



Universidade de Brasília - UnB
Instituto de Ciências Exatas - IE
Departamento de Estatística - EST

Indicadores de testes diagnóstico para dados agrupados

Luiza Dias Bellaguarda

Orientador: Professor Eduardo Freitas da Silva

Brasília
28 de dezembro de 2020

Luiza Dias Bellaguarda

**Indicadores de testes diagnóstico
para dados agrupados**

Relatório final apresentado à disciplina de Trabalho de Conclusão de Curso II de graduação em Estatística, Instituto de Exatas, Universidade de Brasília, como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

Orientador: Professor Eduardo Freitas da Silva

Brasília
28 de dezembro de 2020

Luiza Dias Bellaguarda

Indicadores de testes diagnóstico
para dados agrupados/ Luiza Dias Bellaguarda. – Brasília, 28 de dezembro de 2020-
62 p. : il. (algumas color.) ; 30 cm.

Orientador: Professor Eduardo Freitas da Silva

Relatório Final – Universidade de Brasília

Instituto de Ciências Exatas

Departamento de Estatística

Trabalho de Conclusão de Curso de Graduação, 28 de dezembro de 2020.

1. TCC 1. 2. TCC 2. 3. Pesquisa. 4. \LaTeX . 5. Motivação. 6. Dedicção.

Luiza Dias Bellaguarda

Indicadores de testes diagnóstico para dados agrupados

Relatório final apresentado à disciplina de Trabalho de Conclusão de Curso II de graduação em Estatística, Instituto de Exatas, Universidade de Brasília, como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

Trabalho aprovado. Brasília, 28 de dezembro de 2020:

Professor Eduardo Freitas da Silva
Orientador

Professor George von Borries
Membro da Banca

Professora Maria Teresa Leão
Membro da Banca

Brasília
28 de dezembro de 2020

Agradecimentos

Agradeço, primeiramente, à todos os profissionais do departamento de Estatística que realizam um excelente trabalho e buscam engradecer a profissão. Aos professores que dedicam seu tempo para repassar seu conhecimento sempre da melhor forma para os alunos. Em especial ao professor Eduardo Freitas, que foi um grande orientador, por todo apoio e auxílio durante esse processo.

Agradeço à Empresa Júnior de Estatística, ESTAT, por todos os aprendizados, as amizades, oportunidades e por boa parte do meu crescimento pessoal e profissional.

Agradeço à minha família que sempre me apoiou no momentos mais difíceis com paciência e amor. Em especial, agradeço à minha mãe, Rosângela, e meus irmãos, Lidianne e Rafael, que estiveram torcendo por mim a cada dia.

Agradeço também à Deus por todas as boas oportunidades que tive na vida e por ter colocado as pessoas certas para me acompanhar nessa caminhada.

Resumo

O desempenho de um teste diagnóstico é frequentemente expresso em termos de sensibilidade e especificidade, e outros indicadores, em comparação com o Padrão Ouro. Os cálculos geralmente envolvem várias observações por paciente, o que implica que os dados são agrupados. O objetivo deste estudo foi apresentar e discutir uma metodologia para a estimativa dos indicadores para dados binários não agrupados (análise a nível de paciente) e outra metodologia que considera o paciente como um cluster e as partes do intestino grosso (ou segmentos do intestino) como unidades de diagnóstico do estudo em cada cluster.

Palavras-chave: Dados Categorizados; Dados Binários Agrupados; Teste Diagnóstico.

Abstract

The performance of a diagnostic test is often expressed in terms of sensitivity and specificity, and other indicators, compared to the gold standard. Calculations commonly involve multiple observations per patient, which implies that the data are clustered. The objective of this study was to present and discuss a methodology for estimating indicators of non-clustered binary data (patient level analysis) and another methodology that considers the patient as a cluster and the parts of the large intestine (or intestine segments) as the diagnostic units of the study within each cluster.

Keywords: Categorical Data; Clustered Binary Data; Diagnostic Test.

Lista de ilustrações

Figura 1 – Curva ROC	20
Figura 2 – Diagrama de fluxo da amostra do estudo	26
Figura 3 – Região de interesse no intestino	27
Figura 4 – Curva ROC a nível de indivíduo	29
Figura 5 – Curva ROC a nível de segmento	32

Lista de tabelas

Tabela 1 – Validade de um teste diagnóstico	14
Tabela 2 – Proporção de indivíduos segundo resultado do teste e condição	16
Tabela 3 – Interpretação do Kappa de acordo com Landis e Koch.	18
Tabela 4 – Indicadores do teste diagnóstico (a nível de indivíduo)	27
Tabela 5 – Concordância entre cintilografia e medidas histológicas (a nível de indivíduo)	28
Tabela 6 – Concordância entre colonoscopia e medidas histológicas (a nível de indivíduo)	28
Tabela 7 – Concordância entre cintilografia e colonoscopia (a nível de indivíduo)	29
Tabela 8 – Estatísticas da associação ROC (a nível de indivíduo)	30
Tabela 9 – Estimativa de contraste ROC e resultados de teste por linha (a nível de indivíduo)	30
Tabela 10 – Indicadores do teste diagnóstico (a nível de segmento)	30
Tabela 11 – Concordância entre cintilografia e medidas histológicas (a nível de segmento)	31
Tabela 12 – Concordância entre colonoscopia e medidas histológicas (a nível de segmento)	31
Tabela 13 – Concordância entre cintilografia e colonoscopia (a nível de segmento)	32
Tabela 14 – Estatísticas da associação ROC (a nível de segmento)	33
Tabela 15 – Estimativa de contraste ROC e resultados de teste por linha (a nível de segmento)	33

Lista de abreviaturas e siglas

A	Acurácia
ABCD	Associação Brasileira de Colite Ulcerativa e Doença de Crohn
ASC	Area Swept Out by the Curve (Área Varrida para fora da Curva)
AUC	Área Under the Curve (Área Sob a Curva)
E	Especificidade
EP	Erro Padrão
GEE	Generalized Estimating Equation (Equações de Estimação Generalizadas)
GLMM	Generalized Linear Mixed Model (Modelo Linear Generalizado de Efeitos Mistos)
HIV	Human Immunodeficiency Virus (Vírus da Imunodeficiência Humana)
IC	Intervalo de Confiança
K	Coefficiente Kappa
PLC	Projected Length of the Curve (Comprimento Projetado da Curva)
ROC	Receiver Operating Characteristic Curve (Curva Característica de Operação do Receptor)
RI	Região de Interesse
S	Sensibilidade
SF	Short Form Health Survey (Pesquisa Curta de Saúde)
U	Estatística U de Mann-Whitney
Var	Variância
VP	Valor Preditivo
VPN	Valor Preditivo Negativo
VPP	Valor Preditivo Positivo

Sumário

1	Introdução	11
2	Metodologia	13
2.1	Amostra Independente (a nível de indivíduo)	14
2.1.1	Sensibilidade e Especificidade	14
2.1.2	Valores Preditivos	16
2.1.3	Coeficiente Kappa	17
2.1.4	Acurácia	19
2.1.5	Curva ROC	19
2.2	Dados Agrupados (a nível de segmento)	21
2.2.1	Sensibilidade e Especificidade	21
2.2.2	Valores Preditivos e Acurácia	23
2.2.3	Coeficiente Kappa	24
2.2.4	Curva ROC	25
3	Resultados	26
3.1	Amostra Independente (a nível de indivíduo)	27
3.2	Dados Agrupados (a nível de segmento)	30
4	Conclusão	34
	Referências	35
	Anexos	36
	ANEXO A Código SAS	37

1 INTRODUÇÃO

O diagnóstico precoce de doenças é fundamental na área da saúde. Historicamente, ao se fazer um diagnóstico, a decisão médica era baseada nos sinais e sintomas do paciente [5]. Visando a melhoria de resultados, novos meios de diagnósticos que demonstrem evidências científicas vêm sendo estudados e desenvolvidos.

Teste diagnóstico é uma ferramenta utilizada para identificar a presença ou ausência de uma determinada doença em um indivíduo submetido ao teste. Além disso, são utilizados para avaliar a gravidade do quadro clínico do paciente e monitorar a resposta à uma intervenção clínica.

Alguns critérios para se avaliar aferições são de fundamental importância na área de saúde pelo fato de gerarem informações que constituam a base para se discriminar normalidade e anormalidade.

O desempenho de um teste diagnóstico depende da validade (ou acurácia) e reprodutibilidade (ou precisão) do teste, sendo que a validade refere-se à quanto, em termos quantitativos ou qualitativos, um teste é útil para diagnosticar um evento ou para predizê-lo, e a reprodutibilidade é a habilidade do teste em produzir resultados consistentes (quase os mesmos resultados) quando realizados independentemente e sob as mesmas condições.

Para avaliar essas questões, especialistas da área médica utilizam metodologias baseadas em alguns indicadores, como Sensibilidade e Especificidade, Coeficiente Kappa e outros.

Os pesquisadores geralmente encontram dados agrupados (cluster data). Por exemplo, um estudo pode ser realizado em vários hospitais. Os pacientes do mesmo hospital tendem a ser mais semelhantes (e, portanto, agrupados), porque compartilham as características específicas desse hospital (por exemplo, uma certa seleção de pacientes, experiência do médico, protocolos locais, entre outros). No entanto, além dos indivíduos agrupados, um estudo também pode considerar observações múltiplas por indivíduo, o que implica que as observações dentro de um paciente são agrupadas.

Conjuntos de dados de estudos que avaliam o desempenho diagnóstico de exames de imagem comumente contêm múltiplas observações por paciente. Por exemplo, normalmente temos dados sobre 4 segmentos por paciente ao avaliar o intestino grosso (Cólon Ascendente, Cólon Transverso, Cólon Descendente e Reto).

A escolha do nível de análise deve ser ditada de acordo com a decisão médica. Por exemplo, se a decisão for terapia médica versus não medicação, o nível de análise deve ser o paciente. Por outro lado, se a decisão é realizar uma cintilografia ou uma colonoscopia, o nível lógico de análise seria o segmento. Quando o nível de interesse envolve segmentos, as observações são agrupadas dentro do paciente. As observações dentro dos pacientes (clusters) são frequentemente correlacionadas positivamente, o que pode levar a estimativas viesadas de sensibilidade e especificidade quando o correlacionado a natureza dos dados dentro de cada cluster é ignorada. Mais importante, uma análise que ignora a correlação

geralmente produz um erro padrão estimado equivocadamente pequeno (e intervalo de confiança de 95% [IC]) para a sensibilidade e especificidade, pois considera erroneamente todas as observações como observações independentes. Assim, análises de dados que incluem múltiplas observações por paciente requerem alguma forma de ajuste para explicar a possível correlação entre as observações.

Este trabalho de conclusão de curso tem por objetivo analisar e comparar os indicadores de teste de diagnóstico para amostras independentes (a nível de indivíduo) e para dados agrupados (a nível de segmento) através dos seguintes indicadores: Sensibilidade e Especificidade; Valores Preditivos; Coeficiente Kappa; Acurácia; e Curva ROC.

