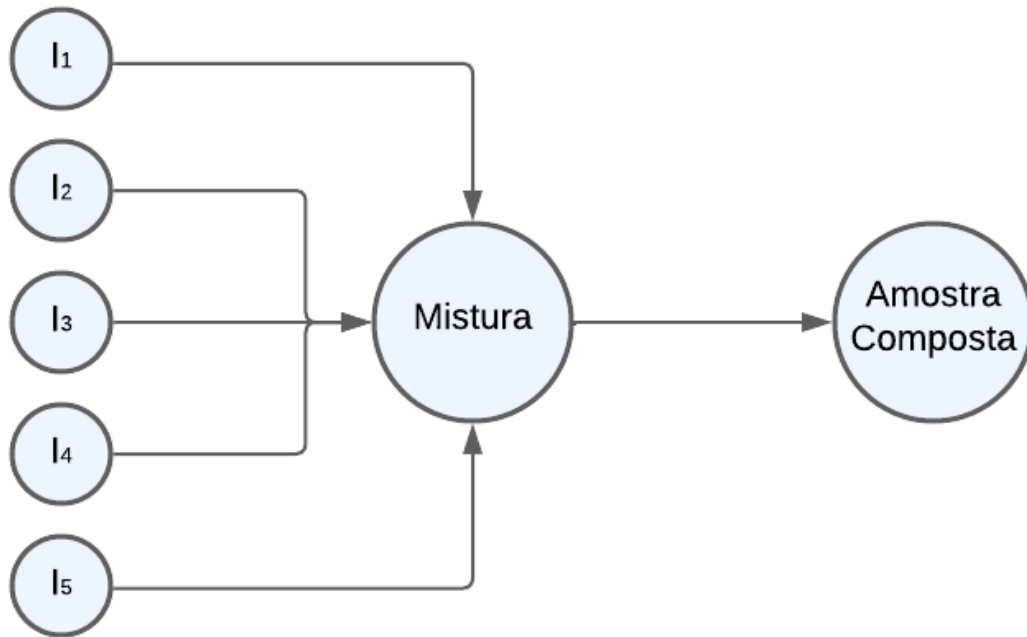


Figura 1 - Fluxograma do processo de testes compostos

Em suma, os testes compostos desempenham um papel vital na melhoria da precisão diagnóstica e na avaliação de modelos de risco biomédico, oferecendo uma abordagem integrada para uma tomada de decisão clínica mais informada e eficaz.

### 2.3. Diluição em testes compostos

A diluição em testes compostos pode impactar a capacidade de detecção da doença na realização do teste diagnóstico. Quando amostras de indivíduos saudáveis são misturadas com amostras de indivíduos doentes, a quantidade do agente causador da doença presente na amostra total pode ser reduzida devido à diluição, o que, em alguns casos, torna mais difícil a identificação precisa do patogénico. Esse efeito ocorre porque a concentração do agente detetável, como um vírus ou bactéria, pode não ser suficiente para ser identificada pelos métodos laboratoriais disponíveis.

Por exemplo, se numa amostra composta por 10 indivíduos apenas um estiver doente, o agente causador da doença será diluído pelas outras 9 amostras não doentes. Dependendo da capacidade de detecção do teste utilizado, essa diluição pode comprometer o resultado, levando a um falso negativo. Portanto, há um equilíbrio entre a economia de testes proporcionada por essa combinação de amostras e o risco de perda de precisão na detecção devido à diluição.

Em cenários de baixa prevalência de doenças, como observado em estratégias de testes durante a pandemia de COVID-19, a diluição de uma amostra positiva com várias amostras negativas pode levar a uma diluição viral significativa. Especificamente, a mistura de uma amostra positiva com sete negativas pode aumentar o número de ciclos de amplificação necessários em testes RT-PCR, elevando o risco de falsos negativos devido à diluição excessiva, como discutido por Grobe *et al.* [3].

Para mitigar os efeitos da diluição, são necessários métodos de detecção capazes de identificar pequenas quantidades do agente, especialmente em cenários de baixa prevalência, onde as chances de uma amostra composta conter indivíduos doentes são menores. Por outro lado, em populações com alta prevalência, a diluição pode ter um impacto menor, já que a presença de mais indivíduos contaminados na amostra aumenta a quantidade total do agente a ser detectado, facilitando a sua identificação.

A diluição, portanto, representa um desafio importante nos testes compostos, exigindo uma escolha cuidadosa da estratégia de combinação de amostras e das tecnologias de detecção empregadas, para garantir que a eficiência na testagem não comprometa a precisão dos resultados [3].

## **2.4. Testes qualitativos versus testes quantitativos**

Os testes qualitativos são métodos de avaliação que fornecem informações sobre a presença ou ausência de uma determinada doença ou característica, sem medir quantitativamente a sua intensidade ou concentração. Em vez de fornecer um valor numérico, os testes qualitativos resultam em respostas binárias, como positivo/negativo, presente/ausente ou sim/não. Estes testes são amplamente utilizados na área da saúde para detectar a presença de doenças, infecções ou outros parâmetros clínicos [4].

Um exemplo comum de teste qualitativo na saúde é o teste de gravidez. Este teste detecta a presença da hormona hCG na urina da mulher. Se a hormona estiver presente, o resultado é positivo, indicando gravidez; se estiver ausente, o resultado é negativo. Outro exemplo é o teste rápido de HIV, que verifica a presença de anticorpos contra o vírus HIV no sangue ou saliva. O resultado é positivo se os anticorpos forem detectados, sugerindo uma doença de HIV, e negativo se não forem detectados.

Os testes qualitativos também são frequentemente utilizados em triagens de infeções, como testes rápidos para COVID-19, onde a presença de antígenos do vírus é verificada. Estes testes são cruciais para identificar rapidamente indivíduos doentes e implementar medidas de contenção adequadas.

De acordo com Santos *et al.* [4], os testes compostos quantitativos são métodos utilizados para analisar a quantidade de uma determinada substância em amostras combinadas de vários indivíduos. Esta abordagem é especialmente útil em contextos de saúde pública, onde é necessário testar grandes populações de forma eficiente e económica. Para realizar um teste composto quantitativo, um grupo de indivíduos é selecionado e é recolhida uma amostra de sangue (ou outro fluido biológico) de cada pessoa. Estas amostras são então misturadas para formar uma amostra homogénea composta. A quantidade total da substância de interesse presente na amostra composta é medida. A quantidade medida é então comparada a um limiar predefinido. Se a quantidade da substância na amostra composta exceder esse limiar, pelo menos um dos indivíduos do grupo é considerado positivo para a doença testada. Caso contrário, todos os indivíduos do grupo são considerados negativos.

Uma das principais vantagens dos testes compostos quantitativos é a redução significativa do número de testes necessários para analisar uma população, o que resulta na economia de tempo e recursos. Esta técnica é particularmente eficaz em situações de baixa prevalência de uma doença, onde a maioria das amostras compostas resultará negativa, permitindo a classificação de múltiplos indivíduos com um único teste. Os testes compostos quantitativos são aplicáveis em diversas áreas da saúde, incluindo triagem de doenças infecciosas como HIV, hepatite e COVID-19. Estes testes também podem ser usados em programas de monitorização ambiental e controlo de qualidade em alimentos e água.

Apesar das vantagens, os testes compostos quantitativos apresentam algumas limitações. Uma preocupação importante é o risco de diluição, onde a mistura de amostras pode reduzir a concentração da substância de interesse a níveis indetetáveis, resultando em falsos negativos. Por isso, a escolha adequada da dimensão do grupo e a definição precisa dos limiares de deteção são cruciais para a eficácia do método. Além disso, a heterogeneidade das amostras, como a variação nos tipos de sangue dos indivíduos, pode afetar a precisão dos resultados. Portanto, é fundamental que os profissionais de saúde avaliem cuidadosamente as condições em que os testes compostos quantitativos serão aplicados para garantir a fiabilidade dos resultados. Para esta análise, é frequente recorrer à modelação da

distribuição da quantidade em análise nos indivíduos doentes e não doentes. A diferença entre estas distribuições irá permitir procurar a metodologia de classificação mais adequada à aplicação pretendida. [5, 6, 7, 8]

É importante mencionar que este trabalho se concentra exclusivamente na análise de testes compostos qualitativos, evitando a complexidade associada à análise da diluição da substância, que permite a classificação e determinação do ponto de corte. Isso ocorre porque, em muitos casos, a complexidade de identificação da presença ou ausência de uma característica específica num teste composto é semelhante à realizada em testes individuais (supondo que o teste consegue sempre identificar a presença da referida característica).

## **2.5. Eficiência e custo relativo**

Num contexto de produção, o conceito de eficiência refere-se à capacidade de uma unidade de produção, como uma empresa ou uma organização, utilizar os seus recursos (como materiais, mão de obra e capital) da melhor maneira possível para produzir uma quantidade desejada de produtos ou serviços. Uma unidade é considerada eficiente se não puder produzir mais produtos ou serviços sem aumentar a quantidade de recursos utilizados, ou se não puder reduzir a quantidade de recursos sem diminuir a produção.

Em termos simples, a eficiência implica alcançar um alto nível de produção com o mínimo de recursos disponíveis, evitando desperdícios e otimizando o uso dos recursos existentes. Assim, será naturalmente uma medida importante para avaliar o desempenho de uma unidade de produção e identificar oportunidades de melhoria na utilização de recursos. [9]

Já a eficiência em testes compostos é fundamental em cenários onde a prevalência de doença é baixa, visando reduzir os custos de classificação e estimação<sup>1</sup>. Testes compostos, como descritos em estudos especializados, empregam estratégias inteligentes de agrupamento para maximizar a utilização eficiente dos recursos disponíveis.

A eficiência é avaliada em termos do número médio esperado de testes necessários por indivíduo para a sua classificação como doente ou não doente. Métodos como a metodologia de Dorfman (que será explicada no próximo capítulo), são empregues para garantir

---

<sup>1</sup> Neste trabalho unicamente iremos utilizar testes compostos com o objetivo de classificar, mas estes testes também podem ser utilizados com o objetivo de estimar a taxa de prevalência da doença, permitindo obter estimadores mais eficiente (com menor variância) realizando o mesmo número de testes de diagnóstico.

resultados precisos e simultaneamente maximizarem a economia de recursos. No entanto, é importante notar que esses métodos podem ser mais eficazes em situações de baixa prevalência da doença, visto que em epidemias com alta prevalência, como no caso do COVID-19, a eficácia pode ser comprometida devido ao grande número de testes necessários para identificar os casos positivos de forma rápida e precisa. Nestas situações, o teste de grupo dará certamente positivo pelo que a sua utilização será irrelevante, provavelmente até aumentando o número de testes necessários para a concretização da classificação.

Portanto, nas metodologias de classificação baseadas em testes compostos, a eficiência é medida pela capacidade de reduzir o número total de testes necessários. Assim, quanto menos testes forem realizados para atingir esses objetivos, maior será a eficiência do método. [4]

Podemos calcular a Eficiência (EF) através da seguinte fórmula:

$$EF (\%) = \left( 1 - \frac{\text{N.}^\circ \text{ de Testes}}{\text{Dimensão da população}} \right) \times 100.$$

O custo relativo nos testes compostos é uma medida que compara o número de testes necessários numa metodologia de testes compostos com o número de testes necessários se fossem realizados testes individuais. Para calcular o custo relativo, é utilizado o número esperado de testes na metodologia aplicada dividido pelo número total de indivíduos, multiplicado por 100 para expressar em percentagem. Assim, o Custo Relativo (CR) é calculado da seguinte forma:

$$CR \% = \left( \frac{\text{N.}^\circ \text{ de Testes}}{\text{Dimensão da população}} \right) \times 100.$$

Esta medida tem uma relação inversa com a eficiência, sendo que, quanto menor o custo relativo, menos testes são necessários realizar e, portanto, maior a eficiência. Logo para avaliar a eficiência dos testes compostos, se o valor do custo relativo for superior a 100, significa que a metodologia de classificação baseada em testes compostos requer mais testes do que aplicando testes individuais, tornando a primeira opção menos eficiente. Por outro lado, um custo relativo inferior a 100 indica que a metodologia de classificação baseada em testes compostos é mais eficiente, necessitando de menos testes para classificar o mesmo número de indivíduos. O objetivo é minimizar o custo relativo, i.e., maximizar a eficiência do processo de classificação da população.

Por exemplo, numa população de 1000 indivíduos, num contexto de baixa prevalência (1%), podemos dividir a população em 100 grupos de 10 indivíduos cada. Com a taxa de prevalência de 1%, aproximadamente 9.56 grupos terão pelo menos um indivíduo positivo. Portanto, seriam necessários 10 testes de grupo positivos mais 100 testes individuais (10 grupos positivos  $\times$  10 indivíduos), totalizando 200 testes. A eficiência pode ser calculada como:

$$EF (\%) = \left(1 - \frac{200}{1000}\right) \times 100 = 80\%.$$

Este valor indica uma economia significativa em relação ao teste individual de todos os 1000 indivíduos, havendo uma diminuição de 80% do número de testes necessários para a classificação de toda a população.

Utilizando os mesmos dados, podemos calcular também o custo relativo:

$$CR \% = \left(\frac{200}{1000}\right) \times 100 = 20\%.$$

Um custo relativo de 20% significa que os testes compostos requerem apenas 20% do número de testes necessários em comparação com os testes individuais, demonstrando uma economia significativa.

## 2.6. Sensibilidade e especificidade

Segundo Van Stralen *et al.* [10], a sensibilidade mede a capacidade de um teste diagnóstico identificar corretamente os indivíduos com a doença alvo, que poderá ser uma doença ou outro desfecho de saúde. Por outras palavras, a sensibilidade é a proporção de resultados positivos no teste entre os pacientes com a doença (verdadeiros positivos) em relação ao total de pacientes com a doença (verdadeiros positivos + falsos negativos).

Por outro lado, a especificidade é definida como a capacidade de um teste originar um resultado negativo quando o paciente não tem a doença (verdadeiros negativos), ou seja, é a proporção de verdadeiros negativos em relação ao número total de pacientes sem a doença (verdadeiros negativos + falsos positivos) [11]. Sensibilidade e especificidade são conceitos fundamentais na avaliação de testes clínicos [12, 13]. Deste modo, a sensibilidade indica a capacidade de o teste identificar as pessoas que têm a doença. Por exemplo, um teste com 100% de sensibilidade identifica todas as pessoas com a doença, enquanto um teste com

80% de sensibilidade identifica 80% das pessoas com a doença. A sensibilidade é especialmente importante em doenças com tratamento, como o cancro, onde um teste sensível pode identificar casos precoces.

Por outro lado, Akobeng [14] aborda a utilidade prática da sensibilidade e especificidade dos testes diagnósticos, destacando a sua importância na deteção ou exclusão de uma pessoa com ou sem doença. Tradicionalmente, os conceitos de sensibilidade e especificidade são explanados com base numa tabela  $2 \times 2$ , representada na Tabela 1.

|      |     | Valor Preditivo          |                          |
|------|-----|--------------------------|--------------------------|
|      |     | Sim                      | Não                      |
| Real | Sim | Verdadeiro Positivo (VP) | Falso Negativo (FN)      |
|      | Não | Falso Positivo (FP)      | Verdadeiro Negativo (VN) |

**Tabela 1 - Matriz de confusão**

Akobeng [14] destaca a utilidade desses parâmetros na prática clínica, explicando que um teste com alta sensibilidade é útil para "descartar" uma doença quando o resultado do teste é negativo, enquanto um teste com alta especificidade é útil para "confirmar" uma doença quando o resultado é positivo.

A sensibilidade dos testes individuais (SI), habitualmente referida unicamente como sensibilidade, é calculada pela seguinte fórmula:

$$SI = \frac{VP}{VP + FN}$$

A especificidade dos testes individuais (EI), habitualmente denominada unicamente por especificidade, é calculada pela seguinte fórmula:

$$EI = \frac{VN}{VN + FP}$$

No entanto, Akobeng [14] também evidencia as limitações da sensibilidade e da especificidade, observando que esses parâmetros não podem ser usados para estimar a probabilidade da doença num paciente individual (taxa de prevalência). Note-se que a sensibilidade é determinada considerando unicamente os doentes, enquanto a especificidade é determinada unicamente pelos não doentes, mantendo-se ambas as medidas constantes

para quaisquer valores da taxa de prevalência. Deste modo, estas duas medidas dependem das características do teste diagnóstico e não da quantidade de doentes na população.

Gallagher [15] traz uma perspectiva adicional ao destacar a confusão conceitual gerada pela formulação contraintuitiva da sensibilidade e especificidade. Enfatiza que essas medidas, ao responderem à pergunta oposta do que muitos médicos esperam, desafiam o raciocínio clínico habitual. No entanto, Gallagher [15] ressalta a importância das propriedades matemáticas da sensibilidade e da especificidade na interpretação dos resultados dos testes diagnósticos. Uma alta sensibilidade indica baixo número de falsos negativos, o que fortalece a confiança em resultados negativos verdadeiros. Da mesma forma, uma alta especificidade indica baixo número de falsos positivos, confirmando a presença da doença com mais confiança.

## 2.7. Acurácia

A acurácia de um teste refere-se à sua capacidade de fornecer resultados corretos, ou seja, identificar corretamente os casos positivos (verdadeiros positivos) e os casos negativos (verdadeiros negativos). A acurácia é uma medida global da precisão do teste e é calculada como a proporção de casos corretamente identificados (verdadeiros positivos + verdadeiros negativos) em relação ao total de casos testados. Uma acurácia alta indica que o teste é eficaz a distinguir indivíduos com a doença-alvo dos que não têm a doença.

A acurácia de um teste diagnóstico é essencial para determinar a sua capacidade de fornecer resultados precisos e relevantes. Embora a sensibilidade e a especificidade sejam métricas comuns para avaliar a acurácia, especialmente em testes simples, quando lidamos com testes compostos que envolvem múltiplas variáveis, a avaliação da acurácia revela-se mais complexa.

A Acurácia é calculada pela seguinte fórmula:

$$\text{Acurácia} = \frac{\text{VP} + \text{VN}}{\text{Dimensão da população}}$$

Conforme discutido por Moons *et al.* [16], a acurácia de um teste diagnóstico é fundamentalmente a sua capacidade de fornecer resultados corretos e precisos em relação ao estado de saúde real de um paciente. Em termos simples, a acurácia reflete o quão bem um teste pode identificar corretamente a presença ou ausência de uma doença específica.

A sensibilidade e a especificidade representam probabilidades inversas que refletem a capacidade do teste de identificar corretamente a presença ou a ausência de uma doença, independentemente da sua prevalência na população. No entanto, ao discutir a acurácia num contexto mais amplo, como em testes compostos que envolvem múltiplas testagens, a avaliação da acurácia revela-se mais complexa. Os testes compostos utilizam múltiplos testes individuais em conjunto para determinar a presença ou a ausência de uma doença. Em vez de realizar um único teste por pessoa, vários testes são aplicados a amostras agrupadas, otimizando os recursos e mantendo a eficácia diagnóstica.

A acurácia de um teste diagnóstico é a sua capacidade de fornecer resultados corretos, ou seja, de identificar com precisão se um paciente possui ou não uma determinada doença médica. Essa medida de acurácia é crucial para avaliar o desempenho de um teste e a sua utilidade na prática clínica [17].

## **2.8. Valor preditivo negativo e valor preditivo positivo**

Além das medidas tradicionais de sensibilidade e especificidade, outras medidas são úteis para avaliar a fiabilidade da classificação em contextos diagnósticos mais complexos. Por exemplo, o valor preditivo positivo (VPP) representa a proporção de casos positivos reais entre todos os casos classificados como positivos pelo teste, enquanto o valor preditivo negativo (VPN) representa a proporção de casos negativos reais entre todos os casos classificados como negativos pelo teste.

Os conceitos de VPN e VPP são fundamentais na avaliação da eficácia de testes de diagnóstico em diversas áreas, incluindo a saúde. Estes valores ajudam a entender a precisão de um teste em identificar corretamente a presença ou ausência de uma doença específica. O VPP é a probabilidade de que um indivíduo tenha realmente uma doença, dado que o teste resultou em positivo. Por outras palavras, é a proporção de resultados positivos verdadeiros entre todos os resultados positivos do teste, sendo calculado pela fórmula:

$$\text{VPP} = \frac{\text{VP}}{\text{VP} + \text{FP}}$$

Um VPP elevado indica que o teste tem uma alta probabilidade de identificar corretamente os indivíduos com a doença quando o resultado é positivo.

O VPN é a probabilidade de que um indivíduo não tenha a doença, dado que o teste resultou negativo. Por outras palavras, é a proporção de resultados negativos verdadeiros entre todos os resultados negativos do teste, sendo calculado pela fórmula:

$$VPN = \frac{VN}{VN + FN}$$

Um VPN elevado indica que o teste tem uma alta probabilidade de identificar corretamente os indivíduos sem a doença quando o resultado é negativo.

Tanto o VPP quanto o VPN são influenciados pela taxa de prevalência da doença na população testada. Em populações onde a doença é rara, o VPP tende a ser menor porque a maioria dos resultados positivos pode ser falso positivo. Inversamente, em populações onde a doença é comum, o VPN tende a ser menor porque a maioria dos resultados negativos pode ser falso negativo. Estes valores são cruciais para a interpretação dos resultados dos testes diagnósticos e para a tomada de decisões clínicas. Por exemplo, num cenário de triagem para uma doença grave ou um tratamento agressivo, um VPP alto é desejável para garantir que os indivíduos identificados como positivos realmente tenham a doença e possam receber o tratamento adequado (garantindo que há poucos indivíduos incorretamente classificados como positivos a serem submetidos a um tratamento agressivo). Da mesma forma, um VPN alto é importante para assegurar que os indivíduos com resultado negativo estejam realmente livres da doença e não precisem de intervenções urgentes, como ocorre em muitas doenças que se agravam rapidamente.

Segundo Van Stralen *et al.* [10], ao testar uma população, e concluir que a sensibilidade e especificidade de cada teste individual era de 85%, quando aplicado a uma população com uma taxa de prevalência de cerca de 6%, o VPP do teste foi de 27%, indicando que, entre os que testaram positivo, 27% realmente tinham a doença. Por outro lado, o VPN foi de 99%, indicando que, entre os que testaram negativo, 99% realmente não tinham a doença.

De forma a poder explicar estes resultados, podemos utilizar o teorema de Bayes, que é uma fórmula fundamental na teoria das probabilidades que permite calcular a probabilidade de uma hipótese H ser verdadeira, dada uma evidência E. Essa probabilidade é conhecida como probabilidade condicional. Utilizando o teorema de Bayes e os valores observados por Van Stralen *et al.* [10], podemos calcular o valor do VPP e do VPN. Para isso, consideramos as seguintes definições:

- $P(D|+)$  é a probabilidade de ter a doença dado um teste positivo (VPP);
- $P(+|D)$  é a probabilidade de um teste positivo dado que a pessoa tem a doença (sensibilidade);
- $P(D)$  é a taxa de prevalência da doença na população;
- $P(+)$  é a probabilidade de um teste positivo, que pode ser encontrada usando a regra da soma ou o teorema da probabilidade total;
- $P(\bar{D}|-)$  é a probabilidade de não ter a doença dado um teste negativo (VPN);
- $P(-|\bar{D})$  é a probabilidade de um teste negativo dado que a pessoa não tem a doença (especificidade);
- $P(\bar{D})$  é a probabilidade de não ter doença na população (inverso da taxa de prevalência);
- $P(-)$  é a probabilidade de um teste negativo, que pode ser encontrada usando a regra da soma.

Para calcular o a probabilidade de um teste positivo:

$$P(+)=P(+|D)\times P(D)+P(+|\bar{D})\times P(\bar{D}).$$

Substituindo os valores observados por Van Stralen *et al.* [10] temos:

$$P(+)=(0.85\times 0.06)+(0.15\times 0.94)=0.051+0.141=0.192.$$

Sendo assim, podemos agora calcular o VPP, utilizando a fórmula (teorema de Bayes):

$$P(D|+)=\frac{P(+|D)\times P(D)}{P(+)}.$$

Substituindo novamente os valores observados por Van Stralen *et al.* [10], obtemos:

$$P(D|+)=\frac{0.85\times 0.06}{0.192}=\frac{0.051}{0.192}\approx 0.27.$$

Já para calcular o VPN usamos:

$$P(\bar{D}|T-)=\frac{P(-|\bar{D})\times P(\bar{D})}{P(-)},$$

onde:

$$P(-) = P(-|D) \times P(D) + P(-|\bar{D}) \times P(\bar{D}).$$

Utilizando estas fórmulas e os valores observados por Van Stralen *et al.* [10] temos:

$$P(-) = (0.15 \times 0.06) + (0.85 \times 0.94) = 0.009 + 0.799 = 0.808.$$

Sendo assim, podemos agora calcular o VPP através de:

$$P(\bar{D}|-) = \frac{0.85 \times 0.94}{0.808} = \frac{0.799}{0.808} \approx 0.99.$$

Este exemplo ilustra como os valores preditivos positivo e negativo podem variar com a taxa de prevalência da doença. Numa população com uma maior taxa de prevalência, o VPP seria maior, enquanto o VPN seria menor (considerando os mesmos valores de sensibilidade e especificidade). Isto destaca a importância de considerar a taxa de prevalência da doença ao interpretar os resultados dos testes de diagnóstico. Em suma, os valores preditivos positivo e negativo são métricas essenciais para avaliar a eficácia dos testes de diagnóstico na prática clínica, ajudando os profissionais de saúde a interpretar os resultados dos testes e tomar decisões informadas sobre o diagnóstico e o tratamento dos pacientes.

## 2.10. Linguagem R e Simulação

De forma a realizar a obter resultados e a analisá-los, as simulações foram realizadas utilizando a linguagem R [18]. O R é uma linguagem de programação livre (grátis) voltada para a análise estatística e gráficos. É amplamente utilizada em pesquisas acadêmicas e profissionais devido à sua vasta gama de *packages*, flexibilidade e capacidade de lidar com grandes volumes de dados.

Especificamente, nesta simulação e análise de resultados, foram utilizados dois *packages* importantes: “*caret*” [19] e “*ggplot2*” [20]. O *package* “*caret*” (*classification and regression training*) é uma ferramenta abrangente que facilita o ajuste e avaliação de modelos de aprendizagem automática (*machine learning*). Oferece funcionalidades que incluem o pré-processamento de dados, seleção de variáveis e ajuste de parâmetros, além de suportar uma ampla gama de algoritmos de classificação e regressão. O “*caret*” é essencial para o desenvolvimento de modelos robustos e a comparação de diferentes técnicas de modelação, utilizando uma interface consistente e intuitiva. Neste caso em específico, foram calculadas métricas como a sensibilidade, especificidade, valor preditivo positivo e valor

preditivo negativo, através da função “*confusionMatrix*” (matriz de confusão). Esta função compara os resultados dos testes simulados com o estado real da doença dos indivíduos, fornecendo uma avaliação detalhada do desempenho dos testes em diferentes condições.

Além disso, o “*ggplot2*” foi utilizado para a visualização dos dados e resultados. Este *package* permite a criação de gráficos sofisticados e personalizáveis. Com o “*ggplot2*” é possível mapear dados para estéticas visuais e construir gráficos como dispersão, *boxplots*, mapas de calor (*heatmaps*), entre outros, de maneira eficiente. A flexibilidade do “*ggplot2*” é ampliada pelo uso de camadas, que permite a construção incremental de gráficos, adicionando ou modificando componentes conforme necessário.

Assim, o uso da linguagem R, juntamente com os *packages* “*caret*” e “*ggplot2*”, foi fundamental para a realização das análises estatísticas e a apresentação clara dos resultados nesta dissertação.

A simulação é uma técnica que permite replicar o comportamento de fenômenos aleatórios, proporcionando uma descrição aproximada das características de um processo, como a aplicação de diferentes metodologias de classificação. Uma das vantagens da simulação é a capacidade de fixar previamente as características da população, como a taxa de prevalência, a sensibilidade e a especificidade dos testes, permitindo gerar aleatoriamente populações que refletem essas condições. Por exemplo, utilizando a função *rbinom* ( $n, l, prob$ ) na linguagem R, é possível criar um vetor de tamanho  $n$  que representa indivíduos doentes (1) ou saudáveis (0), de acordo com a taxa de prevalência estabelecida ( $prob$ ). Essa mesma função pode ser usada para simular os resultados dos testes, considerando a sensibilidade e especificidade pré-definidas.

Ao simular populações de grande dimensão e repetir o processo em várias simulações, os resultados tendem a apresentar pouca variabilidade, aproximando-se dos valores teóricos esperados. Essa precisão é garantida pelo Teorema Limite Central e a Lei dos Grandes Números, que asseguram que as frequências relativas obtidas na simulação convergem para as probabilidades reais à medida que o número de repetições aumenta. O Teorema Limite Central é fundamental na estatística moderna, pois permite que, independentemente da distribuição original da população (desde que tenha variância finita), a distribuição da média de amostras se aproxime de uma distribuição normal à medida que o tamanho da amostra aumenta [21] [22].

Assim, a simulação não só permite avaliar a sensibilidade e a especificidade de uma metodologia de classificação, mas também medir a variabilidade desses valores, garantindo que os resultados sejam precisos, isto é, próximos dos valores teóricos.

Por exemplo, ao simular o lançamento de uma moeda, podemos determinar a frequência relativa de "caras" e "coroas", que se irá aproximar da probabilidade teórica de 50% com o aumento do número de lançamentos. Este princípio pode ser aplicado a testes compostos para avaliar a fiabilidade de metodologias de classificação. Ao variar parâmetros como a dimensão dos grupos, é possível analisar como isso afeta o custo relativo e a fiabilidade da metodologia, permitindo comparar diferentes abordagens e selecionar a mais adequada para a doença em estudo.

## **2.11.Fiabilidade em testes compostos**

Nesta seção será analisada a fiabilidade dos testes compostos, abordando como as métricas de sensibilidade, especificidade e outras medidas são aplicadas no contexto de grupos de indivíduos testados coletivamente. A fiabilidade será discutida considerando dois cenários principais: quando há pelo menos um indivíduo doente no grupo e quando o grupo é composto apenas por indivíduos saudáveis.

### **2.11.1. Grupo doente versus grupo saudável**

Segundo Dorfman [23], podemos entender a aplicação dos testes compostos, especialmente no contexto da diferenciação entre grupos doentes e grupos saudáveis. Dorfman introduziu esta metodologia como uma maneira eficiente de identificar indivíduos doentes em grandes populações, especialmente em situações onde a testagem individual seria impraticável ou muito cara.

No contexto dos testes compostos, um grupo doente é definido como aquele que contém pelo menos um indivíduo doente. Deste modo, quando um grupo apresenta um resultado positivo num teste composto, isso indica que o teste considera a presença da doença em pelo menos um dos membros do grupo.

Por outro lado, um grupo saudável é aquele que não tem nenhum indivíduo doente e, portanto, um resultado negativo no teste composto indica que o teste considera que nenhum dos indivíduos do grupo está doente. Este resultado elimina a necessidade de testes

adicionais para os membros do grupo, proporcionando uma economia substancial de tempo e recursos.

### **2.11.2. Sensibilidade e especificidade em testes compostos**

A sensibilidade e especificidade em testes compostos são métricas cruciais para avaliar a eficácia dessa metodologia, especialmente quando se lida com a triagem de grandes populações. Enquanto a sensibilidade e especificidade individual se concentram na precisão do teste ao nível de cada indivíduo, a sensibilidade e especificidade em testes compostos avaliam o desempenho do teste ao nível de grupo, o que é essencial para entender a eficiência geral dos testes compostos. De acordo com Grobe *et al.* [3], a sensibilidade em testes compostos refere-se à capacidade do teste composto de identificar corretamente um grupo como positivo quando pelo menos um indivíduo dentro desse grupo está realmente doente. Por outras palavras, é a probabilidade de que o teste composto apresente um resultado positivo para um grupo que contém pelo menos um caso positivo real. Esta métrica é particularmente importante em cenários onde a detecção de qualquer caso de doença dentro de um grupo é crítica para evitar a propagação da doença. Por exemplo, se testarmos grupos de 10 alunos para COVID-19 numa escola, e um desses grupos contiver pelo menos um aluno infetado, a sensibilidade em testes compostos diz-nos a probabilidade de que o teste apresente um resultado positivo para esse grupo. Se a sensibilidade for de 95%, isso significa que, em 95% das vezes, o teste composto identificará corretamente o grupo doente como positivo, alertando para a necessidade de testes individuais adicionais nesse grupo.

Por outro lado, a especificidade em testes compostos, segundo Grobe *et al.* [3], refere-se à capacidade do teste composto de identificar corretamente um grupo como negativo quando todos os indivíduos dentro do grupo são saudáveis. Em termos práticos, isso significa que a especificidade em testes compostos é a probabilidade de que um teste composto apresente um resultado negativo para um grupo onde não há nenhum indivíduo doente. Esta métrica é crucial para evitar falsos positivos, que poderiam levar a testes subsequentes desnecessários, aumentando os custos e a carga sobre os recursos de saúde. Por exemplo, se o grupo de 10 alunos na escola não contiver nenhum aluno doente, a especificidade em testes compostos indica a probabilidade de o teste apresentar um resultado negativo, confirmando que o grupo é saudável. Se a especificidade for de 98%, significa que, em 98% das vezes, o teste composto identificará corretamente o grupo como negativo, evitando que os alunos saudáveis sejam submetidos a testes adicionais desnecessários.

Como estamos a utilizar testes qualitativos, assumimos que essas medidas mantêm os valores dos testes individuais, ou seja, a fiabilidade dos testes compostos é considerada igual à dos testes individuais, uma vez que assumimos, por exemplo, que não existem os problemas de diluição da substância de interesse como no caso dos testes quantitativos.

## 3. Metodologias de classificação baseadas em testes compostos

Este capítulo é dedicado à apresentação das diferentes metodologias de classificação baseadas em testes compostos, incluindo metodologias hierárquicas e não hierárquicas. São descritas as características principais de cada tipo, as suas vantagens e desvantagens e são dados exemplos de aplicações práticas. São analisados estudos comparativos que avaliam a eficácia de cada metodologia em diferentes cenários de taxa de prevalência e recursos disponíveis.

### 3.1. Metodologia de Dorfman

Em 1943, Dorfman abordou a questão da identificação de indivíduos doentes numa população, num contexto onde testar cada membro individualmente seria impraticável ou muito dispendioso. Assim, propôs um método alternativo, onde amostras aleatórias são retiradas da população e testadas em conjunto. Com base nos resultados dos testes dessas amostras, é possível deduzir a taxa de prevalência de uma determinada doença na população em estudo. [23]

Deste modo, Dorfman descreveu um procedimento estatístico para combinar os resultados dos testes de grupo e calcular a probabilidade de um indivíduo ao acaso estar doente (taxa de prevalência).

A metodologia de Dorfman, que será igualmente referida como metodologia tradicional, foi desenvolvida inicialmente para encontrar, de forma mais eficiente, quantos e quais os soldados que tinham sífilis, sendo aplicada esta metodologia numa larga escala no sistema de saúde pública dos Estados Unidos durante a segunda guerra mundial. Ao permitir uma abordagem mais eficiente e económica para a deteção de doenças em grandes populações, o trabalho de Dorfman teve um impacto significativo e continua a ser estudado e aplicado até hoje.

Este método propõe uma abordagem eficiente para identificar casos positivos de doença em grandes populações. Inicialmente, retiram-se amostras aleatórias que são agrupadas e testadas em conjunto. Se o teste em grupo resultar em positivo, indica que pelo menos um

indivíduo no grupo está doente. Nesta situação, são realizados testes individuais para identificar quais os indivíduos afetados. Por outro lado, se o teste de grupo der negativo, conclui-se que nenhum indivíduo do grupo está doente, o que resulta na poupança de recursos e de tempo.

Esta estratégia de testes permite uma deteção mais ágil e económica de doenças em populações de grande dimensão. Ao realizar testes em grupo inicialmente e recorrer a testes individuais apenas quando necessário, o método de Dorfman torna-se uma ferramenta valiosa em situações onde a testagem individual seria impraticável ou financeiramente inviável [23].

### **3.2. Metodologia hierárquica**

A metodologia hierárquica em testes compostos é uma estratégia eficaz utilizada na triagem de amostras para identificar indivíduos doentes por uma determinada doença, com o objetivo principal de otimizar a utilização de recursos, reduzindo custos e tempo. Esta abordagem envolve uma série de testes realizados em etapas, começando com testes amplos aplicados a grupos maiores de indivíduos. Se um teste inicial aplicado a um grupo der um resultado negativo, conclui-se que todo o grupo está livre da doença investigada, evitando a necessidade de submeter a mais testes cada membro do grupo. No entanto, se o resultado inicial for positivo, indicando a possível presença da doença no grupo testado, a amostra é então subdividida em subgrupos menores, e cada subgrupo é novamente testado para determinar quais deles contêm indivíduos doentes [24]. O número de vezes em que cada grupo se divide pode ser ajustado, terminando sempre com testes individuais nos casos em que um grupo ou subgrupo der positivo. O número de testes que um indivíduo passa até ser classificado como infetado corresponde ao número de etapas desta metodologia. Deste modo, a metodologia de Dorfman analisada na secção anterior corresponde a uma metodologia hierárquica de duas etapas, onde na primeira etapa são realizados testes a grupos de dimensão  $n$  e na segunda de dimensão 1 (testes individuais) [25].

Os subgrupos que testarem negativo são considerados livres da doença. Este processo de divisão hierárquica e testes subsequentes continua a reduzir a amostra com o objetivo de encontrar indivíduos doentes. Somente os subgrupos que continuam a apresentar resultados positivos avançam para a próxima fase, eventualmente conduzindo a testes individuais para confirmar os casos suspeitos. Por exemplo, ao realizar uma triagem para identificar casos de

COVID-19 numa comunidade com 1000 pessoas, o processo pode passar por dividir as 1000 pessoas em 50 grupos de 20 pessoas cada para o Teste A em conjunto (etapa 1). Suponhamos que 10 desses grupos, totalizando 200 pessoas, testam positivo para COVID-19 (estes grupos têm pelo menos um infetado). De seguida, as 200 pessoas são subdivididas em 20 subgrupos de 10 pessoas cada para o Teste B composto (na prática, cada grupo positivo de 20 pessoas é dividido em dois grupos de 10 pessoas – etapa 2), e suponhamos que 8 dos 20 subgrupos, totalizando 80 pessoas, apresentem resultado positivo no Teste B. Por fim, as 80 pessoas que testaram positivo no Teste B são submetidas a testes individuais (Teste C) para confirmar os casos suspeitos de COVID-19. Esta abordagem hierárquica em três etapas (Teste A em grupos de 20 pessoas, Teste B em grupos de 10 pessoas e Teste C individual) permite identificar eficientemente os casos positivos, concentrando os recursos de testagem nos grupos e indivíduos com maior probabilidade de estarem doentes. Podemos observar o exemplo através do seguinte fluxograma:

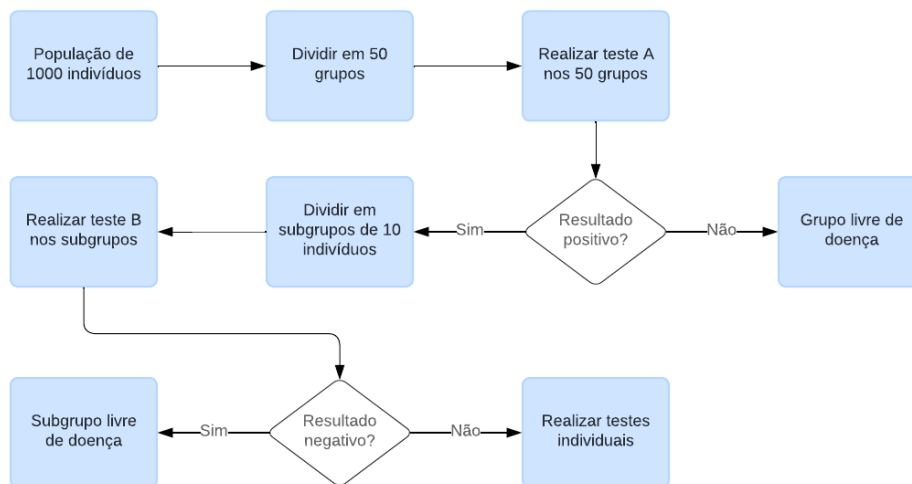


Figura 2 - Fluxograma da metodologia hierárquica em três etapas

Segundo Sherlock *et al.* [24], para otimizar o número de testes, a seleção do número de subgrupos e as suas dimensões em cada etapa é crucial. Têm sido propostos novos procedimentos de reteste informativos, que consideram a heterogeneidade entre os indivíduos, para melhorar essa eficiência. Esses procedimentos utilizam modelos de regressão binária para estimar as probabilidades de positividade individual, e então formam subgrupos de modo a minimizar o número esperado de testes.

Em programas de testes de HIV, como os descritos por Sherlock *et al.* [24], um algoritmo de três etapas é comumente usado. Por exemplo, em Los Angeles, as amostras são

inicialmente agrupadas em dimensões de 90. Se um grupo testar positivo, é dividido em nove subgrupos de 10. Qualquer subgrupo positivo é então direcionado para a realização de testes individuais.