2 METODOLOGIA

O teste diagnóstico é uma ferramenta muito utilizada na área de saúde com as principais finalidades:

- rastreamento de doença em pessoas assintomáticas;
- diagnóstico de doença em pessoa com sintomas ou sinais de doença;
- avaliar a progressão da doença ou resposta ao tratamento.

Validade (acuidade ou exatidão) é um termo muito utilizado na área de saúde e refere-se ao grau em que uma aferição é apropriada para medir o verdadeiro valor daquilo que é medido, observado ou interpretado. Refere-se à capacidade do teste diagnóstico discriminar doentes e sadios ou ainda ao quanto o resultado do teste reflete a real situação do paciente.

O teste diagnóstico ideal é chamado de Padrão Ouro, ligado à validade, pois é capaz de distinguir os pacientes doentes e os sadios. Ou seja, o resultado do teste é positivo quando o paciente tem a doença e negativo quando há a ausência da doença no paciente.

Em alguns casos em que não é possível diagnosticar uma doença através de um teste Padrão Ouro, é necessário fazer outro teste diagnóstico. Os testes mais simples são usados como substitutos dos mais elaborados e mais exatos no estabelecimento da presença de doença, sempre com o entendimento de que algum risco de diagnóstico incorreto existe, mas que é justificado pela factibilidade do teste mais simples. Nesses casos, o teste Padrão Ouro pode ser utilizado como um teste de referência e tem a finalidade de avaliar o desempenho de outros testes.

A reprodutibilidade (confiabilidade, fidedignidade ou precisão) é outro termo muito utilizado na área de saúde e refere-se à consistência ou concordância dos resultados quando a aferição se repete.

Os estudos de reprodutibilidade visam verificar se há concordância entre os resultados obtidos por diferentes estudos (comparação de métodos) ou por um mesmo (repetibilidade), realizado em momentos diferentes. Nesse último caso, o fator tempo não constitui uma variável de confundimento, ou seja, não influencia as principais variáveis do estudo.

Reprodutibilidade e validade relacionam-se da seguinte forma: um teste de baixa reprodutibilidade forçosamente acarreta baixa validade. Por outro lado, um teste de alta reprodutibilidade não assegura alta validade.

2.1 AMOSTRA INDEPENDENTE (A NÍVEL DE INDIVÍDUO)

Esta seção irá apresentar o cálculo dos indicadores, que avaliam a validade e reprodutibilidade, com base na distribuição Binomial uma vez que o tamanho da amostra é pequeno e as variáveis X e Y são binárias, ou seja, as categorias dos resultados se classificam como sucesso e fracasso. Esses rótulos são genéricos e o resultado de sucesso não precisa ser um resultado preferido [1].

A seguir tem-se a definição dos seguintes indicadores: Sensibilidade (S) e Especificidade (E), Valor Preditivo Positivo (VPP), Valor Preditivo Negativo (VPN), o Coeficiente Kappa (K), e a Acurácia (A). Além disso, também será analisada a Curva ROC (Receiver Operating Characteristic Curve).

2.1.1 Sensibilidade e Especificidade

Sensibilidade e especificidade são dois índices tradicionais na avaliação da validade quando as variáveis são categorizadas. Tal avaliação, como já citada, é procedida de maneira que os resultados são comparados com o “padrão ouro”.

Para definir os conceitos de sensibilidade e especificidade, serão utilizados como exemplos testes e doenças com resultados expressos em duas categorias: positivos ou negativos e presente ou ausente, respectivamente.

A Tabela 1 mostra as relações entre os resultados de um teste e o verdadeiro diagnóstico. Assim, na avaliação de um teste diagnóstico existem 4 interpretações possíveis para o resultado, sendo elas:

- **verdadeiro positivo:** quando o teste é positivo na presença da doença;
- **verdadeiro negativo:** quando o teste é negativo na ausência da doença;
- **falso positivo:** quando o teste é positivo na ausência da doença;
- **falso negativo:** quando o teste é negativo quando a doença está presente.

Tabela 1: Validade de um teste diagnóstico

Teste (Resultado)	Doença (Condição)		Total
	Presente (D+)	Ausente (D-)	
Positivo (T+)	Verdadeiro Positivo (a)	Falso Positivo (b)	a+b
Negativo (T-)	Falso Negativo (c)	Verdadeiro Negativo (d)	c+d
Total	a+c	b+d	N

Sensibilidade: é a probabilidade de um teste dar positivo na presença da doença, isto é, avalia a capacidade do teste detectar a doença quando ela está presente.

Especificidade: é a probabilidade de o teste ser negativo, dado que o indivíduo testado realmente não tem a doença.

Essas metodologias podem ser calculadas da seguinte forma:

- $S = P(T+ | D+) = \frac{a}{(a+c)}$
- $E = P(T- | D-) = \frac{d}{(b+d)}$

A variância (Var) e o erro padrão (EP) dessas estimativas podem ser calculados com base na distribuição binomial, como a seguir:

- $\text{Var}(S) = \frac{S(1-S)}{a+c}$ e $\text{Var}(E) = \frac{E(1-E)}{b+d}$
- $\text{EP}(S) = \sqrt{\text{Var}(S)}$ e $\text{EP}(E) = \sqrt{\text{Var}(E)}$

Ou seja, os cálculos são baseados na relação entre resultados positivos e o número de pessoas doentes e entre os resultados negativos e o número de pessoas saudáveis, nessa ordem.

Essas estimativas (sensibilidade e especificidade) estão relacionadas de maneira inversa e isso decorre da continuidade dos resultados dos testes, não havendo um ponto de corte claro e inquestionável para delimitar exames positivos e negativos. Em determinadas situações, valores elevados para um ou para outro indicador são desejáveis.

O teste sensível raramente deixa de encontrar pessoas com a doença e o teste específico raramente cometerá o erro de dizer que pessoas saudáveis são doentes.

Uso dos testes sensíveis e específicos:

Um teste sensível, usualmente usualmente positivo na presença da doença, é o teste de escolha quando não se pode correr o risco de não se diagnosticar uma doença. Eles são úteis quando a probabilidade de doença é relativamente baixa e o propósito do teste é descobrir a doença. Em outras palavras, é mais útil para o médico quando o resultado do teste é negativo.

Já um teste específico é útil para confirmar um diagnóstico sugerido por outros dados. Isto porque um teste altamente específico é raramente positivo na ausência de doença. Os testes altamente específicos são particularmente necessários quando os falsos positivos podem lesar o paciente de forma física, emocional ou financeira. Ou seja, um teste específico é mais útil quando o resultado do teste é positivo.

2.1.2 Valores Preditivos

O valor preditivo (VP) de um teste tem como objetivo verificar a probabilidade do indivíduo ser realmente doente (ou sadio) dado que o teste apresentou resultado positivo (ou negativo), podendo ser positivo (VPP) ou negativo (VPN). Além disso, é determinado pela interação de três variáveis: a sensibilidade (s), a especificidade (e) e a prevalência da doença no grupo de estudo (p).

Tabela 2: Proporção de indivíduos segundo resultado do teste e condição

Teste (Resultado)	População (Condição)		Total
	Doente	Sadia	
Positivo	$p*s$	$(1-p)*(1-e)$	$p*s + (1-p)*(1-e)$
Negativo	$p*(1-s)$	$(1-p)*e$	$p*(1-s) + (1-p)*e$

Preditivos Positivos

É a proporção de verdadeiros positivos entre todos os indivíduos com teste positivo. Expressa a probabilidade de um paciente com o teste positivo ter a doença.

$$VPP = P(D+ | T+)$$

Pelo Teorema de Bayes, temos:

$$P(D+ | T+) = \frac{\overbrace{P(T+ | D+)}^{\text{sensibilidade}} \overbrace{P(D+)}^{\text{prevalência}}}{\underbrace{P(T+ | D+)}_{\text{sensibilidade}} \underbrace{P(D+)}_{\text{prevalência}} + \underbrace{P(T+ | D-)}_{\text{1-especificidade}} \underbrace{P(D-)}_{\text{1-prevalência}}} = \frac{ps}{ps + (1-p)(1-e)} \quad (1)$$

Preditivos Negativos

Já o preditivo negativo representa a probabilidade do indivíduo não estar doente dado que o resultado do teste foi negativo.

$$VPN = P(D- | T-)$$

Da mesma forma, por Bayes obtemos:

$$P(D- | T-) = \frac{\overbrace{P(T- | D-)}^{\text{especificidade}} \overbrace{P(D-)}^{1-\text{prevalência}}}{\underbrace{P(T- | D-)}_{\text{especificidade}} \underbrace{P(D-)}_{1-\text{prevalência}} + \underbrace{P(T- | D+)}_{1-\text{sensibilidade}} \underbrace{P(D+)}_{\text{prevalência}}} = \frac{(1-p)e}{p(1-s) + (1-p)e} \quad (2)$$

Sob o ponto de vista do valor de predição negativa, o teste para se diagnosticar uma doença rara é um sucesso pelo baixo número de falso-negativos. Por outro lado, é um fracasso, pois o valor de predição positiva é muito baixo, acarretando um alto número de falso-positivos. Nesse caso, sugere-se realizar o trabalho com grupos cuja prevalência da doença é alta.

Determinantes do valor preditivo:

O valor preditivo de um teste é propriedade somente do mesmo. É determinado pela sensibilidade e especificidade do teste, e, também pela prevalência da doença na população em que se está sendo testada.

Quanto mais sensível for um teste, melhor será o seu valor preditivo negativo. Em contraposição, quanto mais específico for o teste melhor será seu valor preditivo positivo.

Quando a prevalência da doença na população testada é relativamente alta o teste funciona bem, mas em situações de baixa prevalência, o valor preditivo positivo é próximo de zero e o teste é virtualmente inútil no diagnóstico da doença.

2.1.3 Coeficiente Kappa

É um coeficiente que descreve a intensidade da concordância entre os testes de diagnóstico utilizados. Esse coeficiente se baseia no número de vezes que o resultado é o mesmo entre os testes realizados.

O coeficiente Kappa é calculado por:

$$K = \frac{P_o - P_e}{1 - P_e} \quad (3)$$

em que:

P_o : proporção observada de concordâncias (soma das respostas concordantes dividida pelo total);

P_e : proporção esperada de concordâncias (soma dos valores esperados das respostas concordantes dividida pelo total).

Os valores de Kappa variam de -1 até +1 e quanto maior o valor de kappa, mais forte a concordância. Quando:

- Kappa = 1, a concordância perfeita existe;
- Kappa = 0, a concordância é aquela que seria esperada ao acaso;
- Kappa < 0, a concordância é mais fraca que o esperado pelo acaso.

Várias interpretações para a concordância foram desenvolvidas e a mais utilizada é a de Landis e Koch:

Tabela 3: Interpretação do Kappa de acordo com Landis e Koch.

Kappa	Concordância
$\leq 0,00$	Pobre
0,01 - 0,19	Fraca
0,20 - 0,39	Regular
0,40 - 0,59	Moderada
0,60 - 0,79	Substancial
0,80 - 0,99	Quase Perfeita
1,00	Perfeita

Alguns aspectos devem ser levados em conta na interpretação do Kappa. São estes:

- clara definição do evento, regras inequívocas de mensuração e esquemas de classificação apropriados são meios de evitar que tipos de eventos, examinador, procedimentos em teste e o ambiente onde as observações são feitas interfiram no nível de reprodutibilidade medido;
- apresentar a prevalência, outras medidas de concordância ou dados originais nos resultados sobre nível de reprodutibilidade auxilia na interpretação de baixos valores de confiabilidade. Outra opção para minimizar a influência da prevalência no resultado final é comparar resultados de testes obtidos de populações onde o evento tem mesma prevalência;
- o resultado da repetição de um exame deve ser independente de resultados prévios, obtidos por um mesmo ou outro examinador.

Ao se comparar duas populações, a estatística Kappa possui uma limitação: é a sua dependência na prevalência da verdadeira positividade em cada população. Se essa última tende a 0 (zero) ou a +1 (mais um), Kappa tende a 0 (zero). E ainda, quando a prevalência da positividade entre dois observadores é diferente, o seu valor tende a aumentar.

Com o Kappa, comparações entre duas populações ou dentro de uma mesma em momentos diferentes só podem ser realizadas se a prevalência for parecida e se os observadores ou as situações proporcionarem prevalência de positividade parecidas.

Essa estatística é “a proporção da concordância entre os avaliadores após a concordância ao acaso ter sido removida” e no seu cálculo, deve-se supor:

- os N objetos ou indivíduos categorizados são independentes;
- os avaliadores operam independentemente;
- as categorias são independentes, mutuamente exclusivas e exaustivas.

Além de sua dependência em relação à distribuição marginal,

- a estatística não faz sentido para dados contínuos categorizados, pois a medida que o número de categorias aumenta, o valor de Kappa decresce;
- o coeficiente de correção intraclasse é superior ao Kappa para dados naturalmente ordenados;
- combinações de estatística Kappa Simples podem ser mais úteis que uma medida obtida para os dados politômicos nominais.

2.1.4 Acurácia

Avalia a proporção de todos os testes corretos (verdadeiros positivos e verdadeiros negativos), sobre todos os resultados obtidos.

$$A = \frac{(a+d)}{(a+b+c+d)}$$

2.1.5 Curva ROC

As curvas de característica de operação do receptor (ROC) são usadas como medidas para a precisão dos testes de diagnóstico na medicina e em outros campos quando os resultados do teste são medidas contínuas. As curvas ROC exibem a relação entre a sensibilidade (taxa de positividade verdadeira) e 1-especificidade (taxa de falsos positivos) obtidos ao considerar todos os possíveis "valores de corte" (pontos que separam as duas classes da variável dicotômica) que definem a positividade de uma doença ou condição. Eles mostram o panorama completo da troca entre taxa positiva verdadeira e taxa positiva falsa em diferentes níveis de positividade. Medidas resumidas das curvas ROC, como a

área sob a curva (AUC), o comprimento projetado da curva ROC (PLC) e a área varrida para fora da curva ROC (ASC), podem resumir a capacidade inerente de um teste para discriminar um indivíduo não doente de todos os níveis possíveis de positividade em uma única estatística.

Existem muitas literaturas sobre métodos estatísticos para estimar curvas ROC, calcular a área sob uma curva ROC e até comparar curvas ROC (independentes e correlacionadas) para dados de seção transversal. Embora não seja capaz de testar e comparar automaticamente as curvas ROC, a maioria dos principais pacotes de software estatístico, como SAS, STATA e SPSS, possui procedimentos que podem gerar diretamente uma única curva ROC e calcular a estatística resumida da AUC sob um projeto transversal.

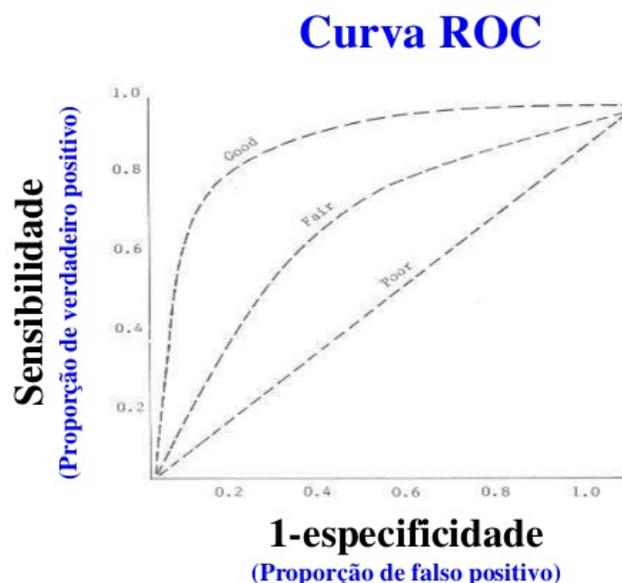
A área sob a curva tem uma propriedade estatística importante: a AUC de um classificador é equivalente à probabilidade de que o classificador classifique uma instância positiva escolhida aleatoriamente mais alta do que uma instância negativa escolhida aleatoriamente. Isso é equivalente à 'estatística U de Mann-Whitney' ou 'estatística de soma de classificação Wilcoxon'. A referência [7] tem uma derivação muito boa.

- $AUC = \frac{U}{n_P * n_N}$
- $U = AUC * n_P * n_N$

Onde, n_P e n_N são o número de exemplos positivos e negativos, respectivamente. Depois de calcular a estatística U, pode-se compará-la com a tabela de valores críticos e obter o nível de significância.

Um teste ideal é aquele que possui a área sob a curva ROC igual a 1. Quando a área é igual a 0.5, o teste não permite distinguir entre os grupos.

Figura 1: Curva ROC



2.2 DADOS AGRUPADOS (A NÍVEL DE SEGMENTO)

Quando há dependência entre múltiplas observações do mesmo paciente (cluster), podemos esperar que haja uma correlação positiva entre as observações. Ao considerar dados binários agrupados, a estimativa pontual da sensibilidade fornece a mesma estimativa que resultaria se tivéssemos ignorado o agrupamento, assumido a independência e usado a estimativa de ponto binomial convencional. No entanto, se a variância da sensibilidade for estimada usando a estimativa de variância binomial convencional, a variância será subestimada se a correlação entre as unidades de diagnóstico for positiva (isto é, a variância estimada for menor que a variância real) ou será superestimada se a correlação entre as unidades de diagnóstico for negativa (isto é, a variância estimada será maior do que a variação verdadeira). Para corrigir esse problema de viés da variância, alguns métodos para dados agrupados com a correção da variância serão utilizados para realizar o cálculo de alguns indicadores.

O esperado para o nosso estudo é que a correlação entre as unidades de diagnóstico seja positiva, pois uma correlação positiva indica que se uma unidade de diagnóstico for testada como um verdadeiro positivo (verdadeiro negativo) em um cluster específico, a probabilidade de testar outras unidades de diagnóstico no mesmo cluster como verdadeiros positivos (verdadeiros negativos) aumentará.

Além disso, a curva ROC é uma importante ferramenta utilizada para diagnosticar e julgar a capacidade de discriminação de diferentes procedimentos. Para isso, é possível utilizar o modelo linear generalizado de efeitos mistos (GLMM)[6] para estimar as probabilidades previstas da positividade de uma doença ou condição, e em seguida a probabilidade estimada é usada como um bio-marcador para construir a curva ROC e calcular a área sob a curva.

2.2.1 Sensibilidade e Especificidade

Neste estudo será utilizado o Estimador de razão para corrigir o viés da variância.

Estimador de razão

Uma estimativa ajustada da variância pode ser calculada usando-se a estimativa de sensibilidade e especificidade específica do paciente ou cluster, pesos proporcionais à proporção de observações de doenças (e não-doenças) em cada paciente e o tamanho médio do cluster para as observações doentes (e não doentes) para sensibilidade e especificidade, respectivamente.

Se S_i e E_i são a sensibilidade e a especificidade calculadas usando apenas os dados do i -ésimo cluster, ou seja, $S_i = a_i / (a_i + c_i)$ e $E_i = d_i / (b_i + d_i)$, e temos D_i e N_i , observações doentes e não-doentes, respectivamente, no i -ésimo cluster (com $D_i = a_i + c_i$ e $N_i = b_i + d_i$), então a sensibilidade e especificidade gerais podem ser calculadas como:

- $S = \frac{\sum D_i S_i}{\sum D_i}$
- $E = \frac{\sum N_i E_i}{\sum N_i}$

Isso produz o mesmo resultado de antes, mas é expresso de maneira ligeiramente diferente. Os erros padrão correspondentes ajustados para armazenamento em cluster são calculados como:

- $EP(S) = \sqrt{\frac{1}{I_D(I_D-1)} \sum \left(\left(\frac{D_i}{\bar{D}} \right)^2 (S_i - S)^2 \right)}$
- $EP(E) = \sqrt{\frac{1}{I_N(I_N-1)} \sum \left(\left(\frac{N_i}{\bar{N}} \right)^2 (E_i - E)^2 \right)}$

onde $\bar{D} = \sum D_i / I_D$ e $\bar{N} = \sum N_i / I_N$ são os tamanhos médios de cluster para as observações com e sem doença, respectivamente, e I_D e I_N são o número de clusters (pacientes) com segmentos doentes e não-doentes, respectivamente.

Outras técnicas sugerias na literatura

Modelo logístico de efeito misto

As características dos testes de sensibilidade e especificidade ajustados para correlação podem ser modeladas usando um modelo de regressão logística de efeitos mistos. Esse método é comparável à realização de uma regressão logística regular, mas agora supõe-se que a sensibilidade e a especificidade exibam alguma variação em torno de uma média comum entre os pacientes (clusters). A abordagem mais simples é considerar apenas uma interceptação aleatória (isto é, um efeito aleatório entre grupos), que permite agrupar as observações nos pacientes. Presume-se que a interceptação aleatória segue uma distribuição normal, o que implica que as sensibilidades ou especificidades específicas dos pacientes devem seguir a distribuição normal de logit. Essa abordagem gera a sensibilidade e especificidade medianas (correspondentes à interceptação aleatória = 0) em todas as estimativas de sensibilidade e especificidade específicas dos pacientes, assumindo, é claro, que o modelo assumido está correto.

Para obter uma estimativa da sensibilidade e especificidade médias em toda a população de pacientes (também chamada de média da população), é necessário calcular a sensibilidade e a especificidade sobre a distribuição normal da interceptação aleatória, e isso envolve uma integração numérica.

Os pontos fortes da abordagem de efeitos aleatórios incluem a eficiência para analisar pequenos conjuntos de dados e a oportunidade de extrair inferências corretas para

dados agrupados em projetos de estudo não pareados e emparelhados.

Equações de estimativas generalizadas

As estimativas de sensibilidade e especificidade ajustadas para correlação nos pacientes também podem ser modeladas usando a abordagem da equação de estimativa generalizada (GEE)[3]. Essa abordagem requer apenas um palpite aproximado da estrutura de correlação entre as observações agrupadas e, portanto, é chamada de "matriz de correlação de trabalho". Quando essa matriz de correlação é especificada como "independente", o que corresponde a nenhuma correlação, as estimativas de GEE são idênticas às estimativas obtidas usando regressão logística e um estimador sanduíche para a variância. Para um tamanho de amostra relativamente grande, onde admitimos que é difícil definir "grande", os intervalos de confiança obtidos com o GEE são válidos, mesmo quando a matriz de correlação não está correta. No entanto, quanto melhor a matriz de correlação for especificada, maior será a precisão dos resultados, ou seja, intervalos de confiança menores. Assim, a especificação incorreta levará a estimativas um tanto conservadoras dos intervalos de 95%.