Segundo o estudo conduzido por Black *et al.* [26], que envolvia testes de clamídia e gonorreia, estados como Idaho implementaram procedimentos de reteste informativos para economizar recursos. Usando dados de fatores de risco, como histórico sexual e observações clínicas, estes métodos ajustam o número de subgrupos e as suas dimensões de forma a minimizar o número de testes necessários, melhorando a eficiência geral.

Os métodos hierárquicos de testes compostos proporcionam uma redução significativa no número de testes necessários em comparação com a utilização de testes individuais. Estudos mostram que esses métodos são particularmente eficazes quando há uma variabilidade significativa nas probabilidades de positividade entre os indivíduos. Os resultados deste estudo mostraram que, para testes de HIV e de clamídia/gonorreia, a aplicação dos novos procedimentos de reteste informativos levou a uma redução significativa no número de testes necessários. No contexto dos programas de testes de HIV nos EUA, observou-se que os métodos propostos poderiam reduzir o número de testes em até 36.8% em comparação com os métodos hierárquicos tradicionais. No Projeto de Prevenção da Infertilidade (IPP), a implementação desses procedimentos também resultou na economia considerável de recursos [26].

Além da eficiência em termos de número de testes, os métodos hierárquicos propostos por Black *et al.* [26] mantiveram altos níveis de fiabilidade na identificação de indivíduos doentes e saudáveis. A consideração das probabilidades individuais de positividade permitiu uma melhor alocação dos recursos de testagem, especialmente em populações com alta variabilidade nas taxas de doença.

Granado [27], obteve resultados que indicam que a metodologia hierárquica, em geral, apresenta uma eficiência muito próxima com e sem erros de classificação. No entanto, tal como na metodologia de Dorfman, a sensibilidade da metodologia hierárquica é sempre inferior à sensibilidade do testes individual e composta considerada (próxima dos 85%), enquanto a especificidade é sempre superior à especificidade do teste individual e composta utilizada. Esta abordagem é particularmente vantajosa em termos de eficiência, reduzindo significativamente o número de testes necessários em comparação com a testagem individual, especialmente em contextos de baixas taxas de prevalência.

Grobe *et al.* [3] demonstraram que a eficiência dos testes compostos é maior em condições de baixa prevalência de doença. Para uma prevalência de 1%, os testes hierárquicos de três etapas mostraram-se mais eficientes do que os testes hierárquicos de duas etapas. Especificamente, no método de duas etapas, para uma prevalência de 1%, a dimensão de grupo ótima foi 13, com uma economia de custos de 83.7%. Na metodologia hierárquica de 3 etapas, para a mesma prevalência, a dimensão de grupo ótima foi 58, 29, 1 com uma economia de custos de 90.6%.

À medida que a prevalência aumenta, a eficiência dos testes compostos diminui. Por exemplo, com uma prevalência de 10%, as dimensões de grupo ótimas foram menores, e a eficiência entre as diferentes metodologias tornou-se comparável, mostrando uma redução significativa na vantagem das metodologias de classificação baseadas em testes compostos em relação aos testes individuais.

A recolha de amostras é uma parte crítica deste processo e pode ser realizada de maneiras que ajudem a balancear a eficiência dos testes compostos e a precisão dos resultados. Uma técnica prática para reduzir a necessidade de recolher amostras novamente em caso de resultados positivos no grupo é o uso de amostras divididas. Com esta técnica, cada amostra é recolhida em dois recipientes separados: a primeira metade é utilizada para o teste de grupo, enquanto a segunda metade é armazenada individualmente e usada apenas se o grupo der positivo, permitindo testes individuais sem a necessidade de nova recolha (usualmente, mesmo em metodologias hierárquicas de 3 ou 4 etapas, só é feita uma recolha que é dividida de forma a ser utilizada, caso necessário, em cada etapa). Ao utilizar amostras divididas, a amostra individual está prontamente disponível para testes corroborativos, reduzindo o tempo e a logística envolvida na recolha de amostras e diminuindo a probabilidade de resultados falsos negativos.

De acordo com Hou *et al.* [28], a utilização de uma metodologia hierárquica com 4 etapas é particularmente eficaz na redução do número de testes necessários em cenários de baixa prevalência de doenças, sem comprometer significativamente a acurácia dos resultados. A escolha por quatro etapas permite uma melhor otimização dos recursos, proporcionando uma redução significativa no número total de testes. Por este motivo, e porque as características das metodologias hierárquicas com 3 etapas são mais conhecidas, neste trabalho optamos por analisar as metodologias de 4 etapas.

Implementar o uso de amostras divididas requer ajustes logísticos, embora o uso dessa técnica possa aumentar os custos iniciais de recolha, melhora a eficiência geral do processo

de teste composto, reduzindo a necessidade de múltiplas recolhas e acelerando a obtenção de resultados precisos.

Grobe *et al.* [3] concluíram que a metodologia de testes compostos baseada em matrizes pode ser uma solução eficaz para aumentar a capacidade de testagem em situações de baixa prevalência de doença. No entanto, a implementação requer uma análise cuidadosa das condições locais de prevalência e a adaptação dos protocolos para equilibrar eficiência e precisão.

### **3.3. Metodologias não hierárquicas**

Os testes compostos organizados em matrizes ou *arrays* representam uma metodologia inovadora para identificar indivíduos doentes numa população, utilizando a organização dos dados em matrizes de dimensões  $n \times n$  (quadradas) ou  $m \times n$ , isto é, não havendo obrigatoriedade de o número de linhas ser igual ao número de colunas, apesar de na prática ser o mais usual. Esta metodologia é usualmente representada por M2, que iremos usar, ou A2<sup>2</sup>. Neste método, os indivíduos são distribuídos em células individuais dentro de uma matriz, e os testes compostos são realizados nos grupos formados pelas linhas e pelas colunas dessas matrizes. Durante a execução dos testes, são identificadas as linhas e colunas com resultados positivos, o que indica a presença de pelo menos um indivíduo doente no grupo. Resultados negativos indicam que nenhum indivíduo naquela linha ou coluna está doente.

A chave para a identificação dos indivíduos doentes está na interseção das linhas e colunas com resultados positivos. Indivíduos localizados nessas interseções são considerados potencialmente doentes e eventualmente precisam de testes individuais para confirmação.

Vamos supor que temos um grupo de 25 indivíduos, todos saudáveis. Nesse caso, não deverá haver linhas ou colunas positivas, como ilustrado na Figura 1.

---

<sup>2</sup> De uma forma mais geral, nesta metodologia os indivíduos podem estar organizados em *arrays* de dimensão  $d$ , metodologia Ad, correspondendo o caso das matrizes ao caso específico de  $d = 2$  (duas dimensões – linhas e colunas). Na realidade, são raros os estudos sobre esta metodologia numa dimensão superior. [32]

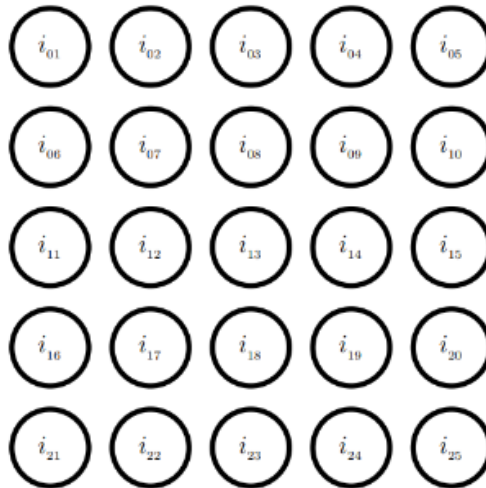


Figura 3 - Matriz quadrada com todos os indivíduos saudáveis

Neste caso, seriam realizados 10 testes compostos e seriam classificados 25 indivíduos. De forma similar, consideremos agora que a linha 2 e a coluna 4 apresentam resultados positivos. Nesse caso, o indivíduo na interseção será considerado doente, como ilustrado na Figura 4, não havendo dúvidas sobre a classificação deste indivíduo e, como tal, sem haver a necessidade de realizar teste individual.

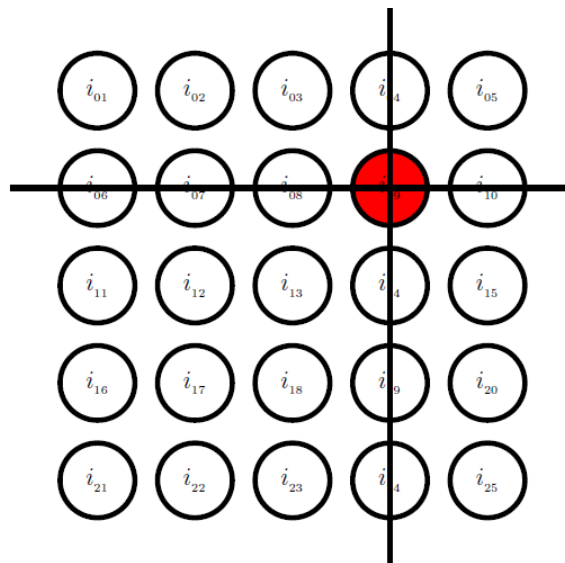


Figura 4 - Matriz quadrática com 1 indivíduo doente

Na mesma matriz, suponhamos que as colunas 2 e 4 resultem positivas, e apenas a linha 2 apresente um resultado positivo. Isso indicaria que os indivíduos nas interseções estão doentes (se não considerarmos a possibilidade de erros de classificação). A Figura 5 ilustra este exemplo.

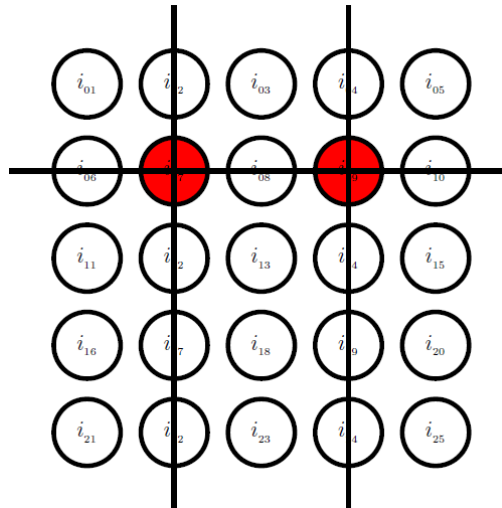


Figura 5 - Matriz quadrada com 2 indivíduos doentes na mesma linha

Se, após os testes, tanto as colunas 2 e 4 quanto as linhas 2 e 4 apresentarem resultados positivos, isso indicaria que os indivíduos nas interseções dessas linhas e colunas são potenciais positivos. Neste caso, mesmo supondo ausência de erros de classificação, os resultados não são suficientes para determinar quais os indivíduos positivos, como ocorria nos cenários previamente apresentados. Assim, esses 4 indivíduos, precisarão de testes individuais para confirmação (até podem estar os 4 doentes). Este exemplo é ilustrado na Figura 6, onde dois dos 4 potenciais doentes são efetivamente doentes e os outros saudáveis, o que ficaria esclarecido com a realização dos testes individuais.

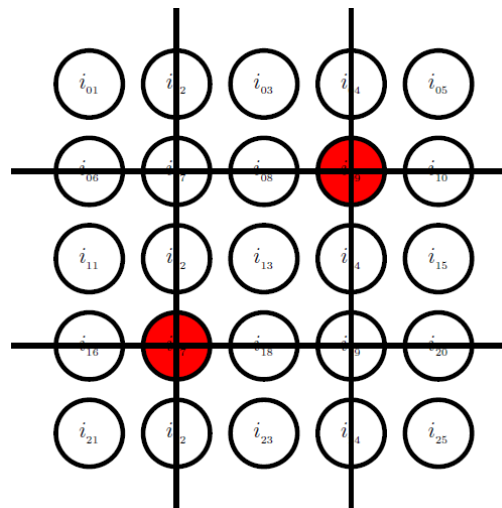


Figura 6 - Matriz quadrada com 2 indivíduos doentes em linhas e colunas diferentes

Como referido, seguindo os mesmos resultados do cenário anterior, também poderá surgir o caso de haver 3 indivíduos doentes na matriz, cf. Figura 7.

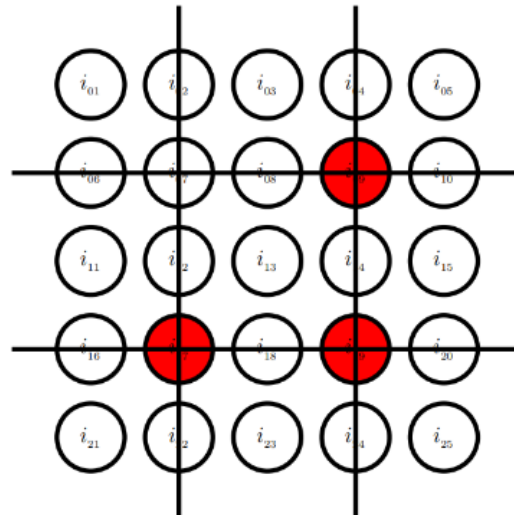


Figura 7 - Matriz quadrada com 3 indivíduos doentes

Agora, imaginemos um novo cenário em que as linhas 1, 3 e 5, assim como as colunas 1, 3 e 5, apresentam resultados positivos. Isso indicaria a presença de 9 possíveis doentes na matriz, necessitando de testes individuais. Na Figura 8, apenas 3 dos indivíduos nas interseções estão doentes, enquanto 6 são saudáveis.

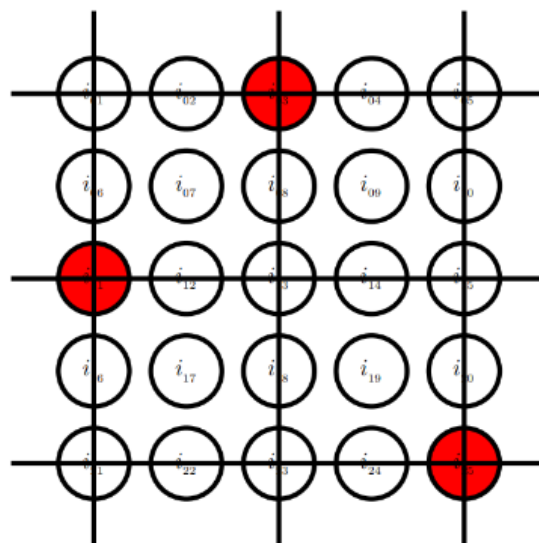


Figura 8 - Matriz quadrada com 3 indivíduos doentes em linhas e colunas distintas

Finalmente, temos um exemplo um pouco mais complexo, mas seguindo o mesmo procedimento: 3 colunas testam positivo e apenas 2 linhas apresentam resultados positivos. Na Figura 9, das seis interseções, apenas 3 indivíduos estão doentes e 3 são saudáveis:

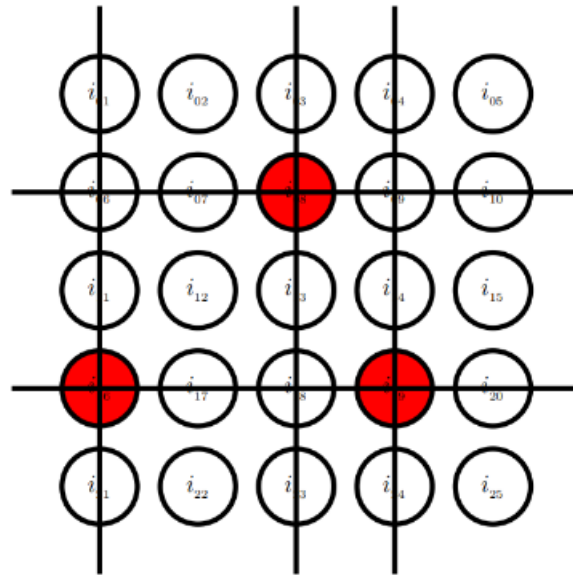


Figura 9 - Matriz Quadrática com 3 indivíduos doentes em 3 colunas distintas e 2 linhas

Esta metodologia oferece vantagens significativas, especialmente em termos de eficiência na detecção de doenças. Ao agrupar indivíduos em matrizes e realizar testes compostos nas linhas e colunas, é possível identificar rapidamente grupos de possíveis doentes, reduzindo assim o número total de testes necessários em comparação com as abordagens tradicionais, desde que a taxa de prevalência seja diminuta.

Segundo Grobe *et al.* [3], a metodologia de testes compostos não hierárquica pode incluir, antes da realização dos testes às linhas e colunas, um teste composto com todos os indivíduos da matriz (*master pool*). Esta abordagem é especialmente útil em situações onde há necessidade de aumentar a capacidade de testagem e reduzir custos em taxas de prevalência muito reduzidas.

Assim, para a realização desta metodologia, são utilizadas *pools* e *master pools*. As *pools* consistem na combinação de várias amostras individuais numa única amostra composta, ou seja, uma *pool* é equivalente a um grupo de teste (a uma linha ou a uma coluna da matriz). Já as *master pools* correspondem a uma amostra composta maior, formada pela combinação de várias *pools* menores (todos os indivíduos da matriz). Se a *master pool* testar negativo, todas as *pools* menores (e, portanto, todas as amostras individuais dentro dessas *pools*) são consideradas negativas (todos os indivíduos da matriz são considerados saudáveis).

Em suma, no teste em matriz (M2), as amostras são organizadas em matrizes, com *pools* formadas por todas as amostras na mesma linha e coluna. Cada amostra é testada duas vezes

na primeira etapa (linha e coluna correspondente). Se uma *pool* testar positivo, as amostras na interseção das linhas e colunas positivas são testadas individualmente para identificar os casos positivos (caso haja dúvidas sobre quais os indivíduos doentes). Uma variação do M2 é o M2m, que adiciona uma *master pool* antes de organizar as matrizes. Se a *master pool* testar positivo, a abordagem matricial é aplicada para identificar os positivos. Se a *master pool* testar negativo, todos os indivíduos da matriz são classificados como saudáveis. Esta estratégia pode melhorar a eficiência, especialmente em prevalências mais baixas.

Segundo Granado [27], os resultados das simulações mostram que a eficiência dos testes compostos é significativa, especialmente em populações com baixa prevalência. Para grupos de 10 indivíduos com uma prevalência de 0.05, o custo relativo foi de 37.94%, resultando numa economia de 62.06% dos testes. Em grupos de 25 indivíduos com prevalência de 0.01, o custo relativo foi de 13.57%, gerando uma economia de 86.43% dos testes. Em grupos de 106 indivíduos com prevalência de 0.001, o custo relativo foi de 2.97%, proporcionando uma economia de 97.03% dos testes.

Além da sua eficiência, mesmo com uma taxa de erro de classificação de 5% (sensibilidade e especificidade de 0.95), as metodologias de classificação baseadas em testes compostos permanecem como uma alternativa vantajosa. No entanto, esses erros podem dificultar a interpretação dos resultados, especialmente quando há divergências entre os testes de linhas e colunas. Por exemplo, se uma linha testar positivo e nenhuma coluna testar positiva, há certamente um erro de classificação, ou a linha positiva é um falso positivo ou (pelo menos) uma das colunas é um falso negativo. Nestes casos, considera-se que todos os elementos da linha positiva são potenciais doente e aplicam-se testes individuais. Ainda assim, os resultados demonstram que a metodologia é sólida, mantendo níveis adequados de sensibilidade e de especificidade.

Granado [27] conclui que as metodologias de classificação baseadas em testes compostos são recomendadas principalmente em cenários de baixa prevalência, onde a economia de recursos é maximizada sem comprometer significativamente a fiabilidade dos resultados.

### **3.4. Fiabilidade de uma metodologia de classificação**

Nesta secção, serão definidas várias métricas para avaliar a fiabilidade de uma metodologia de classificação, sendo elas: sensibilidade, especificidade, acurácia, VPP e VPN da metodologia [29].

### 3.4.1. Sensibilidade, especificidade e acurácia de uma metodologia

A sensibilidade de uma metodologia de classificação refere-se à probabilidade de que um indivíduo doente ser corretamente classificado como doente (positivo) aplicando a metodologia de classificação na população em análise. Conforme definido por Dhamnetiya *et al.* [30], a sensibilidade é a proporção de verdadeiros positivos (indivíduos corretamente identificados como doentes) sobre o total de indivíduos que realmente têm a doença. Esta métrica é crucial para avaliar a capacidade do teste de detetar corretamente os casos de doença. Por exemplo, suponhamos que uma metodologia de classificação é aplicada num grupo de 200 indivíduos, dos quais 80 são realmente doentes. Se a metodologia, independentemente do número de testes e de etapas necessárias para realizar a classificação, classificar corretamente 69 desses 80 indivíduos como doentes, a sensibilidade seria de 86,25%. Isso significa que a metodologia tem uma alta capacidade de identificar corretamente indivíduos doentes, minimizando a ocorrência de falsos negativos.

Por outro lado, a especificidade mede a capacidade da metodologia identificar corretamente os indivíduos saudáveis, classificando-os como negativos. Especificidade, conforme explicado por Dhamnetiya *et al.* [30], é a proporção de verdadeiros negativos (indivíduos saudáveis corretamente identificados como saudáveis) sobre o total de indivíduos que realmente não têm a doença. Esta medida é essencial para evitar a sobrecarga de falsos positivos, que podem levar a testes desnecessários e aumentar os custos. Por exemplo, no mesmo grupo de 200 indivíduos, se 120 são saudáveis e a metodologia classifica corretamente 95 desses 120 como saudáveis, a especificidade seria de 79.17%. Isso indica que a metodologia é razoavelmente eficaz em evitar falsos positivos, um ponto crucial abordado no artigo, onde se discute o impacto de uma baixa especificidade em contextos de triagem populacional.

A acurácia é uma medida global da fiabilidade da metodologia de classificação, refletindo a proporção de todas as classificações corretas (tanto verdadeiros positivos quanto verdadeiros negativos) em relação ao número total de casos avaliados. Esta acurácia pode ser definida como uma métrica que combina sensibilidade e especificidade para fornecer uma visão geral do desempenho do teste. Por exemplo, se, no grupo de 200 indivíduos, 164 foram corretamente classificados (69 verdadeiros positivos + 95 verdadeiros negativos), a acurácia da metodologia seria de 82%. Uma acurácia elevada sugere que a metodologia de classificação é, em geral, bastante fiável, como explicado por Dhamnetiya *et al.* [30].

### 3.4.2. VPP e VPN da metodologia

O VPP refere-se à probabilidade de que um indivíduo classificado como doente pela metodologia realmente tenha a doença. O VPP é calculado como a proporção de verdadeiros positivos sobre o total de positivos. Esta métrica é crucial para entender a fiabilidade de um teste em contextos clínicos. Por exemplo, se a metodologia classifica 94 indivíduos como doentes, mas apenas 69 desses realmente têm a doença, o VPP seria de 73.40%. Isso significa que, quando a metodologia identifica um indivíduo como doente, há 73.40% de chance de que essa classificação esteja correta.

Finalmente, VPN, conforme discutido no referido artigo, é a probabilidade de que um indivíduo classificado como saudável pela metodologia realmente seja saudável. O VPN é calculado como a proporção de verdadeiros negativos sobre o total de negativos. O VPN é essencial para avaliar a capacidade de uma metodologia de classificação em contextos onde a minimização dos falsos negativos é crítica. Por exemplo, se a metodologia classifica 106 indivíduos como saudáveis, dos quais 95 realmente são saudáveis, o VPN seria de 89.62%. Isso significa que, quando a metodologia classifica um indivíduo como saudável, há 89.62% de chance de que essa classificação esteja correta, ilustrando a importância do VPN.

Sublinhamos que os conceitos sensibilidade, especificidade, acurácia e valores preditivos positivos e negativos são iguais, quer sejam no contexto dos testes individuais, quer seja nos testes compostos (onde precisamos da definição de grupo doente e saudável), quer seja na aplicação de uma metodologia de classificação.

## 4. Análise de resultados

Neste capítulo, serão comparados os resultados de três metodologias de testes compostos: a metodologia hierárquica de Dorfman (doravante designada apenas por Dorfman ou tradicional), a hierárquica com 4 etapas (doravante designada apenas por hierárquica) e a metodologia não hierárquica de duas dimensões, baseada em matrizes, com e sem *master pool*. Todas as metodologias foram aplicadas (via simulação) em populações de 100000 indivíduos, com o objetivo de avaliar a eficiência, acurácia, sensibilidade e especificidade das metodologias sob diferentes condições de sensibilidade, especificidade e taxa de prevalência. Os *scripts* da linguagem R utilizados para a concretização das simulações estão apresentados no Anexo 37 para o método tradicional, no Anexo 38 para a metodologia hierárquica com 4 etapas e no Anexo 39 para a metodologia não hierárquica (com e sem *master pool*).

### 4.1. Configuração das simulações

Nas metodologias hierárquica e de Dorfman, foram utilizadas taxas de prevalência ( $p$ ) de 0.005, 0.01 e 0.025. Para cada taxa de prevalência, estudaram-se vários níveis de sensibilidade (individual e composta, que, conforme explicado, serão consideradas com valor igual) iguais à especificidade (individual e composta), com valores de 0.9, 0.95, 0.99 e 1. Além disso, foram analisados quatro grupos onde a sensibilidade é diferente da especificidade: (0.90 e 1), (1 e 0.90), (0.95 e 1) e (1 e 0.95). Nas combinações desta metodologia, foram também testadas várias dimensões de grupo, de forma a discernir quanto à dimensão ideal para cada taxa de prevalência, com o intuito de reduzir custos, tempo e recursos nas testagens.

Recordamos que a metodologia hierárquica é recomendada para taxas de prevalência muito baixas, pelo que unicamente se considerou valores inferiores ou iguais a 2.5%. Nesta metodologia, foram testadas três dimensões de grupo diferentes na primeira etapa: 64, 81 e 125. Em cada grupo inicial, foi implementada uma divisão em quatro etapas:

- No grupo de dimensão 64, a primeira etapa consistia num único teste ao grupo de 64 indivíduos. Caso este teste inicial fosse positivo, na segunda etapa, os indivíduos seriam divididos em 4 subgrupos de dimensão 16. Os 4 grupos seriam então testados,

e em caso de resultado positivo, o grupo em causa seria novamente subdividido em 4 grupos de dimensão 4 (terceira etapa). Finalmente, se algum desses subgrupos de 4 testasse positivo, seriam realizados testes individuais em cada membro desse subgrupo.

- No grupo de dimensão 81, o primeiro estágio consistia num subgrupo de 81 indivíduos. Caso este subgrupo testasse positivo, na segunda etapa, seria dividido em 3 grupos de 27 indivíduos. Se algum desses subgrupos de 27 testasse positivo, na terceira etapa, seria subdividido em 3 subgrupos de 9 indivíduos. Finalmente, se algum desses subgrupos de 9 testasse positivo, seriam realizados testes individuais em cada membro desse subgrupo.
- No grupo de dimensão 125, a primeira etapa consistia num subgrupo de 125 indivíduos. Caso este subgrupo testasse positivo, na segunda etapa, seria dividido em 5 subgrupos de 25 indivíduos. Se algum desses subgrupos de 25 testasse positivo, na terceira etapa, seria subdividido em 5 subgrupos de 5 indivíduos. Finalmente, se algum desses subgrupos de 5 testasse positivo, seriam realizados testes individuais em cada membro desse subgrupo.

Cada cenário de teste foi repetido 100 vezes em populações simuladas de 100000 indivíduos, para assegurar a robustez dos resultados, avaliando a variabilidade dos resultados obtidos. Os gráficos para a obtenção dos resultados apresentados estão disponíveis nos Anexos 1 a 12 (metodologia hierárquica e de Dorfman quando  $SI = EI$ , isto é, quando a sensibilidade e especificidade são iguais), Anexos 13 a 24 (metodologia hierárquica e de Dorfman quando  $SI$  e  $EI$  são distintas) e Anexos 25 a 36 (metodologia não hierárquica baseada em matrizes com e sem *master pool* quando  $SI = EI$ ).

O objetivo desta comparação é analisar as diferentes metodologias com base nos seguintes critérios: média total de testes (que é proporcional ao custo relativo), eficiência (inverso ao custo relativo), acurácia, sensibilidade e especificidade da metodologia em avaliação. Para tal, inicialmente são comparadas as metodologias de Dorfman e a hierárquica, primeiro considerando sensibilidade e especificidade assumindo iguais valores (Secção 4.2) e depois considerando valores distintos (Secção 4.3). Por fim, são analisados os resultados associados à metodologia não hierárquica (Secção 4.5). Na discussão dos resultados (Secção 5) é feita uma comparação entre todas as metodologias de classificação analisadas.

## 4.2. Sensibilidade igual à especificidade

Nesta secção iremos analisar os impactos de quando o valor de SI é igual ao valor da EI nos resultados obtidos, considerando como esta simetria influencia os parâmetros críticos da metodologia de classificação.

### 4.2.1. Média total de testes e eficiência

Os gráficos apresentados no Anexo 1 comparam a média do total de testes em diferentes condições de sensibilidade e especificidade para um valor de  $p = 0.025$  (Figura 10 no Anexo 1). Utilizam duas metodologias distintas de testes compostos: a hierárquica, no primeiro gráfico, e a metodologia de Dorfman, no segundo gráfico.

Conforme se pode verificar na Tabela 2, na metodologia hierárquica, a média do total de testes aumenta conforme a sensibilidade aumenta, especialmente em grupos iniciais maiores, apesar da diferença não ser significativa em muitos casos. Para todas as dimensões de grupo iniciais (64, 81 e 125), a média do número total de testes realizados para classificar toda a população é maior quando a sensibilidade é unitária, e menor quando a sensibilidade é igual a 0.9 (valor mais baixo analisado). A diferença entre os diferentes valores da sensibilidade torna-se mais evidente com o aumento da dimensão do grupo inicial, sendo o impacto mais significativo para a dimensão do grupo inicial de 81.

Deste modo, a eficiência, diminui ligeiramente à medida que  $SI = EI$  aumenta. Para uma taxa de prevalência  $p = 0.025$ , conforme apresentado na Tabela 2, a eficiência é maior quando  $SI = EI = 0.9$  (76.55%) e diminui gradualmente até 75.49% quando  $SI = EI = 1$ .

| Prevalência ( $p$ ) | SI = EI | Média de testes | Eficiência (%) | Dimensão de grupo ( $n$ ) |
|---------------------|---------|-----------------|----------------|---------------------------|
| 0.025               | 0.90    | 23445           | 76.55%         | 64,16,4,1                 |
| 0.025               | 0.95    | 23829           | 76.17%         | 64,16,4,1                 |
| 0.025               | 0.99    | 24379           | 75.62%         | 64,16,4,1                 |
| 0.025               | 1       | 24507           | 75.49%         | 64,16,4,1                 |

**Tabela 2 – Dimensão de grupo com melhor média de testes e eficiência usando a metodologia hierárquica numa taxa de prevalência de 2.5% quando  $SI = EI$**

Na metodologia de Dorfman, a média do total de testes (Figura 11 no Anexo 1) diminui de forma menos pronunciada até ao grupo de dimensão 8. A variação entre as diferentes sensibilidades é notável, especialmente para dimensões de grupo pequenas, e estabiliza a

partir do grupo de dimensão 4. Para as dimensões de grupo de 11 a 20, a média do total de testes aumenta gradualmente com o aumento da dimensão do grupo. Para todas as dimensões de grupo, a sensibilidade (e especificidade) de  $SI = EI = 0.9$  resulta numa média de testes mais alta comparada às outras sensibilidades.

Comparando as duas metodologias, a metodologia hierárquica é mais eficiente em termos do número de testes necessários. Os resultados obtidos na aplicação da metodologia de Dorfman para o grupo com a menor média de testes, para a taxa de prevalência  $p = 0.025$ , estão apresentados na Tabela 3.

| Prevalência ( $p$ ) | SI = EI | Média de testes | Eficiência (%) | Dimensão de grupo ( $n$ ) |
|---------------------|---------|-----------------|----------------|---------------------------|
| 0.025               | 0.90    | 37140           | 62.86%         | 8                         |
| 0.025               | 0.95    | 33923           | 65.13%         | 7                         |
| 0.025               | 0.99    | 31207           | 68.79%         | 7                         |
| 0.025               | 1       | 30575           | 69.43%         | 7                         |

**Tabela 3 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia de Dorfman numa taxa de prevalência de 2.5% quando  $SI = EI$**

Estes dados indicam que, para uma taxa de prevalência  $p = 0.025$ , o grupo de dimensão 7 é consistentemente eficiente para sensibilidades de 0.95, 0.99 e 1, com a menor média de testes ocorrendo com  $SI = EI = 1$ . Além disso, ao analisar a eficiência, verifica-se que esta aumenta conforme a  $SI = EI$  cresce. Para esta taxa de prevalência, a eficiência é menor quando a  $SI = EI = 0.9$  (62.86%) e aumenta progressivamente até 69.43% para  $SI = EI = 1$ , conforme apresentado na Tabela 3.

Concluindo, a metodologia hierárquica requer menos testes que a metodologia de Dorfman nesta taxa de prevalência. A metodologia de Dorfman é mais eficiente para grupos menores e médios, com um aumento significativo na média do total de testes à medida que a dimensão do grupo aumenta. O aumento da sensibilidade (e especificidade, que assume igual valor) resulta num aumento na média do total de testes na metodologia hierárquica, enquanto na metodologia de Dorfman ocorre o inverso.

Observando a Figura 12 (Anexo 2), é possível verificar que para uma taxa de prevalência de  $p = 0.01$  na metodologia hierárquica, quando a amostra é dividida inicialmente, em grupos de 64 e 125 indivíduos, há uma tendência para a média total de testes diminuir quando  $SI = EI$  aumentam. O mesmo não acontece quando a amostra é dividida inicialmente em grupos

de 81 indivíduos, visto que a média total de testes aumenta com o aumento de  $SI = EI$ . Nesta taxa de prevalência, na metodologia hierárquica, a melhor dimensão de grupo inicial é 64.

Em relação à aplicação da metodologia hierárquica podemos observar a dimensão de grupo com menos média total de testes, na Tabela 4.

| Prevalência ( $p$ ) | $SI = EI$ | Média de testes | Eficiência (%) | Dimensão de grupo ( $n$ ) |
|---------------------|-----------|-----------------|----------------|---------------------------|
| 0.01                | 0.90      | 12562           | 87.44%         | 64,16,4,1                 |
| 0.01                | 0.95      | 12273           | 87.73%         | 64,16,4,1                 |
| 0.01                | 0.99      | 12201           | 87.80%         | 64,16,4,1                 |
| 0.01                | 1         | 12182           | 87.82%         | 64,16,4,1                 |

**Tabela 4 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia hierárquica numa taxa de prevalência de 1% quando  $SI = EI$**

Assim sendo, a análise da eficiência, confirma que, para um grupo de 64 indivíduos, a eficiência melhora ligeiramente à medida que  $SI = EI$  aumentam, passando de 87.44% para  $SI = EI = 0.9$  até 87.82% para  $SI = EI = 1$  (note-se que a diferença é pequena).

Já na metodologia tradicional (Figura 13 no Anexo 2), podemos observar que, tal como geralmente acontece na metodologia hierárquica, a média do total de testes diminui, para  $n = 12$ , quando  $SI = EI$  aumentam. Nos valores da Tabela 5 seguinte podemos observar o  $n$  ótimo que minimiza a média total de testes para cada valor de  $SI$ .

| Prevalência ( $p$ ) | $SI = EI$ | Média de testes | Eficiência (%) | Dimensão de grupo ( $n$ ) |
|---------------------|-----------|-----------------|----------------|---------------------------|
| 0.01                | 0.90      | 27426           | 72.57%         | 12                        |
| 0.01                | 0.95      | 23541           | 76.46%         | 11                        |
| 0.01                | 0.99      | 20331           | 79.59%         | 10                        |
| 0.01                | 1         | 19525           | 80.47%         | 11                        |

**Tabela 5 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia de Dorfman numa taxa de prevalência de 1% quando  $SI = EI$**

Relativamente à eficiência, observa-se que aumenta de forma consistente com o aumento de  $SI = EI$ . Sendo que sobe de 72.57% para  $SI = EI = 0.9$  até 80.47% para  $SI = EI = 1$ , como se pode observar na Tabela 5.

Assim conclui-se que com uma taxa de prevalência de 0.01 a metodologia hierárquica obtém uma média de testes consideravelmente baixa relativamente à metodologia tradicional, o que

faz com que a metodologia hierárquica seja melhor na poupança de recursos e tempo utilizado nas testagens.

De acordo com a Figura 14 (Anexo 3) referentes a uma taxa de prevalência de 0.005, pode-se observar que na metodologia hierárquica quando o grupo inicial é 64 ou 125 a média total de testes diminui quando  $SI = EI$  aumentam, quando o grupo inicial é de 81 indivíduos a média total de testes tende a aumentar quando  $SI = EI$  aumentam.

É possível concluir ainda que em todos os níveis de  $SI = EI$ , o grupo inicial com menor número médio de testes é utilizando grupos de 64 indivíduos, com as médias de testes para cada valor de  $SI$  apresentados na Tabela 6:

| Prevalência ( $p$ ) | $SI = EI$ | Média de testes | Eficiência (%) | Dimensão de grupo ( $n$ ) |
|---------------------|-----------|-----------------|----------------|---------------------------|
| 0.005               | 0.90      | 7866            | 92.13%         | 64,16,4,1                 |
| 0.005               | 0.95      | 7435            | 92.57%         | 64,16,4,1                 |
| 0.005               | 0.99      | 7205            | 92.79%         | 64,16,4,1                 |
| 0.005               | 1         | 7166            | 92.83%         | 64,16,4,1                 |

**Tabela 6 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia hierárquica numa taxa de prevalência de 0.5% quando  $SI = EI$**

Ao analisar os diferentes valores de  $SI = EI$ , conclui-se que o grupo inicial de 64 indivíduos é consistentemente o que apresenta a menor média total de testes. A Tabela 6 mostra que, para esse grupo, a eficiência apresenta uma leve melhoria, aumentando de 92.13% para 92.83% à medida que  $SI = EI$  aumentam.

Na metodologia tradicional (Figura 15 no Anexo 3), observa-se uma tendência geral de diminuição na média do total de testes à medida que a dimensão do grupo aumenta. Esta tendência é mais acentuada nos grupos menores e estabiliza nos grupos maiores, de 11 a 20. As sensibilidades e especificidades de 0.99 e 1 apresentam consistentemente as menores médias de testes. Em contraste,  $SI = EI = 0.95$  e  $0.9$  tendem a ter médias de testes mais altas.

Nos grupos de dimensão 2 a 10, a redução na média do total de testes é significativa, especialmente até ao grupo de dimensão 6. A partir do grupo de dimensão 7, a redução continua, mas de forma menos acentuada. Nos grupos de dimensão 11 a 20, a média do total de testes estabiliza, mostrando menos variação entre as diferentes dimensões de grupo. A tendência de menor média total de testes para  $SI = EI = 1$  persiste, mesmo nas dimensões de

grupo maiores. Podemos observar as dimensões de grupos com menos total de testes na Tabela 7:

| Prevalência ( $p$ ) | SI = EI | Média de Testes | Eficiência (%) | Dimensão de Grupo ( $n$ ) |
|---------------------|---------|-----------------|----------------|---------------------------|
| 0.005               | 0.90    | 22429           | 77.57%         | 17                        |
| 0.005               | 0.95    | 18138           | 82.62%         | 16                        |
| 0.005               | 0.99    | 14750           | 85.25%         | 14                        |
| 0.005               | 1       | 13898           | 86.10%         | 15                        |

**Tabela 7 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia tradicional de Dorfman numa taxa de prevalência de 0.5% quando SI = EI**

Deste modo, conforme apresentado na Tabela 7, verifica-se que a eficiência aumenta à medida que SI = EI crescem. Para uma taxa de prevalência de  $p = 0.005$ , a eficiência varia de 77.57% para SI = EI = 0.90 até 86.10% para SI = EI = 1. A tabela também destaca que a dimensão de grupo mais eficiente, variando entre 14 e 17 indivíduos, depende dos valores de SI = EI, sendo o grupo de 15 indivíduos o mais eficiente com SI = EI = 1.

Podemos concluir que em todas as taxas de prevalências testadas a metodologia hierárquica obteve os melhores resultados uma vez que obteve sempre uma média total de testes menor que a observada na metodologia tradicional. Conclui-se ainda que apesar de terem sido testados 4 valores de SI = EI diferentes em 4 taxas de prevalência diferentes, na metodologia hierárquica o melhor grupo foi sempre o de 64 indivíduos. Já na metodologia tradicional é possível concluir que à medida que a taxa de prevalência e SI = EI aumentam, a dimensão do grupo ideal, com menor média total de testes, diminui.

#### 4.2.2. Acurácia da metodologia

A acurácia da metodologia refere-se à capacidade do teste de classificar corretamente os indivíduos como doentes ou saudáveis (conforme definido na Secção 3.4.1). Assim, é a proporção de resultados verdadeiros (tanto verdadeiros positivos como verdadeiros negativos) no total de indivíduos testados.

Ao longo dos gráficos desta secção, a escala em Y (eixo das ordenadas) é alterada livremente para sublinhar as diferenças existentes entre os diversos métodos utilizados, mesmo que na realidade estas sejam pouco significativas.

Para uma taxa de prevalência de  $p = 0.025$  (Figura 16 no Anexo 4), na metodologia hierárquica, é possível observar que, para  $SI = EI = 0.9$ , a acurácia da metodologia média é ligeiramente inferior, mas ainda muito alta, próxima de 0.98 para todas as dimensões de grupo inicial (64, 81 e 125). Não há diferença significativa entre as dimensões de grupo inicial, com todos a apresentarem valores de acurácia muito próximos. Para  $SI = EI = 0.95$ , a acurácia média aumenta ligeiramente, ficando assim, próxima de 0.99 para todas as dimensões de grupo inicial. Novamente, não há diferença significativa entre a acurácia as dimensões de grupo inicial, todos com valores muito próximos.

Com  $SI = EI = 0.99$ , a acurácia da metodologia continua muito alta, próxima de 1 para todas as dimensões de grupo analisadas. Todas as dimensões de grupo apresentam valores de acurácia muito próximos, o que indica uma alta acurácia. Finalmente, para  $SI = EI = 1$ , a acurácia média atinge o seu valor máximo, 1, para todos os dimensões de grupo inicial. Todas as dimensões de grupo inicial (64, 81 e 125) apresentam valores de acurácia igual a 1, indicando uma acurácia perfeita. De facto, não há erros de classificação em qualquer das metodologias aplicadas neste trabalho caso a sensibilidade e a especificidade dos testes individuais e composto não tenham erros de classificação. A forma como estas metodologias foram arquitetadas garante esta propriedade.

Para a taxa de prevalência  $p = 0.025$ , a acurácia da metodologia é extremamente alta para todas as combinações de sensibilidade e especificidade, independentemente da dimensão do grupo inicial. A acurácia aumenta ligeiramente com a sensibilidade e especificidade, atingindo o valor máximo de 1 quando  $SI = EI = 1$ , como seria esperado. Além disso, não há diferença significativa na acurácia da metodologia entre as diferentes dimensões de grupo inicial, sugerindo que todos são igualmente eficazes em termos de acurácia da metodologia para esta taxa de prevalência. Ainda assim a dimensão de grupo inicial 81 continua a ser inferior em todas as configurações de  $SI = EI$  menores que 1.

No que diz respeito à metodologia de Dorfman (Figura 17 no Anexo 4), observa-se que, conforme a dimensão do grupo aumenta, a acurácia da metodologia tende a diminuir ligeiramente para valores de  $SI = EI = 0.9$  e  $0.95$ , enquanto para valores de  $SI = EI = 0.99$  e  $1$ , a acurácia da metodologia se mantém estável e próxima de 1. O que indica que a acurácia da metodologia tem uma relação inversa com o aumento da dimensão de grupo.

Comparando as duas metodologias, na metodologia hierárquica, a acurácia é extremamente alta para todas as dimensões de grupo inicial (64, 81 e 125) e combinações de sensibilidade

e especificidade. Não há diferença significativa entre as diferentes dimensões de grupo inicial, sugerindo que todos são igualmente eficazes em termos de acurácia da metodologia. Na metodologia tradicional, a acurácia da metodologia também é muito alta para todas as dimensões de grupo e combinações de sensibilidade e especificidade. Todas as dimensões de grupo apresentam acurácia inferior para a  $SI = EI = 0.9$ , mas para  $SI = EI = 0.95, 0.99$  e  $1$ , a acurácia é muito alta, próxima de  $1$ .

Portanto, ambas as metodologias, hierárquica e tradicional, apresentam uma acurácia extremamente alta para a taxa de prevalência  $p = 0.025$ . No entanto, a metodologia hierárquica mostra uma consistência ligeiramente melhor na acurácia entre as diferentes dimensões de grupo inicial.

Passando para uma taxa de prevalência  $p = 0.01$  (Figura 18 no Anexo 5), relativamente à metodologia hierárquica, observa-se que a acurácia da metodologia é extremamente alta para todas as combinações de sensibilidade e especificidade, independentemente da dimensão do grupo inicial (64, 81 e 125). Para  $SI = EI = 0.9$ , a acurácia é ligeiramente inferior, mas ainda muito alta, próxima de  $0.98$ . À medida que a sensibilidade e especificidade aumentam, a acurácia também aumenta, atingindo o valor máximo de  $1$  para  $SI = EI = 1$ . Não há diferença significativa na acurácia da metodologia entre as diferentes dimensões de grupo inicial, sugerindo que todos são, tal como na taxa de prevalência  $p = 0.025$ , igualmente eficazes em termos de acurácia para esta taxa de prevalência.

Na metodologia tradicional (Figura 19 no Anexo 5), observa-se uma tendência semelhante. A acurácia da metodologia é muito alta para todos os dimensões de grupo e  $SI = EI$ . Para grupos menores ( $n = 2$  e  $3$ ) e para a  $SI = EI = 0.9$ , a acurácia é superior, mas diminui à medida que a dimensão de grupo aumenta, ficando abaixo de  $0.975$  para grupos de dimensão  $20$ . À medida que  $SI = EI$  aumenta, a acurácia também aumenta, chegando muito próxima de  $1$  para  $SI = EI = 0.99$ . Para grupos de dimensão  $11$  a  $20$ , a acurácia mantém-se alta para todas as dimensões de grupo e sensibilidades, com valores consistentemente altos, especialmente para sensibilidades de  $0.95, 0.99$  e  $1$ .

Comparando as duas metodologias, na metodologia hierárquica, a acurácia da metodologia é extremamente alta para todas as dimensões de grupo inicial (64, 81 e 125) e combinações de sensibilidade e especificidade. Por outro lado, não há diferença significativa entre as diferentes dimensões de grupo inicial, sugerindo que todos são igualmente eficazes em termos de acurácia da metodologia. Na metodologia tradicional, a acurácia também é muito

alta para todos os dimensões de grupo e combinações de sensibilidade e especificidade. Grupos maiores apresentam acurácia ligeiramente inferior para  $SI = EI = 0.9$ , mas para  $SI = EI = 0.95, 0.99$  e  $1$ , a acurácia é muito alta, próxima de  $1$ .

Portanto, ambas as metodologias, hierárquica e tradicional, apresentam uma acurácia extremamente alta para a taxa de prevalência  $p = 0.01$ . No entanto, a metodologia hierárquica mostra, mais uma vez, uma consistência ligeiramente melhor na acurácia entre as diferentes dimensões de grupo inicial.

Quanto a taxa de prevalência  $p = 0.005$  (Figura 20 no Anexo 6), observamos, na metodologia hierárquica, que a acurácia da metodologia é extremamente alta para todas as combinações de sensibilidade e especificidade, independentemente da dimensão do grupo inicial (64, 81 e 125). Para  $SI = EI = 0.9$ , a acurácia é ligeiramente inferior, mas ainda muito alta, próxima de  $0.99$ . À medida que  $SI = EI$  aumenta, a acurácia também aumenta, atingindo o valor máximo de  $1$  para  $SI = EI = 1$ . Não há diferença significativa na acurácia entre as diferentes dimensões de grupo inicial, o que sugere que todos são igualmente eficazes em termos de acurácia para esta taxa de prevalência.

Relativamente à metodologia tradicional (Figura 21 no Anexo 6), observa-se uma tendência semelhante, apesar de não haver aumentos ou diminuição significativos em comparação com os resultados da taxa de prevalência  $p = 0.01$ .

Visto isto, na metodologia hierárquica, a acurácia da metodologia é extremamente alta para todas as dimensões de grupo inicial (64, 81 e 125) e combinações de sensibilidade e especificidade. Não há diferença significativa entre as diferentes dimensões de grupo inicial, sugerindo que todos são igualmente eficazes em termos de acurácia. Na metodologia tradicional, a acurácia também é muito alta para todos os dimensões de grupo e combinações de sensibilidade e especificidade. Grupos maiores apresentam acurácia ligeiramente inferior para  $SI = EI = 0.9$ , mas para  $SI = EI = 0.95, 0.99$  e  $1$ , a acurácia é muito alta, próxima de  $1$ .

Portanto, ambas as metodologias, hierárquica e tradicional, apresentam uma acurácia extremamente alta para a taxa de prevalência  $p = 0.005$ . No entanto, a metodologia hierárquica mostra uma consistência ligeiramente melhor na acurácia entre as diferentes dimensões de grupo inicial.

Em suma, após análise de ambas as metodologias, hierárquica e tradicional, em todas as taxas de prevalência, é possível concluir mais uma vez que a metodologia hierárquica

oferece melhores resultados em comparação com a metodologia tradicional. Além disso, é de reforçar que, é possível observar a melhoria substancial dos resultados obtidos na metodologia hierárquica à medida que a taxa de prevalência baixava, provando que esta metodologia é mais apropriada ao estudar amostras com baixa taxa de prevalência de doença. Contudo, quando a taxa de prevalência diminui, a sensibilidade perde peso na determinação da acurácia, pelo que bons resultados na acurácia não implicam necessariamente valores elevados nas sensibilidades nestes cenários.

### 4.2.3. Sensibilidade da metodologia

A sensibilidade de uma metodologia é a capacidade da metodologia identificar corretamente os indivíduos que possuem a doença em questão (conforme definido na Secção 3.4.1). Em termos estatísticos, é a proporção de verdadeiros positivos entre os indivíduos que realmente possuem a doença.

Seguindo para análise da sensibilidade da metodologia numa taxa de prevalência  $p = 0.025$  (Figura 22 no Anexo 7),  $p = 0.01$  (Figura 24 no Anexo 8),  $p = 0.005$  (Figura 26 no Anexo 9), na metodologia hierárquica, observa-se para  $SI = EI = 0.9$ , a sensibilidade da metodologia é bastante baixa, ficando próxima de 0.65. Uma vez que a metodologia hierárquica é dividida em 4 etapas, a sensibilidade irá diminuir a cada etapa. Para calcular a sensibilidade desta metodologia quando  $SI = EI = 0.9$ , deve ter-se em conta que  $SI = EI$  de cada etapa equivale a 0.9, neste caso, da  $SI$  e  $EI$  da etapa anterior, ou seja, a  $SI = EI$  da segunda etapa será 90% da  $SI = EI$  da primeira e assim consequentemente para as seguintes etapas, fazendo com que, quantas mais etapas houver menor a sensibilidade da metodologia ( $SM$ ):

$$SM = SI_{e1} \times SI_{e2} \times SI_{e3} \times SI_{e4}$$

$$SM = 0.9 \times 0.9 \times 0.9 \times 0.9 = 0.6561$$

Isto ocorre uma vez que, para um indivíduo ser classificado como doente, terá de obter testes positivos em todas as etapas. À medida que  $SI = EI$  aumenta, a sensibilidade da metodologia também aumenta significativamente, chegando a ultrapassar 0.80 para  $SI = EI = 0.95$  e 0.95 para  $SI = EI = 0.99$ , e chegando a 1 para  $SI = EI = 1$ . Não há diferença significativa na sensibilidade da metodologia entre as diferentes dimensões de grupo inicial, sugerindo que todos são igualmente eficazes em termos de sensibilidade para esta taxa de prevalência. Deste modo, a sensibilidade de uma metodologia hierárquica parece unicamente depender de  $SI = EI$  e do número de etapas.

Já na metodologia tradicional numa taxa de prevalência  $p = 0.025$  (Figura 23 no Anexo 7),  $p = 0.01$  (Figura 25 no Anexo 8),  $p = 0.005$  (Figura 27 no Anexo 9), observa-se a mesma tendência. Apesar de que para  $SI = EI = 0.9$ , a sensibilidade está próxima de 0.83 e para  $SI = EI = 0.95$  ultrapassa 0.9. Relativamente a  $SI = EI = 0.99$  e 1, permanecem os resultados observados na metodologia hierárquica.

Conforme referido anteriormente, quanto mais etapas, menor a sensibilidade, uma vez que a metodologia tradicional, apenas possui duas etapas, terá uma sensibilidade melhor que a hierárquica de 4 etapas. Podemos confirmar através dos seguintes cálculos para  $SI = EI = 0.9$ :

$$SM = SI_{e1} \times SI_{e2}$$

$$SM = 0.9 \times 0.9 = 0.81$$

Visto isto, é possível concluir que, quando  $SI = EI$  é mais baixa, a metodologia tradicional, tem melhor capacidade para identificar corretamente os indivíduos que possuem a doença, enquanto quando a  $SI = EI$  é mais alta, ambas as metodologias possuem excelente capacidade de identificação dos indivíduos doentes (apesar de se manter a desigualdade, esta deixa de ser tão significativa). Podemos também concluir que a metodologia hierárquica ganha eficiência, mas perde sensibilidade.

Será importante referir uma vez mais que, uma vez que este estudo se concentra em testes qualitativos, os resultados sobre sensibilidade e especificidade associados aos testes compostos e às metodologias, não serão afetados pela dimensão do grupo, o que seria provável de ocorrer caso se estivesse a lidar com testes quantitativos.

#### 4.2.4. Especificidade da metodologia

A especificidade da metodologia refere-se à capacidade de a metodologia identificar corretamente os indivíduos que não têm a doença (conforme definido na Secção 3.4.1.). Ou seja, é a proporção de verdadeiros negativos entre aqueles que realmente não têm a doença. Alta especificidade significa que o teste é eficaz em minimizar falsos positivos, identificando corretamente a ausência da doença na maioria dos casos em que ela não está presente.

Seguindo para análise da especificidade da metodologia numa taxa de prevalência  $p = 0.025$  (Figura 28 no Anexo 10),  $p = 0.01$  (Figura 30 no Anexo 11),  $p = 0.005$  (Figura 32 no Anexo 12), na metodologia hierárquica, observa-se para  $SI = EI = 0.9$ , a especificidade é a mais

baixa de todos valores de  $SI = EI$  observados, apesar de ter um valor muito próximo de 1. À medida que  $SI = EI$  aumenta, a especificidade também aumenta significativamente, aproximando-se de 1 para  $SI = EI = 0.99$  e sendo igual a 1 quando  $SI = EI = 1$ . É possível observar que a dimensão de grupo  $n = 81$  é constantemente a pior.

Já na metodologia tradicional, numa taxa de prevalência  $p = 0.025$  (Figura 29 no Anexo 10),  $p = 0.01$  (Figura 31 no Anexo 11),  $p = 0.005$  (Figura 33 no Anexo 12), observa-se a mesma tendência. Apesar de que para  $SI = EI = 0.9$ , a especificidade está próxima de 0.99 e para  $SI = EI = 0.95$  estar muito próxima de 1. Relativamente a  $SI = EI = 0.99$  e 1, permanecem os resultados observados na metodologia hierárquica.

Visto isto, é possível concluir que, quando  $SI = EI$  é mais baixa, a metodologia hierárquica, tem melhor capacidade para identificar corretamente os indivíduos que não possuem a doença, enquanto quando a  $SI = EI$  é mais alta, ambas as metodologias possuem boa capacidade de identificação dos indivíduos não doentes (a desigualdade mantém-se, com a metodologia hierárquica a apresentar maiores valores de especificidade do modelo, mas a diferença é cada vez menos significativa).

Notemos que os resultados parecessem confirmar o esperado, que quanto maior for o número de etapas da metodologia hierárquica menor será a sensibilidade (um indivíduo tem teste positivo se tiver teste positivo em todas as etapas, pelo que esta probabilidade diminui com o aumento do número de etapas). Por outro lado, quanto maior for o número de etapas da metodologia hierárquica maior será o valor da especificidade (um indivíduo tem teste negativo se obtiver um teste negativo em qualquer uma das etapas, pelo que esta probabilidade aumenta com o aumento do número de etapas).

### **4.3.Sensibilidade e especificidade com valores distintos**

#### **4.3.1. Média total de testes**

Na metodologia hierárquica, observa-se que a média do total de testes varia conforme a sensibilidade e a especificidade mudam. Para todas as dimensões de grupo iniciais (64, 81 e 125), a média do total de testes é maior quando a sensibilidade é de 1 e a especificidade é de 0.9, e menor quando a sensibilidade é de 0.9 e a especificidade é de 1. Note-se que esta relação já seria esperada, pois a poupança de teste decorre da ocorrência de testes negativos nos quais todo o grupo é classificado como saudável. Deste modo, quanto maior for a

probabilidade de um teste ser positivo (e menor a de um teste ser negativo) maior será o número esperado de testes a realizar. Naturalmente, a probabilidade de um teste ser positivo aumenta com a sensibilidade e diminui com a especificidade.

É possível observar que, na metodologia hierárquica, numa taxa de prevalência de  $p = 0.025$  (Figura 34 no Anexo 13), a dimensão de grupo inicial de 81 obtém uma maior média total de testes e a dimensão de grupo inicial de 64 é constantemente o mais eficiente. Os resultados obtidos estão apresentados na Tabela 8.

| Prevalência ( $p$ ) | SI   | EI   | Média de testes | Eficiência (%) | Dimensão do grupo ( $n$ ) |
|---------------------|------|------|-----------------|----------------|---------------------------|
| 0.025               | 0.90 | 1    | 19799           | 80.20%         | 64,16,4,1                 |
| 0.025               | 0.95 | 1    | 22056           | 77.94%         | 64,16,4,1                 |
| 0.025               | 1    | 0.90 | 28638           | 71.36%         | 64,16,4,1                 |
| 0.025               | 1    | 0.95 | 26469           | 73.53%         | 64,16,4,1                 |

**Tabela 8 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia hierárquica numa taxa de prevalência de 2.5% quando SI e EI são distintas**

Observa-se que a eficiência é mais alta quando a  $SI = 0.90$  e  $EI = 1$ , alcançando 80.20%, o que indica uma utilização mais eficaz dos recursos disponíveis. À medida que  $SI$  aumenta para 1 e  $EI$  diminui para 0.90, a eficiência cai para 71.36%, o que é consistente com o aumento no número total de testes realizados.

Na metodologia tradicional (Figura 35 no Anexo 13), a média do total de testes diminui até ao grupo de dimensão  $n = 7$ . Para as dimensões de grupo de  $n = 8$  a 20, a média do total de testes aumenta gradualmente com o aumento da dimensão do grupo. Para todas as dimensões de grupo, a combinação  $SI = 0.9$ ,  $EI = 1$  resulta numa média de testes mais baixa comparada às outras combinações. Os resultados estão apresentados na Tabela 9.

| Prevalência ( $p$ ) | SI   | EI   | Média de testes | Eficiência (%) | Dimensão do grupo ( $n$ ) |
|---------------------|------|------|-----------------|----------------|---------------------------|
| 0.025               | 0.90 | 1    | 28907           | 71.09%         | 7                         |
| 0.025               | 0.95 | 1    | 29678           | 70.32%         | 7                         |
| 0.025               | 1    | 0.90 | 38943           | 61.06%         | 7                         |
| 0.025               | 1    | 0.95 | 34708           | 65.29%         | 7                         |

**Tabela 9 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia de Dorfman numa taxa de prevalência de 2.5% quando SI e EI são distintas**

Estes dados indicam que, para  $p = 0.025$ , o grupo de dimensão  $n = 7$  é consistentemente mais eficiente para as combinações de sensibilidade e especificidade analisadas, com a menor média de testes ocorrendo na combinação  $SI = 0.9$  e  $EI = 1$ , tendo uma eficiência de 71.09%.

Concluindo, a metodologia hierárquica requer menos testes em todas as combinações de  $SI$  e  $EI$  abordadas, quando comparada com a metodologia de Dorfman. A metodologia tradicional é mais eficiente para grupos pequenos e médios, com um aumento significativo na média do total de testes à medida que a dimensão do grupo aumenta. O aumento da sensibilidade resulta num aumento na média do total de testes em ambas as metodologias.

Para uma taxa de prevalência  $p = 0.01$  na metodologia hierárquica (Figura 36 no Anexo 14), observam-se os mesmos resultados em termos de dimensão de grupo inicial, porém, verifica-se que a média total de testes necessária para testar a população diminui consideravelmente em comparação com a média total de testes necessária de  $p = 0.025$ , na metodologia hierárquica. Para todas as dimensões de grupo, a média do total de testes é maior quando  $SI = 1$  e  $EI = 0.9$ , e menor quando a  $SI = 0.9$  e  $EI = 1$ . Nesta taxa de prevalência, a dimensão de grupo inicial que necessita de menos testes continua a ser  $n = 64$ , como é possível observar na Tabela 10:

| Prevalência ( $p$ ) | SI   | EI   | Média de testes | Eficiência (%) | Dimensão do grupo ( $n$ ) |
|---------------------|------|------|-----------------|----------------|---------------------------|
| 0.01                | 0.90 | 1    | 10097           | 89.90%         | 64,16,4,1                 |
| 0.01                | 0.95 | 1    | 11107           | 88.89%         | 64,16,4,1                 |
| 0.01                | 1    | 0.90 | 14895           | 85.11%         | 64,16,4,1                 |
| 0.01                | 1    | 0.95 | 13407           | 86.59%         | 64,16,4,1                 |

**Tabela 10 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia hierárquica numa taxa de prevalência de 1% quando  $SI$  e  $EI$  são distintas**

Assim, em termos de eficiência, esta, é mais alta (89.90%) quando  $SI = 0.90$  e  $EI = 1$ . Esta combinação mostra-se a mais eficaz, maximizando a economia de recursos ao minimizar o número total de testes. Em contrapartida, quando a  $SI$  aumenta para 1 e  $EI$  diminui para 0.90, a eficiência cai para 85.11%.

Na metodologia tradicional (Figura 37 no Anexo 14), a média do total de testes diminui de forma acentuada até ao grupo de dimensão  $n = 12$ . Para as dimensões de grupo  $n = 13$  a 20, a média do total de testes aumenta gradualmente com o aumento da dimensão do grupo. Para todas as dimensões de grupo, a combinação  $SI = 0.9$  e  $EI = 1$  resulta numa média de testes

mais baixa comparada às outras combinações. Os resultados da melhor dimensão de grupo desta metodologia na taxa de prevalência  $p = 0.01$  são apresentados na Tabela 11.

| Prevalência ( $p$ ) | SI   | EI   | Média de testes | Eficiência (%) | Dimensão do grupo ( $n$ ) |
|---------------------|------|------|-----------------|----------------|---------------------------|
| 0.01                | 0.90 | 1    | 18488           | 81.51%         | 11                        |
| 0.01                | 0.95 | 1    | 19030           | 80.97%         | 11                        |
| 0.01                | 1    | 0.90 | 28477           | 71.52%         | 12                        |
| 0.01                | 1    | 0.95 | 23997           | 76.00%         | 11                        |

**Tabela 11 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia de Dorfman numa taxa de prevalência de 1% quando SI e EI são distintas**

Estes dados indicam que, para  $p = 0.01$ , o grupo de dimensão  $n = 11$  é consistentemente eficiente para as combinações de sensibilidade e especificidade analisadas, com a menor média de testes a surgir quando  $SI = 0.9$  e  $EI = 1$ . Observa-se que a combinação de  $SI = 0.90$  e  $EI = 1$  alcança a maior eficiência (81.51%), acompanhada pela menor média de testes necessários, especialmente para grupos com dimensão  $n = 11$ , que demonstram ser constantemente mais eficientes. Em contraste, quando  $SI$  atinge o valor máximo ( $SI = 1$ ) e  $EI$  diminui para 0.90, a eficiência cai para 71.52%, refletindo um aumento significativo no número de testes realizados.

Sendo assim, tal como observado na taxa de prevalência  $p = 0.025$ , a metodologia hierárquica continua a ser mais eficiente que a metodologia tradicional, com uma média total de testes bastante inferior à da metodologia tradicional.

Para uma taxa de prevalência  $p = 0.005$  (Figura 38 no Anexo 15) continuamos com os mesmos resultados na metodologia hierárquica, sendo  $n = 64$  a melhor dimensão de grupo inicial e uma descida na média total de testes, como se pode observar na Tabela 12:

| Prevalência ( $p$ ) | SI   | EI   | Média de testes | Eficiência (%) | Dimensão do grupo ( $n$ ) |
|---------------------|------|------|-----------------|----------------|---------------------------|
| 0.005               | 0.90 | 1    | 6095            | 93.91%         | 64,16,4,1                 |
| 0.005               | 0.95 | 1    | 6613            | 93.39%         | 64,16,4,1                 |
| 0.005               | 1    | 0.90 | 9148            | 90.85%         | 64,16,4,1                 |
| 0.005               | 1    | 0.95 | 8023            | 91.98%         | 64,16,4,1                 |

**Tabela 12 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia hierárquica numa taxa de prevalência de 0.5% quando SI e EI são distintas**

Observa-se que a combinação  $SI = 0.90$  e  $EI = 1$  resulta na maior eficiência (93.91%), indicando que essa configuração maximiza a economia de recursos ao necessitar do menor número de testes. Mesmo quando  $EI = 1$ , um aumento na sensibilidade para 0.95 causa uma ligeira diminuição na eficiência, com esta atingindo o seu mínimo em 90.85% quando  $SI = 1$  e  $EI = 0.90$ .

Na metodologia tradicional (Figura 39 no Anexo 15), observa-se uma estabilização na média total de testes até  $n = 16$ , e uma discrepância de valores entre as diferentes combinações de  $SI$  e  $EI$  para cada  $n$ . Sendo que nesta metodologia, de  $n = 17$  a  $n = 20$ , observa-se um aumento muito ligeiro do número médio de testes. Na Tabela 13, estão representadas as dimensões de grupo com a menor média de testes para  $p = 0.005$ .

| Prevalência ( $p$ ) | SI   | EI   | Média de testes | Eficiência (%) | Dimensão do grupo ( $n$ ) |
|---------------------|------|------|-----------------|----------------|---------------------------|
| 0.005               | 0.90 | 1    | 18488           | 81.51%         | 16                        |
| 0.005               | 0.95 | 1    | 19030           | 80.97%         | 15                        |
| 0.005               | 1    | 0.90 | 28477           | 71.52%         | 15                        |
| 0.005               | 1    | 0.95 | 23997           | 76.00%         | 16                        |

**Tabela 13 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia de Dorfman numa taxa de prevalência de 0.5% quando  $SI$  e  $EI$  são distintas**

Assim podemos concluir que para esta taxa de prevalência, as melhores dimensões de grupo, na metodologia tradicional, são  $n = 15$  e  $n = 16$ . Visto que a melhor média total de testes ocorre quando  $SI = 0.9$  e  $EI = 1$ ,  $n = 16$  é a melhor dimensão de grupo para esta taxa de prevalência. Como tal, a combinação  $SI = 0.90$  e  $EI = 1$  oferece a maior eficiência (81.51%), especialmente para grupos de dimensão  $n = 16$ . Quando  $SI$  aumenta para 1 e  $EI$  é mantida em 0.90, a eficiência diminui significativamente para 71.52%, indicando que essa combinação exige mais testes, tornando-se menos eficaz.

Ainda assim, a metodologia hierárquica, continua a possuir os melhores resultados, sendo que necessita de menos de metade dos testes que a metodologia tradicional.

#### 4.3.2. Acurácia da metodologia

Relativamente à acurácia da metodologia hierárquica, para a taxa de prevalência de  $p = 0.025$  (Figura 40 no Anexo 16), observa-se que, no grupo inicial de 64 indivíduos, a combinação  $SI = 0.95$  e  $EI = 1$  apresenta a melhor acurácia, enquanto a combinação  $SI = 1$  e  $EI = 0.9$

apresenta a pior. Deste modo, é notória a maior influência na acurácia do valor da especificidade em relação ao valor da sensibilidade, pois estamos a considerar unicamente taxas de prevalência diminutas. No grupo de 81 indivíduos, a acurácia da metodologia tende a diminuir em comparação ao grupo de 125 indivíduos, nas combinações  $SI = 1$  e  $EI = 0.9$  e  $SI = 1$  e  $EI = 0.95$ . O grupo de 81 indivíduos mostra a menor acurácia nas combinações  $SI = 1$  e  $EI = 0.9$  e  $SI = 1$  e  $EI = 0.95$ , sendo a combinação  $SI = 1$  e  $EI = 0.9$  a menor acurácia.

Embora o padrão de acurácia da metodologia se mantenha o mesmo para todas as restantes taxas de prevalência analisadas ( $p = 0.01$  (Figura 42 no Anexo 17) e  $p = 0.005$  (Figura 44 no Anexo 18)), observa-se que o valor da acurácia da metodologia diminui conforme a taxa de prevalência aumenta. Portanto, apesar dos resultados serem consistentes, a eficácia da acurácia tende a ser maior em taxas de prevalência mais baixas e diminui gradualmente à medida que a prevalência aumenta.

Este comportamento ocorre devido à relação entre sensibilidade e especificidade e a prevalência da doença. Quando a prevalência é baixa, um teste com alta especificidade ( $EI = 1$ ) reduz significativamente o número de falsos positivos, melhorando a acurácia. À medida que a prevalência aumenta, mesmo testes com alta sensibilidade ( $SI = 1$ ) começam a gerar mais falsos positivos, uma vez que o valor de  $EI$  é inferior a 1, o que compromete a acurácia geral. Além disso, uma especificidade menor ( $EI = 0.9$ ) em condições de alta prevalência resulta num aumento dos falsos positivos, reduzindo ainda mais a acurácia. Deste modo, quanto menor for a taxa de prevalência maior será o peso da especificidade (e menor da sensibilidade) na determinação da acurácia. Como os testes compostos são aplicados unicamente em problemas com taxas de prevalência diminuta, a especificidade terá neste contexto um peso sempre significativo na determinação da acurácia. Portanto, a combinação de alta especificidade e sensibilidade ligeiramente inferior ( $SI = 0.95$ ,  $EI = 1$ ) tende a ter maior acurácia, especialmente em prevalências mais baixas.

Passando para a metodologia tradicional, para a prevalência de  $p = 0.025$  (Figura 41 no Anexo 16), observa-se que a combinação  $SI = 0.95$  e  $EI = 1$  apresenta a melhor acurácia da metodologia em todas as dimensões de grupo, enquanto a combinação  $SI = 1$  e  $EI = 0.9$  apresenta a pior acurácia. A acurácia da metodologia mantém-se alta e constante para as combinações  $SI = 0.9$  e  $EI = 1$  e  $SI = 0.95$  e  $EI = 1$ , com leve diminuição conforme a dimensão do grupo aumenta. Já as combinações  $SI = 1$  e  $EI = 0.9$  e  $SI = 1$  e  $EI = 0.95$ , têm

as piores acurácias com a agravante destas diminuir com o aumento da dimensão de grupo.

Para as outras taxas de prevalência ( $p = 0.01$  (Figura 43 no Anexo 17) e  $p = 0.005$  (Figura 45 no Anexo 18)), observa-se o mesmo padrão geral. A combinação  $SI = 0.95$  e  $EI = 1$  continua a apresentar a melhor acurácia em todas as dimensões de grupo, enquanto  $SI = 1$  e  $EI = 0.9$  apresenta a pior acurácia. As combinações  $SI = 0.9$  e  $EI = 1$  e  $SI = 0.95$  e  $EI = 1$  mantêm uma acurácia alta e constante, com uma leve diminuição quando a dimensão do grupo aumenta. Por outro lado, as combinações  $SI = 1$  e  $EI = 0.9$  e  $SI = 1$  e  $EI = 0.95$ , têm sempre a pior acurácia e mostram uma tendência para diminuir à medida que a dimensão do grupo aumenta.

Além disso, observa-se que os níveis de acurácia da metodologia aumentam conforme a taxa de prevalência diminui. Por outras palavras, para as prevalências menores ( $p = 0.01$  e  $p = 0.005$ ), a acurácia é geralmente mais alta em todas as combinações de  $SI$  e  $EI$ , quando comparadas à prevalência de  $p = 0.025$ . Isso ocorre porque, com uma prevalência mais baixa, há uma menor probabilidade de falsos negativos, o que melhora a precisão geral dos testes. A alta especificidade ( $EI = 1$ ) é particularmente eficaz em reduzir o número de falsos positivos, o que é crucial para manter uma alta acurácia, especialmente em prevalências mais baixas. Por outro lado, uma especificidade menor ( $EI = 0.9$ ) em condições de alta prevalência resulta em mais falsos positivos, reduzindo a acurácia da metodologia.

Comparando as duas metodologias, observa-se que ambas apresentam padrões consistentes de acurácia em função das combinações de sensibilidade ( $SI$ ) e especificidade ( $EI$ ), bem como da dimensão do grupo e da taxa de prevalência. No entanto, a metodologia hierárquica tende a ter maior acurácia para todas as taxas de prevalência analisadas.

Na metodologia hierárquica, a acurácia da metodologia permanece elevada mesmo com o aumento da dimensão do grupo, especialmente para as combinações  $SI = 0.9$  e  $EI = 1$  e  $SI = 0.95$  e  $EI = 1$ . Em contraste, na metodologia tradicional, há uma tendência de diminuição da acurácia com o aumento da dimensão do grupo, para as combinações  $SI = 1$  e  $EI = 0.9$  e  $SI = 1$  e  $EI = 0.95$ . Além disso, a metodologia hierárquica parece ser menos sensível às variações na prevalência, mantendo níveis de acurácia mais elevados à medida que a prevalência diminui. Na metodologia tradicional, a acurácia também melhora com a diminuição da prevalência, mas não de forma tão consistente quanto na hierárquica. É possível observar que em ambas as metodologias, as piores combinações são as mesmas.

### 4.3.3. Sensibilidade da metodologia

Relativamente à sensibilidade da metodologia hierárquica, para a taxa de prevalência  $p = 0.025$  (Figura 46 no Anexo 19), observa-se que, as combinações  $SI = 1$  e  $EI = 0.9$  e  $SI = 1$  e  $EI = 0.9$  apresentam a maior sensibilidade, alcançando 100% em todas as dimensões de grupo (notemos que a sensibilidade do teste individual unitária implica a sensibilidade da metodologia igualmente unitária em qualquer das metodologias). Por outro lado, a combinação  $SI = 0.9$  e  $EI = 1$  mostra a menor sensibilidade, ficando em torno de 65% (cerca de  $0.9^4 = 0.6561$ ) para todas as dimensões de grupo. A combinação  $SI = 0.95$  e  $EI = 1$  apresenta uma sensibilidade intermédia, próxima de 81% (cerca de  $0.95^4 = 0.8145$ ).

Para a prevalência de  $p = 0.01$  (Figura 48 no Anexo 20), os padrões observados são similares. As combinações  $SI = 1$  e  $EI = 0.9$  e  $SI = 1$  e  $EI = 0.95$  mantêm uma sensibilidade da metodologia de 100% para todas as dimensões de grupo. A combinação  $SI = 0.9$  e  $EI = 1$  continua a apresentar a menor sensibilidade, em torno de 65% para todas as dimensões de grupo. A combinação  $SI = 0.95$  e  $EI = 1$  mantêm uma sensibilidade próxima de 81%, independentemente da dimensão do grupo.

Para a prevalência de  $p = 0.005$  (Figura 50 no Anexo 21), os resultados seguem a mesma tendência.

A queda na sensibilidade da metodologia hierárquica quando  $SI$  é menor que 1 deve-se principalmente ao aumento dos falsos negativos. Tal como referido na secção 4.2.3., nesta abordagem, os testes ocorrem em múltiplas etapas, e se um falso negativo acontecer numa etapa superior, os indivíduos positivos podem ser ignorados nas etapas subsequentes. Como cada etapa depende dos resultados anteriores, um erro inicial propaga-se, reduzindo a deteção de casos positivos. Assim, a sensibilidade da metodologia cai à medida que os falsos negativos se acumulam. Podemos explicar melhor através da seguinte fórmula, utilizando  $SI = 0.95$  como exemplo:

$$SM = SI_{e1} \times SI_{e2} \times SI_{e3} \times SI_{e4}$$

$$SM = 0.95 \times 0.95 \times 0.95 \times 0.95 = 0.8145$$

Para a prevalência de  $p = 0.025$  aplicando a metodologia tradicional (Figura 47 no Anexo 19), os resultados seguem a mesma tendência. As combinações  $SI = 1$  e  $EI = 0.9$  e  $SI = 1$  e  $EI = 0.95$  apresentam sensibilidade de 100% para todos os dimensões de grupo. A

combinação  $SI = 0.9$  e  $EI = 1$  continua a ter a menor sensibilidade, em torno de 81% (cerca de  $0.9^2 = 0.81$ ), enquanto a combinação  $SI = 0.95$  e  $EI = 1$  apresenta uma sensibilidade de cerca de 90% (cerca de  $0.95^2 = 0.9025$ ).

No que diz respeito à metodologia tradicional, para a prevalência de  $p = 0.01$  (Figura 49 no Anexo 20), a combinação  $SI = 1$  e  $EI = 0.9$  apresenta a maior sensibilidade, alcançando 100% em todas as dimensões de grupo. A combinação  $SI = 1$  e  $EI = 0.95$  também mantém uma sensibilidade de 100% em todos os dimensões de grupo. Por outro lado, a combinação  $SI = 0.9$  e  $EI = 1$  mostra a menor sensibilidade da metodologia, variando em torno de 81% para todos os dimensões de grupo. A combinação  $SI = 0.95$  e  $EI = 1$  apresenta uma sensibilidade intermediária, próxima de 90%.

Para a taxa de prevalência de  $p = 0.005$  (Figura 51 no Anexo 21), os padrões observados são semelhantes. A combinação  $SI = 1$  e  $EI = 0.9$  mantém uma sensibilidade de 100% para todas as dimensões de grupo, assim como a combinação  $SI = 1$  e  $EI = 0.95$ . A combinação  $SI = 0.9$  e  $EI = 1$  continua a apresentar a menor sensibilidade da metodologia, em torno de 80% para todas as dimensões de grupo. A combinação  $SI = 0.95$  e  $EI = 1$  mantém uma sensibilidade próxima de 90%. Sendo assim a sensibilidade da metodologia hierárquica é independente da taxa de prevalência e da dimensão de grupo, ou seja, só depende do número de etapas realizadas.

Comparando a metodologia hierárquica com a metodologia tradicional em termos de sensibilidade da metodologia, observa-se que ambas apresentam variações significativas dependendo das combinações de sensibilidade e especificidade utilizadas. Na metodologia hierárquica, as combinações  $SI = 1$  e  $EI = 0.9$  e  $SI = 1$  e  $EI = 0.95$  atingem uma sensibilidade de 100% para todas as taxas de prevalência e dimensões de grupo, enquanto a combinação  $SI = 0.9$  e  $EI = 1$  apresenta uma sensibilidade significativamente menor, em torno de 65%. A combinação  $SI = 0.95$  e  $EI = 1$  na metodologia hierárquica apresenta uma sensibilidade intermediária, próxima de 81%. Já na metodologia tradicional, as combinações  $SI = 1$  e  $EI = 0.9$  e  $SI = 1$  e  $EI = 0.95$  também atingem uma sensibilidade de 100% para todas as taxas de prevalência e dimensões de grupo, mas a combinação  $SI = 0.9$  e  $EI = 1$  apresenta uma sensibilidade ligeiramente superior, em torno de 81%. A combinação  $SI = 0.95$  e  $EI = 1$  na metodologia tradicional apresenta uma sensibilidade intermediária, próxima de 90%.

A principal diferença entre as duas metodologias é que a hierárquica tem mais etapas, o que resulta numa queda mais acentuada na sensibilidade da metodologia quando a sensibilidade

não é perfeita ( $SI < 1$ ). Na metodologia tradicional, apesar de também haver uma queda na sensibilidade quando  $SI$  é menor que 1, essa queda é menos pronunciada. Isto deve-se ao facto de que a metodologia tradicional faz apenas dois testes no máximo, e a hierárquica faz até 4. No final, só é positivo (classificado como doente) se for positivo nos 4 testes. Em resumo, enquanto ambas as metodologias se beneficiam de altas combinações de  $SI$  e  $EI$ , a metodologia hierárquica tende a ser mais suscetível à diminuição da sensibilidade em casos de menor sensibilidade do teste individual, enquanto a tradicional mantém uma sensibilidade ligeiramente superior nessas situações.

#### 4.3.4. Especificidade da metodologia

Para a prevalência de  $p = 0.025$  (Figura 52 no Anexo 22), a especificidade da metodologia hierárquica apresenta-se elevada para todas as combinações de sensibilidade e especificidade. A combinação  $SI = 0.95$  e  $EI = 1$  e  $SI = 0.9$  e  $EI = 1$  atinge uma especificidade de 100% para todas as dimensões de grupo (especificidade individual unitária implica especificidade da metodologia igual a 1). As combinações  $SI = 1$  e  $EI = 0.9$  e  $SI = 1$  e  $EI = 0.95$  apresentam especificidades inferiores, mas ainda bastante altas, superiores a 97.5% e 99%, respetivamente.

Para a prevalência de  $p = 0.01$  (Figura 54 do Anexo 23) e  $p = 0.005$  (Figura 56 do Anexo 24), a especificidade da metodologia hierárquica segue um padrão semelhante ao observado para a taxa de prevalência  $p = 0.025$ . A combinação  $SI = 0.95$  e  $EI = 1$  e  $SI = 0.9$  e  $EI = 1$  continuam a apresentar especificidade máxima, de 100%, independentemente da dimensão do grupo. As combinações  $SI = 1$  e  $EI = 0.95$  e  $SI = 1$  e  $EI = 0.9$  mantêm especificidades altas, próximas de 100%, contudo mais elevadas do que em  $p = 0.025$ . É possível observar que nas combinações  $SI = 1$  e  $EI = 0.95$  e  $SI = 1$  e  $EI = 0.9$  a especificidade vai aumentando, conforme a taxa de prevalência diminui.