Em contraste com os modelos de efeitos mistos, a correlação em si não é de interesse primário na abordagem GEE. Além disso, a sensibilidade e a especificidade estimada é interpretada como a média ponderada na população do estudo ("população média") na abordagem GEE. Essa abordagem, portanto, fornece a sensibilidade e especificidade médias em todas as estimativas de sensibilidade e especificidade específicas do paciente. Isso contrasta com a sensibilidade e especificidade em um modelo de efeitos mistos, que tem uma interpretação condicional. Portanto, as estimativas pontuais obtidas com um GEE são, em geral, diferentes das estimativas pontuais obtidas com um modelo logístico de efeitos mistos. Finalmente, a abordagem GEE é menos eficiente que a abordagem de efeitos mistos, porque se baseia em menos suposições (não se baseia em um modelo de probabilidade totalmente especificado como a abordagem de efeitos mistos). Do lado positivo, podemos dizer que, pelo mesmo motivo, a abordagem GEE é mais robusta contra a especificação incorreta do modelo.

2.2.2 Valores Preditivos e Acurácia

O cálculo para Valores Preditivos (Positivos e Negativos) e Acurácia segue os mesmos métodos utilizados para calcular os indicadores de Sensibilidade e Especificidade citados acima.

2.2.3 Coeficiente Kappa

A estatística Kappa tem sido amplamente aplicada às classificações cruzadas encontradas em epidemiologia, diagnóstico por imagem e outras. Várias extensões dessa estatística foram desenvolvidas, incluindo Kappas ponderados para categorias e Kappas permitindo múltiplas observações por indivíduo e categorizações múltiplas.

Na prática, existem muitos casos em que unidades de amostragem são apresentadas em grupos, e nesses casos, as unidades do mesmo cluster tendem a ter resultados semelhantes.

Para estimar Kappa para os dados de pares combinados em cluster, a tarefa mais desafiadora é determinar a variação assintótica da estatística Kappa enquanto considera a correlação nos clusters. Entretanto, muitas vezes pouco se sabe sobre a adequação de quaisquer estruturas de correlação assumidas, portanto, são desejados métodos de análise que ajustam várias unidades em um cluster, mas evitam suposições específicas sobre estruturas de correlação dentro do cluster. Com base em técnicas de amostragem e para testar a homogeneidade marginal, foi criado um método para ajustar o efeito de agrupamento sem assumir estruturas específicas de distribuição ou correlação dentro do agrupamento. Essa estratégia concede flexibilidade em situações de tamanhos variados de cluster e situações em que é inviável assumir uma correlação homogênea dentro do cluster. O método proposto não postula nenhuma estrutura de correlação dentro do cluster ou premissas de distribuição.

Suponha que haja K clusters independentes na população de estudo em que n_k ($k = 1, \dots, K$) é o tamanho fixo da amostra para o k -ésimo cluster com $\sum_{k=1}^K n_k = N$. Para o k -ésimo cluster, denotamos o número correspondente de observações como $n_{ij,k}$, n_{i+} e $n_{+j,k}$, a probabilidade de cair na i -ésima linha e j -ésima coluna como $p_{ij,k}$, e as probabilidades marginais como $p_{i+,k} = \sum_j p_{ij,k}$ e $p_{+j,k} = \sum_i p_{ij,k}$.

Além disso, denotamos:

$$\omega_k = n_k \left(\sum_{k=1}^K n_k \right)^{-1} = \frac{n_k}{N}, \quad P_{o,k} = \frac{1}{n_k} \sum_{i=1}^g n_{ii,k}, \quad p_{i+,k} = \frac{n_{i+,k}}{n_k}, \quad p_{+i,k} = \frac{n_{+i,k}}{n_k}$$

Assumindo homogeneidade entre os clusters K , isto é, $E(n_{ij,k}/n_k) = p_{ij}$ para $k = 1, \dots, K$, esperamos ter os mesmos parâmetros p_{ij} , p_{i+} , p_{+i} , P_o , P_e para todos os K clusters. Para os dados em cluster podemos calcular o Kappa K_c usando a equação (3) com P_o e P_e sendo:

$$P_o = \sum_{k=1}^K \omega_k P_{o,k} = \frac{1}{N} \sum_{k=1}^K \sum_{i=1}^g n_{ii,k}, \quad P_e = \sum_{i=1}^g p_{i+} p_{+i}$$

onde,

$$p_{i+} = \sum_{k=1}^K \omega_k p_{i+,k} = \frac{1}{N} \sum_{k=1}^K n_{i+,k} \quad , \quad p_{+i} = \sum_{k=1}^K \omega_k p_{+i,k} = \frac{1}{N} \sum_{k=1}^K n_{+i,k}$$

Com isso, concluímos que o coeficiente Kappa K_c é idêntico ao K em (3) calculado usando todos os dados ignorando a correlação entre elementos de um mesmo grupo. Além disso, a interpretação dos resultados segue conforme a tabela 3.

2.2.4 Curva ROC

Segundo Honghu Liu e Tongtong Wu [4], o desenho de medidas repetidas é amplamente utilizado na pesquisa de serviços biológicos, médicos e de saúde, mas faltam métodos estatísticos para estimar curvas ROC e suas estatísticas resumidas para tais desenhos, especialmente abordagens baseadas em modelos de regressão, que oferecem a oportunidade de avaliar o impacto de co-variáveis sobre a potência e precisão de um teste. Na pesquisa de serviços médicos e de saúde, por exemplo, os pesquisadores geralmente precisam observar a alteração dos resultados do paciente ao longo do tempo em função do tratamento, do tempo e de outras co-variáveis. Essas medidas de resultado geralmente estão em formato contínuo, mas podem ser dicotomizadas, escolhendo algum valor crítico clinicamente significativo c e definindo o teste como positivo se a medida de resultado for maior que c . Por exemplo, podemos precisar analisar o nível de colesterol do paciente ao longo do tempo em função do tratamento, dieta e outras características do paciente; podemos precisar observar o estado de saúde do paciente medido pelo Short Form Health Survey 36 (SF-36) ou Short Form Health Survey 12 (SF-12) ao longo do tempo como uma função do processo de atendimento e estrutura das organizações médicas; e podemos precisar estudar o resultado virológico de pacientes com vírus da imunodeficiência humana (HIV) como uma função da adesão ao tratamento com medicamentos antirretrovirais e outras co-variáveis. O poder discriminador longitudinal inerente dos tratamentos e outras co-variáveis nos resultados medidos ao longo do tempo pode ser analisado e avaliado por curvas ROC adequadamente calculadas que levam em consideração a dependência das observações dentro de um sujeito. Métodos de computação estatística e software de computador para essas análises são poucos e necessários.

Além disso, utiliza-se o modelo misto linear generalizado (GLMM)[6] para calcular curvas ROC e a área sob uma curva ROC para o projeto de medidas repetidas. Um programa estatístico escrito em SAS, é especificamente utilizado para implementar o algoritmo proposto e realizar os cálculos. Com este programa, será possível ajustar as curvas ROC e calcular a área sob uma curva ROC, considerando a mesma estatística U de Mann-Whitney descrita na seção 3.1.5, para dados de projetos de medidas repetidas.

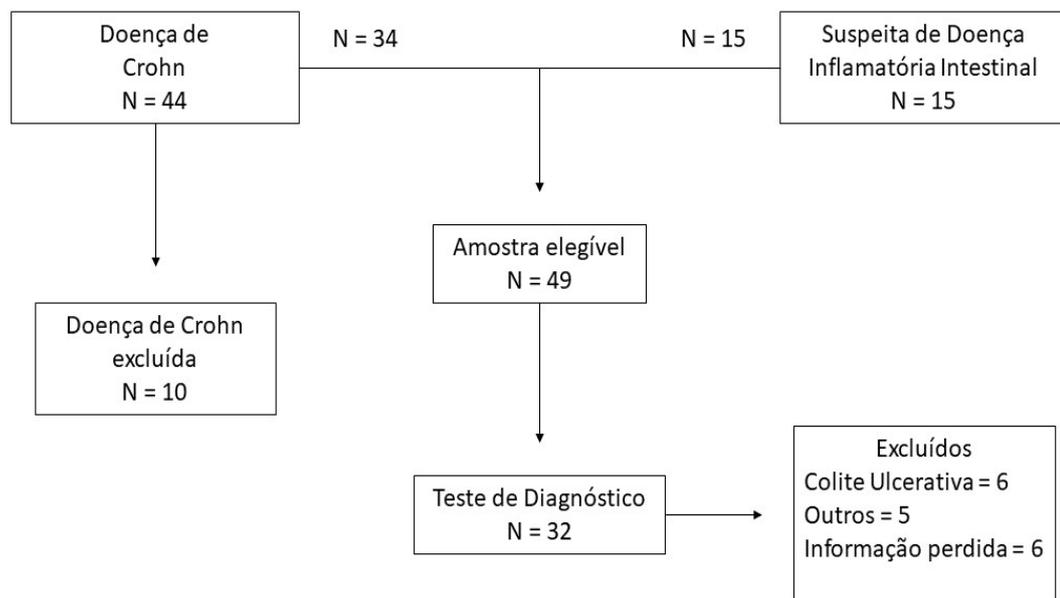
3 RESULTADOS

Segundo a Associação Brasileira de Colite Ulcerativa e Doença de Crohn (ABCD)[2], a Doença de Crohn é uma doença inflamatória séria do trato gastrointestinal. Ela afeta predominantemente a parte inferior do intestino delgado (íleo) e intestino grosso (cólon), mas pode afetar qualquer parte do trato gastrointestinal.

Os dados a serem utilizados neste trabalho foram obtidos através de um estudo transversal realizado por 6 médicos do Hospital de Base do Distrito Federal.

A população do estudo foi derivada de indivíduos elegíveis que tinham a Doença de Crohn em tratamento ou pacientes com suspeita de doença inflamatória intestinal. Esses indivíduos [n = 59] foram incluídos em um estudo transversal consecutivo prospectivo de janeiro de 2018 a junho de 2019. Um total de 32 pacientes, variando de 18 a 70 anos, com doença de Crohn de acordo com critérios clínicos, endoscópicos e histopatológicos estabelecidos foram incluídos e analisados. Nove homens e 23 mulheres foram incluídos. Os critérios de exclusão contém gravidez, colectomia, uso de corticosteroide, idade inferior a 18 anos, doença exclusivamente anoperineal, mulheres lactantes e colite ulcerativa. Seis pacientes não completaram o protocolo. A amostra foi estabelecida para critérios clínicos, endoscópicos e histopatológicos.