Para a prevalência de  $p = 0.025$ , a especificidade da metodologia tradicional (Figura 53 no Anexo 22) mantém-se bastante elevada, com pequenas variações entre as diferentes combinações de sensibilidade e especificidade. Observa-se que a combinação  $SI = 0.95$  e  $EI = 1$  e  $SI = 0.9$  e  $EI = 1$  apresentam especificidade unitária, para todas as dimensões de grupo. As combinações  $SI = 1$  e  $EI = 0.95$  e  $SI = 1$  e  $EI = 0.9$  também apresentam altas especificidades, próximas de 100%, sendo que a combinação  $SI = 1$  e  $EI = 0.95$  é constantemente mais elevada que  $SI = 1$  e  $EI = 0.9$ . É importante referir que é possível

observar um decréscimo na especificidade à medida que a dimensão de grupo aumenta, nestas combinações.

Tal como acontece em  $p = 0.025$ , o mesmo ocorre em  $p = 0.01$  (Figura 55 no Anexo 23) e  $p = 0.005$  (Figura 57 no Anexo 24), podendo observar-se um aumento na especificidade à medida que a taxa de prevalência diminui.

Ao comparar a metodologia hierárquica com a tradicional, observamos que ambas mantêm uma especificidade alta para todas as combinações de sensibilidade e especificidade. Contudo, algumas diferenças subtis destacam-se: em ambas as metodologias, as combinações  $SI = 0.95$  e  $EI = 1$  e  $SI = 0.9$  e  $EI = 1$  alcançam especificidade unitária, indicando uma perfeita eficácia na deteção de verdadeiros negativos (se a especificidade individual for 1, então a especificidade de cada metodologia também será unitária). As combinações  $SI = 1$  e  $EI = 0.95$  e  $SI = 1$  e  $EI = 0.9$  mantêm especificidades altas, mas inferiores àquelas com  $EI = 1$ . Na metodologia hierárquica, a especificidade não parece ser afetada pelo aumento da dimensão de grupo, o que contrasta com a metodologia de Dorfman, na qual a especificidade tende a diminuir à medida que a dimensão de grupo aumenta. Isso deve-se ao efeito acumulativo dos testes em várias etapas, que pode introduzir variações na deteção de verdadeiros negativos. A metodologia tradicional mostra uma consistência maior na especificidade independentemente da dimensão do grupo, destacando-se como uma abordagem mais estável em termos de manutenção de especificidade alta. Ambas as metodologias demonstram que as combinações com  $EI = 1$  mantêm especificidade de grupo de 100% em todas as prevalências analisadas ( $p = 0.01$ ,  $p = 0.005$ , e  $p = 0.025$ ), indicando que a especificidade da metodologia é robusta às mudanças na taxa de prevalência.

Observando os resultados, podemos concluir que a metodologia hierárquica tem maior capacidade para identificar os indivíduos sem doença do que a metodologia tradicional em todas as taxas de prevalência e combinações de SI e EI.

#### **4.4. Metodologia não hierárquica**

Nesta secção, iremo-nos focar na comparação de resultados utilizando duas variantes: com e sem a implementação de *master pool*. As dimensões de grupo utilizadas serão organizadas em matrizes quadradas, ou seja,  $4 \times 4$ ,  $5 \times 5$ ,  $6 \times 6$ ,  $7 \times 7$ ,  $8 \times 8$ ,  $9 \times 9$ ,  $10 \times 10$ ,  $25 \times 25$ ,  $36 \times 36$ ,  $49 \times 49$  e  $64 \times 64$  para ambas as metodologias. As taxas de prevalência utilizadas são 0.005, 0.01 e 0.025, as mesmas aplicadas anteriormente nas metodologias hierárquica e de Dorfman.

#### 4.4.1. Média de total de testes

Na metodologia não hierárquica com *master pool*, observa-se que a média do total de testes varia conforme a sensibilidade e especificidade mudam. Esta variação, a partir da dimensão 10 x 10, torna-se mais evidente com o aumento da dimensão do grupo, sendo as dimensões de grupo de 10 x 10 e 4 x 4 consistentemente as mais eficientes. Para uma taxa de prevalência de  $p = 0.025$  (Figura 58 no Anexo 25) os resultados são os ilustrados na Tabela 14:

| Prevalência ( $p$ ) | SI = EI | Média de testes | Eficiência (%) | Dimensão de grupo ( $n \times n$ ) |
|---------------------|---------|-----------------|----------------|------------------------------------|
| 0.025               | 0.90    | 25574           | 74.43%         | 10 x 10                            |
| 0.025               | 0.95    | 25641           | 74.36%         | 10 x 10                            |
| 0.025               | 0.99    | 24303           | 75.69%         | 4 x 4                              |
| 0.025               | 1       | 23890           | 76.11%         | 4 x 4                              |

**Tabela 14 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia não hierárquica com *master pool* numa taxa de prevalência de 2.5% quando SI = EI**

A eficiência aumenta à medida que a sensibilidade se aproxima de 1, com a eficiência mais alta (76.11%) a ocorrer quando SI = EI = 1. Essa combinação resulta na menor média de testes necessários, tornando-a a configuração mais eficaz para a dimensão de grupo 4 x 4. Em contraste, quando a sensibilidade é dada por SI = EI = 0.90, a eficiência cai para 74.43%, indicando um maior número de testes necessários e, portanto, uma menor eficiência na utilização dos recursos.

Na metodologia não hierárquica sem *master pool* (Figura 59 no Anexo 25), a média do total de testes diminui de forma menos pronunciada até ao grupo de dimensão 10 x 10. A variação entre os diferentes valores de SI = EI diminui até à dimensão de grupo 10 x 10 e torna a aumentar na dimensão de grupo 25 x 25. Para todas as dimensões de grupo até 10 x 10, SI = EI = 1 resulta numa média de testes mais baixa comparada às restantes SI = EI. A partir de 25 x 25, nota-se um aumento considerável da média total de testes e uma inversão nos valores de SI = EI, passando SI = EI = 0.9 a ter a menor média total de testes, como se pode observar na Tabela 15.

| Prevalência ( $p$ ) | SI = EI | Média de testes | Eficiência (%) | Dimensão de grupo ( $n \times n$ ) |
|---------------------|---------|-----------------|----------------|------------------------------------|
| 0.025               | 0.90    | 29267           | 70.73%         | 10 x 10                            |
| 0.025               | 0.95    | 27823           | 72.18%         | 10 x 10                            |
| 0.025               | 0.99    | 26554           | 73.45%         | 10 x 10                            |
| 0.025               | 1       | 26244           | 73.76%         | 10 x 10                            |

**Tabela 15 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia não hierárquica sem *master pool* numa taxa de prevalência de 2.5% quando SI = EI**

Estes dados indicam que, para a taxa de prevalência  $p = 0.025$ , o grupo de dimensão 10 x 10 é o mais eficiente em todos os valores analisados de SI = EI, sendo a eficiência mais alta (73.76%) observada quando SI = EI = 1, o que naturalmente coincide com a menor média de testes necessários. À medida que os valores de SI = EI diminuem, a eficiência também cai, atingindo 70.73% para SI = EI = 0.90, reflexo do aumento no número total de testes necessários.

Concluindo, ambas as metodologias têm, para valores de SI = EI iguais a 0.9 e 0.95, uma dimensão de grupo ideal igual, neste caso, 10 x 10, sendo que na metodologia com *master pool* a dimensão de grupo 4 x 4 consegue obter melhores resultados para valores de SI = EI iguais a 0.99 e 1. Dependendo do valor de SI = EI, poderá haver uma ligeira diferença na média do número de testes necessários entre a utilização ou não de *master pool*, mas não é uma diferença significativa. Assim sendo, para uma taxa de prevalência  $p = 0.025$ , ambas as metodologias são igualmente adequadas.

Para uma taxa de prevalência  $p = 0.01$  (Figura 60 no Anexo 26), verifica-se que a média do número total de testes necessários para testar a população diminui consideravelmente em comparação com a taxa de prevalência  $p = 0.025$  na metodologia com *master pool*. Na taxa de prevalência  $p = 0.01$ , a dimensão de grupo ideal (entre as dimensões analisadas) passa a ser 25 x 25 para SI = EI mais baixas e 5 x 5 para SI = EI mais altas, como é possível observar na Tabela 16.

| Prevalência ( $p$ ) | SI = EI | Média de testes | Eficiência (%) | Dimensão de grupo ( $n \times n$ ) |
|---------------------|---------|-----------------|----------------|------------------------------------|
| 0.01                | 0.90    | 14600           | 85.40%         | 25 x 25                            |
| 0.01                | 0.95    | 14194           | 85.81%         | 25 x 25                            |
| 0.01                | 0.99    | 13516           | 86.48%         | 5 x 5                              |
| 0.01                | 1       | 13214           | 86.79%         | 5 x 5                              |

**Tabela 16 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia não hierárquica com *master pool* numa taxa de prevalência de 1% quando SI = EI**

Considerando as dimensões ótimas, nota-se que a eficiência aumenta à medida que SI = EI se aproxima de 1, alcançando o seu valor máximo (86.79%) quando SI = EI = 1. Comparando com a prevalência de  $p = 0.025$ , a redução para  $p = 0.01$  leva a uma diminuição significativa na média total de testes necessários.

Na metodologia não hierárquica sem *master pool* (Figura 61 no Anexo 26), a média do número total de testes diminui de forma acentuada até ao grupo de dimensão 25 x 25, observando-se uma tendência para o seu aumento a partir desta dimensão. A menor média de testes continua a ser quando SI = EI = 1, tal como observado na Tabela 17.

| Prevalência ( $p$ ) | SI = EI | Média de testes | Eficiência (%) | Dimensão de grupo ( $n \times n$ ) |
|---------------------|---------|-----------------|----------------|------------------------------------|
| 0.01                | 0.90    | 16123           | 83.88%         | 25 x 25                            |
| 0.01                | 0.95    | 14738           | 85.26%         | 25 x 25                            |
| 0.01                | 0.99    | 13774           | 86.23%         | 25 x 25                            |
| 0.01                | 1       | 13537           | 86.46%         | 25 x 25                            |

**Tabela 17 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia não hierárquica sem *master pool* numa taxa de prevalência de 1% quando SI = EI**

Ao contrário do observado na taxa de prevalência  $p = 0.025$ , a metodologia não hierárquica com *master pool* revela ser mais eficiente que a metodologia sem *master pool*, com uma média do número total de testes ligeiramente inferior.

Considerando a dimensão ótima de 25 x 25, à medida que os valores de SI = EI diminuem, a eficiência também cai ligeiramente, com a eficiência mais baixa (83.88%) a ocorrer quando SI = EI = 0.90.

Para uma taxa de prevalência  $p = 0.005$  (Figura 62 no Anexo 27), continuamos com os mesmos resultados observados na taxa de prevalência  $p = 0.01$ , sendo que há uma descida na média do número total de testes, como se pode observar na Tabela 18.

| Prevalência ( $p$ ) | SI = EI | Média de testes | Eficiência (%) | Dimensão de grupo ( $n \times n$ ) |
|---------------------|---------|-----------------|----------------|------------------------------------|
| 0.005               | 0.90    | 10046           | 89.95%         | 36 x 36                            |
| 0.005               | 0.95    | 9432            | 90.57%         | 36 x 36                            |
| 0.005               | 0.99    | 8680            | 91.32%         | 7 x 7                              |
| 0.005               | 1       | 8385            | 91.62%         | 6 x 6                              |

**Tabela 18 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia não hierárquica com *master pool* numa taxa de prevalência de 0.5% quando SI = EI**

Conforme se pode verificar na Tabela 18, para uma taxa de prevalência de  $p = 0.005$ , a metodologia não hierárquica com *master pool* atinge a sua maior eficiência (91.62%) quando SI = EI = 1. À medida que os valores de SI = EI diminuem, a eficiência também diminui, sendo a menor eficiência (89.95%) observada quando SI = EI = 0.90. A dimensão de grupo 36 x 36 é a mais eficiente quando SI = EI é baixa, sendo que quando SI = EI = 1 a dimensão de grupo ideal é 6 x 6.

Na metodologia não hierárquica sem *master pool* (Figura 63 no Anexo 27), observa-se uma diminuição na média do número total de testes até à dimensão 25 x 25 e, em cada dimensão, com o aumento dos valores de SI = EI. Nesta metodologia, existe uma estabilização entre os diferentes valores de 25 x 25 e 36 x 36, mas também se verifica uma diminuição da média do número total de testes quando comparado com a taxa de prevalência anterior. Na Tabela 19 estão representadas as dimensões de grupo com a menor média de testes para taxa de prevalência  $p = 0.005$ .

| Prevalência ( $p$ ) | SI = EI | Média de testes | Eficiência (%) | Dimensão de grupo ( $n \times n$ ) |
|---------------------|---------|-----------------|----------------|------------------------------------|
| 0.005               | 0.90    | 11125           | 88.88%         | 36 x 36                            |
| 0.005               | 0.95    | 9761            | 90.24%         | 36 x 36                            |
| 0.005               | 0.99    | 8845            | 91.15%         | 36 x 36                            |
| 0.005               | 1       | 8612            | 91.39%         | 36 x 36                            |

**Tabela 19 - Dimensão de grupo com melhor média de testes e eficiência usando a metodologia não hierárquica sem *master pool* numa taxa de prevalência de 0.5% quando SI = EI**

Assim, podemos concluir que para esta taxa de prevalência, a melhor dimensão de grupo na metodologia não hierárquica sem *master pool* é 36 x 36 e o melhor valor de SI continua a ser  $SI = EI = 1$ . Assim, conforme apresentado na Tabela 19, na metodologia não hierárquica sem *master pool* com uma taxa de prevalência de  $p = 0.005$ , a eficiência máxima de 91.39% é obtida quando  $SI = EI = 1$ , correspondendo naturalmente ao menor número médio de testes necessários. Em relação à dimensão ótima (entre os valores analisados), à medida que os valores de  $SI = EI$  diminuem, observa-se uma ligeira redução na eficiência, atingindo o valor mais baixo de 88.88% quando  $SI = EI = 0.90$ . Deste modo, a configuração de grupo de dimensão 36 x 36 sobressai-se como a mais eficaz, oferecendo a maior eficiência, superando inclusive os resultados observados nas taxas de prevalência anteriormente analisadas.

Em suma, a metodologia não hierárquica sem *master pool* mostra uma maior redução no total de testes para grupos muito grandes, enquanto a metodologia com *master pool* mantém uma boa eficiência em grupos menores. Esta observação pode ser confirmada pelo facto de que à medida que a taxa de prevalência diminui, a dimensão de grupo ideal aumenta na metodologia sem *master pool*, o que não ocorre necessariamente na metodologia com *master pool*. Mesmo com esta menor dimensão, a eficiência da metodologia com *master pool* é ligeiramente superior do que a obtida sem *master pool*.

#### 4.4.2. Acurácia da metodologia

Relativamente à acurácia da metodologia não hierárquica com *master pool*, observam-se algumas tendências claras nos gráficos para diferentes taxas de prevalência. Para a taxa de prevalência  $p = 0.025$  (Figura 64 no Anexo 28), observa-se um valor constante da acurácia à medida que a dimensão do grupo aumenta até  $n = 10 \times 10$ , sendo que a partir daí a acurácia diminui com o aumento da dimensão de grupo para  $SI = EI < 1$ , mas permanece constante para  $SI = EI = 1$ .

Quando se reduz a taxa de prevalência  $p = 0.01$  (Figura 66 do Anexo 29), o padrão de acurácia da metodologia mantém-se semelhante. A acurácia mostra uma ligeira melhoria com o aumento do tamanho do grupo para combinações menos precisas, especialmente  $SI = EI = 0.9$ .

Para a taxa de prevalência  $p = 0.005$  (Figura 68 do Anexo 30), o padrão observado nas taxas de prevalência anteriores persiste. Quando  $SI = EI = 1$  continua a fornecer a maior acurácia da metodologia, com  $SI = EI = 0.9$  permanecendo a menos precisa. A acurácia parece

constante até  $n = 10 \times 10$  (neste caso parece manter-se até uma dimensão superior  $25 \times 25$ ), mas diminui com o aumento de grupo a partir daí, para  $SI = EI$  inferior à unidade.

Apesar do padrão ser o mesmo em todas as taxas de prevalência, podemos observar um aumento da acurácia geral à medida que a taxa de prevalência diminui, quando  $SI = EI$  é inferior à unidade. Conforme previamente referido, sempre que  $SI = EI = 1$ , as metodologias de classificação baseadas em testes compostos também não têm erros de classificação.

Na metodologia não hierárquica sem *master pool* (Figura 65 do Anexo 28), para a taxa de prevalência  $p = 0.025$ , a acurácia da metodologia apresenta uma ligeira diminuição com o aumento da dimensão do grupo a partir da dimensão  $10 \times 10$  para  $SI = EI$  inferior a 1, sendo a diminuição mais acentuada para valores mais baixos de  $SI = EI$ . Além disso, para os valores mais baixos de  $SI = EI$  (0.9 e 0.95) nota-se uma tendência crescente da acurácia com a dimensão do grupo até à dimensão  $10 \times 10$ .

Para a taxa de prevalência  $p = 0.01$  (Figura 67 do Anexo 29), o padrão é semelhante ao da taxa de prevalência  $p = 0.025$ . Há uma tendência de aumento da acurácia para valores baixos de  $SI = EI$  com o aumento da dimensão do grupo, mas essa tendência inverte-se em grupos muito grandes, sendo que nesta taxa de prevalência a acurácia começa a diminuir a partir da dimensão de grupo  $25 \times 25$ .

Finalmente, para taxa de prevalência  $p = 0.005$  (Figura 69 no Anexo 30), a acurácia da metodologia comporta-se de forma semelhante aos casos anteriores. Assim, a acurácia aumenta inicialmente quando  $SI = EI$  é igual a 0.9 ou 0.95, mas há uma diminuição a partir da dimensão de grupo  $25 \times 25$ .

Concluindo, conforme esperado, quanto maior o valor de  $SI = EI$  maior é o valor da acurácia. Em particular, a combinação de  $SI = EI = 1$  proporciona consistentemente a maior acurácia em ambas as metodologias, uma vez que não tem associado qualquer erro de classificação independentemente da dimensão utilizada. Por outro lado, combinações  $SI = EI$  com valores mais baixos apresentam uma acurácia crescente até determinada dimensão a partir da qual a acurácia começa a diminuir.

#### 4.4.3. Sensibilidade da metodologia

Relativamente à sensibilidade da metodologia não hierárquica com *master pool* para a taxa de prevalência  $p = 0.025$  (Figura 70 do Anexo 31), a sensibilidade é maior quanto maior for

o valor da combinação  $SI = EI$ . Por outro lado, para combinações  $SI = EI$  diferentes da unidade, a acurácia diminui com o aumento da dimensão do grupo, sendo a diminuição mais acentuada quanto menor for o valor da combinação  $SI = EI$ . Por fim, os valores da sensibilidade tendem a estabilizar a partir da dimensão  $25 \times 25$ .

Quando a taxa de prevalência é reduzida para  $p = 0.01$  (Figura 72 do Anexo 32) e para  $p = 0.005$  (Figura 74 do Anexo 33), o padrão permanece muito semelhante, sem diferenças significativas a relevar.

Na metodologia não hierárquica sem *master pool*, a sensibilidade da metodologia naturalmente diminui com a diminuição do valor da combinação  $SI = EI$  (Figura 71 do Anexo 31). Tal como na metodologia não hierárquica com *master pool*, observa-se uma tendência decrescente na sensibilidade com o aumento da dimensão do grupo quando  $SI = EI$  é inferior a 1, especialmente nos valores mais baixos da combinação  $SI = EI$ . Esta tendência é alterada a partir da dimensão  $25 \times 25$ , a partir da qual parece estabilizar.

Para as taxas de prevalência  $p = 0.01$  (Figura 73 do Anexo 32) e  $p = 0.005$  (Figura 75 do Anexo 33) o padrão é muito semelhante, sem qualquer diferença significativa.

Concluindo, as combinações de  $SI = EI$  mais elevadas proporcionam consistentemente maiores valores da sensibilidade em ambas as metodologias. Por outro lado, a sensibilidade da metodologia diminui até certa altura com o aumento da dimensão dos grupos em combinações  $SI = EI$  com valores mais baixos. Em geral, a sensibilidade da metodologia tende a ser ligeiramente maior em prevalências mais baixas e diminui com o aumento da prevalência.

#### **4.4.4. Especificidade da metodologia**

Relativamente à especificidade da metodologia não hierárquica com *master pool*, numa taxa de prevalência de  $p = 0.025$  (Figura 76 do Anexo 34), a especificidade mantém-se consistentemente alta, especialmente com  $SI = EI = 1$ , que se mantém estável em 100% para todas as dimensões de grupo e prevalências. As combinações de  $SI = EI = 0.9$  e  $0.95$  mostram uma diminuição na especificidade com o aumento da dimensão do grupo a partir da dimensão  $10 \times 10$ , sendo esta diminuição mais pronunciada para valores mais baixos da combinação  $SI = EI$ . A combinação  $SI = EI = 0.99$  mantém uma especificidade praticamente constante e alta.

Para a taxa de prevalência  $p = 0.01$  (Figura 78 do Anexo 35), o padrão é semelhante ao observado para taxa de prevalência  $p = 0.025$ . As combinações de  $SI = EI = 0.9$  e  $0.95$  continuam a mostrar uma diminuição na especificidade a partir de determinada dimensão, que neste caso passa a ser da dimensão  $25 \times 25$ , sendo esta diminuição menos acentuada do que quando a taxa de prevalência assumia o valor  $p = 0.025$ . A especificidade para  $SI = EI = 0.99$  permanece alta e sem variações significativas.

No caso de a taxa de prevalência assumir  $p = 0.005$  (Figura 80 do Anexo 36), o comportamento da especificidade da metodologia mantém-se semelhante aos casos anteriores, sendo a referida diminuição para combinações de  $SI = EI = 0.9$  e  $0.95$  a partir da dimensão  $25 \times 25$  menor nesta taxa de prevalência. Por fim, note-se que apesar do comportamento da especificidade ser semelhante nas três taxas de prevalência analisadas, a especificidade da metodologia é maior quanto menor for a taxa de prevalência.

Na metodologia não hierárquica sem *master pool* (Figura 77 do Anexo 34), na taxa de prevalência  $p = 0.025$ , a especificidade da metodologia é mais elevada para valores da combinação  $SI = EI$  mais elevados. Para valores da combinação  $SI = EI$  inferiores à unidade, a especificidade da metodologia aumenta com o aumento da dimensão de grupo até à dimensão  $10 \times 10$ , diminuindo a partir desta dimensão. Estas variações são mais acentuadas quanto menor for o valor da combinação  $SI = EI$ .

Para taxa de prevalência  $p = 0.01$  (Figura 79 do Anexo 35) o padrão da evolução da especificidade da metodologia mantém-se semelhante ao de taxa de prevalência  $p = 0.025$ . No entanto, ao contrário do observado anteriormente, nesta taxa de prevalência a especificidade apenas começa a diminuir a partir da dimensão de grupo  $25 \times 25$ , havendo um aumento mais considerável em relação ao observado nas dimensões anteriores. O mesmo se pode observar na evolução da especificidade da metodologia para a taxa de prevalência  $p = 0.005$  (Figura 81 do Anexo 36). Ao observar as 3 taxas de prevalência, conclui-se que a especificidade aumenta quando a taxa de prevalência diminui, esta conclusão é especialmente observável quando  $SI = EI = 0.9$  e  $0.95$ .

Em conclusão, os valores da combinação de  $SI = EI$  mais elevados consistentemente proporcionam uma maior especificidade em ambas as metodologias, mantendo-se praticamente constante e próxima de 100%. As combinações de  $SI = EI$  com valores mais baixos têm especificidades mais baixas, especialmente na metodologia não hierárquica sem

*master pool*, e com maior variabilidade (aumentando e depois diminuindo), sendo relevante, nestes casos, escolher uma dimensão adequada.

## 5. Discussão de resultados

Neste capítulo, exploramos uma análise mais aprofundada e comparativa das diferentes metodologias de classificação baseadas em testes compostos, ou seja, nas metodologias de Dorfman, hierárquica e não hierárquica com e sem uso de *master pool*.

A análise dos resultados para uma taxa de prevalência de 0.01 (1%) revelou que os resultados obtidos estão em consonância com as conclusões de Dorfman [23], que demonstrou que para essa prevalência, o tamanho ideal do grupo era 11, maximizando a eficiência ao minimizar o número total de testes necessários, assumindo sensibilidade e especificidade perfeitas ( $SI = EI = 1$ ). De facto, nos resultados obtidos com esta metodologia, o grupo de tamanho 11 também apresentou a menor média de número de testes, com um total de 19525 testes, o que confirma a adequação da abordagem proposta por Dorfman.

Na análise da metodologia hierárquica para as taxas de prevalência de 0.01 e 0.005, observou-se que, com sensibilidade e especificidade perfeitas ( $SI = 1, EI = 1$ ), a metodologia hierárquica demonstrou uma eficiência superior em comparação à metodologia tradicional de Dorfman. Para uma taxa de prevalência  $p = 0.01$ , a média do número de testes foi de 12182 para um grupo inicial de 64 indivíduos, enquanto, para a taxa de prevalência de 0.005, a média foi de 7166 testes. Estes resultados confirmam as conclusões de Granado [27], que destacou a maior eficiência das metodologias hierárquicas em cenários de baixa prevalência.

Quando a sensibilidade e a especificidade foram reduzidas para 0.95, houve um ligeiro aumento na média de testes necessários, com 12273 testes para uma taxa de prevalência  $p = 0.01$  e 7435 testes para  $p = 0.005$ . Embora essa redução de prevalência tenha tido impacto na eficiência, a metodologia hierárquica continuou a ter a melhor eficiência em relação à metodologia tradicional, corroborando as observações de Granado [27] sobre estas metodologias mesmo quando há variação na precisão dos testes.

Na análise das metodologias não hierárquica com e sem *master pool* para as taxas de prevalência de 0.01 e 0.005, a metodologia não hierárquica sem *master pool* mostrou-se menos eficiente, resultando no maior número total de testes necessários. Para a prevalência de 0.01, a metodologia sem *master pool* alcançou uma média de 13537 testes com  $SI$  e  $EI$  iguais a 1, ligeiramente mais que os 13214 testes da metodologia com *master pool*. A mesma tendência foi observada para a prevalência de 0.005, com 8612 testes na metodologia sem

*master pool* versus 8385 testes na versão com *master pool*. Mesmo com a redução de SI e EI para 0.95, a metodologia com *master pool* manteve a sua superioridade.

Os resultados obtidos neste trabalho, relativamente às metodologias não hierárquicas, não são consistentes com as observações de Grobe *et al.* [3], que concluíram que a metodologia sem *master pool* tende a ser mais eficiente em termos de economia de testes, especialmente em cenários de baixa prevalência. Grobe *et al.* destacaram que, em situações com prevalência de 0.01, a metodologia sem *master pool* resultou numa menor média total de testes em comparação à versão com *master pool*. No entanto, tal não foi observado nos resultados das simulações efetuadas no presente trabalho, o que relança a discussão sobre a relevância do *master pool* em termos de eficiência metodológica.

Neste trabalho, os resultados obtidos com a metodologia hierárquica de quatro etapas coincidem com os obtidos por Hou *et al.* [28], que também apontam para uma redução significativa no número de testes necessários e uma manutenção de alta especificidade. Enquanto a sensibilidade tende a diminuir com o aumento das etapas, os ganhos em eficiência e a redução de falsos positivos podem justificar o uso desta metodologia em cenários de baixa prevalência, dependendo da sensibilidade inicial da metodologia e da quantidade de etapas. Todavia, apesar da quantidade de testes diminuir significativamente, se a probabilidade de identificar indivíduos doentes for próxima de 70% esta não será a metodologia ideal, pois neste cenário a sensibilidade da metodologia seria inferior a 25%. Hou *et al.* (2017) destacam que a metodologia hierárquica, especialmente em contextos de múltiplas etapas, oferece uma abordagem eficiente para a triagem de doenças em larga escala, o que está em linha com os resultados observados.

Comparando as várias metodologias estudadas, num contexto em que a sensibilidade se equipara à especificidade, é possível observar que todas elas obtêm melhores resultados, em termos de eficiência, à medida que a taxa de prevalência diminui. Além disso, a metodologia hierárquica destaca-se como sendo a mais eficiente em todas as taxas de prevalência e independentemente de SI e EI.

Em termos de acurácia, todas as metodologias mostram resultados ótimos, sendo que apresentam acurácias iguais ou superiores a 93% em todas as taxas de prevalência estudadas. Observa-se também uma relação inversa entre a acurácia e a taxa de prevalência. No entanto, é importante destacar que, mesmo com acurácia alta, a sensibilidade pode ser muito baixa

em contextos de baixa prevalência, o que compromete a capacidade de identificar corretamente os indivíduos com a doença.

De forma a completar a análise e comparação das 4 metodologias estudadas, podemos verificar nas Tabelas 20 e 21, o desempenho das quatro metodologias em termos de VPN, VPP, SM, EM e eficiência, utilizando as dimensões de grupo ideais referentes às taxas de prevalência  $p = 0.01$  e  $p = 0.005$ , para  $SI = EI = 0.95$  e  $SI = EI = 0.99$ . Além das métricas utilizadas para a comparação das metodologias, é referido também o erro padrão<sup>3</sup> de cada métrica, entre parêntesis, em cada uma das metodologias. Serão também identificadas as metodologias com a melhor eficiência, SM, EM, VPP e VPN a verde e as piores a laranja.

| Metodologia                            | VPN                | VPP                | SM                 | EM                 | Eficiência (%)     | Dimensão de grupo |
|--|--------------------|--------------------|--------------------|--------------------|--------------------|-------------------|
| <b>Dorfman</b>                         | 0.999<br>(0.00001) | 0.571<br>(0.00103) | 0.902<br>(0.00093) | 0.993<br>(0.00003) | 76.46%<br>(0.033%) | 11                |
| <b>Hierárquica</b>                     | 0.998<br>(0.00001) | 0.841<br>(0.00117) | 0.815<br>(0.00123) | 0.998<br>(0.00002) | 87.73%<br>(0.031%) | 64,16,4,1         |
| <b>Não hierárquica com master pool</b> | 0.998<br>(0.00002) | 0.743<br>(0.00134) | 0.819<br>(0.00216) | 0.997<br>(0.00002) | 85.81%<br>(0.043%) | 25 x 25           |
| <b>Não hierárquica sem master pool</b> | 0.999<br>(0.00001) | 0.744<br>(0.00109) | 0.858<br>(0.00106) | 0.997<br>(0.00002) | 85.26%<br>(0.039%) | 25 x 25           |

Tabela 20 - Comparação das metodologias estudadas, numa taxa de prevalência  $p = 0.01$  e  $SI = EI = 0.95$

Observando a Tabela 20, em termos de VPN, todas as metodologias apresentam valores bastante próximos, variando entre 0.998 e 0.999. Isso significa que, independentemente da metodologia escolhida, a fiabilidade dos resultados negativos é extremamente alta. Essa

<sup>3</sup> O erro padrão da estimativa (EP), que permite avaliar a variabilidade dos valores obtidos na simulação, é definido pela fórmula:

$$EP = \frac{s}{\sqrt{n}}$$

sendo:

- $s$  – desvio padrão da amostra;
- $n$  – dimensão da amostra.

semelhança no VPN indica que todas as metodologias são equivalentes em termos de acurácia na previsão de resultados negativos. Já o VPP apresenta maiores variações. A metodologia de Dorfman tem o VPP mais baixo, com um valor de 0.571, sugerindo que essa metodologia é menos eficaz a garantir que os indivíduos classificados como doentes realmente o são. Por outro lado, a metodologia hierárquica apresenta o maior VPP, com 0.841, indicando um desempenho superior na fiabilidade dos resultados positivos. As metodologias não hierárquicas situam-se numa posição intermediária, com VPP de 0.743 e 0.744.

É possível observar uma relação direta positiva entre eficiência, EM e VPP, bem como entre SM e VPN, ou seja, podemos verificar que a metodologia com maior eficiência é a metodologia com maior valor de EM e de VPP e a metodologia com maior sensibilidade é a metodologia com maior VPN. Deste modo, como seria espectável, na escolha entre metodologias há um *trade-off* entre a sensibilidade e a especificidade, bem como entre o VPN e o VPP.

Relativamente ao erro padrão nos valores de VPN, são muito similares entre todas as metodologias, variando entre 0.00001 e 0.00002. Isso sugere que a incerteza na estimativa do VPN é praticamente a mesma e quase nula em todas as metodologias, reforçando a ideia de que as estimativas apresentadas para o VPN (médias) são extremamente precisas.

O mesmo ocorre nas restantes medidas, com erros padrão máximos de 0.00134 para o VPP, 0.00216 para a sensibilidade da metodologia, 0.00003 para a especificidade da metodologia, e 0.043 para a eficiência. Por conseguinte, tendo em consideração a diminuta variabilidade das 100 estimativas calculadas (cada cenário foi repetido 100 vezes, conforme previamente referido, obtendo-se uma estimativa em cada repetição), a precisão das estimativas apresentadas é elevada. Recorde-se que, como estamos a trabalhar com proporções, é expectável que as estimativas mais próximas de 0.5 apresentem maior variabilidade, enquanto as estimativas mais próximas de 1 (supondo unicamente valores superiores a 0.5) tendem a apresentar menor variabilidade (consequência de a variância de uma prova binária ser máxima para  $p = 0.5$ ). Por outro lado, se recorrermos a intervalos com 95 por cento de confiança gerados com recurso ao Teorema Limite Central, teremos:

$$] \hat{p} - 1,96 EP; \hat{p} + 1,96 EP [$$

onde  $\hat{p}$  representa a estimativa da proporção  $p$  em análise. Deste modo, valores do erro padrão diminutos geram intervalos de confiança com amplitude reduzida (amplitude do intervalo igual a 3.92 vezes o valor do erro padrão), o que confirma a precisão das estimativas apresentadas. [31]

Na Tabela 21, podem ser observados os resultados obtidos para as mesmas métricas analisadas anteriormente, mas para uma taxa de prevalência de  $p = 0.005$  utilizando a combinação  $SI = SI = 0.99$ .

| Metodologia                            | VPN                 | VPP                | SM                 | EM                  | Eficiência (%)     | Dimensão de grupo |
|--|---------------------|--------------------|--------------------|---------------------|--------------------|-------------------|
| <b>Dorfman</b>                         | 0.999<br>(0.000003) | 0.873<br>(0.00141) | 0.979<br>(0.00063) | 0.999<br>(0.000009) | 85.25%<br>(0.035%) | 14                |
| <b>Hierárquica</b>                     | 0.999<br>(0.000005) | 0.969<br>(0.00073) | 0.960<br>(0.00094) | 0.999<br>(0.000004) | 92.79%<br>(0.024%) | 64,16,4,1         |
| <b>Não hierárquica com master pool</b> | 0.999<br>(0.000003) | 0.915<br>(0.00108) | 0.982<br>(0.00059) | 0.999<br>(0.000007) | 91.32%<br>(0.028%) | 7 x 7             |
| <b>Não hierárquica sem master pool</b> | 0.999<br>(0.000004) | 0.945<br>(0.00092) | 0.971<br>(0.00076) | 0.997<br>(0.000005) | 91.15%<br>(0.027%) | 36 x 36           |

Tabela 21 - Comparação das metodologias estudadas, numa taxa de prevalência  $p = 0.005$  e  $SI = EI = 0.99$

Em termos de VPN, todas as metodologias apresentam resultados praticamente idênticos, com valores aproximadamente de 0.999. Isso mostra que todas as metodologias de classificação baseadas em testes compostos são extremamente fiáveis quando classificam como saudáveis, sem distinção relevante entre elas nesse aspeto.

Já no VPP, há algumas diferenças, pois a metodologia de Dorfman apresenta um VPP de 0.873, o mais baixo entre as quatro metodologias. Isso indica que tem uma menor fiabilidade quando o resultado do teste é positivo em comparação com as demais metodologias. A metodologia hierárquica, por outro lado, apresenta o maior VPP, com 0.969, sugerindo que é a mais fiável quando a classificação é positiva (classificação como doente). As

metodologias não hierárquicas situam-se entre essas duas, com a abordagem com *master pool* apresentando um VPP de 0.915 e a abordagem sem *master pool* com um valor ligeiramente superior de 0.945. Isso indica que ambas as metodologias não hierárquicas têm uma fiabilidade razoável quando apresentam resultados positivos, sendo mais eficazes do que a metodologia de Dorfman, mas menos do que a metodologia hierárquica. É de realçar que a metodologia não hierárquica com *master pool* consegue ser a melhor em termo de SM e EM, sem ser a pior em qualquer outra métrica.

Para o cenário em análise, os erros padrão máximos são: 0.000005 para o VPN, 0.00141 para o VPP, 0.00094 para a sensibilidade da metodologia, 0.000009 para a especificidade da metodologia, e 0.03478 para a eficiência. Uma vez mais, os valores observados dos erros padrão são extremamente baixos, confirmando a alta precisão das estimativas apresentadas. Apesar de, ao longo do trabalho, não ter sido apresentada a variabilidade de cada estimativa apresentada, em todas as análises apresentadas foi utilizada a mesma metodologia (100 repetições de cada cenário, sendo avaliada uma população de 100000 indivíduos em cada repetição), tendo os valores obtidos sido muito semelhantes ao que são apresentados nas Tabelas 20 e 21.

Com base na análise comparativa realizada sobre as quatro metodologias de testes compostos, conclui-se que a metodologia hierárquica é a mais eficiente, destacando-se em cenários de baixa prevalência devido à sua capacidade de minimizar o número total de testes necessários. A metodologia não hierárquica com *master pool*, apesar de não ter a maior eficiência, tem a melhor capacidade para identificar os indivíduos doentes (apesar da metodologia de Dorfman apresentar maior sensibilidade para valores mais baixos de SI). Além disso, a metodologia hierárquica apresenta um excelente VPP, o que a torna particularmente fiável quando classifica indivíduos como doentes. Considerando estes resultados, é recomendável a adoção da metodologia hierárquica em situações onde o custo e o tempo de testagem são críticos e a prevalência da doença é baixa.

Tal como observado na Tabela 20, observamos as mesmas relações entre eficiência, especificidade e VPP e entre SM e VPN. Sendo assim, podemos concluir que quanto mais eficiente for uma metodologia, maior tenderá a ser a probabilidade de identificar corretamente os indivíduos saudáveis e mais fiável serão as classificações positivas (isto é, tenderá a haver poucos falsos positivos, sendo os erros de classificação centrados nos falsos negativos). Da mesma forma, quanto maior for a sensibilidade da metodologia, maior será a

probabilidade de identificar os indivíduos doentes e a fiabilidade das classificações negativas (isto é, tenderá a ocorrer poucos falsos negativos, sendo os erros de classificação centrados nos falsos positivos).

Por conseguinte, a escolha da metodologia ideal não é óbvia, acabando por depender do cenário em que será necessária a utilização de metodologias de classificação baseadas em testes compostos, bem como dos objetivos dessa classificação, nomeadamente das características que se considerem mais relevantes. Por exemplo, numa aplicação cujo objetivo seja minimizar os recursos utilizados, a eficiência será a principal medida a ter em consideração, mesmo que isso implique alguma diminuição da sensibilidade da classificação. Por outro lado, caso a identificação de todos os doentes seja o principal objetivo, então a melhor escolha recairá numa metodologia com elevada sensibilidade, mesmo que isso implique menor eficiência. Por exemplo, ao testar uma doença infecciosa de baixa prevalência e baixo risco, o foco é minimizar o número de testes individuais realizados para economizar recursos e tempo. Assim, a eficiência é fundamental, pois a maior parte dos grupos de teste será negativa. Mas, se o cenário for o diagnóstico de uma doença grave, por exemplo a triagem de doenças genéticas raras em recém-nascidos, será melhor ser utilizada uma metodologia com maior sensibilidade, pois pretende-se que todos os indivíduos doentes sejam bem classificados, evitando a existência de falsos negativos.

## 6. Conclusão

Neste trabalho foi realizada uma análise crítica sobre a eficiência e eficácia de quatro diferentes metodologias de classificação baseadas em testes compostos, comparando uma abordagem tradicional de testes compostos (metodologia de Dorfman) com três algoritmos baseados em metodologias desenvolvidas posteriormente. Foi realizada uma comparação dessas metodologias sob variados cenários, considerando métricas críticas como o custo relativo, a sensibilidade e a especificidade de cada abordagem. O objetivo primordial foi identificar a melhor metodologia para a classificação de indivíduos quanto à presença de uma doença, de acordo com as métricas indicadas, com a análise a ser realizada recorrendo a simulações estatísticas efetuadas em linguagem R.

As simulações permitiram uma análise detalhada da performance de cada metodologia, facilitando a escolha da dimensão ótima do(s) grupo(s) que minimiza o custo de classificação, enquanto mantém a fiabilidade dos resultados dentro de parâmetros aceitáveis. A investigação destacou que, embora as metodologias de classificação baseadas em testes compostos ofereçam vantagens substanciais em termos de eficiência de custos em cenários de baixa prevalência, a escolha da dimensão do(s) grupo(s) deve ser cuidadosamente realizada para evitar uma deterioração significativa na sensibilidade, o que poderia comprometer a fiabilidade do processo de diagnóstico.

Os resultados mostram que não existe uma "melhor" metodologia universal; a escolha da metodologia apropriada depende intrinsecamente do cenário na qual serão utilizadas, sendo que as principais métricas a ter em conta são a sensibilidade e a eficiência, uma vez que são as mais relevantes. Notavelmente, mesmo com variações nestes parâmetros, as simulações realizadas mostraram que as metodologias de classificação baseadas em testes compostos tendem a manter uma performance razoavelmente estável, sugerindo que podem ser adaptadas de maneira flexível a diferentes contextos epidemiológicos sem perda significativa de eficácia.

Os resultados, consistentes através de múltiplas simulações caracterizadas por reduzida variabilidade (baixos desvios padrões), reforçam a robustez das conclusões e a fiabilidade dos dados gerados para suportar decisões clínicas e políticas de saúde.

Por fim, o presente trabalho pretende contribuir para a melhoria do diagnóstico médico, ao mostrar que é possível otimizar a prática diagnóstica com o enfoque no aumento da eficiência, enquanto mantém a integridade dos resultados. Neste sentido, este estudo não só sublinha a importância da inovação contínua em metodologias de testes de diagnóstico, como também ressalva a necessidade de uma avaliação rigorosa e contínua dessas inovações, para garantir que a implementação em contextos reais seja tanto eficaz quanto eficiente.

Deste modo, consideramos que esta investigação contribui para um conhecimento mais profundo na área dos testes diagnósticos e oferece uma base que pode ser útil para futuras pesquisas, com o objetivo de ampliar a aplicação das metodologias de classificação baseadas em testes compostos, especialmente em ambientes com recursos limitados ou onde a rapidez e a eficiência são essenciais para a saúde pública.

## Bibliografia

- [1] M. S. Pepe, H. Janes, G. Longton, W. Leisenring e P. Newcomb, “Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker,” *American Journal of Epidemiology*, pp. 882-890, 2004.
- [2] C. M. Florkowski, “Sensitivity, Specificity, Receiver-Operating Characteristic (ROC) Curves and Likelihood Ratios: Communicating the Performance of Diagnostic Tests,” *The Clinical Biochemist Reviews*, pp. 83-87, 2008.
- [3] N. Grobe, A. Cherif, X. Wang, Z. Dong e P. Kotanko, “Sample pooling: Burden or Solution?,” *Clinical Microbiology and Infection*, vol. 27, pp. 1212 - 1220, 2021.
- [4] R. Santos, J. P. Martins e M. Felgueiras, “An Overview of Quantitative Continuous,” pp. 627-641, 2015.
- [5] F. K. Hwang, “Group testing with a dilution effect,” *Biometrika*, vol. 63, pp. 671-673, 1976.
- [6] J. P. Martins, R. Santos e R. Sousa, “Testing the Maximum by the Mean in Quantitative Group Tests,” *Studies in Theoretical and Applied Statistics, Selected Papers of the Statistical Societies*, pp. 55-63, 2014.
- [7] R. Santos, J. P. Martins e M. Felgueiras, “Known Mean, Unkonwn Maxima? Testing the Maximum Knowing only the Mean,” *Communications in Statistics: Simulation and Computation*, vol. 44, pp. 2473-2491, 2015.
- [8] M. Felgueiras, J. P. Martins e R. Santos, “Distributions families in counting bacteria for compound sampling,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8581 LNCS (PART 3), pp. 539-551, 2014.

- [9] R. D. Banker, A. Charnes e W. W. Cooper, “Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis,” *Management Science*, vol. 30, pp. 1078-1092, 1984.
- [10] K. J. Van Stralen, V. S. Stel, J. B. Reitsma, F. W. Dekker, C. Zoccali e K. J. Jager, “Diagnostic methods I: sensitivity, specificity, and other measures of accuracy.,” *Kidney International*, vol. 75, pp. 1257-1263, 2009.
- [11] A. G. Lalkhen e A. McCluskey, “Clinical tests: sensitivity and specificity,” *Continuing Education in Anaesthesia Critical Care & Pain*, vol. 8, nº 6, pp. 221-223, 2008.
- [12] R. Trevethan, “Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice,” *Frontiers in Public Health*, vol. 5, p. 307, 2017.
- [13] R. Santos, M. Felgueiras, J. P. Martins e L. Ferreira, “Accuracy Measures for Binary Classification Based on a Quantitive Variable,” *REVSTAT - Statistical Journal*, vol. 17, pp. 223-244, 2019.
- [14] A. K. Akobeng, “Understanding diagnostic tests 1: sensitivity, specificity and predictive values,” *Acta Pædiatrica*, vol. 94, pp. 1418-1423, 2005.
- [15] E. J. Gallagher, “The Problem With Sensitivity and Specificity...,” *Annals of Emergency Medicine*, vol. 42, pp. 298-303, 2003.
- [16] K. G. M. Moons e F. E. Harrell, “Sensitivity and specificity should be de-emphasized in diagnostic accuracy studies,” *Academic Radiology*, vol. 10, pp. 670-672, 2003.
- [17] D. L. Sackett e R. B. Haynes, “The architecture of diagnostic research,” *British Medical Journal*, vol. 324, pp. 539-541, 2002.
- [18] R. C. Team, “R: A language and environment for statistical computing,” *R Project for Statistical Computing*, 2024.
- [19] M. Khun, J. Wing, S. Weston, A. Williams, C. Keefer, A. Engelhardt, T. Cooper, Z. Mayer, B. Kenkel, R. C. Team, M. Benesty, R. Lescarbeau, A. Ziem, L. Scrucca, Y.

- Tang, C. Candan e T. Hunt, “Package ‘caret’: Classification and Regression Training, R package version 6.0-94,” 2023.
- [20] H. Wickham, W. Chang, L. Henry, T. L. Pedersen, K. Takahashi, C. Wilke, K. Woo, H. Yutani, D. Dunnington, T. v. d. Brand e P. PBC, “Package ‘ggplot2’: Create Elegant Data Visualisations Using the Grammar of Graphics, R package version 3.5.1,” 2024.
- [21] S. G. Kwak e J. H. Kim, “Central limit theorem: The cornerstone of modern statistics,” *Korean Journal of Anesthesiology*, vol. 70(2), pp. 144-156, 2017.
- [22] S. Pedro, R. Santos e L. Cotrim, “Simulação - uma aplicação ao problema da ruína do jogador,” *Boletim da Sociedade Portuguesa de Matemática*, vol. 70, pp. 85-103, 2014.
- [23] R. Dorfman, “The Detection of Defective Members of Large Populations,” *Annals of Mathematical Statistics*, vol. 14, pp. 436-440, 1943.
- [24] M. Sherlock, N. M. Zetola e J. D. Klausner, “Routine Detection of Acute HIV Infection Through RNA Pooling: Survey of Current Practice in the United States,” *Sexually Transmitted Diseases*, vol. 34, pp. 314-316, 2007.
- [25] R. Santos, D. Pestana e J. P. Martins, “Extensions of Dorfman’s Theory,” *Studies in Theoretical and Applied Statistics, Selected Papers of the Statistical Societies*, pp. 179-189, 2013.
- [26] M. S. Black, C. R. Bilder e J. M. Tebbs, “Optimal Retesting Configurations for Hierarchical Group Testing,” *Journal of the Royal Statistical Society Series C: Applied Statistics*, vol. 64, n° 4, p. 693–710, 2015.
- [27] A. Granado, “Análises clínicas compostas: Um estudo crítico via simulação [Dissertação de Mestrado],” *Politécnico de Leiria*, 2014.
- [28] P. Hou, J. M. Tebbs, C. R. Bilder e C. S. McMahan, “Hierarchical group testing for multiple infections,” *Biometrics*, vol. 73(2), pp. 656-665, 2017.

- [29] R. Santos, J. P. Martins e M. Felgueiras, “Discrete Compound Tests and Dorfman's Methodology in the Presence of Misclassification,” *Springer Proceedings in Mathematics and Statistics*, vol. 136, pp. 85-98, 2015.
- [30] D. Dhamnetiya, R. P. Jha, S. e K. Bhattacharyya, “How to Analyze the Diagnostic Performance of a New Test? Explained with Illustrations,” *Journal of Laboratory Physicians*, vol. 14, pp. 90-98, 2022.
- [31] “Introdução à Probabilidade e à Estatística,” *Fundação Calouste Gulbenkian*, vol. 4<sup>a</sup> ed, 2010.
- [32] J. P. Martins, M. Felgueiras e R. Santos, “Three-dimensional array-based group testing algorithms with on-stage,” *AIP Conference Proceedings*, vol. 1702, p. 030004, 2015.

## Anexos

### Anexo 1: Total de testes para $p = 0.025$ utilizando a metodologia hierárquica e de Dorfman quando $SI = EI$

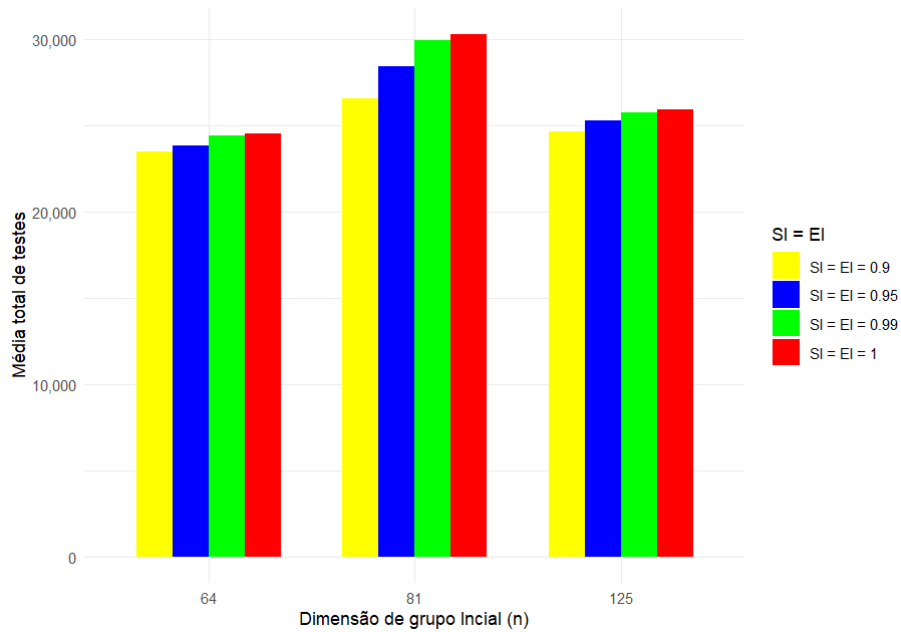


Figura 10 - Gráfico da média total de testes usando a metodologia hierárquica, para uma taxa de prevalência de 2.5% quando  $SI = EI$

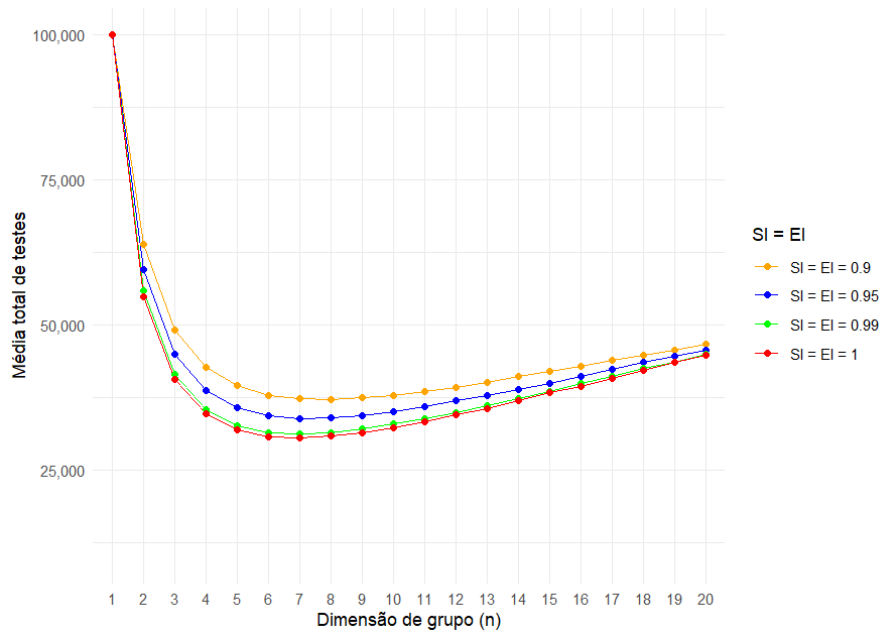
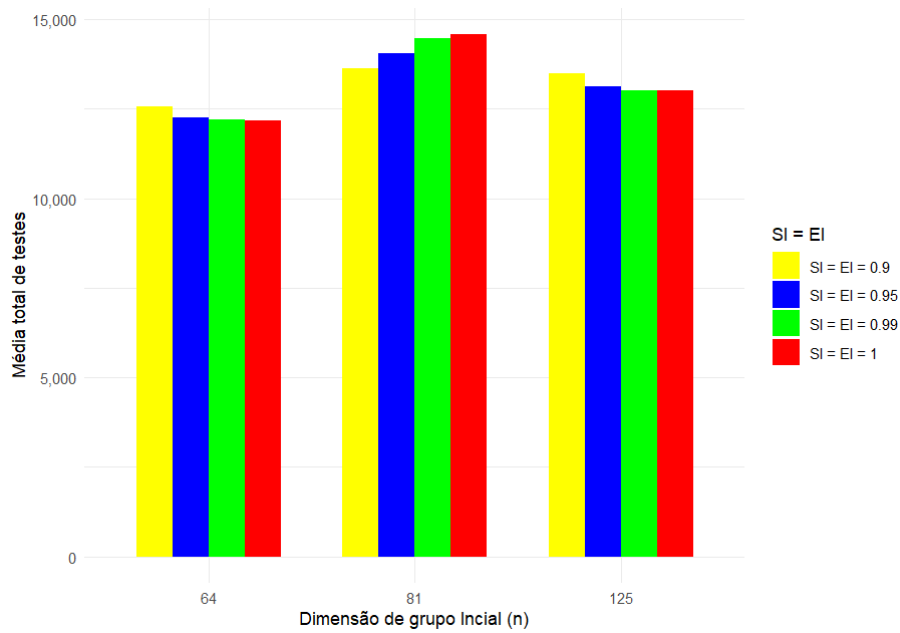
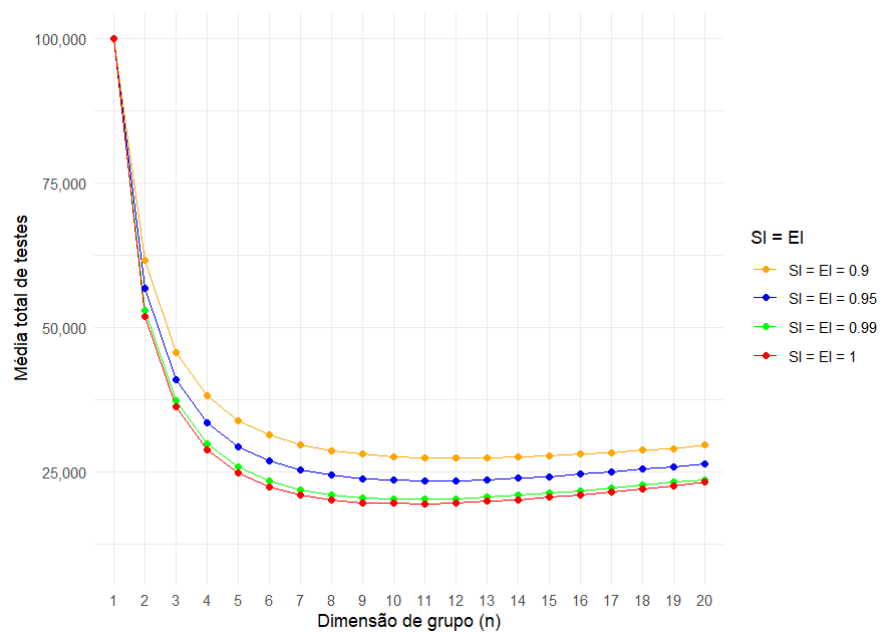


Figura 11 - Gráfico da média total de testes usando a metodologia de Dorfman, para uma taxa de prevalência de 2.5% quando  $SI = EI$

**Anexo 2: Total de testes para  $p = 0.01$  utilizando a metodologia hierárquica e de Dorfman quando  $SI = EI$**

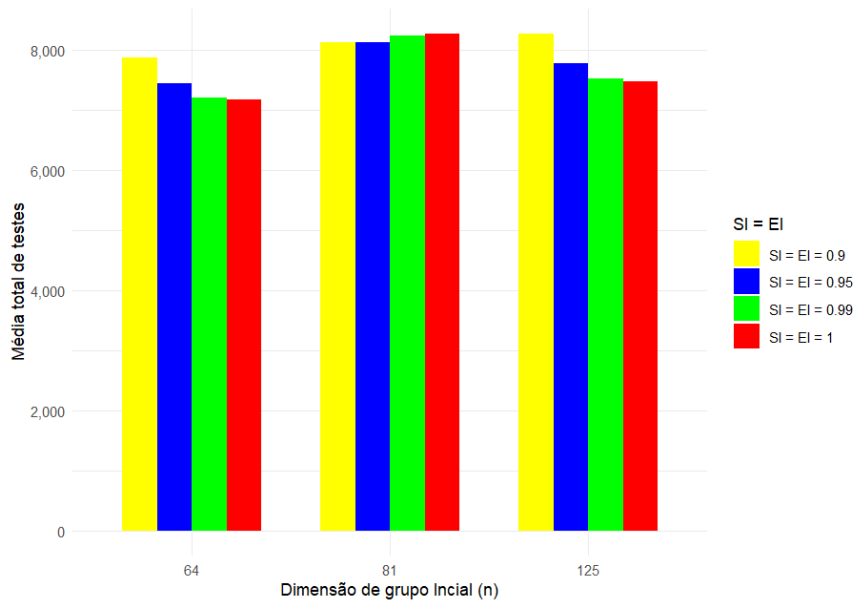


**Figura 12 - Gráfico da média total de testes usando a metodologia hierárquica, para uma taxa de prevalência de 1% quando  $SI = EI$**

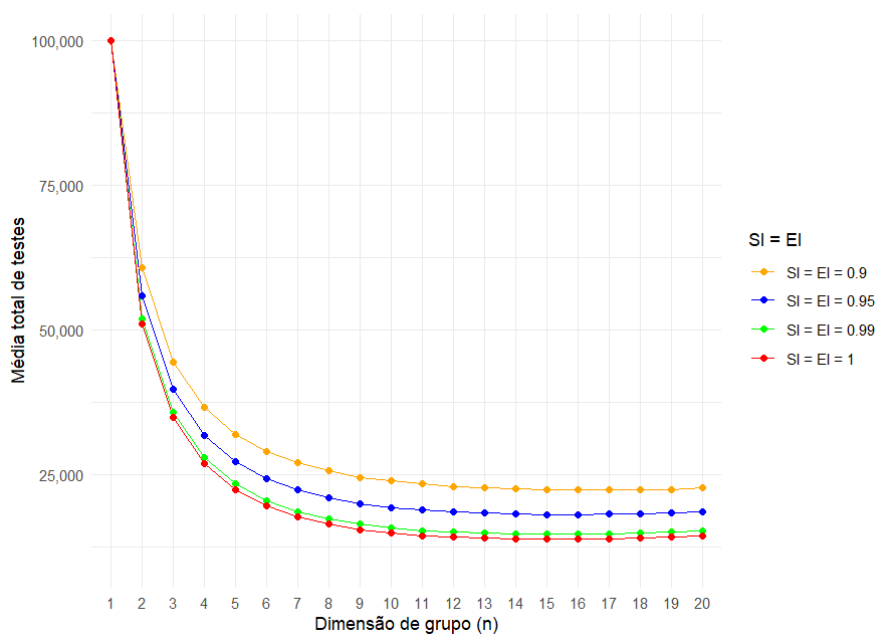


**Figura 13 - Gráfico da média total de testes usando a metodologia de Dorfman, para uma taxa de prevalência de 1% quando  $SI = EI$**

**Anexo 3: Total de testes para  $p = 0.005$  utilizando a metodologia hierárquica e de Dorfman quando  $SI = EI$**



**Figura 14 - Gráfico da média total de testes usando a metodologia hierárquica, para uma taxa de prevalência de 0.5% quando  $SI = EI$**



**Figura 15 - Gráfico da média total de testes usando a metodologia de Dorfman, para uma taxa de prevalência de 0.5% quando  $SI = EI$**

### Anexo 4: Acurácia da metodologia hierárquica e de Dorfman para $p = 0.025$ quando $SI = EI$

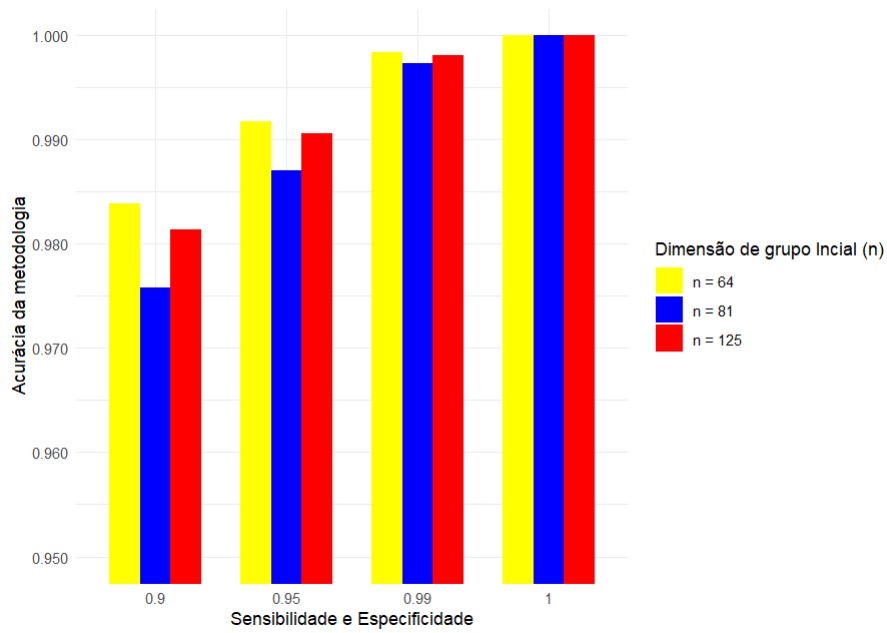


Figura 16 - Gráfico da média da acurácia da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 2.5% quando  $SI = EI$

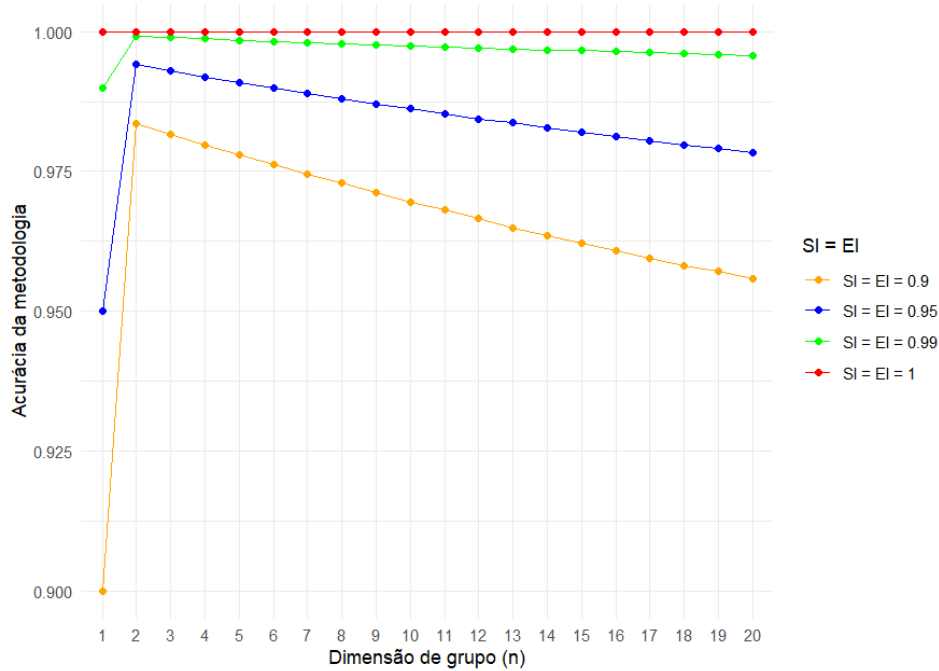
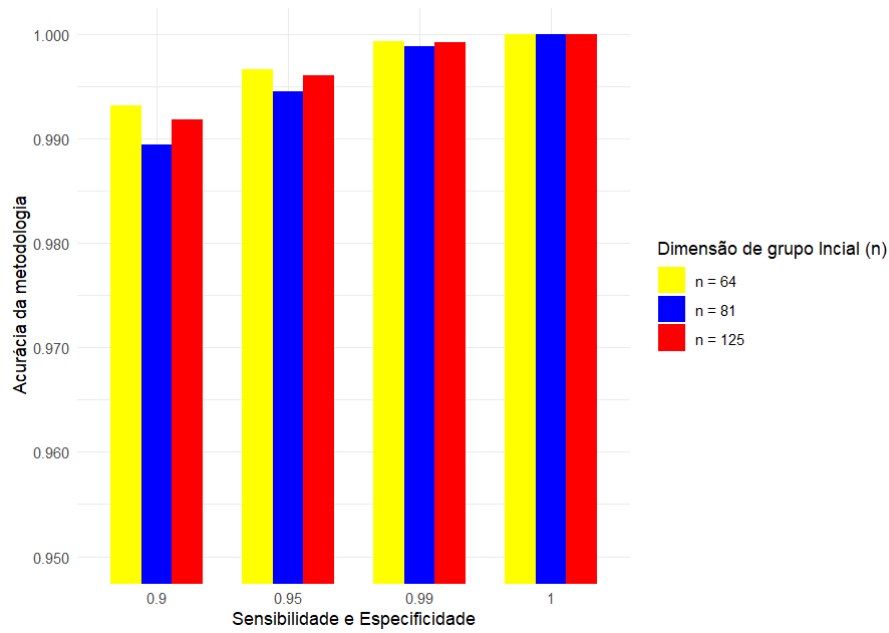
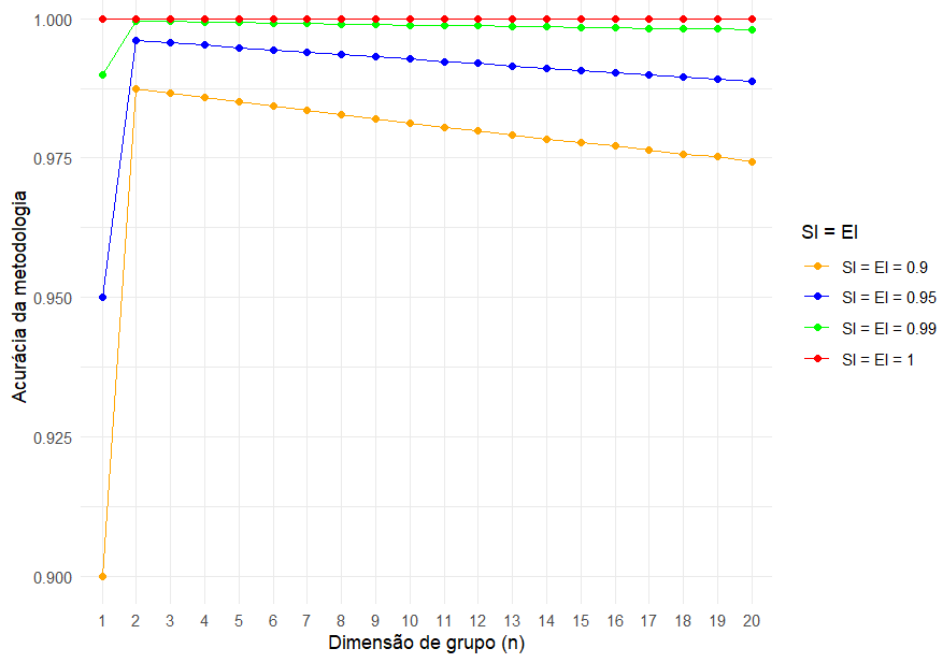


Figura 17 - Gráfico da média da acurácia da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 2.5% quando  $SI = EI$

**Anexo 5: Acurácia da metodologia hierárquica e de Dorfman para  $p = 0.01$  quando  $SI = EI$**

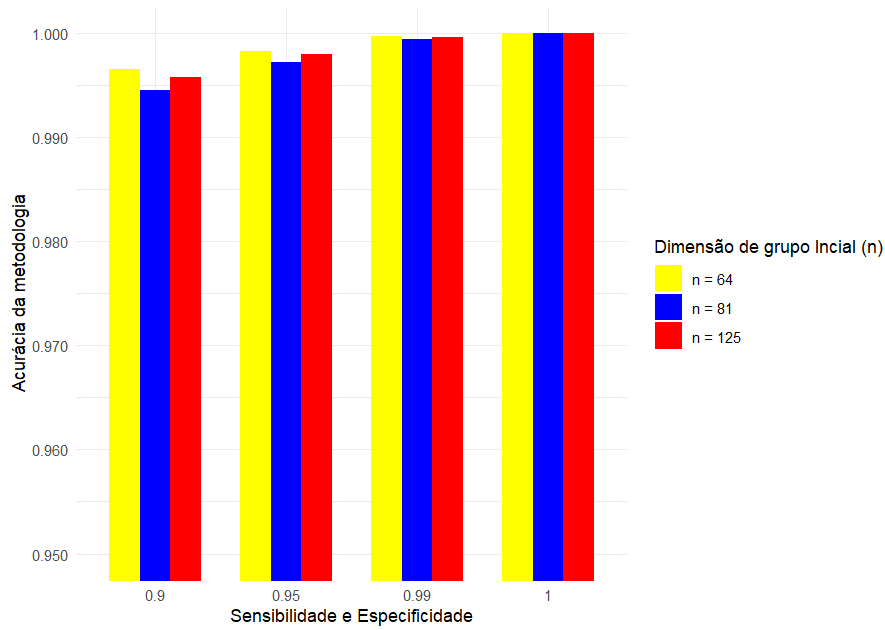


**Figura 18 - Gráfico da média da acurácia da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 1% quando  $SI = EI$**

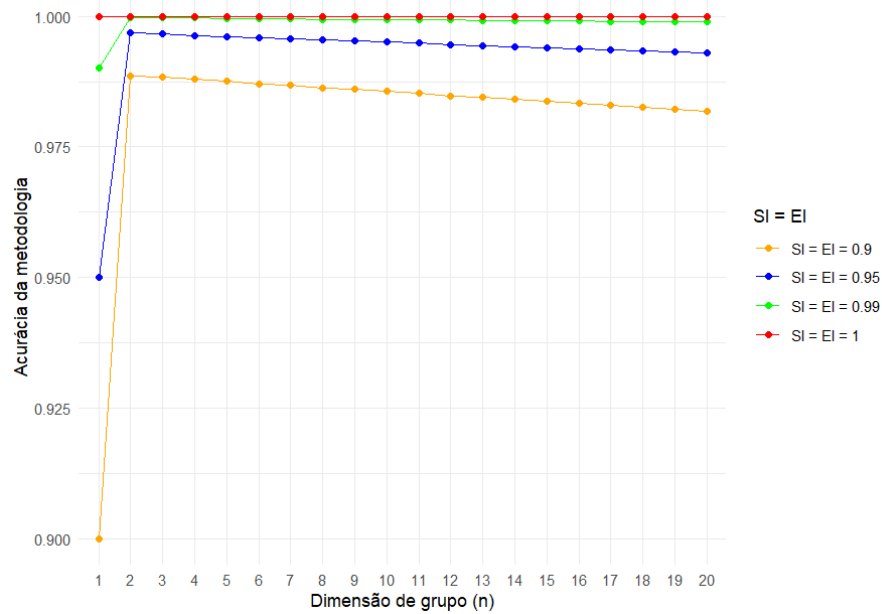


**Figura 19 - Gráfico da média da acurácia da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 1% quando  $SI = EI$**

**Anexo 6: Acurácia da metodologia hierárquica e de Dorfman para  $p = 0.005$  quando  $SI = EI$**

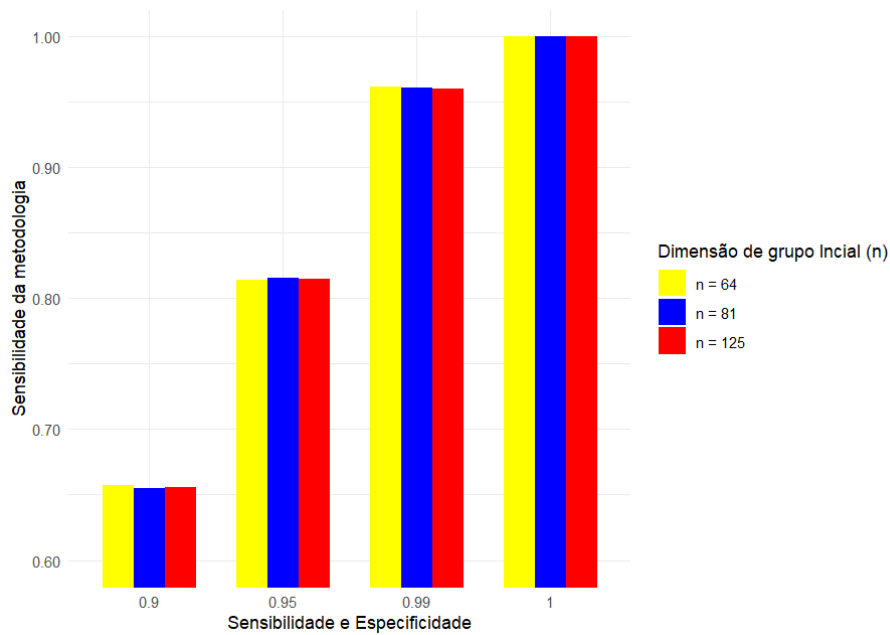


**Figura 20 - Gráfico da média da acurácia da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 0.5% quando  $SI = EI$**

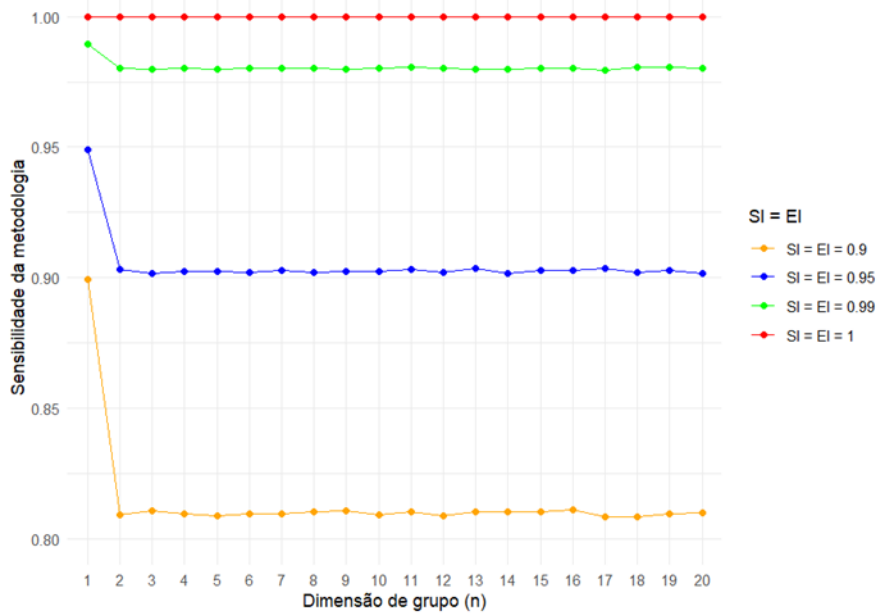


**Figura 21 - Gráfico da média da acurácia da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 0.5% quando  $SI = EI$**

**Anexo 7: Sensibilidade da metodologia hierárquica e de Dorfman para  $p = 0.025$  quando  $SI = EI$**



**Figura 22 - Gráfico da média da sensibilidade da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 2.5% quando  $SI = EI$**



**Figura 23 - Gráfico da média da sensibilidade da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 2.5% quando  $SI = EI$**

## Anexo 8: Sensibilidade da metodologia hierárquica e de Dorfman para $p = 0.01$ quando $SI = EI$

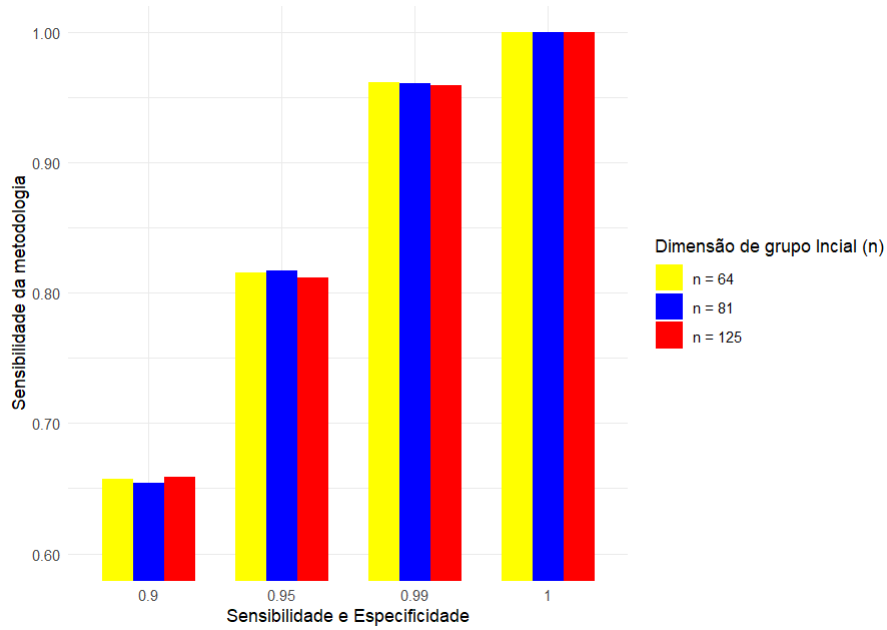


Figura 24 - Gráfico da média da sensibilidade da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 1% quando  $SI = EI$

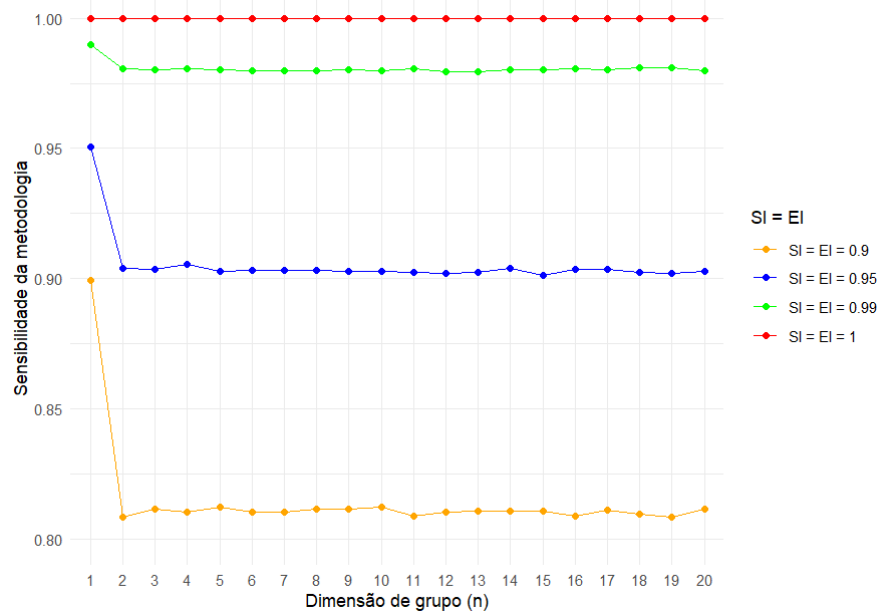
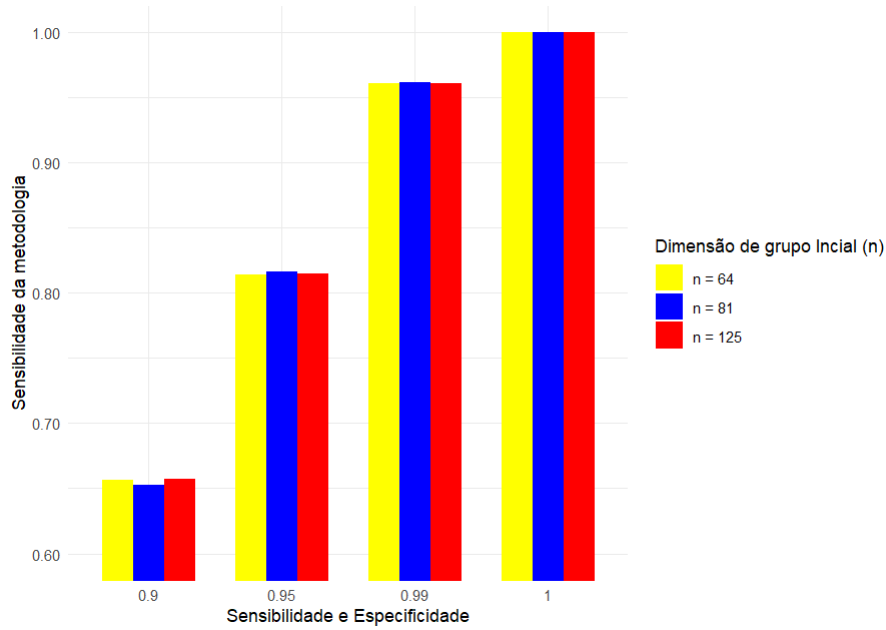
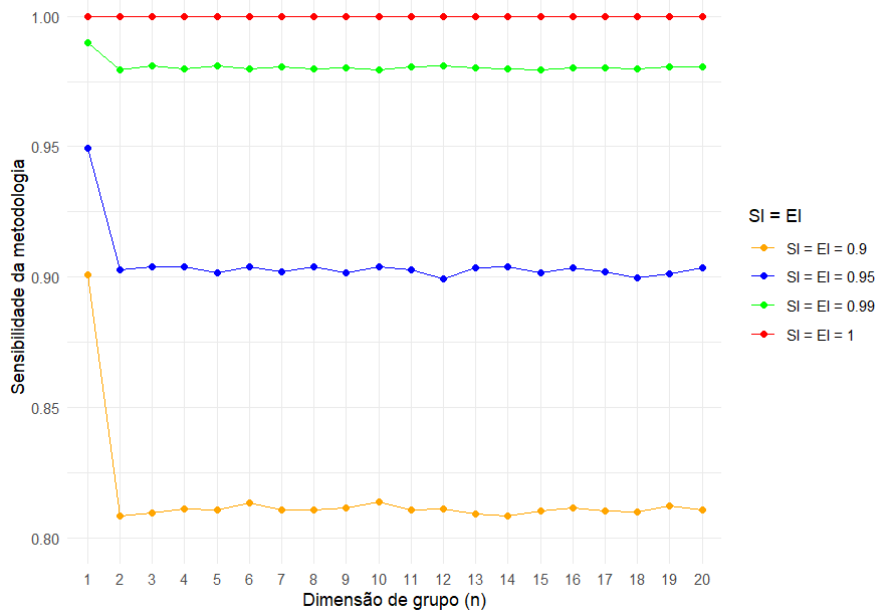