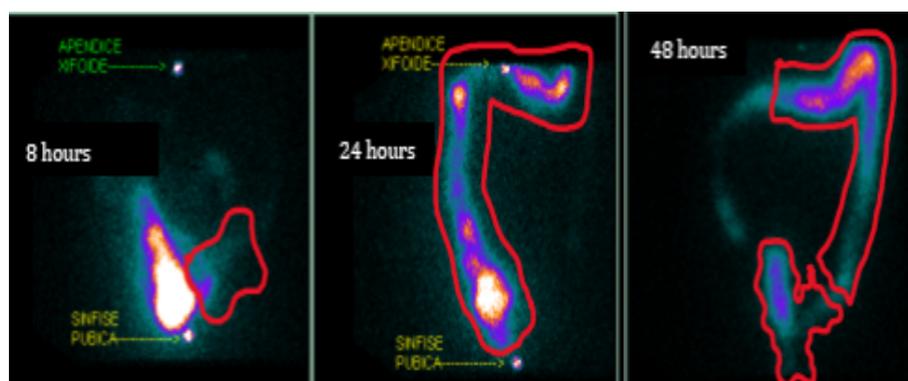
Figura 2: Diagrama de fluxo da amostra do estudo



Referência: Tarjra, João Batista M. et. al. Annals of Nuclear Medicine. February 2020.
Endereço eletrônico: DOI: 10.1007/s12149-020-01447-w [8]

A zona intestinal com doença de Crohn dos 32 pacientes foi dividida em quatro segmentos por região de interesse (RI) no intestino. Um total de 128 segmentos foram analisados. Cada segmento iniciou a análise por cintilografia seguida de colonoscopia e biópsias. O grau de atividade foi dado segmento por segmento.

Figura 3: Região de interesse no intestino



Referência: Tarjra, João Batista M. et. al. *Annals of Nuclear Medicine*. February 2020. Endereço eletrônico: DOI: 10.1007/s12149-020-01447-w [8]

Oito horas RI 1 = íleo; 24 horas RI 2 = cólon direito e meio cólon transverso; 48 horas RI 3 = cólon esquerdo e metade do cólon transverso; e RI 4 = sigmoide/reto.

3.1 AMOSTRA INDEPENDENTE (A NÍVEL DE INDIVÍDUO)

Tabela 4: Indicadores do teste diagnóstico (a nível de indivíduo)

Estatística (%)	Colonoscopia	Cintilografia
Sensibilidade	62,50	87,50
IC (95%)	40,59 - 81,20	67,64 - 97,34
Especificidade	62,50	37,50
IC (95%)	24,49 - 91,48	8,52 - 75,51
Valores Preditivos Positivos	83,33	80,77
IC (95%)	58,58 - 96,42	60,65 - 93,45
Valores Preditivos Negativos	35,71	50,00
IC (95%)	12,76 - 64,86	11,81 - 88,19
Acurácia	62,50	75,00
IC (95%)	43,69 - 78,90	56,50 - 88,54

De acordo com os indicadores da tabela 4, é possível dizer que a cintilografia possui uma alta sensibilidade, ou seja, raramente deixa de encontrar pessoas doentes. Já a colonoscopia apresentou a mesma proporção moderada para a sensibilidade e especificidade,

enquanto que a cintilografia apresentou uma baixa especificidade, o que pode gerar várias classificações erradas de pessoas sadias como doentes.

Em relação aos valores preditivos, os dois exames apontam VPP acima de 80% e VPN de 50% ou menos, isso significa que, se o teste for positivo, existe mais de 80% de chance do paciente realmente ser doente, apesar do resultado ser positivo. Se o teste for negativo, existe até 50% de chance do paciente ser normal, apesar do resultado ser negativo.

Por último, vale ressaltar que a cintilografia possui acurácia maior que o exame de colonoscopia.

Tabela 5: Concordância entre cintilografia e medidas histológicas (a nível de indivíduo)

Cintilografia	Medidas Histológicas		
	Cicatrização (%)	Atividade (%)	Total (%)
Cicatrização (%)	3 9,38	3 9,38	6 18,75
Atividade (%)	5 15,63	21 65,63	26 81,25
Total (%)	8 25,00	24 75,00	32 100,00

$$\text{Kappa} = 0,2727$$

$$\text{IC} = [-0,1078 - 0,6533]$$

Tabela 6: Concordância entre colonoscopia e medidas histológicas (a nível de indivíduo)

Colonoscopia	Medidas Histológicas		
	Cicatrização (%)	Atividade (%)	Total (%)
Cicatrização (%)	5 15,63	9 28,13	14 43,75
Atividade (%)	3 9,38	15 46,88	18 56,25
Total (%)	8 25,00	24 75,00	32 100,00

$$\text{Kappa} = 0,2000$$

$$\text{IC} = [-0,1185 - 0,5185]$$

Tabela 7: Concordância entre cintilografia e colonoscopia (a nível de indivíduo)

Cintilografia	Colonoscopia		
	Cicatrização (%)	Atividade (%)	Total (%)
Cicatrização (%)	5 15,63	1 3,13	6 18,75
Atividade (%)	9 28,13	17 53,13	26 81,25
Total (%)	14 43,75	18 56,25	32 100,00

Kappa = 0,3220

IC = [0,0332 - 0,6108]

Ao analisar as informações acima, pode-se dizer que a concordância entre as medidas histológicas e o exame de cintilografia, seguido de colonoscopia, é regular uma vez que o Coeficiente Kappa foi de 0,27 e 0,20, respectivamente.

No que diz respeito a concordância entre a cintilografia e a colonoscopia, também podemos afirmar que é regular ao observar o kappa = 0,32.

Figura 4: Curva ROC a nível de indivíduo

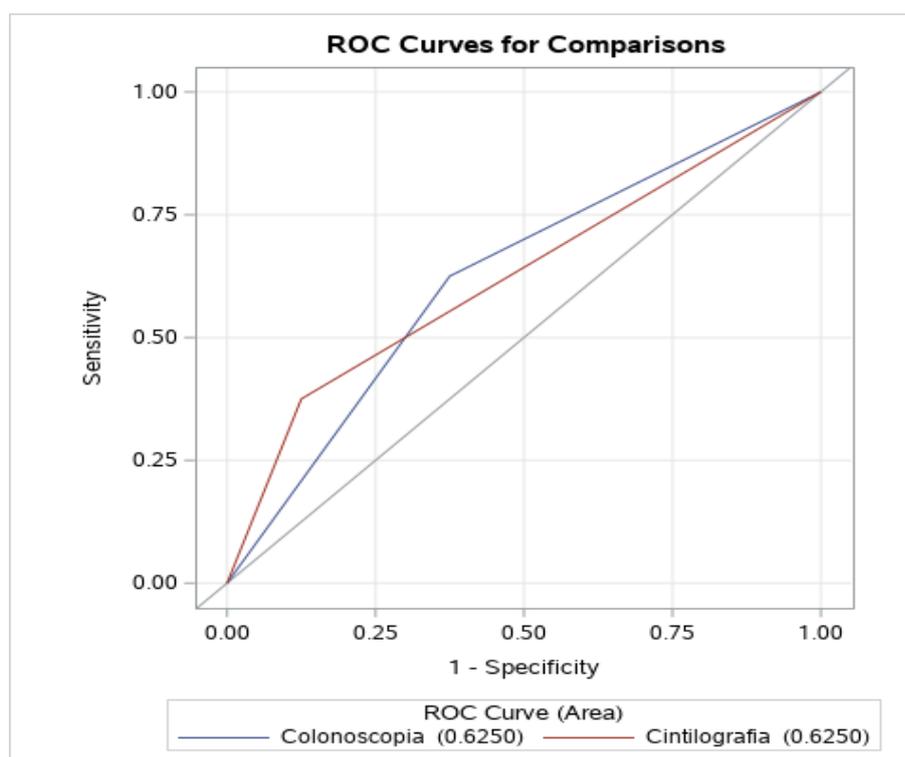


Tabela 8: Estatísticas da associação ROC (a nível de indivíduo)

Modelo ROC	Mann-Whitney		
	Área	Erro Padrão	IC de Wald (95%)
Colonoscopia	0,6250	0,1045	0,4202 0,8298
Cintilografia	0,6250	0,0978	0,4334 0,8166

Tabela 9: Estimativa de contraste ROC e resultados de teste por linha (a nível de indivíduo)

Contraste	Estimativa	Erro Padrão	IC de Wald (95%)		Qui-Quadrado	P-valor
Colonoscopia - Cintilografia	0	0,0982	-0,1924	0,1924	0	1

A curva ROC foi executada para o exame de colonoscopia e cintilografia (Figura 4). A área sob a curva para colonoscopia e cintilografia foi a mesma (0,62) e sugere que o modelo não é bom uma vez que o resultado não é tão distante de 0,5.

Além disso, o p-valor apresentado na tabela 9 indica que não há diferença significativa entre a colonoscopia e o exame cintilográfico.

3.2 DADOS AGRUPADOS (A NÍVEL DE SEGMENTO)

Para os cálculos dos indicadores de Sensibilidade; Especificidade; Valores Preditivos; e Acurácia foi utilizado o Estimador de Razão descrito na seção 3.2.1. Além disso, foram utilizados os códigos em SAS das referências [9] e [10] para o cálculo de todos os indicadores e gerar a curva ROC.

Tabela 10: Indicadores do teste diagnóstico (a nível de segmento)

Estatística (%)	Colonoscopia	Cintilografia
Sensibilidade	48,00	64,00
IC (95%)	32,20 - 63,80	49,73 - 78,27
Especificidade	98,71	96,15
IC (95%)	95,95 - 101,49	92,23 - 100,07
Valores Preditivos Positivos	96,00	91,42
IC (95%)	88,41 - 103,59	82,03 - 100,83
Valores Preditivos Negativos	74,75	80,64
IC (95%)	64,39 - 85,13	70,09 - 91,20
Acurácia	78,90	83,59
IC (95%)	70,35 - 87,45	75,78 - 91,40

Observando a tabela 10, podemos afirmar que a sensibilidade da cintilografia a nível do segmento foi de 64%, maior do que a colonoscopia com apenas 48%. A análise da especificidade de ambos os testes mostrou valores semelhantes com 96% para cintilografia e 98% para colonoscopia. Logo, a sensibilidade e especificidade dos exames de colonoscopia e cintilografia foram semelhantes na identificação de áreas de atividades inflamatórias detectadas na análise histológica.

Já os valores preditivos positivos acima de 90% indicam que existe mais de 90% de chance do paciente com teste positivo ter a doença. Os valores preditivos negativos, de 74,75% para colonoscopia e 80,64% para cintilografia, indicam a probabilidade do indivíduo ser realmente sadio dado que o teste apresentou resultado negativo.

Também vale ressaltar que, assim como a análise a nível de indivíduo, nessa análise a nível de segmento a cintilografia apresentou maior acurácia do teste quando comparada com a colonoscopia.