Figura 25 - Gráfico da média da sensibilidade da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 1% quando  $SI = EI$

**Anexo 9: Sensibilidade da metodologia hierárquica e de Dorfman para  $p = 0.005$  quando  $SI = EI$**



**Figura 26 - Gráfico da média da sensibilidade da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 0.5% quando  $SI = EI$**



**Figura 27 - Gráfico da média da sensibilidade da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 0.5% quando  $SI = EI$**

## Anexo 10: Especificidade da metodologia hierárquica e de Dorfman para $p = 0.025$ quando $SI = EI$

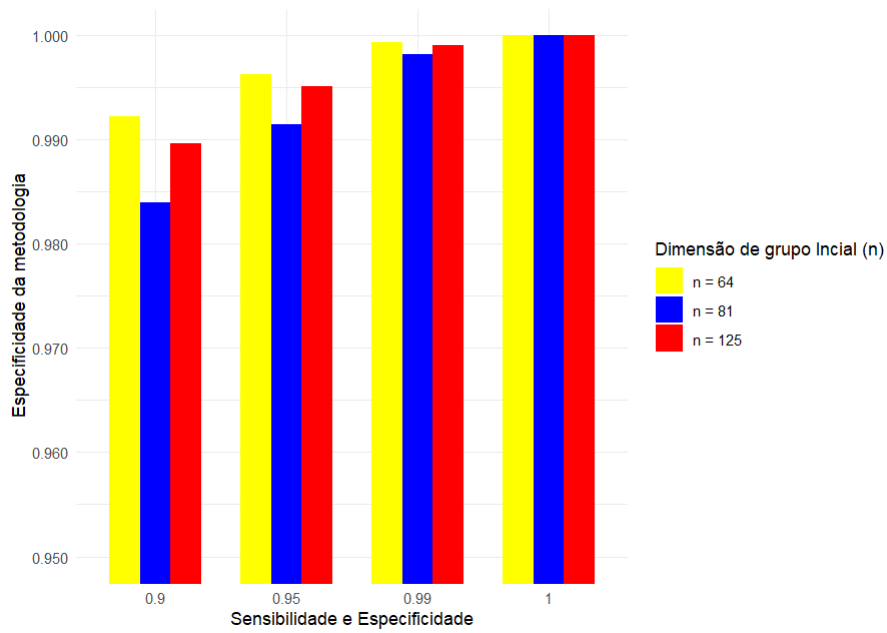


Figura 28 - Gráfico da média da especificidade da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 2.5% quando  $SI = EI$

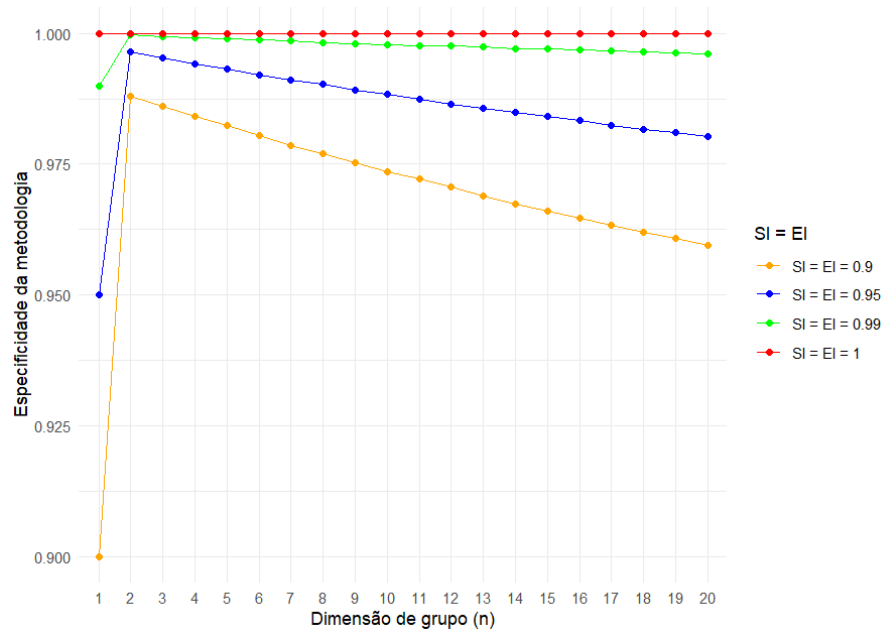
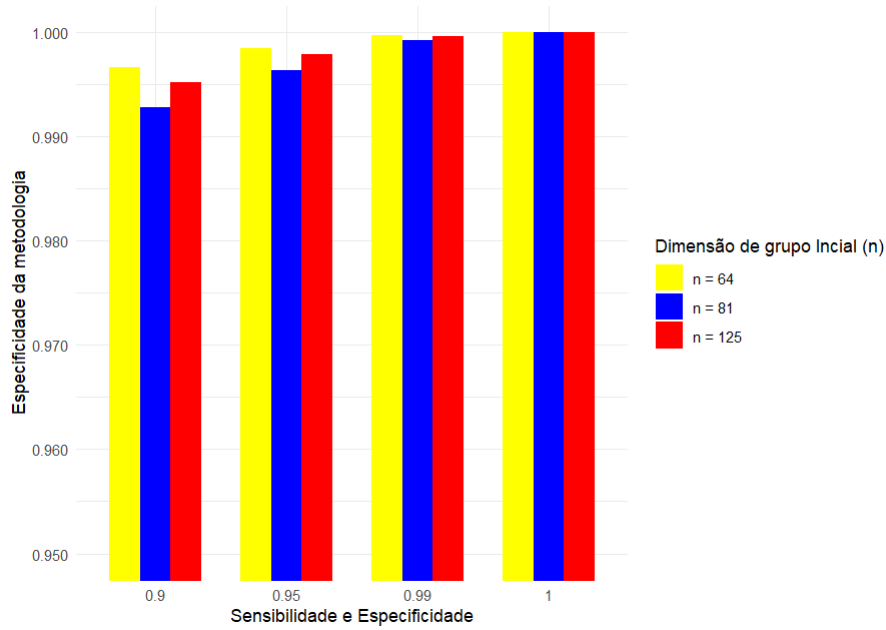
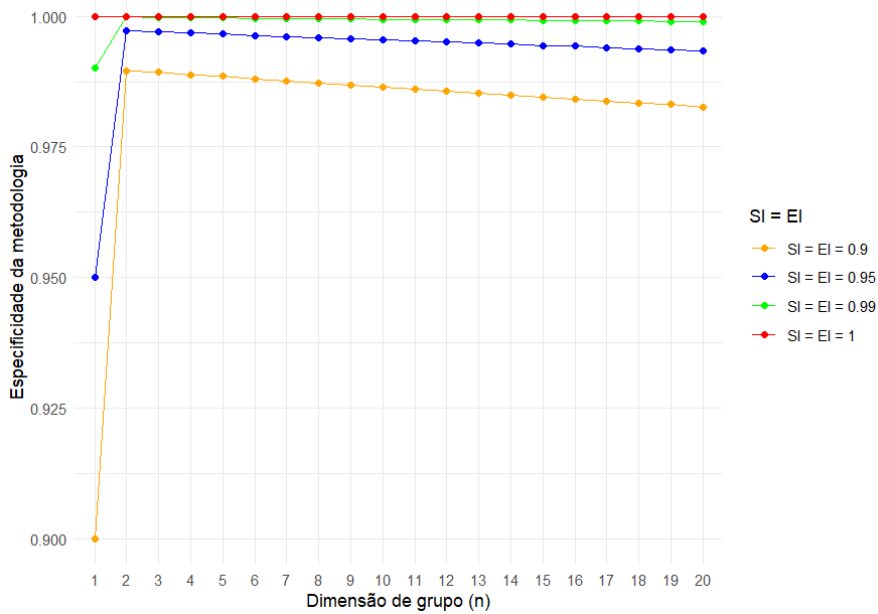


Figura 29 - Gráfico da média da especificidade da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 2.5% quando  $SI = EI$

**Anexo 11: Especificidade da metodologia hierárquica e de Dorfman para  $p = 0.01$  quando  $SI = EI$**



**Figura 30 - Gráfico da média da especificidade da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 1% quando  $SI = EI$**



**Figura 31 - Gráfico da média da especificidade da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 1 % quando  $SI = EI$**

## Anexo 12: Especificidade da metodologia hierárquica e de Dorfman para $p = 0.005$ quando $SI = EI$

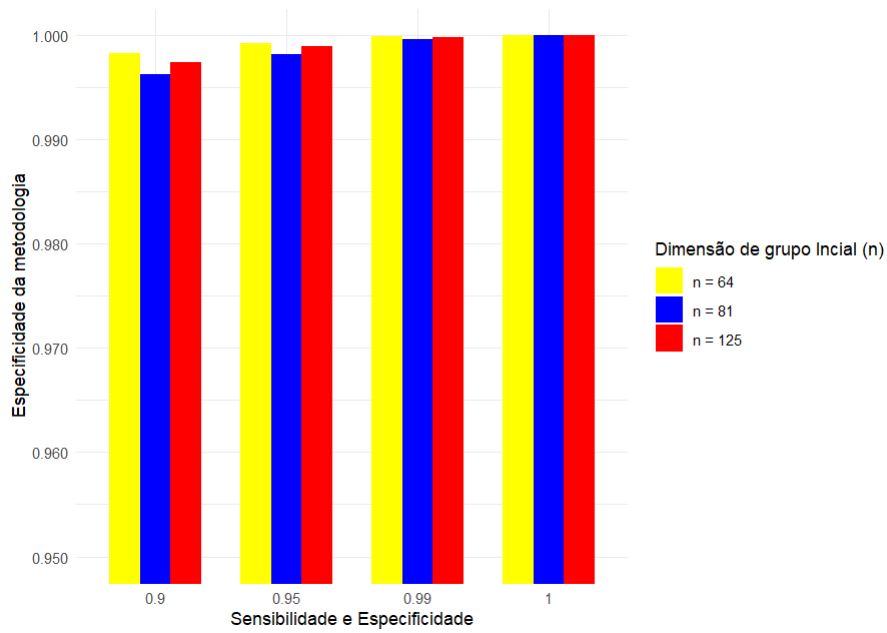


Figura 32 - Gráfico da média da especificidade da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 0.5% quando  $SI = EI$

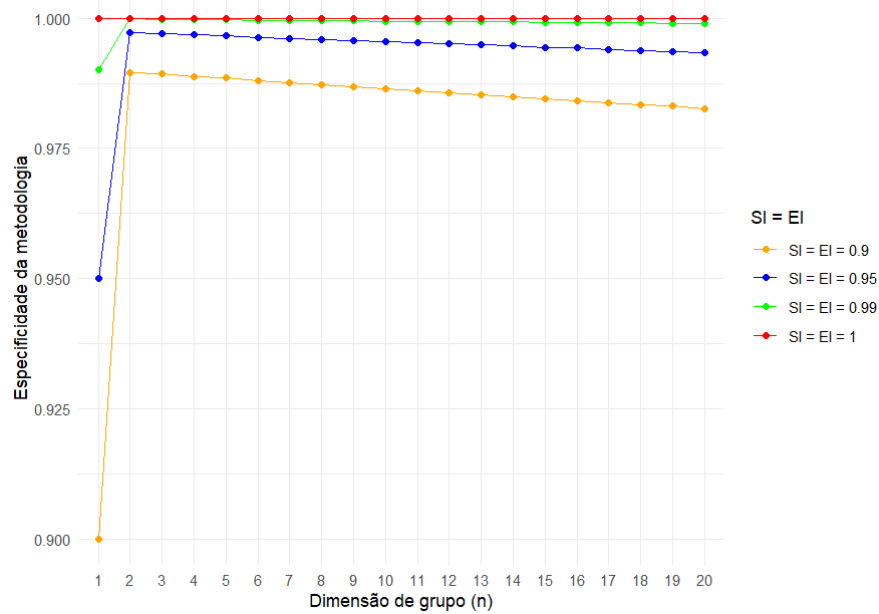
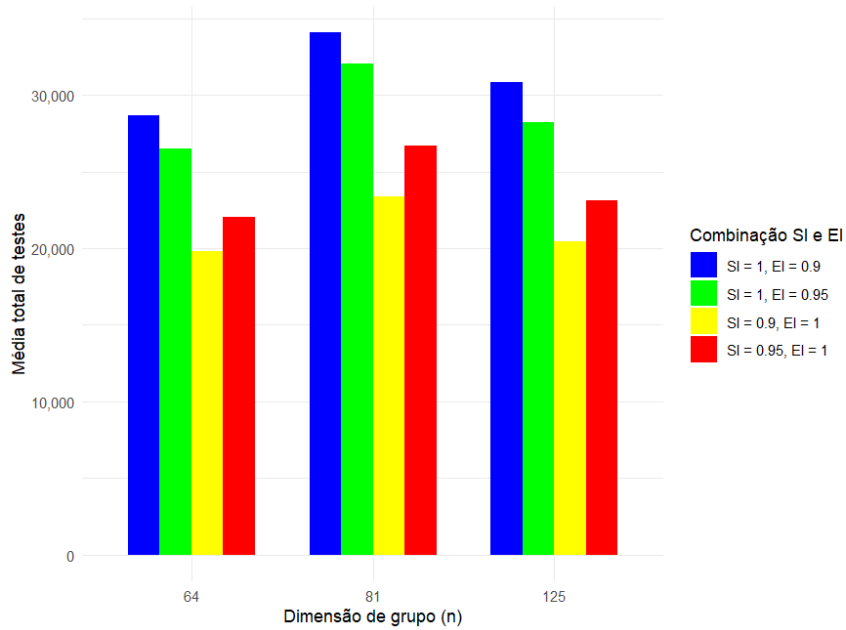
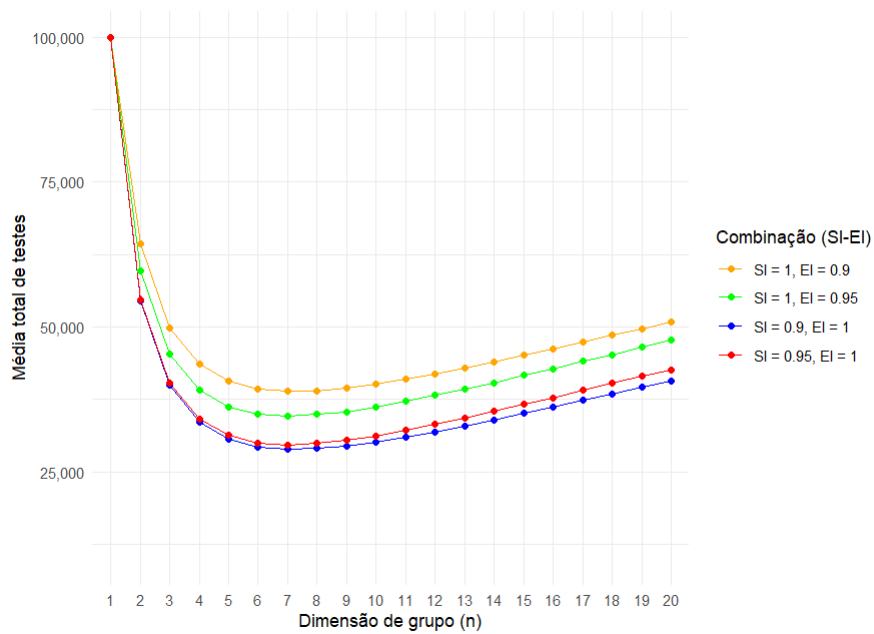


Figura 33 - Gráfico da média da especificidade da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 0.5 % quando  $SI = EI$

**Anexo 13: Total de testes para  $p = 0.025$  utilizando a metodologia hierárquica e de Dorfman quando SI e EI são distintos**



**Figura 34 - Gráfico da média total de testes usando a metodologia hierárquica, para uma taxa de prevalência de 2.5% quando SI e EI são distintas**



**Figura 35 - Gráfico da média total de testes usando a metodologia de Dorfman, para uma taxa de prevalência de 2.5% quando SI e EI são distintas**

### Anexo 14: Total de testes para $p = 0.01$ utilizando a metodologia hierárquica e de Dorfman quando SI e EI são distintos

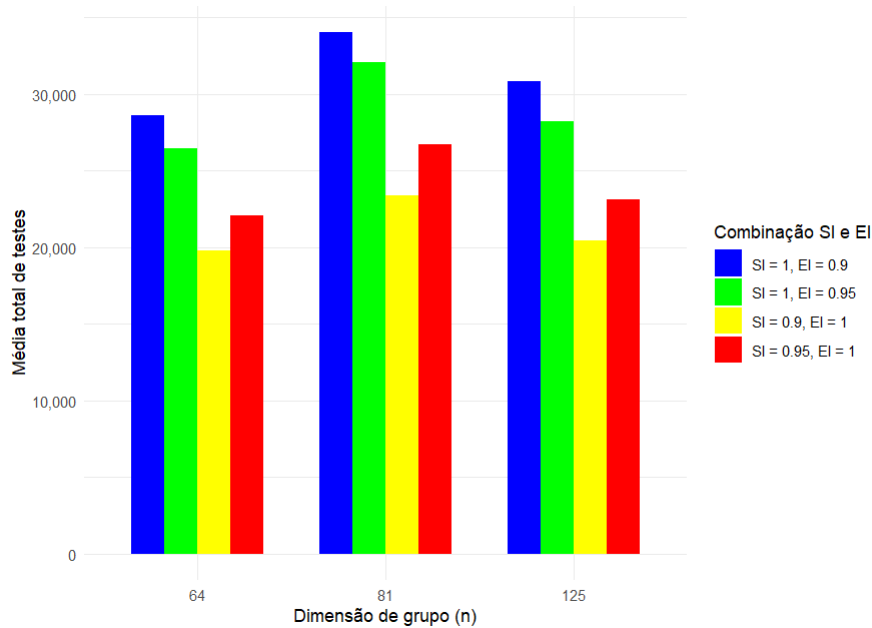


Figura 36 - Gráfico da média total de testes usando a metodologia hierárquica, para uma taxa de prevalência de 1% quando SI e EI são distintas

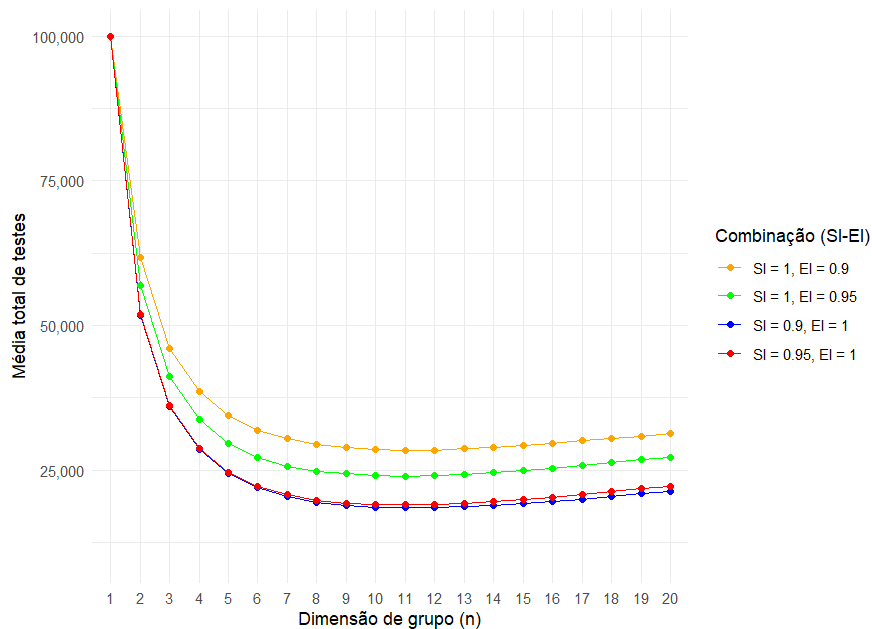
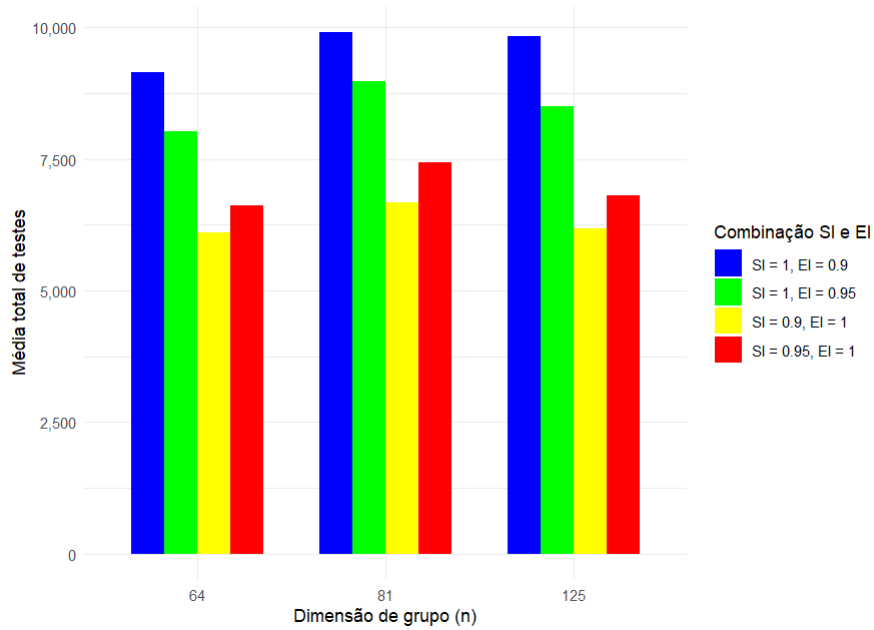
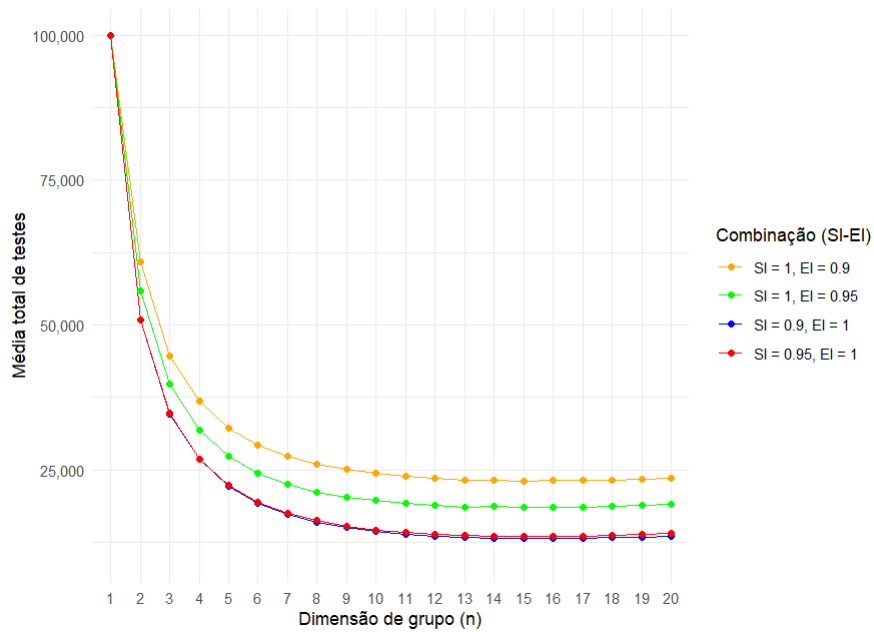


Figura 37 - Gráfico da média total de testes usando a metodologia de Dorfman, para uma taxa de prevalência de 1% quando SI e EI são distintas

**Anexo 15: Total de testes para  $p = 0.005$  utilizando a metodologia hierárquica e de Dorfman quando SI e EI são distintos**



**Figura 38 - Gráfico da média total de testes usando a metodologia hierárquica, para uma taxa de prevalência de 0.5% quando SI e EI são distintas**



**Figura 39 - Gráfico da média total de testes usando a metodologia de Dorfman, para uma taxa de prevalência de 0.5% quando SI e EI são distintas**

## Anexo 16: Acurácia da metodologia hierárquica e de Dorfman para $p = 0.025$ quando SI e EI são distintos

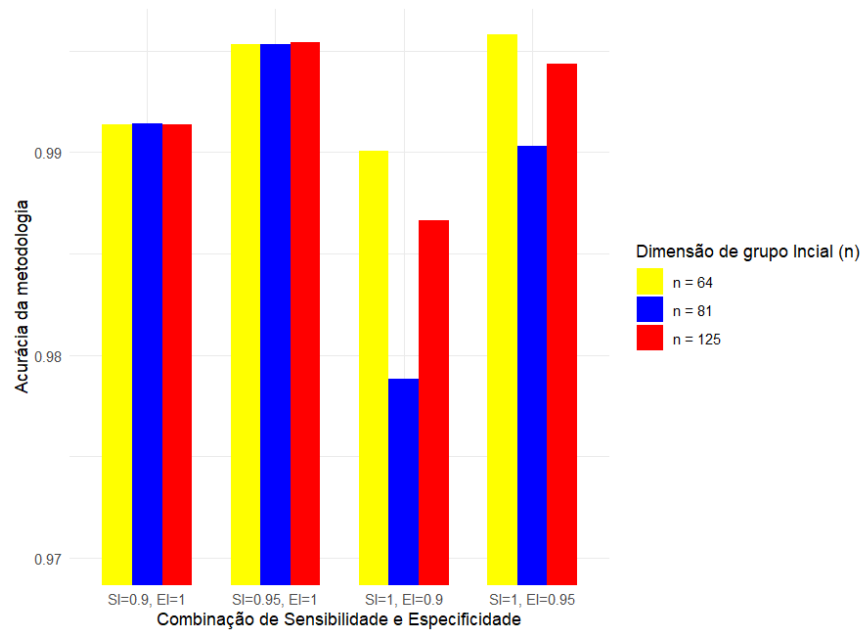


Figura 40 - Gráfico da média da acurácia da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 2.5% quando SI e EI são distintas

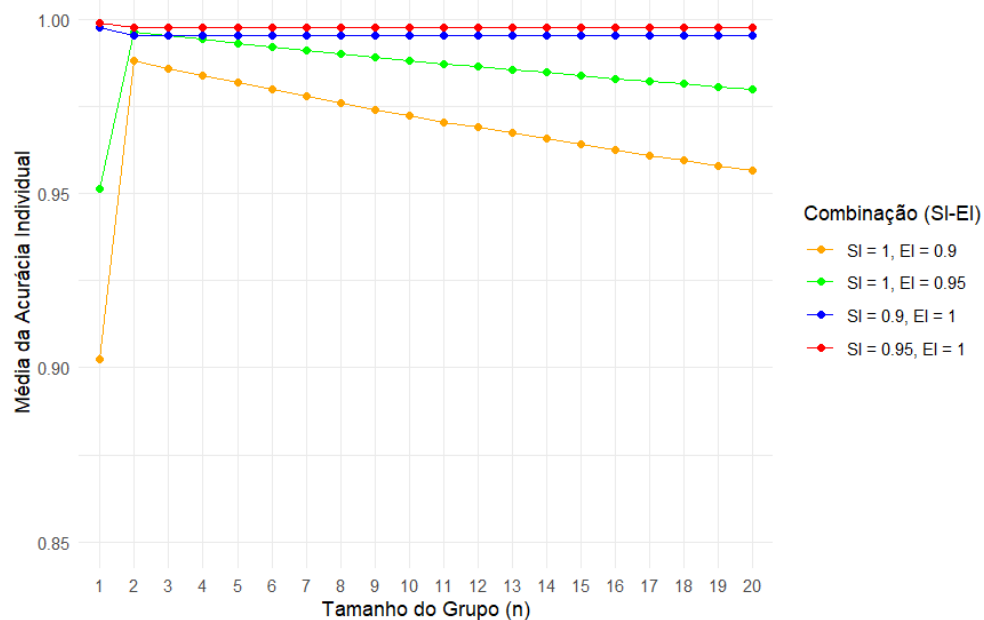


Figura 41 - Gráfico da média da acurácia da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 2.5% quando SI e EI são distintas

## Anexo 17: Acurácia da metodologia hierárquica e de Dorfman para $p = 0.01$ quando SI e EI são distintos

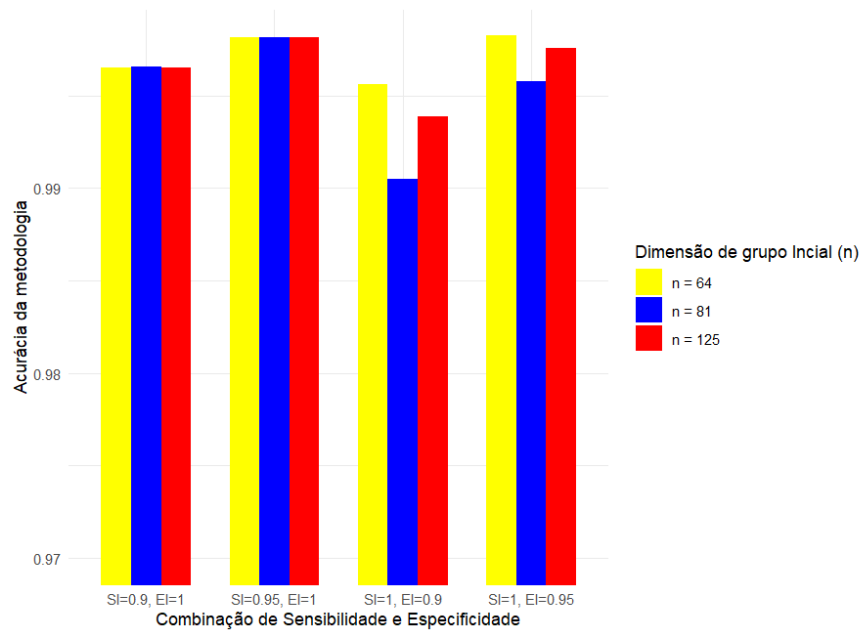


Figura 42 - Gráfico da média da acurácia da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 1% quando SI e EI são distintas

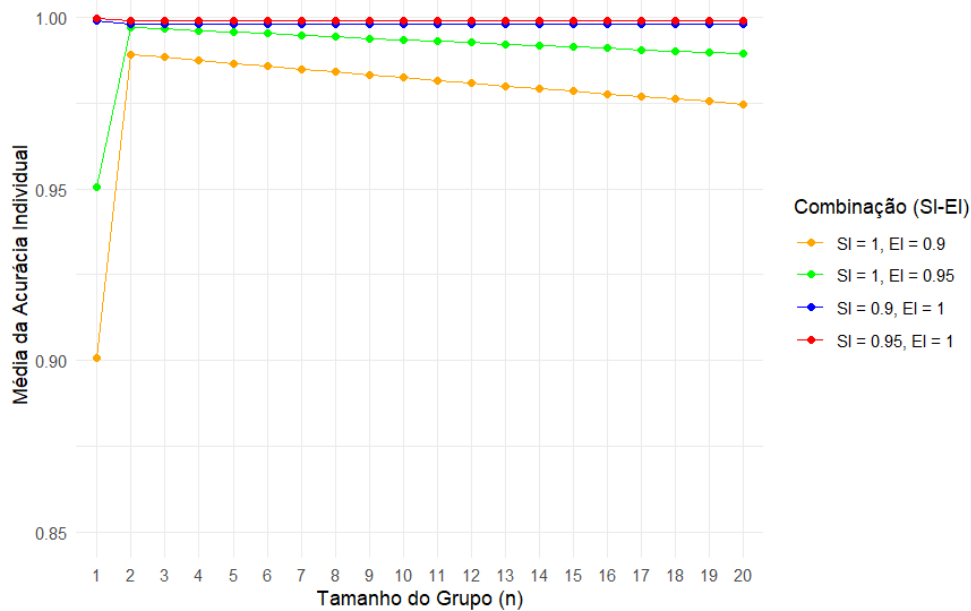


Figura 43 - Gráfico da média da acurácia da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 1% quando SI e EI são distintas

## Anexo 18: Acurácia da metodologia hierárquica e de Dorfman para $p = 0.005$ quando SI e EI são distintos

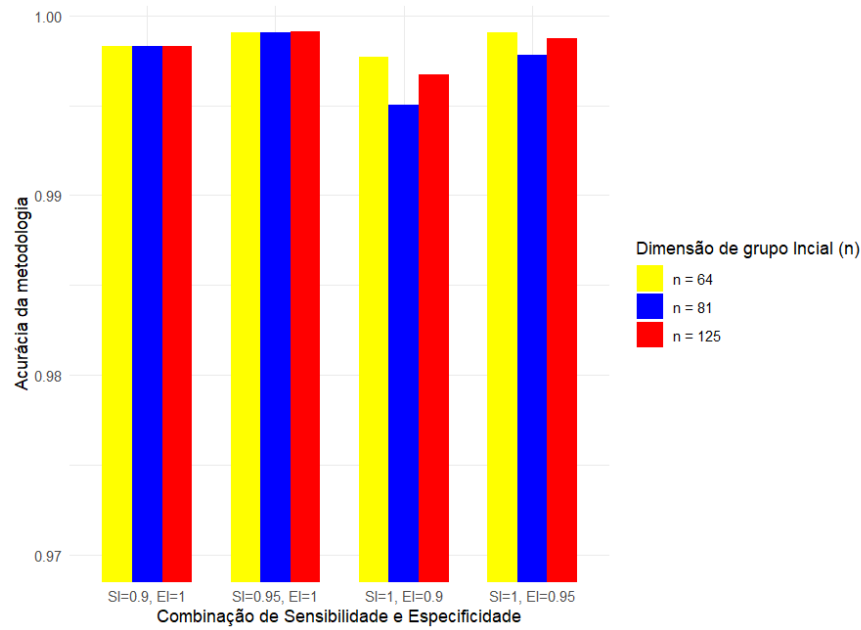


Figura 44 - Gráfico da média da acurácia da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 0.5% quando SI e EI são distintas

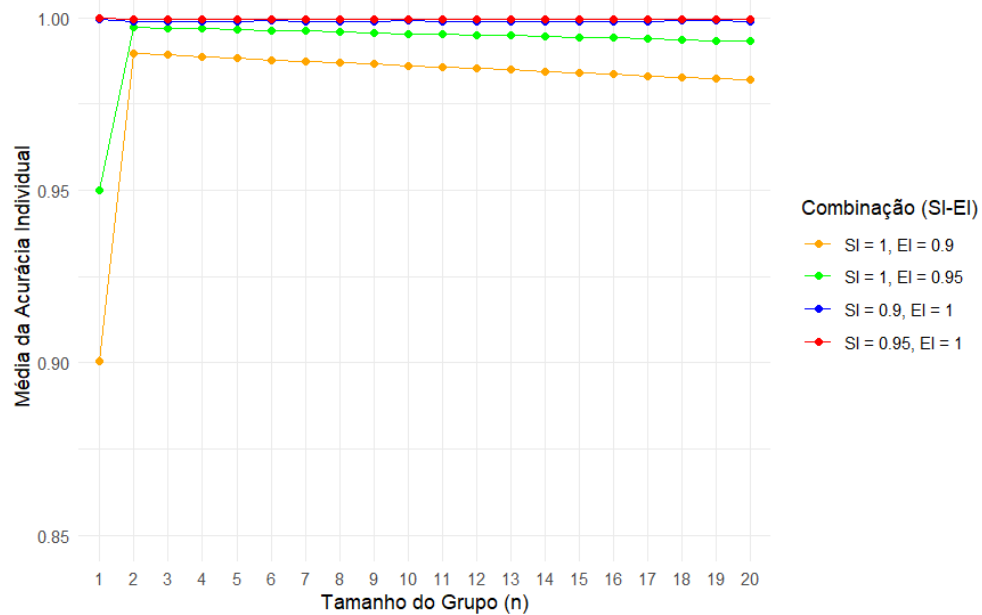


Figura 45 - Gráfico da média da acurácia da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 0.5% quando SI e EI são distintas

## Anexo 19: Sensibilidade da metodologia hierárquica e de Dorfman para $p = 0.025$ quando SI e EI são distintos

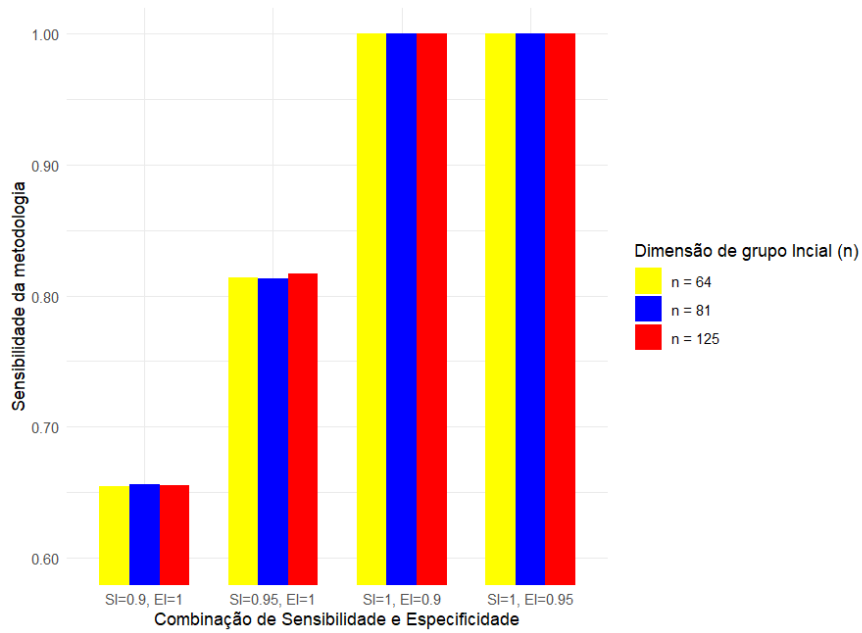


Figura 46 - Gráfico da média da sensibilidade da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 2.5% quando SI e EI são distintas

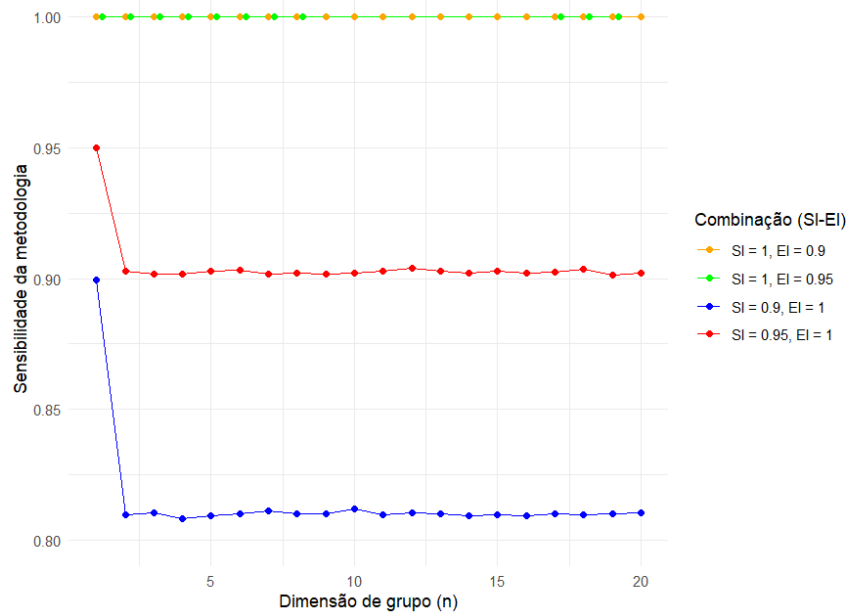
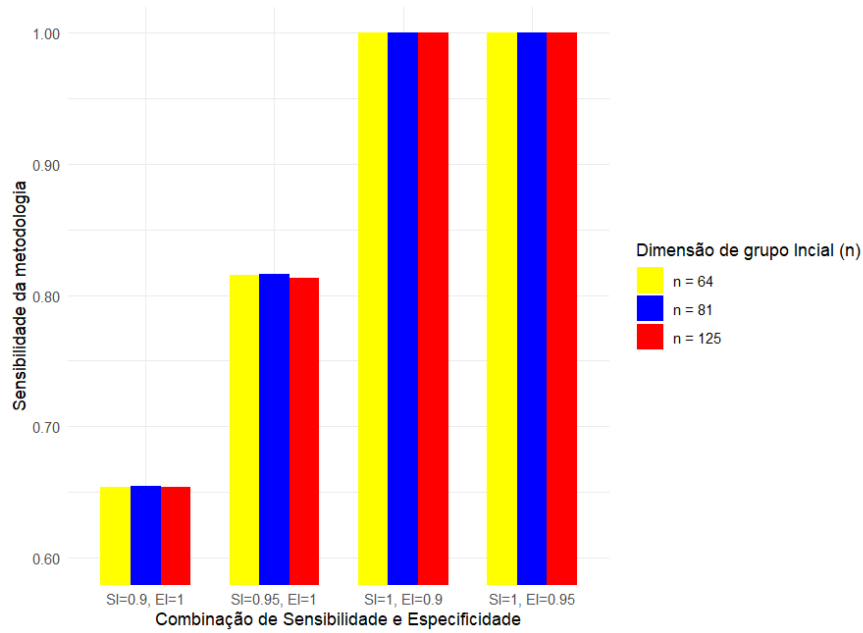
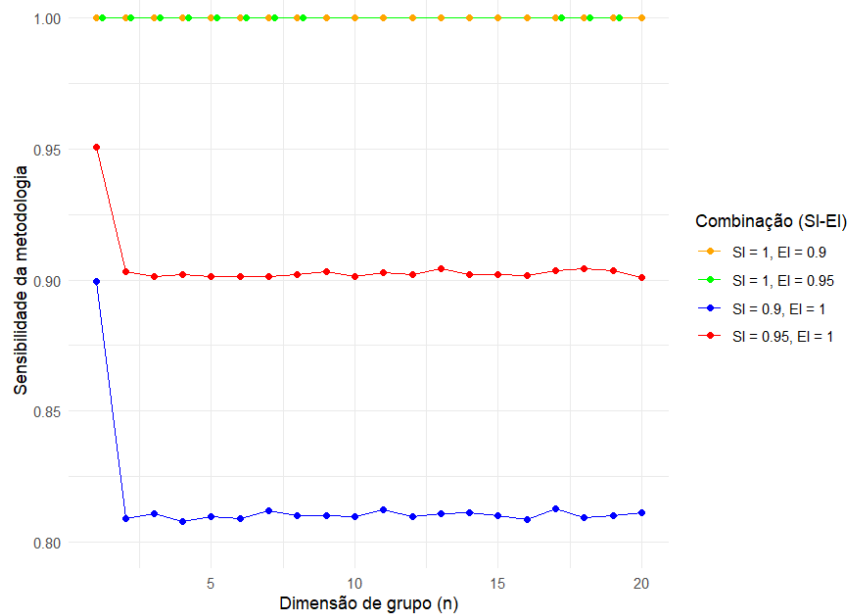


Figura 47 - Gráfico da média da sensibilidade da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 2.5% quando SI e EI são distintas

**Anexo 20: Sensibilidade da metodologia hierárquica e de Dorfman para  $p = 0.01$  quando SI e EI são distintos**

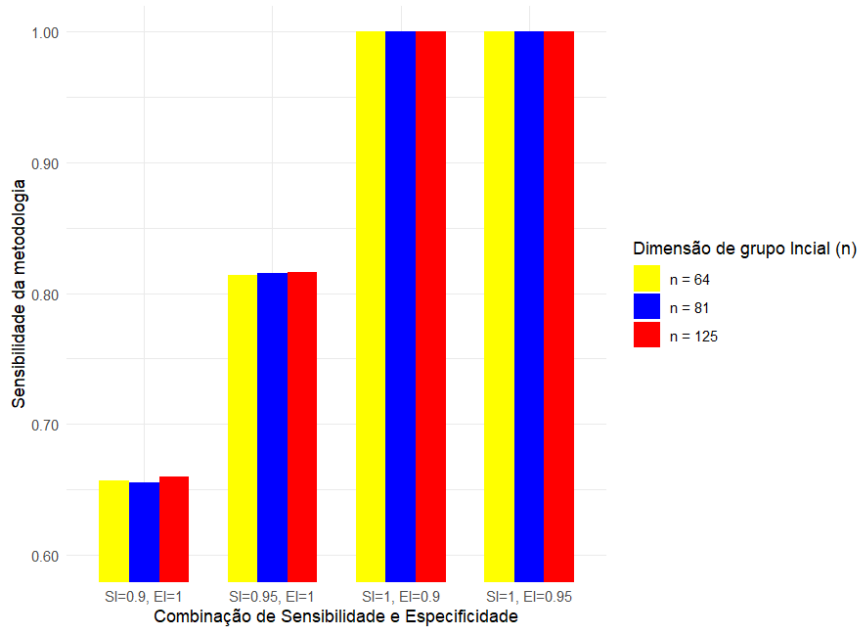


**Figura 48 - Gráfico da média da sensibilidade da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 1% quando SI e EI são distintas**

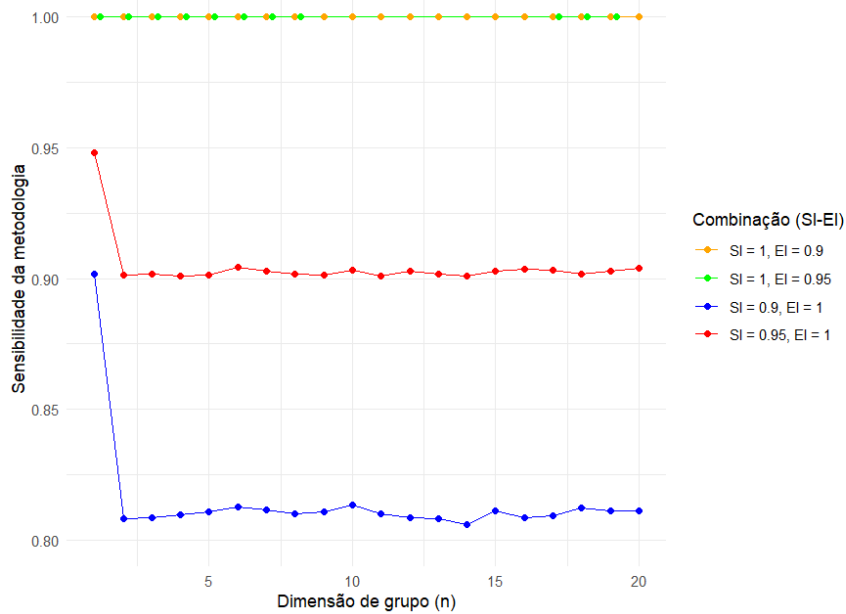


**Figura 49 - Gráfico da média da sensibilidade da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 1% quando SI e EI são distintas**

**Anexo 21: Sensibilidade da metodologia hierárquica e de Dorfman para  $p = 0.005$  quando SI e EI são distintos**



**Figura 50 - Gráfico da média da sensibilidade da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 0.5% quando SI e EI são distintas**



**Figura 51 - Gráfico da média da sensibilidade da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 0.5% quando SI e EI são distintas**

## Anexo 22: Especificidade da metodologia hierárquica e de Dorfman para $p = 0.025$ quando SI e EI são distintos

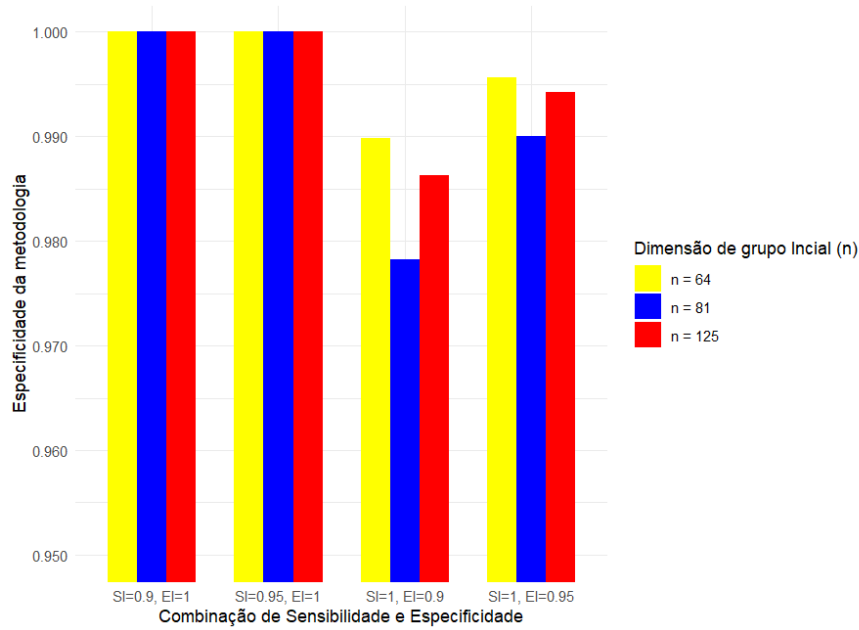


Figura 52 - Gráfico da média da especificidade da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 2.5% quando SI e EI são distintas

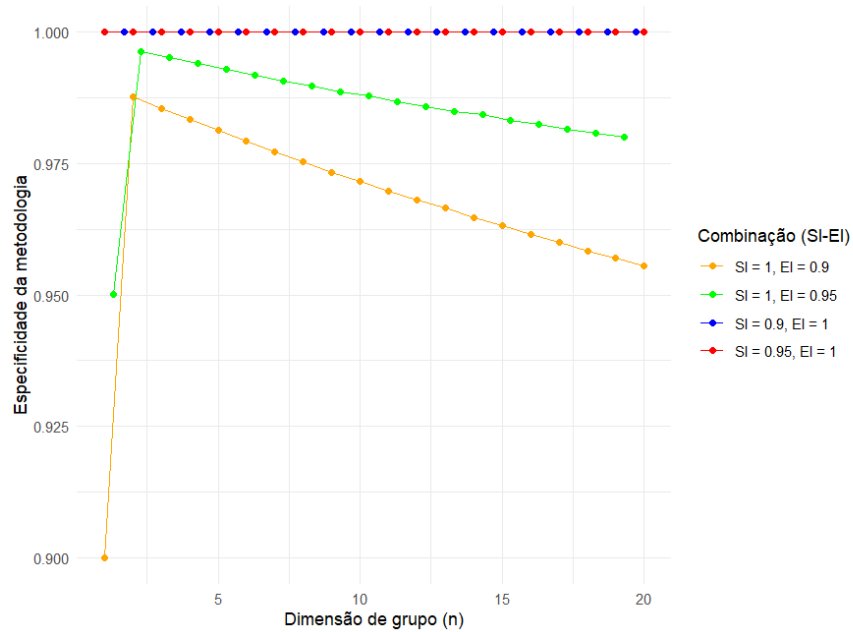


Figura 53 - Gráfico da média da especificidade da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 2.5% quando SI e EI são distintas

## Anexo 23: Especificidade da metodologia hierárquica e de Dorfman para $p = 0.01$ quando SI e EI são distintos

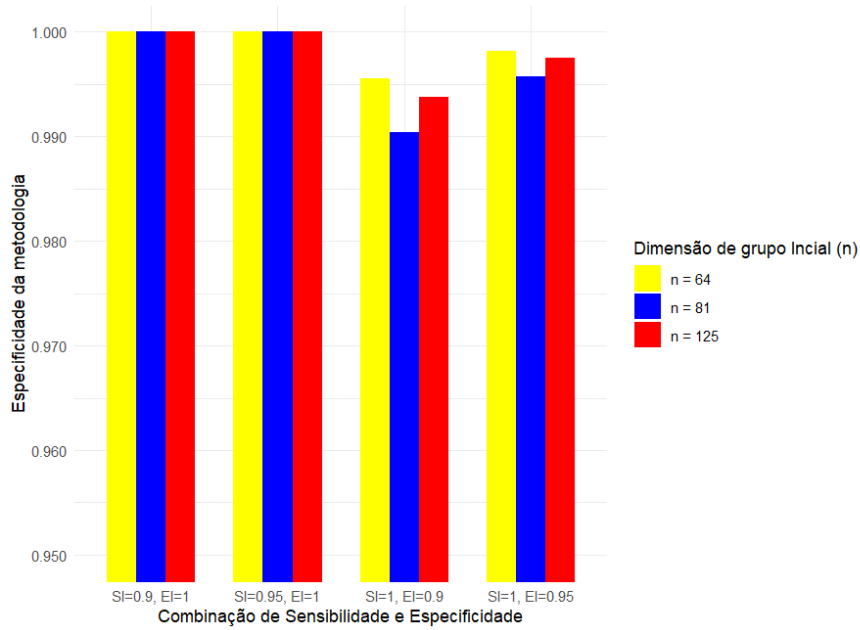


Figura 54 - Gráfico da média da especificidade da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 1% quando SI e EI são distintas

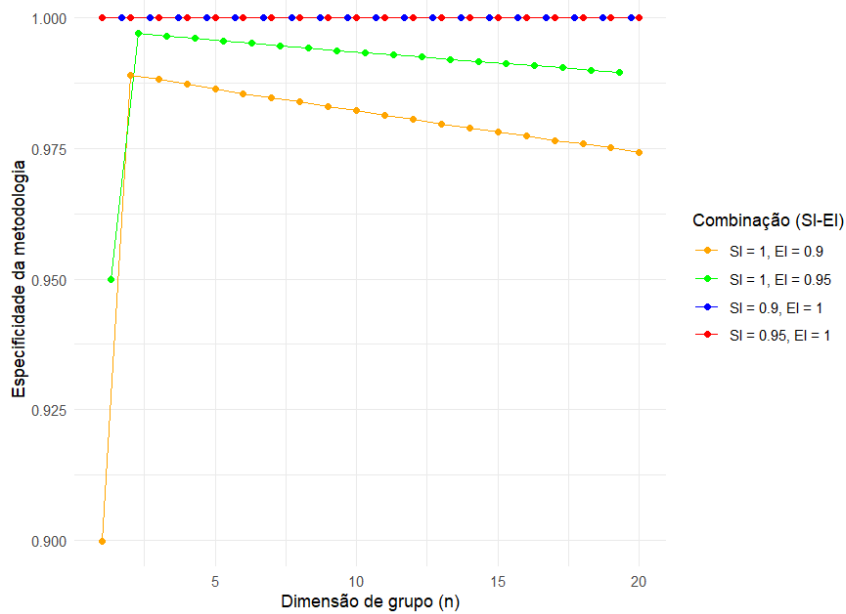


Figura 55 - Gráfico da média da especificidade da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 1 % quando SI e EI são distintas

## Anexo 24: Especificidade da metodologia hierárquica e de Dorfman para $p = 0.005$ quando SI e EI são distintos

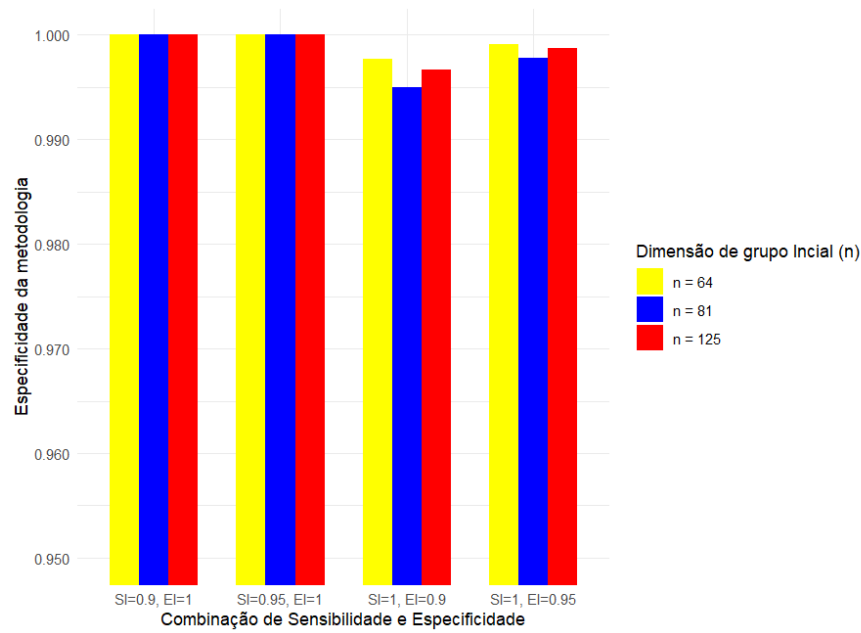


Figura 56 - Gráfico da média da especificidade da metodologia usando a metodologia hierárquica, para uma taxa de prevalência de 0.5% quando SI e EI são distintas

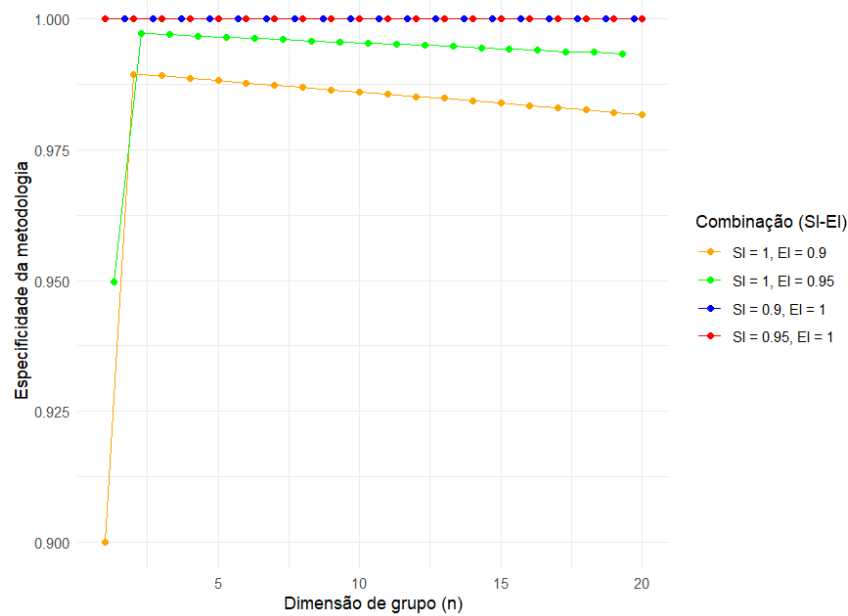


Figura 57 - Gráfico da média da especificidade da metodologia usando a metodologia de Dorfman, para uma taxa de prevalência de 0.5 % quando SI e EI são distintas

**Anexo 25: Total de testes para  $p = 0.025$  utilizando a metodologia não hierárquica com e sem *master pool* quando  $SI = EI$**

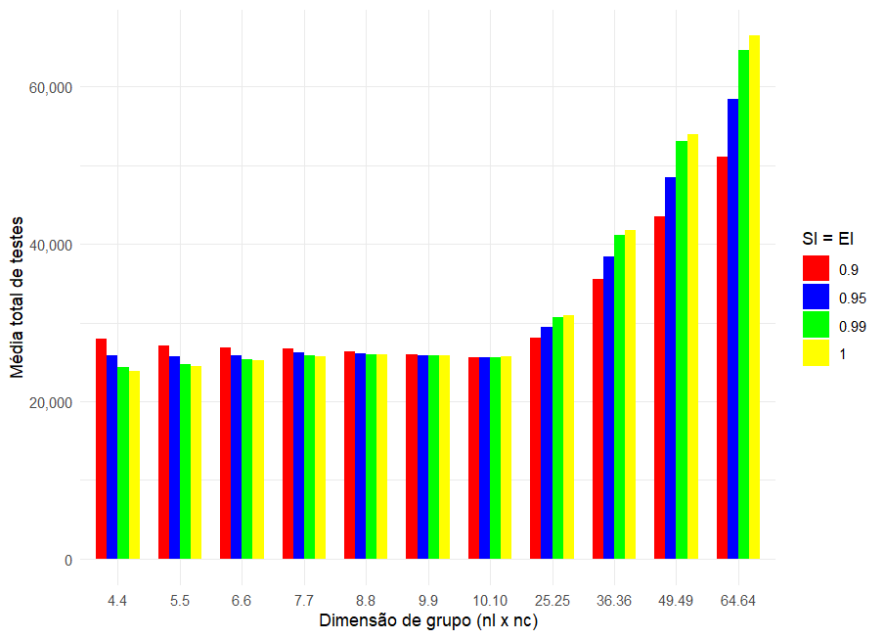


Figura 58 - Gráfico da média total de testes usando a metodologia não hierárquica com *master pool*, para uma taxa de prevalência de 2.5% quando  $SI = EI$

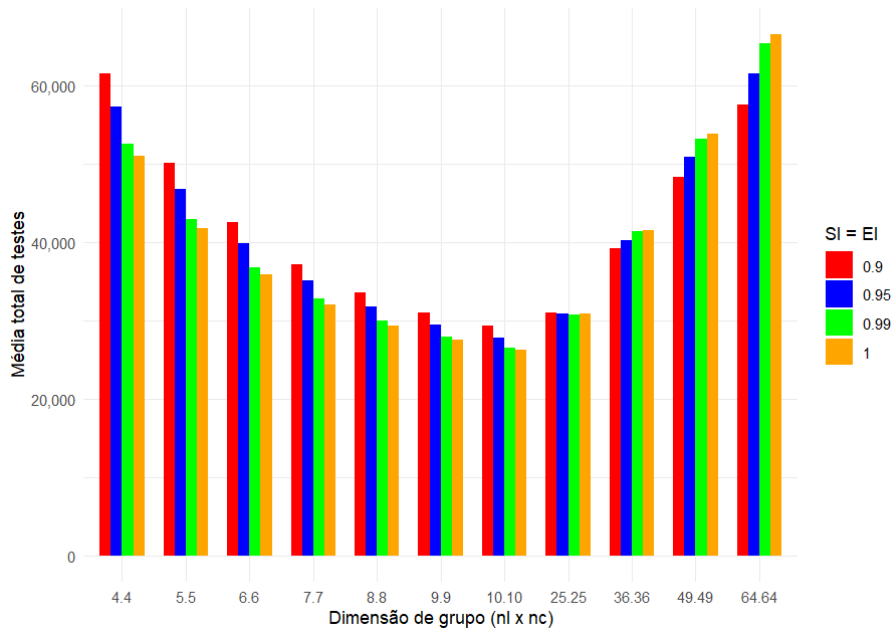
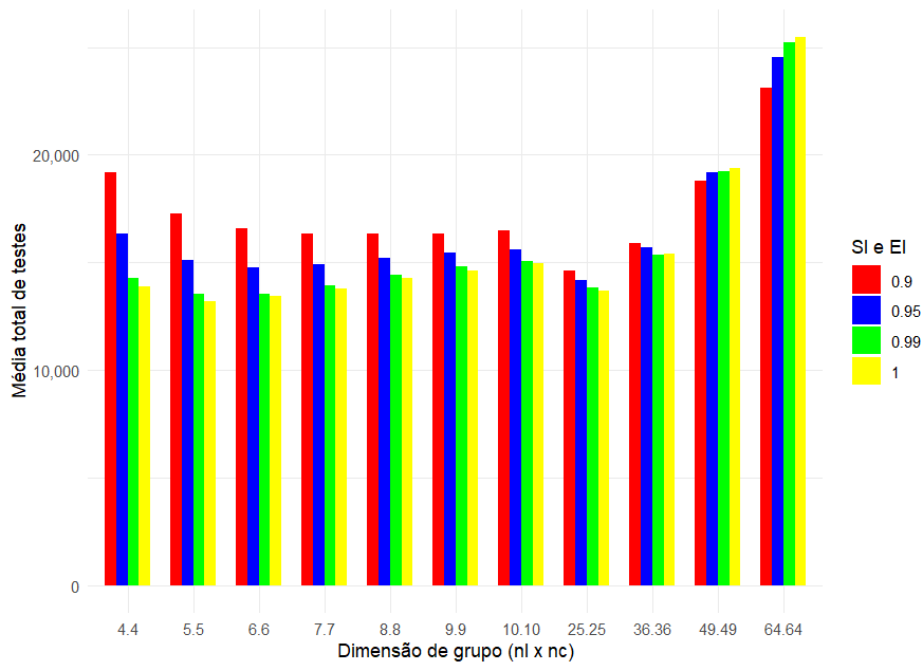
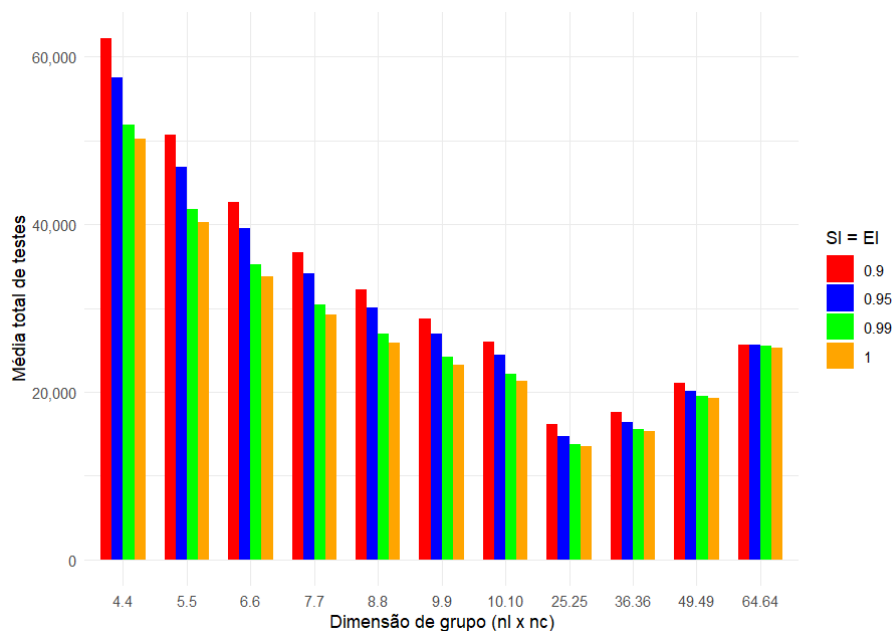


Figura 59 - Gráfico da média total de testes usando a metodologia não hierárquica sem *master pool*, para uma taxa de prevalência de 2.5% quando  $SI = EI$

**Anexo 26: Total de testes para  $p = 0.01$  utilizando a metodologia não hierárquica com e sem *master pool* quando  $SI = EI$**

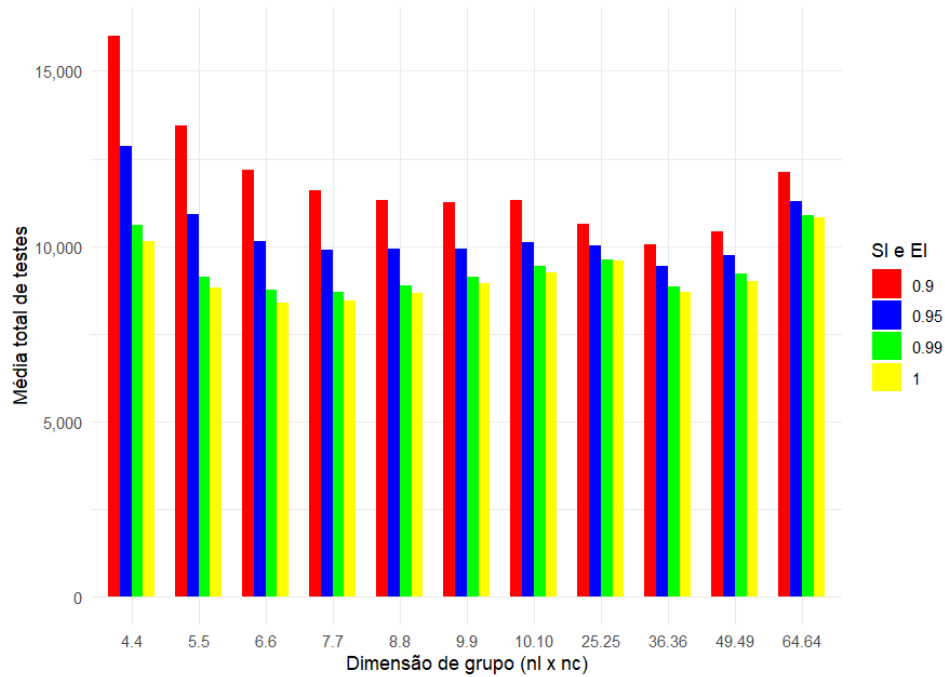


**Figura 60 - Gráfico da média total de testes usando a metodologia não hierárquica com *master pool*, para uma taxa de prevalência de 1% quando  $SI = EI$**

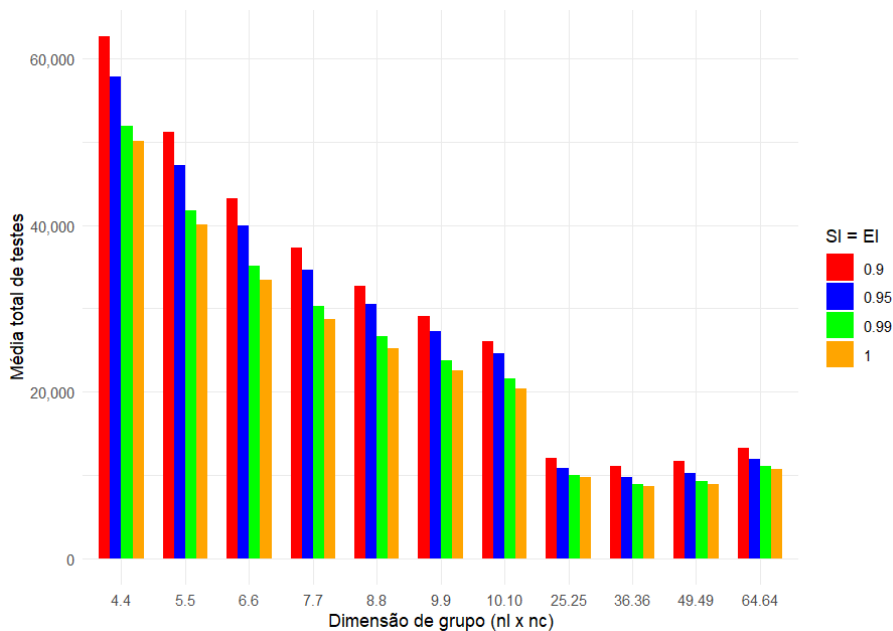


**Figura 61 - Gráfico da média total de testes usando a metodologia não hierárquica sem *master pool*, para uma taxa de prevalência de 1% quando  $SI = EI$**

**Anexo 27: Total de testes para  $p = 0.05$  utilizando a metodologia não hierárquica com e sem *master pool* quando  $SI = EI$**

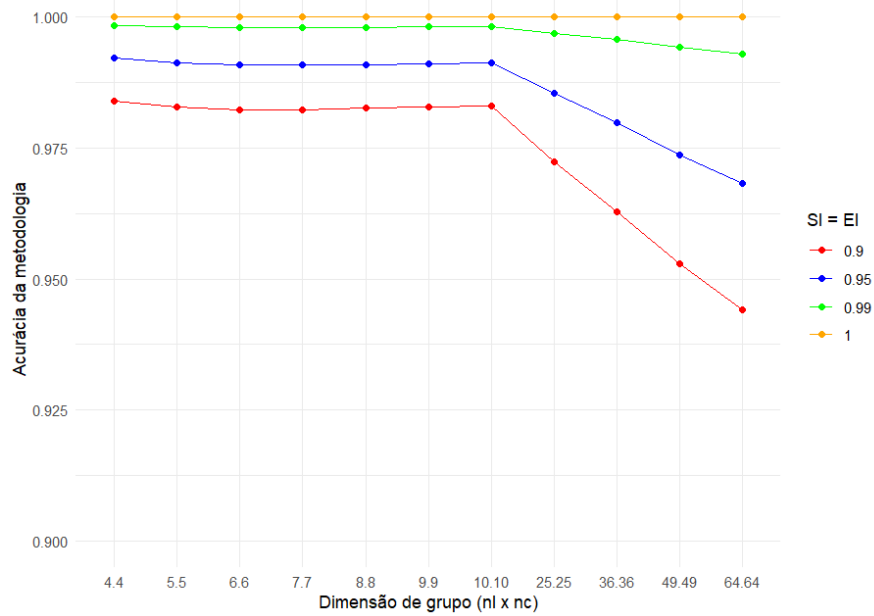


**Figura 62 - Gráfico da média total de testes usando a metodologia não hierárquica com *master pool*, para uma taxa de prevalência de 0.5% quando  $SI = EI$**

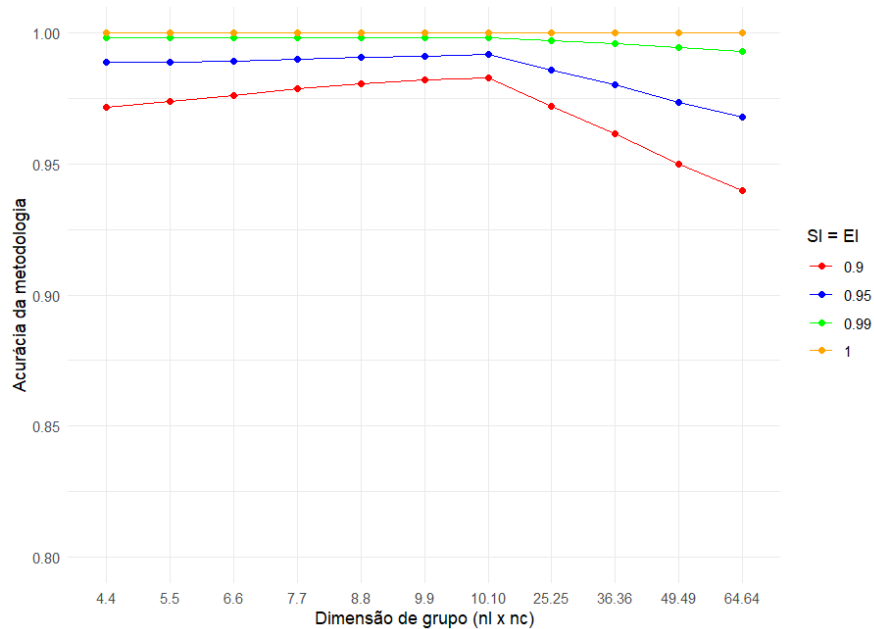


**Figura 63 - Gráfico da média total de testes usando a metodologia não hierárquica sem *master pool*, para uma taxa de prevalência de 0.5% quando  $SI = EI$**

**Anexo 28: Acurácia da metodologia não hierárquica com e sem *master pool* para  $p = 0.025$  utilizando quando  $SI = EI$**

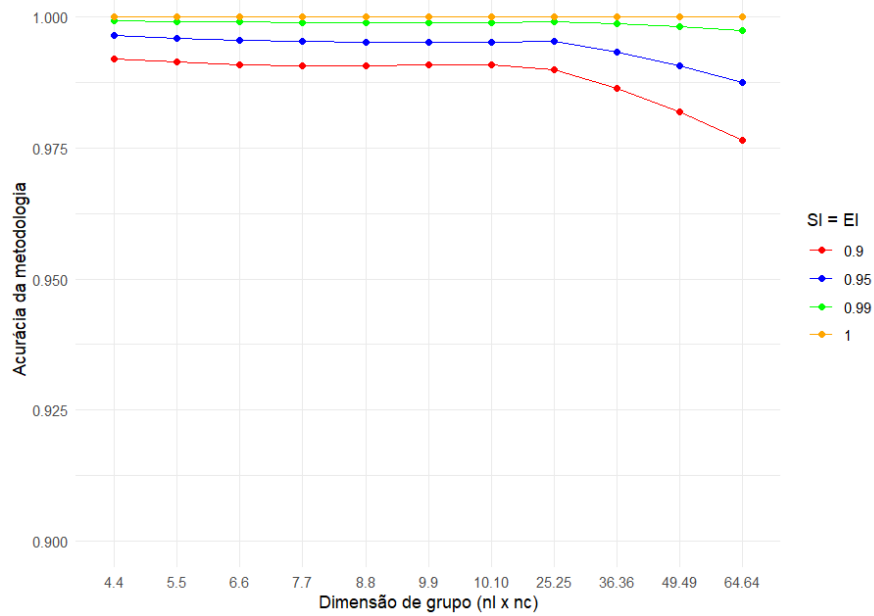


**Figura 64 - Gráfico da média da acurácia da metodologia usando a metodologia não hierárquica com *master pool*, para uma taxa de prevalência de 2.5% quando  $SI = EI$**

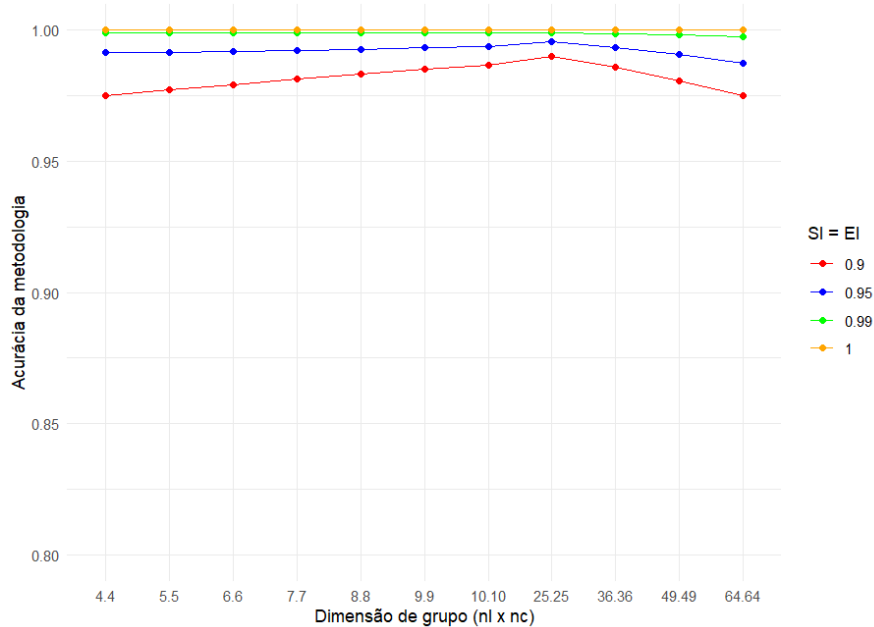


**Figura 65 - Gráfico da média da acurácia da metodologia usando a metodologia não hierárquica sem *master pool*, para uma taxa de prevalência de 2.5% quando  $SI = EI$**

**Anexo 29: Acurácia da metodologia não hierárquica com e sem *master pool* para  $p = 0.025$  utilizando quando  $SI = EI$**

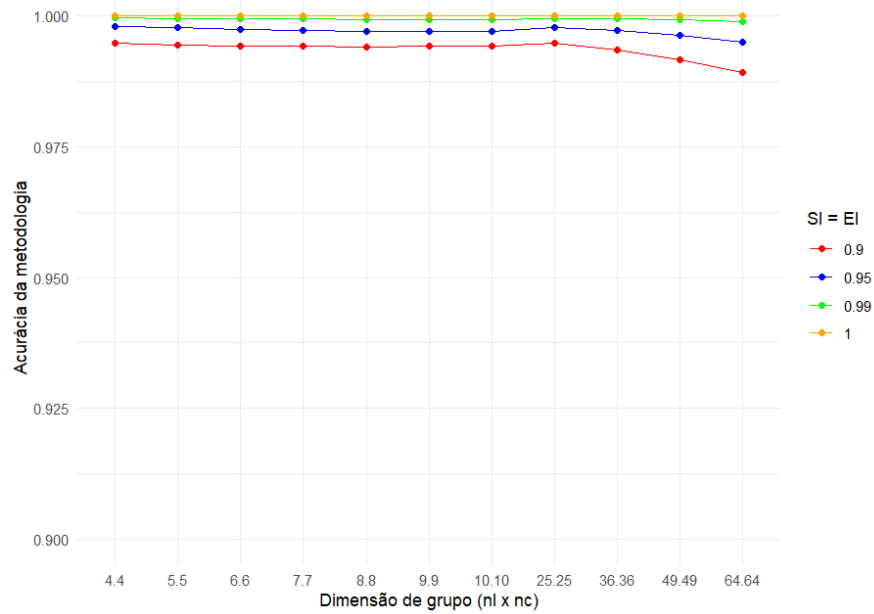


**Figura 66 - Gráfico da média da acurácia da metodologia usando a metodologia não hierárquica com *master pool*, para uma taxa de prevalência de 1% quando  $SI = EI$**