Tabela 11: Concordância entre cintilografia e medidas histológicas (a nível de segmento)

Cintilografia	Medidas Histológicas		
	Cicatrização (%)	Atividade (%)	Total (%)
Cicatrização (%)	75 58,59	18 14,06	93 72,66
Atividade (%)	3 2,34	32 25,00	35 27,34
Total (%)	78,00 60,94	50,00 39,06	128 100,00

Kappa = 0,6357

IC = [0,4797 - 0,7917]

Tabela 12: Concordância entre colonoscopia e medidas histológicas (a nível de segmento)

Colonoscopia	Medidas Histológicas		
	Cicatrização (%)	Atividade (%)	Total (%)
Cicatrização (%)	77 70,16	26 20,31	103 80,47
Atividade (%)	1 0,78	24 18,75	25 19,53
Total (%)	78,00 60,94	50,00 39,06	128 100,00

Kappa = 0,5132

IC = [0,3466 - 0,6798]

Tabela 13: Concordância entre cintilografia e colonoscopia (a nível de segmento)

Cintilografia	Colonoscopia		
	Cicatrização (%)	Atividade (%)	Total (%)
Cicatrização (%)	87 67,97	6 4,69	93 72,66
Atividade (%)	16 12,50	19 14,84	35 27,34
Total (%)	103 80,47	25 19,53	128 100,00

Kappa = 0,5251

IC = [0,3544 - 0,6958]

Os níveis de concordância foram analisados por meio da estatística kappa não ponderada. Houve concordância substancial entre a cintilografia e a histologia com kappa = 0,63 (Tabela 11), porém a concordância com a colonoscopia foi moderada com Kappa = 0,52. Para fins comparativos, a concordância entre colonoscopia e histologia foi estudada com kappa = 0,51 (Tabela 12).

Figura 5: Curva ROC a nível de segmento

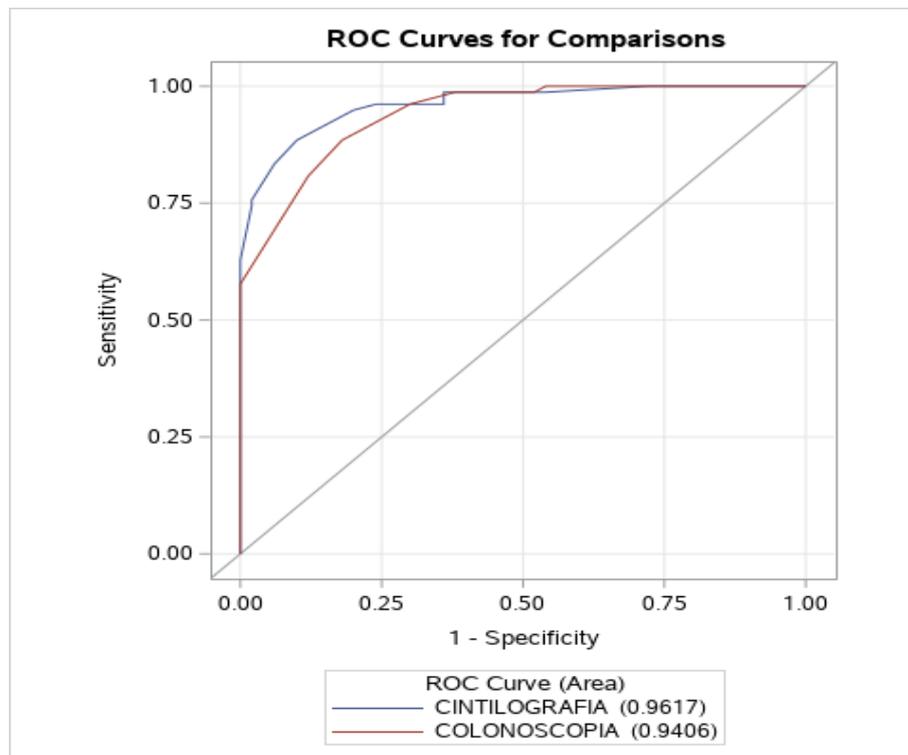


Tabela 14: Estatísticas da associação ROC (a nível de segmento)

Modelo ROC	Mann-Whitney			
	Área	Erro Padrão	IC de Wald (95%)	
Colonoscopia	0,9406	0,0184	0,9047	0,9766
Cintilografia	0,9617	0,0143	0,9336	0,9897

Tabela 15: Estimativa de contraste ROC e resultados de teste por linha (a nível de segmento)

Contraste	Estimativa	Erro Padrão	IC de Wald (95%)		Qui-Quadrado	P-valor
Colonoscopia - Cintilografia	-0,0210	0,0214	-0,0631	0,0210	0,9616	0,3268

A análise da curva ROC foi executada para colonoscopia e cintilografia e foram comparadas entre si (Figura 5). A área sob a curva para colonoscopia foi de 0,94 e para cintilografia de 0,96.

Com o p-valor de 0,3268 é possível dizer que a cintilografia não revelou significativa a partir da área sob a curva ROC em comparação com a colonoscopia para auxiliar na identificação de inflamação.

4 CONCLUSÃO

Este estudo apresentou os resultados dos indicadores de teste diagnóstico analisados primeiramente entre os pacientes e depois por segmento de atividade inflamatória conforme descrito anteriormente (Figura 3). O padrão ouro para comparação analítica foram a atividade histológica e os exames de colonoscopia realizados nos segmentos de análise.

Ao comparar a sensibilidade e especificidade nas duas metodologias, verificamos que os exames são mais sensíveis do que específicos a nível de paciente, enquanto são poucos sensíveis e muito específicos a nível de segmento.

Com os valores preditivos vimos que tanto para amostra independente quanto para dados agrupados, a cintilografia apresenta o melhor VPN visto que é mais sensível (redução de falsos negativos) e a colonoscopia tem o melhor VPP por ser um exame mais específico (redução de falsos positivos). Entretanto, apesar do estudo a nível de segmento ter mais especificidade do que sensibilidade, ao contrário da análise a nível de indivíduo, temos altos valores preditivos (positivos e negativos) nos dois exames.

Observamos que o nível de concordância entre histologia e cintilografia a nível de segmento foi $\kappa = 0,63$ (IC 95%, 0,47 - 0,79) e a nível de indivíduo foi $\kappa = 0,27$ (IC 95%, -0,10 - 0,65), o que indica concordância substancial entre os dois procedimentos. Quando analisamos o nível de concordância com a colonoscopia, a taxa foi um pouco menor. O resultado no segmento indicou um coeficiente κ de 0,52 (IC 95%, 0,35 - 0,69). Esta avaliação indica uma concordância moderada entre os exames de colonoscopia e cintilografia em comparação com a histologia como padrão-ouro de diagnóstico.

A área sob a curva ROC com cintilografia não difere significativamente da área sob a curva ROC com colonoscopia na análise a nível de indivíduo uma vez que ambos os exames apresentaram o resultado de 0,62 (Figura 4). Para o caso da análise a nível de segmento, a AUC da cintilografia foi de 0,96 e a AUC da colonoscopia foi de 0,94 (Figura 5). Além disso, o p-valor obtido foi de 0,32, o que sugere que a cintilografia e a colonoscopia são igualmente precisas na detecção de inflamação ativa.

Neste trabalho, resumimos os métodos estatísticos comumente usados para calcular os indicadores de teste diagnóstico para dados agrupados, ajustando para a possível correlação entre as observações dentro de cada paciente (cluster). Em nosso estudo, os métodos produziram resultados comparáveis. No entanto, para outras aplicações, os resultados podem variar mais dependendo do método de análise usado, a magnitude da correlação e o número de observações por cluster.

Vale ressaltar que se o efeito de agrupamento não for levado em consideração, a variância (limite de confiança) será enviesada visto que o cálculo ignora a correlação positiva entre os segmentos do intestino grosso dentro do paciente.

Referências

- [1] Agresti, A. An Introduction to Categorical Data Analysis 3rd edn. John Wiley Sons, 2019.
- [2] Associação Brasileira de Colite Ulcerativa e Doença de Crohn. ABCD, 2017. Sobre a Doença de Crohn. Disponível em <<https://www.abcd.org.br/sobre-a-doenca-de-crohn/>>. Acesso em: 04/09/2020.
- [3] Hedeker, D; Gibbons, R. D. Longitudinal Data Analysis. John Wiley Sons, 2006.
- [4] Honghu Liu, Tongtong Wu. Estimating the Area under a Receiver Operating Characteristic (ROC) Curve For Repeated Measures Design. Journal of Statistical Software, Vol. 8, 2003. Endereço eletrônico: DOI: 10.18637/jss.v008.i12
- [5] Jama. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA, v. 268, n. 17, 1992.
- [6] Jiang, J. Linear and Generalized Linear Mixed Models and Their Applications. New York, Springer, 2007.
- [7] Mason, S. J., Graham, N. E.(2002). Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. Quarterly Journal of the Royal Meteorological Society (128): 2145–2166.
- [8] Tarjra, João Batista M. et. al. Accuracy of Oral ⁶⁷Gallium Citrate Scintigraphy in assessment of inflammatory activity of Crohn's disease. Annals of Nuclear Medicine, Fevereiro 2020. Endereço eletrônico: DOI: 10.1007/s12149-020-01447-w
- [9] William F. McCarthy, Nan Guo. The Estimation of Sensitivity and Specificity of Clustered Binary Data. SAS SUGI proceedings: Statistics and Data Analysis, Paper 206-31.
- [10] Zhao Yanga, Ming Zhou. Kappa statistic for clustered matched-pair data. Statistics in Medicine, 2014. Endereço eletrônico: DOI: 10.1002/sim.6113