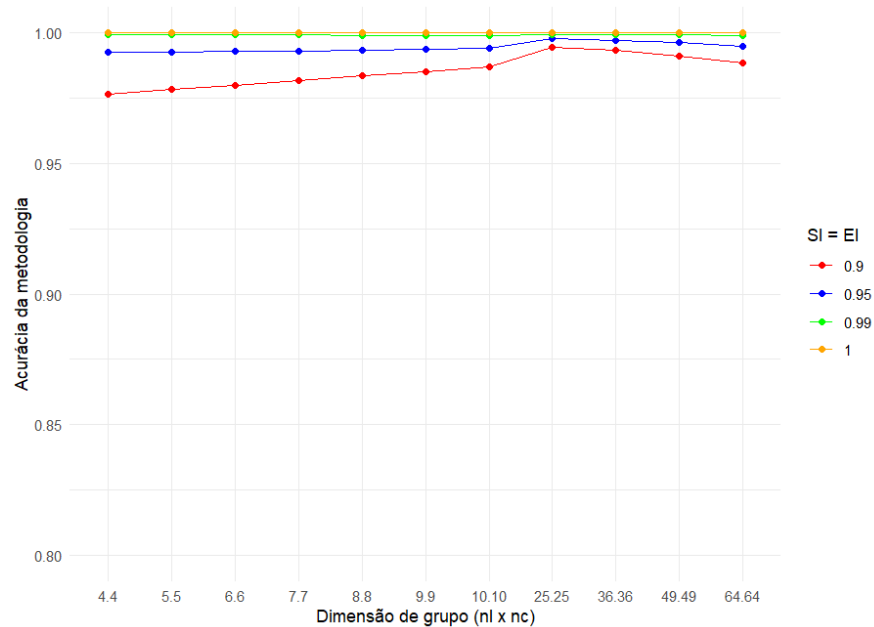


**Figura 67 - Gráfico da média da acurácia da metodologia usando a metodologia não hierárquica sem *master pool*, para uma taxa de prevalência de 1% quando  $SI = EI$**

**Anexo 30: Acurácia da metodologia não hierárquica com e sem *master pool* para  $p = 0.005$  utilizando quando  $SI = EI$**

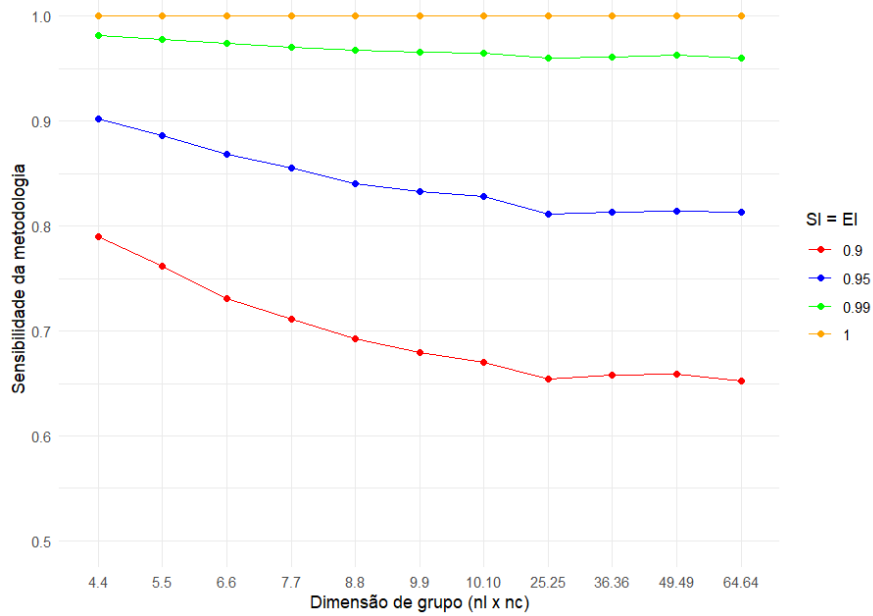


**Figura 68 - Gráfico da média da acurácia da metodologia usando a metodologia não hierárquica com *master pool*, para uma taxa de prevalência de 0.5% quando  $SI = EI$**

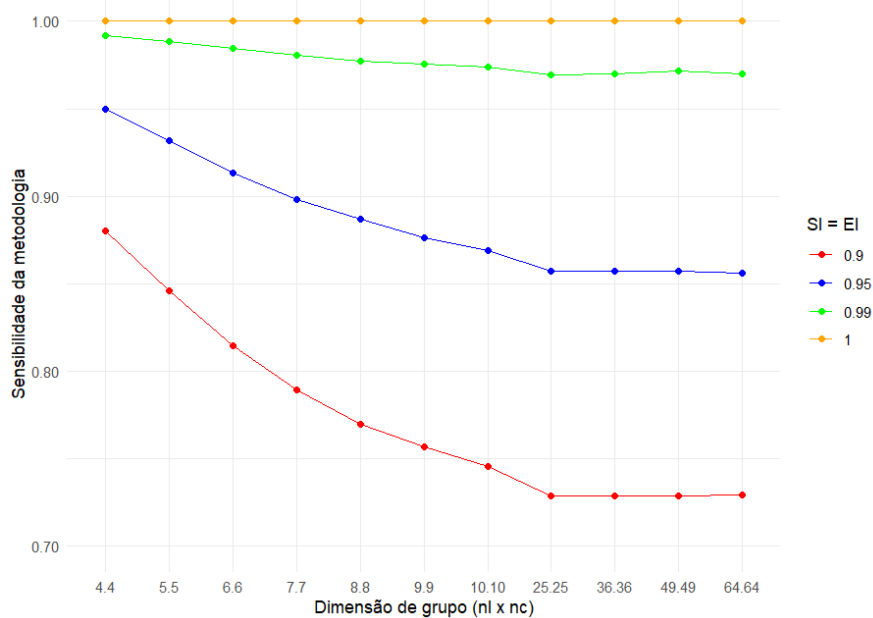


**Figura 69 - Gráfico da média da acurácia da metodologia usando a metodologia não hierárquica sem *master pool*, para uma taxa de prevalência de 0.5% quando  $SI = EI$**

**Anexo 31: Sensibilidade da metodologia não hierárquica com e sem *master pool* para  $p = 0.025$  utilizando quando  $SI = EI$**

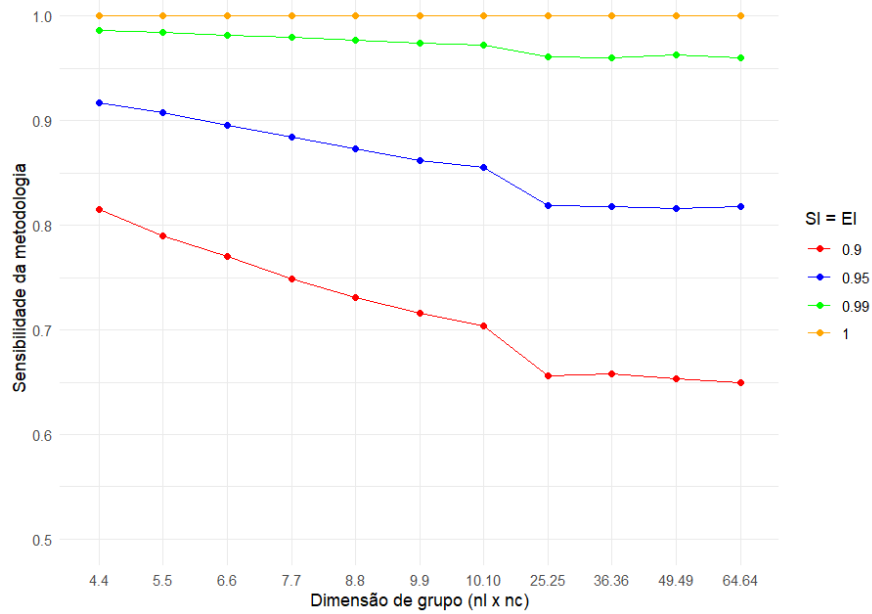


**Figura 70 - Gráfico da média da sensibilidade da metodologia usando a metodologia não hierárquica com *master pool*, para uma taxa de prevalência de 2.5% quando  $SI = EI$**

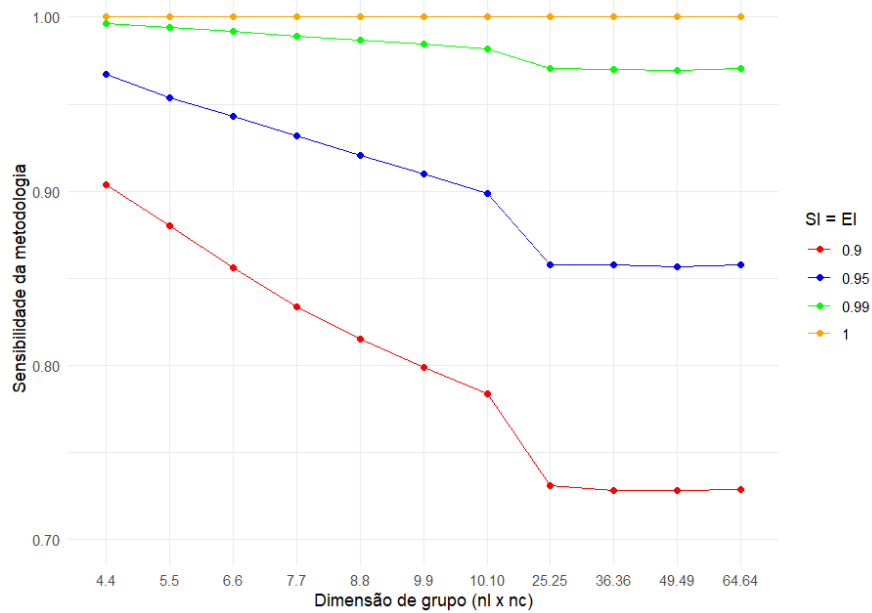


**Figura 71 - Gráfico da média da sensibilidade da metodologia usando a metodologia não hierárquica sem *master pool*, para uma taxa de prevalência de 2.5% quando  $SI = EI$**

**Anexo 32: Sensibilidade da metodologia não hierárquica com e sem *master pool* para  $p = 0.01$  utilizando quando  $SI = EI$**

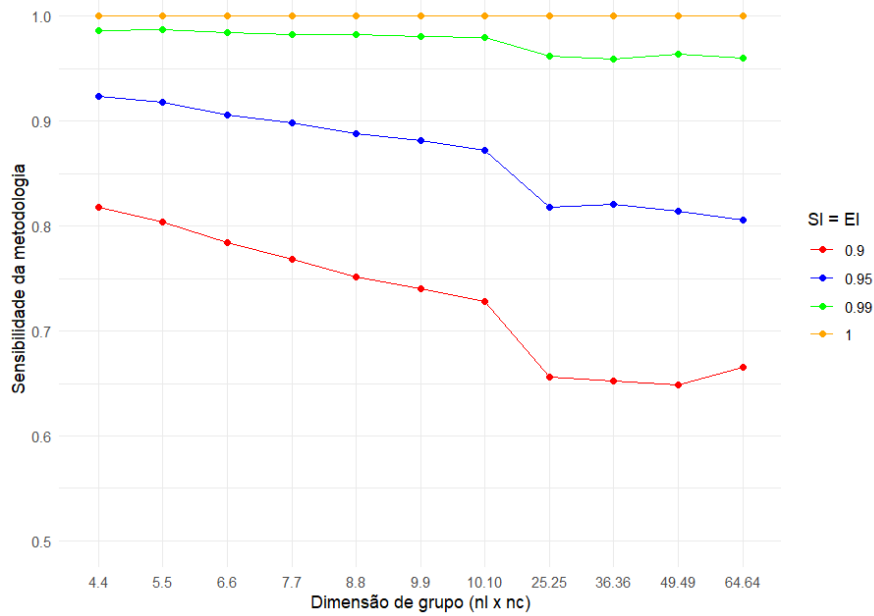


**Figura 72 - Gráfico da média da sensibilidade da metodologia usando a metodologia não hierárquica com *master pool*, para uma taxa de prevalência de 1% quando  $SI = EI$**

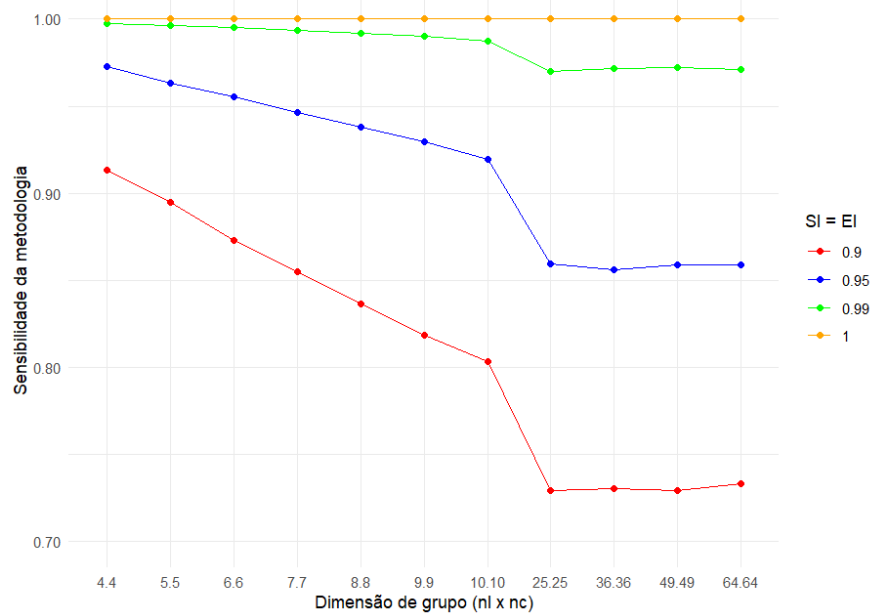


**Figura 73 - Gráfico da média da sensibilidade da metodologia usando a metodologia não hierárquica sem *master pool*, para uma taxa de prevalência de 1% quando  $SI = EI$**

**Anexo 33: Sensibilidade da metodologia não hierárquica com e sem *master pool* para  $p = 0.005$  utilizando quando  $SI = EI$**

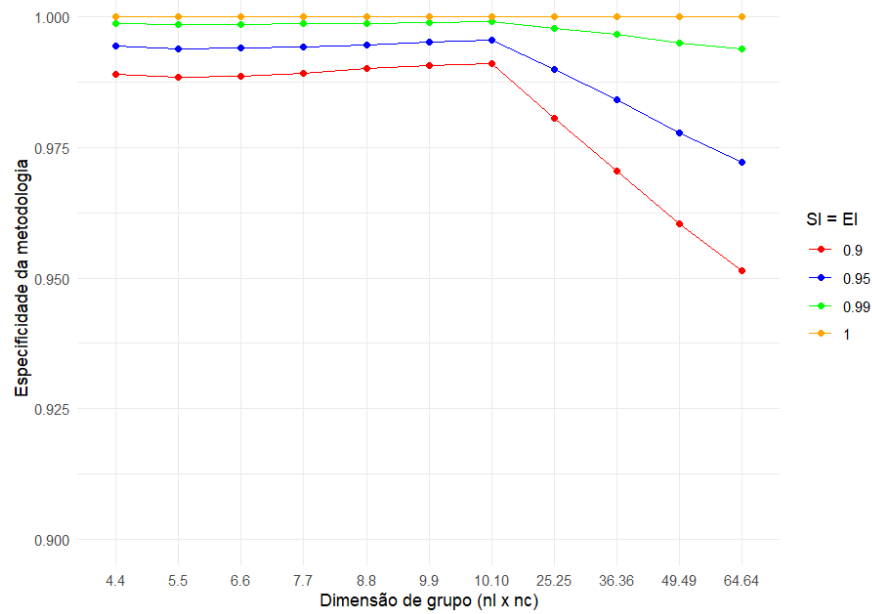


**Figura 74 - Gráfico da média da sensibilidade da metodologia usando a metodologia não hierárquica com *master pool*, para uma taxa de prevalência de 0.5% quando  $SI = EI$**

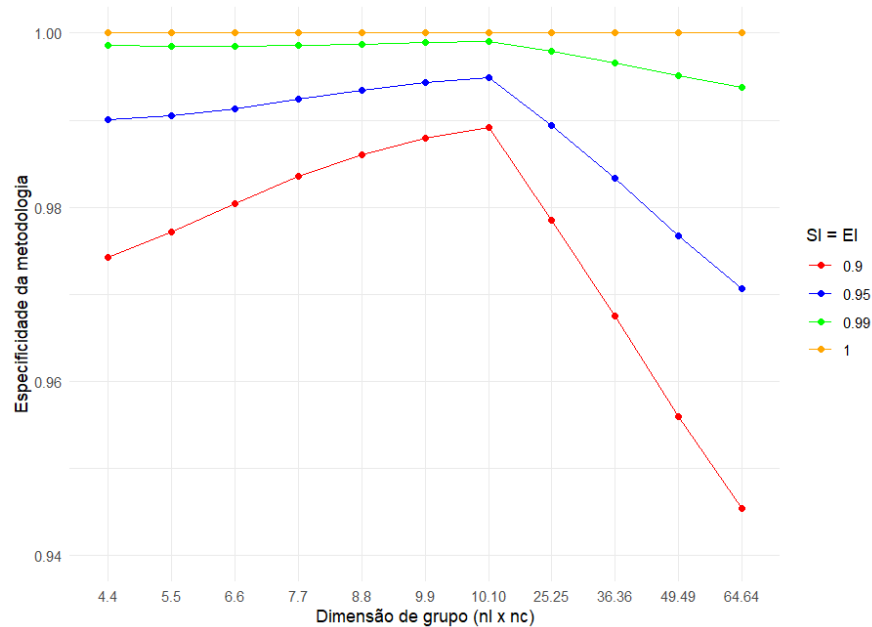


**Figura 75 - Gráfico da média da sensibilidade da metodologia usando a metodologia não hierárquica sem *master pool*, para uma taxa de prevalência de 0.5% quando  $SI = EI$**

**Anexo 34: Especificidade da metodologia não hierárquica com e sem *master pool* para  $p = 0.025$  utilizando quando  $SI = EI$**



**Figura 76 - Gráfico da média da especificidade da metodologia usando a metodologia não hierárquica com *master pool*, para uma taxa de prevalência de 2.5% quando  $SI = EI$**



**Figura 77 - Gráfico da média da especificidade da metodologia usando a metodologia não hierárquica sem *master pool*, para uma taxa de prevalência de 2.5% quando  $SI = EI$**

**Anexo 35: Especificidade da metodologia não hierárquica com e sem *master pool* para  $p = 0.01$  utilizando quando  $SI = EI$**

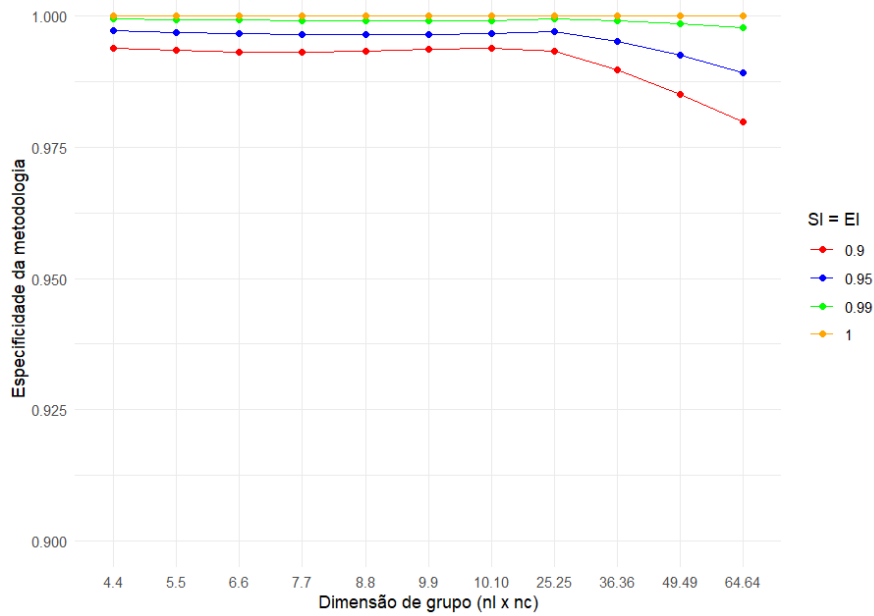


Figura 78 - Gráfico da média da especificidade da metodologia usando a metodologia não hierárquica com *master pool*, para uma taxa de prevalência de 1% quando  $SI = EI$

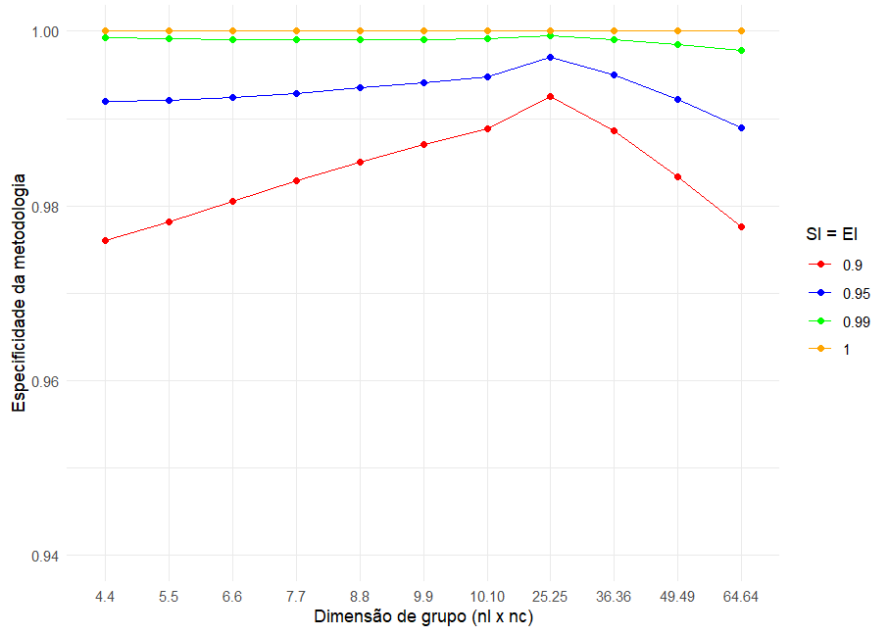
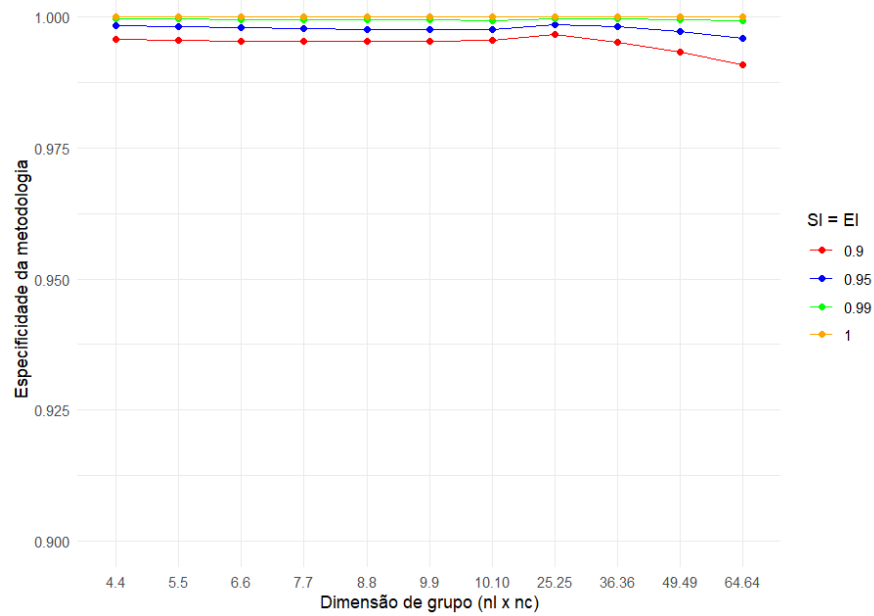
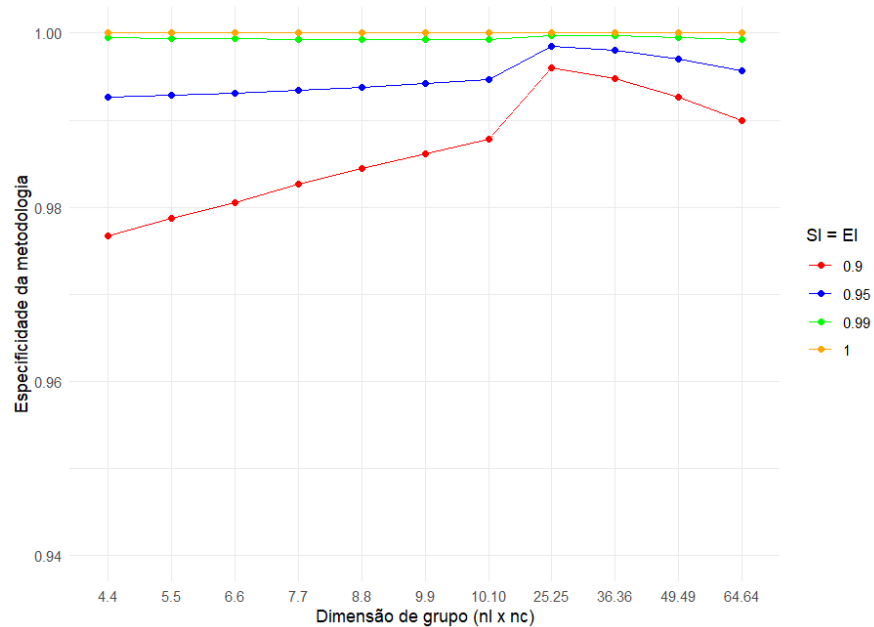


Figura 79 - Gráfico da média da especificidade da metodologia usando a metodologia não hierárquica sem *master pool*, para uma taxa de prevalência de 1% quando  $SI = EI$

**Anexo 36: Especificidade da metodologia não hierárquica com e sem *master pool* para  $p = 0.005$  utilizando quando  $SI = EI$**



**Figura 80 - Gráfico da média da especificidade da metodologia usando a metodologia não hierárquica com *master pool*, para uma taxa de prevalência de 0.5% quando  $SI = EI$**



**Figura 81 - Gráfico da média da especificidade da metodologia usando a metodologia não hierárquica sem *master pool*, para uma taxa de prevalência de 0.5% quando  $SI = EI$**

### Anexo 37: Programação elaborada através da linguagem R para a realização de testes compostos utilizando a metodologia de Dorfman para uma taxa de prevalência $p = 0.9$

```

library(caret)
library(dplyr)
library(openxlsx)
set.seed(1)
Rep <- 100
N <- 10000 # Número de indivíduos

SI = 0.9 # SI (Sensibilidade do teste individual)
EI = 0.9 # EI (Especificidade do teste individual)
SG = SI # SG (Sensibilidade composta)
EG = EI # EG (Especificidade composta)

Resultados <- data.frame(matrix(nrow = 0, ncol = 20))

p_valores <- c(0.005,0.01,0.025) #(valores das taxas de prevalência)
n_valores <- 2:20 #(Dimensões de grupo)

for(n in n_valores){
  for(p in p_valores){
    for(r in 1:Rep){
      dados <- data.frame(
        doenca = rbinom(N, 1, p),
        resultados_teste_grupo = NA
      )
      Grupo <- Grupo_Teste <- c()
      contador_testes <- 0
      cat(n,p,r)
      print(sys.time())
      for (i in 1:floor(N/n)){
        grupo_indices <- (n * (i - 1) + 1):(n * i)
        Grupo <- c(Grupo,max(dados$doenca[grupo_indices]))
        if(rbinom(1, 1, ifelse(any(dados$doenca[grupo_indices] == 1), SG, 1- EG)) == 1){
          Grupo_Teste <- c(Grupo_Teste,1)
          contador_testes <- contador_testes + 1 + length(grupo_indices)
          dados$resultados_teste_grupo[grupo_indices] <- rbinom(length(grupo_indices), 1, ifelse(dados$doenca[grupo_indices] == 1, SI, 1 - EI))
        } else {contador_testes <- contador_testes + 1
          Grupo_Teste <- c(Grupo_Teste,0)
          dados$resultados_teste_grupo[grupo_indices] <- rep(0, length(grupo_indices))
        }
      }
    }
  }
}

```

```

MCG <- confusionMatrix(as.factor(Grupo_Teste),
                      as.factor(Grupo),
                      positive = "1")
MCI <- confusionMatrix(as.factor(dados$resultados_teste_grupo),
                      as.factor(dados$doenca),
                      positive = "1")
Resultados <- rbind(Resultados,c(N,r,n,p,SI,EI,SG,EG,contador_testes,round(contador_testes*100/N,3),
MCI$overall["Accuracy"],MCI$byClass["Sensitivity"],MCI$byClass["Specificity"],
MCI$byClass["Pos Pred Value"],MCI$byClass["Neg Pred Value"],
MCG$overall["Accuracy"],MCG$byClass["Sensitivity"],MCG$byClass["Specificity"],
MCG$byClass["Pos Pred Value"],MCG$byClass["Neg Pred Value"]))
}}}

n=1
for(p in p_valores){
  for(r in 1:Rep){
    dados <- data.frame(
      doenca = rbinom(N, 1, p),
      resultados_teste = NA
    )
    cat(n,p,r)
    print(Sys.time())
    for (i in 1:N){
      ifelse(dados$doenca[i]==1,
            dados$resultados_teste[i] <- rbinom(1,1,SI),
            dados$resultados_teste[i] <- rbinom(1,1,1-EI))
    }
    MCI <- confusionMatrix(as.factor(dados$resultados_teste),
                          as.factor(dados$doenca),
                          positive = "1")
    Resultados <- rbind(Resultados,c(N,r,n,p,SI,EI,SG,EG,N,100,
                                   MCI$overall["Accuracy"],MCI$byClass["Sensitivity"],MCI$byClass["Specificity"],
                                   MCI$byClass["Pos Pred Value"],MCI$byClass["Neg Pred Value"],
                                   NA,NA,NA,NA,NA))
  }}

colnames(Resultados) <- c("N","replicas","n","p","SI","EI","SG","EG","Total Testes", "Eficiência",
                        "Acurácia Ind", "Sensibilidade Ind", "Especificidade Ind", "VPP Ind", "VPN Ind",
                        "Acurácia TG", "Sensibilidade TG", "Especificidade TG", "VPP TG", "VPN TG")

save(Resultados, file="Resultados.RData")
write.xlsx(Resultados, file="Resultados.xlsx")

```

## Anexo 38: Programação elaborada através da linguagem R para a realização de testes compostos utilizando a metodologia hierárquica para uma taxa de prevalência $p = 1$ e $n = 125$

```

library(caret)
library(dplyr)
library(openxlsx)
set.seed(1)
Rep <- 100
N <- 100000

SI = 1
EI = 1
SG = SI
EG = EI

Resultados <- data.frame(matrix(nrow = 0, ncol = 20))
p_valores <- c(0.005,0.01,0.025)

n=125
for(p in p_valores){
  for(r in 1:Rep){
    dados <- data.frame(
      doenca = rbinom(N, 1, p),
      classifica = NA
    )
    Grupo <- Grupo_Teste <- c()
    contador_testes <- 0
    cat(n,p,r)
    print(sys.time())
    for (i in 1:floor(N/n)){
      grupo_indices <- (n*(i-1)+1):(n*i)
      Grupo <- c(Grupo,max(dados$doenca[grupo_indices]))
      contador_testes <- contador_testes + 1
      if(rbinom(1, 1, ifelse(any(dados$doenca[grupo_indices] == 1), SG, 1- EG)) == 1){
        Grupo_Teste <- c(Grupo_Teste,1)
        for(j in 1:5){
          grupo_indices2 <- (n*(i-1)+25*(j-1)+1):(n*(i-1)+25*j)
          Grupo <- c(Grupo,max(dados$doenca[grupo_indices2]))
          contador_testes <- contador_testes + 1
          if(rbinom(1, 1, ifelse(any(dados$doenca[grupo_indices2] == 1), SG, 1- EG)) == 1){
            Grupo_Teste <- c(Grupo_Teste,1)
            for(k in 1:5){
              grupo_indices3 <- (n*(i-1)+25*(j-1)+5*(k-1)+1):(n*(i-1)+25*(j-1)+5*k)
              Grupo <- c(Grupo,max(dados$doenca[grupo_indices3]))
              contador_testes <- contador_testes + 1
              if(rbinom(1, 1, ifelse(any(dados$doenca[grupo_indices3] == 1), SG, 1- EG)) == 1){
                Grupo_Teste <- c(Grupo_Teste,1)
                contador_testes <- contador_testes + length(grupo_indices3)
                dados$classifica[grupo_indices3] <- rbinom(length(grupo_indices3), 1, ifelse(dados$doenca[grupo_indices3] == 1, SI, 1 - EI))}
            }
          } else{
            Grupo_Teste <- c(Grupo_Teste,0)
            dados$classifica[grupo_indices3] <- rep(0, length(grupo_indices3))}}
        }
      } else {
        Grupo_Teste <- c(Grupo_Teste,0)
        dados$classifica[grupo_indices2] <- rep(0, length(grupo_indices2))}}
    }
  }
}

```

```

else {
  Grupo_Teste <- c(Grupo_Teste,0)
  dados$classifica[grupo_indices] <- rep(0, length(grupo_indices))}}
MCG <- confusionMatrix(as.factor(Grupo_Teste),
  as.factor(Grupo),
  positive = "1")
MCI <- confusionMatrix(as.factor(dados$classifica),
  as.factor(dados$doenca),
  positive = "1")
Resultados <- rbind(Resultados,c(N,r,n,p,SI,EI,SG,EG,contador_testes,round(contador_testes*100/N,3),
MCI$overall["Accuracy"],MCI$byClass["Sensitivity"],MCI$byClass["Specificity"],
MCI$byClass["Pos Pred Value"],MCI$byClass["Neg Pred Value"],
MCG$overall["Accuracy"],MCG$byClass["Sensitivity"],MCG$byClass["Specificity"],
MCG$byClass["Pos Pred Value"],MCG$byClass["Neg Pred Value"]))
}}

colnames(Resultados) <- c("N","replicas","n","p","SI","EI","SG","EG","Total Testes", "Eficiência",
  "Acurácia Ind", "sensibilidade Ind", "Especificidade Ind", "VPP Ind", "VPN Ind",
  "Acurácia TG", "sensibilidade TG", "Especificidade TG", "VPP TG", "VPN TG")
save(Resultados, file="HC_n125_SI1_EI1.RData")
write.xlsx(Resultados, file="HC_125_SI1_EI1.xlsx")

```

## Anexo 39: Programação elaborada através da linguagem R para a realização de testes compostos utilizando a metodologia não hierárquica com e sem *master pool*

```

library(caret)
library(dplyr)
library(openxlsx)

Resultados <- data.frame(matrix(nrow = 0, ncol = 22))

set.seed(12345)
Rep <- 100
p_valores <- c(0.005, 0.01, 0.025)
SI_valores <- c(0.9, 0.95, 0.99, 1)
n_valores <- c(16, 25, 36, 49, 64, 81, 100, 36^2, 25^2, 49^2, 64^2)

# with or without masterpool
MP_valores = c(0,1)

for(MP in MP_valores){
  for(p in p_valores){
    for(n in n_valores){
      N <- 100000
      n2 = sqrt(n)
      for(SI in SI_valores){
        EI = SG = EG = SI
        for(r in 1:Rep){
          dados <- data.frame(
            doenca = rbinom(N, 1, p),
            classifica = NA)
          Grupo <- Grupo_Teste <- c()
          contador_testes <- 0
          cat(n,p,r)
          print(Sys.time())
          for(i in 1:floor(N/n)){
            grupo_indices <- (n*(i-1)+1):(n*i)
            if(MP == 1){
              Grupo <- c(Grupo,max(dados$doenca[grupo_indices]))
              contador_testes <- contador_testes + 1
              if(rbinom(1, 1, ifelse(any(dados$doenca[grupo_indices] == 1), SG, 1- EG)) == 1){
                Grupo_Teste <- c(Grupo_Teste,1)
                A1 <- matrix(dados$doenca[grupo_indices], nrow=n2,ncol=n2)
                linhas <- c()
                for(l in 1:n2){
                  Grupo <- c(Grupo,max(A1[l,]))
                  if(rbinom(1, 1, ifelse(any(A1[l,]), SG, 1- EG)) == 1){
                    Grupo_Teste <- c(Grupo_Teste,1)
                    linhas <- c(linhas,1)}
                  else {
                    Grupo_Teste <- c(Grupo_Teste,0)
                    linhas <- c(linhas,0)}}
                colunas <- c()
                for(c in 1:n2){

```



```

else {Grupo_Teste <- c(Grupo_Teste,0)
      dados$classifica[grupo_indices] <- rep(0, length(grupo_indices))
    }
  }
else{
  A1 <- matrix(dados$doenca[grupo_indices], nrow=n2,ncol=n2)
  linhas <- c()
  for(l in 1:n2){
    Grupo <- c(Grupo,max(A1[l,]))
    if(rbinom(1, 1, ifelse(any(A1[l,]), SG, 1- EG)) == 1){
      Grupo_Teste <- c(Grupo_Teste,1)
      linhas <- c(linhas,1)}
    else {
      Grupo_Teste <- c(Grupo_Teste,0)
      linhas <- c(linhas,0)}}
  colunas <- c()
  for(c in 1:n2){
    Grupo <- c(Grupo,max(A1[,c]))
    if(rbinom(1, 1, ifelse(any(A1[,c]), SG, 1- EG)) == 1){
      Grupo_Teste <- c(Grupo_Teste,1)
      colunas <- c(colunas,1)}
    else {
      Grupo_Teste <- c(Grupo_Teste,0)
      colunas <- c(colunas,0)}}
  contador_testes <- contador_testes + n2 + n2
  lp <- sum(linhas)
  cp <- sum(colunas)
  dados$classifica[grupo_indices] <- rep(0, length(grupo_indices))
  if(lp==0 & cp!=0){
    for(c in which(colunas!=0)){
      for(j in 1:n2){
        ii=(i-1)*n+(c-1)*n2+j
        contador_testes <- contador_testes + 1
        if(rbinom(1, 1, ifelse(dados$doenca[ii]==1, SG, 1- EG)) == 1){
          dados$classifica[ii] <- 1}
      }
    }
  }
  if(cp==0 & lp!=0){
    for(l in which(linhas!=0)){
      for(j in 1:n2){
        ii=(i-1)*n+(j-1)*n2+l
        contador_testes <- contador_testes + 1
        if(rbinom(1, 1, ifelse(dados$doenca[ii]==1, SG, 1- EG)) == 1){
          dados$classifica[ii] <- 1}
      }
    }
  }
  if((cp==1 | lp==1) & (cp * lp!=0)){
    for(l in which(linhas!=0)){
      for(j in which(colunas!=0)){
        ii=(i-1)*n+(j-1)*n2+l
        dados$classifica[ii] <- 1}
      }
    }
  }
  if(cp>1 & lp>1){
    for(jl in which(linhas!=0)){
      for(jc in which(colunas!=0)){
        ii=(i-1)*n+(jc-1)*n2+jl

```

```

        contador_testes <- contador_testes + 1
        if(rbinom(1, 1, ifelse(dados$doenca[ii]==1, SG, 1- EG)) == 1){
            dados$classifica[ii] <- 1}
        }}}}
MCG <- confusionMatrix(as.factor(Grupo_Teste),
                      as.factor(Grupo),
                      positive = "1")
MCI <- confusionMatrix(as.factor(dados$classifica),
                      as.factor(dados$doenca),
                      positive = "1")
Resultados <- rbind(Resultados,c(N,r,n1,n2,MP,p,SI,EI,SG,EG,contador_testes,round(contador_testes/N,3),
MCI$overall["Accuracy"],MCI$byClass["Sensitivity"],MCI$byClass["Specificity"],
MCI$byClass["Pos Pred Value"],MCI$byClass["Neg Pred Value"],
MCG$overall["Accuracy"],MCG$byClass["Sensitivity"],MCG$byClass["Specificity"],
MCG$byClass["Pos Pred Value"],MCG$byClass["Neg Pred Value"]))
}}}}

colnames(Resultados) <- c("N","replicas","n1","nc","MP","p","SI","EI","SG","EG","Total Testes", "Eficiência",
                        "Acurácia Ind", "Sensibilidade Ind", "Especificidade Ind", "VPP Ind", "VPN Ind",
                        "Acurácia TG", "Sensibilidade TG", "Especificidade TG", "VPP TG", "VPN TG")
save(Resultados, file="M2.RData")
write.xlsx(Resultados, file="M2.xlsx")}

colnames(Resultados) <- c("N","replicas","n1","nc","MP","p","SI","EI","SG","EG","Total Testes", "Eficiência",
                        "Acurácia Ind", "Sensibilidade Ind", "Especificidade Ind", "VPP Ind", "VPN Ind",
                        "Acurácia TG", "Sensibilidade TG", "Especificidade TG", "VPP TG", "VPN TG")
save(Resultados, file="M2.RData")
write.xlsx(Resultados, file="M2.xlsx")

summary(Resultados)

```