Anexos

ANEXO A – Código SAS

Programa SAS com Macro para computação de sensibilidade e especificidade a nível de segmento.

```
data dslkup;
  format err 5.3 tail1 tail2 5.3;
  input err tail1 tail2 @@;
cards;
0.5 0 0.674 0.4 0.253 0.842 0.3 0.524 1.036 0.2 0.842 1.282
0.1 1.282 1.645 0.05 1.645 1.96 0.025 1.96 2.248 0.01 2.326 2.576
0.005 2.576 2.813;
run;

*****
* Segment-Level. Sensitivity & Specificity *
*****

proc import out=mysasdata
datafile="/home/luizacaest0/TCC/Segmento/dados s.xlsx"
dbms=xlsx replace;
sheet='Planilha1';
run;

data sorted; set mysasdata;
ptid=id;
artery=segmento;
invasive=histologia;
mdct=CINTILOGRAFIA;
output;
run;

proc print;
run;

proc sort data=sorted;by ptid;
run;
```

```
proc freq;
tables histologia*CINTILOGRAFIA/nopct;
run;

data segm_tbl2a (keep=ptid TPi segmEvnts ) /* Sensitivity data set */
  segm_tbl2b (keep=ptid NotTPi NotsegmEvnts) /* Specificity data set */;
retain TPi segmEvnts NotTPi NotsegmEvnts ;
set sorted; /* We assume that data set has been sorted already */
by ptid;
if first.ptid then
do;
TPi =0; /* No of TPi */
segmEvnts =0; /* No. of segment (Ni) with Event */
NotTPi =0; /* No of TNi */
NotsegmEvnts=0; /* No. of segment (Ni) without Event*/
end;
segmEvnts =sum(segmEvnts,(invasive=1) );
TPi =sum(TPi, (invasive=1 & MDct=1));
NotsegmEvnts=sum(NotsegmEvnts, (invasive=0));
NotTPi =sum(NotTPi, (invasive=0 & MDct=0));
if last.ptid then
do;
if TPi>=0 & segmEvnts>=0 then output segm_tbl2a; * Senticitivity *;
if NotTPi>=0 & NotsegmEvnts>=0 then output segm_tbl2b; * Specificity *; end;
label
ptid = "Patient ID"
TPi = "No of TPi"
segmEvnts = "No. of segment (Ni) with Event"
NotTPi = "No of TNi"
NotsegmEvnts = "No. of segment (Ni) without Event";
run;

/* Cálculo Acurácia e IC 95 % */
data b;merge segm_tbl2a segm_tbl2b;
TPi=TPi+NotTPi;
Ni=segmEvnts + NotsegmEvnts;
Sei=TPi/Ni;
data c;set b;
SE_numerador = Ni*Sei;
run;
```

```
proc print;
var SE_numerador Ni;
sum SE_numerador Ni;
run;

data d;set c;
SE = 0.883928571428571;
N_barra = 4;
run;

data e;set d;
Numerador = (Ni/N_barra)**2 * (Sei - SE)**2;
proc print;
var Numerador;
sum Numerador;
run;

data f;set e;
variancia_ACC = 0.93527/(28*27);
run;

data g;set f;
if _N_ = 1;
li = SE - 1.96*sqrt (variancia_ACC);
ls = SE + 1.96*sqrt(variancia_ACC);
run;

proc print;
var li SE ls;
run;

%macro Calc_sum(ds,Dxac);
%global sumT sumN I &Dxac meanN;
data _null_;
retain sum&Dxac.T sum&Dxac.N;
set &ds end=EOF;
%if _N_ eq 1 %then
%do;sum&Dxac.T=0;sum&Dxac.N=0;%end;
%if %upcase(&Dxac)=SE %then
```

```
%do;sum&Dxac.T =sum(sum&Dxac.T,TPi);sum&Dxac.N =sum(sum&Dxac.N,SegmEvnts);%end;
%else %if %upcase(&Dxac)=SP %then
%do;sum&Dxac.T=sum(sum&Dxac.T,NotTPi);sum&Dxac.N=sum(sum&Dxac.N,NotSegmEvnts);
%end;
if EOF then
do;
call symputx('sumT',sum&Dxac.T);
call symputx('sumN',sum&Dxac.N);
call symputx('I',_N_);
call symputx("&Dxac",sum&Dxac.T/sum&Dxac.N);
call symputx('meanN',sum&Dxac.N/_N_);
end;
run;

%mend calc_sum;
* --- Sensitivity --- *;
%calc_sum(segm_tbl2a,se)
data segm_se;
retain totpart3;
set segm_tbl2a end=EOF;
if _N_ eq 1 then totpart3=0;
Sei =TPi/SegmEvnts;
NidivN =SegmEvnts/&meanN;
part3 =(NiDivN)**2*(Sei-&se)**2;
totpart3=sum(totpart3,part3);
if EOF then call symputx('Varse',put(1/(&I*(&I-1))*totpart3,7.4));
*if _N_ le 5 then put _all_;
label ptid = "Patient ID"
TPi = "No. of TPi"
SegmEvnts = "No. of Segments (Ni) with Event"
Sei = "Sei"
NidivN = "Ni/Mean N"
part3 = "(Ni/mean N)**2 * (Sei-Se)**2";
run;

* --- Specificity --- *;
%calc_sum(segm_tbl2b,sp)
data segm_sp;
retain totpart3;
set segm_tbl2b end=EOF;
```

```

if _N_ eq 1 then totpart3=0;
Spi =NotTPi/NotSegmEvnts;
NidivN =NotSegmEvnts/&meanN;
part3 =(NiDivN)**2*(Spi-&sp)**2;
totpart3=sum(totpart3,part3);
if EOF then call symputx('Varsp',put(1/(&I*(&I-1))*totpart3,7.4));
label ptid = "Patient ID"
NotTPi = " No of TNi"
NotsegmEvnts = "No. of Segments (Ni) without Event"
Spi = "Spi"
NidivN = "Ni/Mean N"
part3 = "(Ni/mean N)**2 * (Spi-Sp)**2";
run;

%macro cau_Z(conf=95);
%global ci Za sen sep spn spp;
%let ci=&conf; %let Za=; %let sen=;%let sep=;%let spn=;%let spp=;
data _null_;
set dslookup (where=(err=%sysevalf(1-&conf/100))); /* Look-up Table */
call symputx('Za',tail2);
run;

data temp;
sqrtsen=sqrt(input("&varse",8.));
sen=%sysevalf(&se)-%sysevalf(&Za)*sqrtsen;
sep=%sysevalf(&se)+%sysevalf(&Za)*sqrtsen;
call symputx('sen',put(sen,7.4));
call symputx('sep',put(sep,7.4));
sqrtsep=sqrt(input("&varsp",8.));
spn=%sysevalf(&sp)-%sysevalf(&Za)*sqrtsep;
spp=%sysevalf(&sp)+%sysevalf(&Za)*sqrtsep;
call symputx('spn',put(spn,7.4));
call symputx('spp',put(spp,7.4));
run;

%mend cau_Z;
%cau_Z; /* Default macro variable is 95 */
options nodate nonumber orientation=landscape;
proc template; /* Creating a template */
define style myhtml;

```

```
parent=Styles.html;
replace Body from Document
"Controls the Body file." /
bottommargin = 0.25in
topmargin = 0.10in
rightmargin = 0.10in
leftmargin = 0.25in;
end;
run;

ods listing close; /* ODS starts */
proc print data=segm_se noobs label /*data=all*/
  style(header)={font_face=Arial font_weight=bold font_size=8pt
  just=c background=white} style(data)={just=center font_face=Arial font_size=8pt};
var ptid Tpi SegmEvnts Sei NidivN part3;
run;

ods escapechar='\'';
ods html text='\R/html"\ul" Segment-Level ';
ods html text="\S={just=c font_size=10pt} Sensitivity &se";
ods html text="\S={just=c font_size=10pt} Variance of Sensitivity &varse";
ods escapechar='^';
ods html text="^S={just=c font_size=10pt} &CI% confidence interval for sensitivity
[&sen, &sep]";

proc print data=segm_sp noobs label /*data=all*/
  style(header)={font_face=Arial font_weight=bold font_size=8pt
  just=c background=white} style(data)={just=center font_face=Arial font_size=8pt};
var ptid NotTPi NotsegmEvnts Spi NidivN part3;
run;

ods escapechar='\'';
ods html text='\R/html"\ul" Segment-Level ';
ods html text="\S={just=c font_size=10pt} Specificity &sp";
ods html text="\S={just=c font_size=10pt} Variance of Specificity &varsp";
ods escapechar='^';
ods html text="^S={just=c font_size=10pt} &CI% confidence interval for specificity
[&spn, &sp]";
ods html close;
ods listing; /* ODS Ends */
```

Programa SAS com Macro para computação de valor preditivo positivo e negativo a nível de segmento.

```
data dslkup;
  format err 5.3 tail1 tail2 5.3;
  input err tail1 tail2 @@;
cards;
0.5 0 0.674 0.4 0.253 0.842 0.3 0.524 1.036 0.2 0.842 1.282
0.1 1.282 1.645 0.05 1.645 1.96 0.025 1.96 2.248 0.01 2.326 2.576
0.005 2.576 2.813
;run;

*****
* Segment-Level. Sensitivity & Specificity *
*****;

proc import out=mysasdata
  datafile="/home/luizacaest0/TCC/Segmento/dados s.xlsx"
  dbms=xlsx replace;
  sheet='Planilha1';
run;

data sorted; set mysasdata;
ptid=id;
artery=segmento;
invasive=histologia;
mdct=CINTILOGRAFIA;
output;
run;

proc sort data=sorted;by ptid;
run;

proc freq;
tables histologia*CINTILOGRAFIA;
run;

data segm_tbl2a (keep=ptid TPi segmEvnts ) /* Sensitivity data set */
  segm_tbl2b (keep=ptid NotTPi NotsegmEvnts) /* Specificity data set */;
```

```
retain TPi segmEvnts NotTPi NotsegmEvnts ;
set sorted; /* We assume that data set has been sorted already */
by ptid;
if first.ptid then
do;
TPi =0; /* No of TPi */
segmEvnts =0; /* No. of segment (Ni) with Event */
NotTPi =0; /* No of TNi */
NotsegmEvnts=0; /* No. of segment (Ni) without Event*/
end;
segmEvnts =sum(segmEvnts,(MDct=1) );
TPi =sum(TPi, (invasive=1 & MDct=1));
NotsegmEvnts=sum(NotsegmEvnts, (MDct=0));
NotTPi =sum(NotTPi, (invasive=0 & MDct=0));
if last.ptid then
do;
if TPi >=0 & segmEvnts >=0 then output segm_tbl2a; * Senticivity *;
if NotTPi >= 0 & NotsegmEvnts >=0 then output segm_tbl2b; * Specificity *;
end;
label
ptid = "Patient ID"
TPi = "No of TPi"
segmEvnts = "No. of segment (Ni) with Event"
NotTPi = "No of TNi"
NotsegmEvnts = "No. of segment (Ni) without Event";
run;

%macro Calc_sum(ds,Dxac);
%global sumT sumN I &Dxac meanN;
data _null_;
retain sum&Dxac.T sum&Dxac.N;
set &ds end=EOF;
%if _N_ eq 1 %then
%do;sum&Dxac.T=0;sum&Dxac.N=0;%end;
%if %upcase(&Dxac)=SE %then
%do;sum&Dxac.T =sum(sum&Dxac.T,TPi);sum&Dxac.N =sum(sum&Dxac.N,SegmEvnts);%end;
%else %if %upcase(&Dxac)=SP %then
%do;sum&Dxac.T=sum(sum&Dxac.T,NotTPi);sum&Dxac.N=sum(sum&Dxac.N,NotSegmEvnts);
%end;
if EOF then
```

```

do;
call symputx('sumT',sum&Dxac.T);
call symputx('sumN',sum&Dxac.N);
call symputx('I',_N_);
call symputx("&Dxac",sum&Dxac.T/sum&Dxac.N);
call symputx('meanN',sum&Dxac.N/_N_);
end;
run;

%mend calc_sum;

* --- Sensitivity --- *;
%calc_sum(segm_tbl2a,se)
data segm_se;
retain totpart3;
set segm_tbl2a end=EOF;
if _N_ eq 1 then totpart3=0;
Sei =TPi/SegmEvnts;
NidivN =SegmEvnts/&meanN;
part3 =(NiDivN)**2*(Sei-&se)**2;
totpart3=sum(totpart3,part3);
if EOF then call symputx('Varse',put(1/(&I*(&I-1))*totpart3,7.4));
*if _N_ le 5 then put _all_;
label ptid = "Patient ID"
TPi = "No. of TPi"
SegmEvnts = "No. of Segments (Ni) with Event"
Sei = "Sei"
NidivN = "Ni/Mean N"
part3 = "(Ni/mean N)**2 * (Sei-Se)**2";
run;

* --- Specificity --- *;
%calc_sum(segm_tbl2b,sp)
data segm_sp;
retain totpart3;
set segm_tbl2b end=EOF;
if _N_ eq 1 then totpart3=0;
Spi =NotTPi/NotSegmEvnts;
NidivN =NotSegmEvnts/&meanN;
part3 =(NiDivN)**2*(Spi-&sp)**2;

```

```
totpart3=sum(totpart3,part3);
if EOF then call symputx('Varsp',put(1/(&I*(&I-1))*totpart3,7.4));
label ptid = "Patient ID"
NotTPi = " No of TNi"
NotsegmEvnts = "No. of Segments (Ni) without Event"
Spi = "Spi"
NidivN = "Ni/Mean N"
part3 = "(Ni/mean N)**2 * (Spi-Sp)**2";
run;

%macro cau_Z(conf=95);
%global ci Za sen sep spn spp;
%let ci=&conf; %let Za=; %let sen=;%let sep=;%let spn=;%let spp=;
data _null_;
set dslookup (where=(err=%sysevalf(1-&conf/100))); /* Look-up Table */
call symputx('Za',tail2);
run;

data temp;
sqrtsen=sqrt(input("&varse",8.));
sen=%sysevalf(&se)-%sysevalf(&Za)*sqrtsen;
sep=%sysevalf(&se)+%sysevalf(&Za)*sqrtsen;
call symputx('sen',put(sen,7.4));
call symputx('sep',put(sep,7.4));
sqrtsep=sqrt(input("&varsp",8.));
spn=%sysevalf(&sp)-%sysevalf(&Za)*sqrtsep;
spp=%sysevalf(&sp)+%sysevalf(&Za)*sqrtsep;
call symputx('spn',put(spn,7.4));
call symputx('spp',put(spp,7.4));
run;

%mend cau_Z;
%cau_Z; /* Default macro variable is 95 */
options nodate nonumber orientation=landscape;
proc template; /* Creating a template */
define style myhtml;
parent=Styles.html;
replace Body from Document
"Controls the Body file." /
bottommargin = 0.25in
```

```
topmargin = 0.10in
rightmargin = 0.10in
leftmargin = 0.25in;
end;
run;

ods listing close; /* ODS starts */
proc print data=segm_se noobs label /*data=all*/
  style(header)={font_face=Arial font_weight=bold font_size=8pt
  just=c background=white} style(data)={just=center font_face=Arial font_size=8pt};
var ptid Tpi SegmEvnts Sei NidivN part3;
run;

ods escapechar='\';
ods html text='\R/html"\ul" Segment-Level ';
ods html text="\S={just=c font_size=10pt} VPP &se";
ods html text="\S={just=c font_size=10pt} Variance of VPP &varse";
ods escapechar='^';
ods html text="^S={just=c font_size=10pt} &CI% confidence interval for VPP
[&sen, &sep]";
proc print data=segm_sp noobs label /*data=all*/
  style(header)={font_face=Arial font_weight=bold font_size=8pt
  just=c background=white} style(data)={just=center font_face=Arial font_size=8pt};
var ptid NotTPi NotsegmEvnts Spi NidivN part3;
run;

ods escapechar='\';
ods html text='\R/html"\ul" Segment-Level ';
ods html text="\S={just=c font_size=10pt} VPN &sp";
ods html text="\S={just=c font_size=10pt} Variance of VPN &varsp";
ods escapechar='^';
ods html text="^S={just=c font_size=10pt} &CI% confidence interval for VPN
[&spn, & spp]";
ods html close;
ods listing; /* ODS Ends */
```



```
%macro ClusterKappa (dsin = , /* the input dataset */
  clusV = , /* the cluster variable*/
  unitV = , /* the unit variable*/
  Resp1V = , /* Response variable for procedure 1, Row variable
  need to be numeric integer and > 0 */
  Resp2V = /* Response variable for procedure 2, Column Variable
  need to be numeric integer and > 0 */);

proc sql noprint;
create table clustDS as
select distinct &clusV from &dsin
order by &clusV; quit;

data clustDS;
  set clustDS; clusterID + 1;
proc sort data = &dsin;
  by &clusV &unitV;
data resp_;
  merge &dsin clustDS;
  by &clusV;
proc sort data = resp_;
  by clusterID &unitV;
data F_Ana_DS (drop = &clusV &unitV rename = (clusterID = cluster));
  set resp_;
  by clusterID &unitV;
  if first.clusterID then subject = 0;
  subject + 1;
data anal_D;
  set F_Ana_DS (rename = (&Resp1V = resp1__ &Resp2V = resp2__ ));
  maxR = max(resp1__, resp2__);
run;

proc sql noprint;
  select max(maxR) into: N_Cat /*** to get the # of response category */
  from anal_D;
  select max(cluster) into: nclus /*** to get the # of clusters ***/
  from anal_D;
create table count1 as /*this is the calculate the marginal total for each
  cluster of treatment 1 */
```

```
select distinct cluster, resp1__, count(*) as count1
from anal_D
group by cluster, resp1__;
create table count2 as /*this is the calculate the marginal total for each
cluster of treatment 2 */
select distinct cluster, resp2__, count(*) as count2
from anal_D
group by cluster, resp2__;
quit;

*** this is to construct the dataset with 0 in each combination;
data constr1; /*** generate a dataset with full combination of tabulation */
do cluster = 1 to &nclus;
do resp1__ = 1 to &N_Cat;
count1 = 0; output;
end;
end;

proc sort data = constr1;
by cluster resp1__;
data count1; /** merge the datasets to get the full combination of tabulation */
merge constr1 count1 ;
by cluster resp1__;
data constr2;
do cluster = 1 to &nclus;
do resp2__ = 1 to &N_Cat;
count2 = 0; output;
end;
end;

proc sort data = constr2;
by cluster resp2__;
data count2; /** merge the datasets to get the full combination of tabulation */
merge constr2 count2 ;
by cluster resp2__;
run; quit;

proc summary data = count1; /*** contains the # of total subjects */
class cluster;
var count1;
```

```

output out = total (drop = _freq_ where = (_TYPE_ = 0) ) sum = total ;
data _null_;
set total (keep = total);
call symput("total", total);
run;

proc summary data = count1; /*contains the # of total subjects for each cluster;*/
class cluster;
var count1;
output out = n_k (drop = _freq_ where = (_TYPE_ = 1) ) sum = n_k ;
proc sort data = n_k (drop = _type_) out = n_k;
by cluster;
run;

data concordA;
set anal_D;
if resp1__ = resp2__ then freq = 1; else freq = 0;
proc summary data = concordA; /*to calculate the concordant pair at the diagonal
of the k x k table;*/
class cluster;
var freq;
output out = concord (drop = _freq_ where = (_TYPE_ = 1) ) sum = n_ii_Sk ;
proc sort data = concord (drop = _type_) out = concord;
by cluster;
run;

%macro outputM(countds = );
%do i = 1 %to &N_Cat;
%let seqD = %sysfunc( substr( &countds, 6, 1));
data margin&seqD.&i (rename=(count&seqD=%if &seqD=1 %then %do; n&i._k %end;
%else %do; n_&i.k %end;) drop=resp&seqD.__);
set &countds;
if resp&seqD.__ = &i then output;
proc sort data = margin&seqD.&i;
by cluster;
run;
%end;

%mend;

%outputM(countds = Count1); /** output the marginal total for treatment 1*/;

```

```
%outputM(countds = Count2); /*** output the marginal total for treatment 2*/;

data finalD;
merge N_k
%do i = 1 %to &N_Cat;
margin1&i margin2&i
%end; concord;
by cluster;
data finalD;
set finalD;
BigN = &total;
run;

data finalD_A (drop = i);
set finalD;
array varlistA {*} %do i = 1 %to &N_Cat; n&i._k n_&i.k %end; n_ii_Sk;
array varlistB {*} %do i = 1 %to &N_Cat; p&i._k p_&i.k %end; p_ii_Sk;
do i = 1 to dim(varlistA);
varlistB {i} = varlistA {i}/n_k;
end;
omega_k = n_k/BigN;
run;

*** calculation using the Kappa under the independence (non-clustered situation);
data fakestr;
do resp1__ = 1 to &N_Cat;
do resp2__ = 1 to &N_Cat;
count = 0; output;
end;
end;
proc sort data = fakestr;
by resp1__ resp2__;
run;

proc summary data = anal_D;
class resp1__ resp2__;
output out = interm_D (where = (_TYPE_ = 3) rename = (_FREQ_ = count) );
proc sort data = interm_D (drop = _TYPE_) out = interm_D;
by resp1__ resp2__;
data A_Response;
```

```

merge fakestr interm_D;
by resp1__ resp2__;
run;

title2 "*** Part 1: Kappa statistics assuming independence *** ";
proc freq data = A_Response;
tables resp1__ * resp2__/agree;
weight count/zeros;
run;

proc iml;
* Read data into IML ;
use finalD_A; read all ;
kappOD = J( 1, 6, .);
*** this is used to calculate the point estimation;
%do i = 1 %to &N_Cat;
p&i._ = omega_k' * p&i._k;
p_&i = omega_k' * p_&i.k;
%end;

P_0 = omega_k' * p_ii_Sk;
P_E = P1_ * P_1 %do i = 2 %to &N_Cat; + P&i._ * P_&i %end; ;

kappa = (P_0 - P_E) / (1 - P_E); ** this is the calculated Kappa value;

*** end of point estimation calculation;

*** this part is for the calculation of the variance based on Delta method;
PO_V = p_ii_Sk;
P_V = p1_k %do i = 2 %to &N_Cat; || p&i._k %end; ;
PV_ = p_1k %do i = 2 %to &N_Cat; || p_&i.k %end; ;

omega_k2 = omega_k # omega_k;

BigOmega = (I(&nclus) - omega_k * J(1, &nclus, 1)) * diag(omega_k2) * (I(&nclus)
- J(&nclus,1,1)*t(omega_k));

V11 = PO_V' * BigOmega * PO_V;
V12 = PO_V' * BigOmega * (P_V * PV_' + PV_ * P_V') * omega_k;
V21 = V12;

```

```

V22 = omega_k' * (PV_ * P_V' * BigOmega * P_V * PV_' + P_V * PV_' * BigOmega *
  PV_ * P_V' + PV_ * P_V' * BigOmega * PV_ * P_V' + P_V * PV_' * BigOmega *
  P_V * PV_') * omega_k;

VMatrix = J( 2, 2, .);
LVector = J( 1, 2, .);
VMatrix[1,1] = V11;
VMatrix[1,2] = V12;
VMatrix[2,1] = V21;
VMatrix[2,2] = V22;
VMatrix = (&nclus # &nclus) # VMatrix/(&nclus - 1);

LVector[1] = 1 / (1 - P_E);
LVector[2] = (P_0 - 1) / ( (1 - P_E) # (1 - P_E) );

ASE_Del = sqrt(LVector * VMatrix * LVector' / &nclus); *variance using Delta method;

lower = kappa - 1.96 # ASE_Del;
upper = kappa + 1.96 # ASE_Del;

* print V11, V12, V21, V22 VMatrix Var_Del P_0 kappa;

* end of the calculation of the variance based on Delta method;

* this part is for pooling the final results information;

title2 "*** Part 2: Kappa statistics for clustered matched-pair data *** ";

print P_0 P_E kappa ASE_Del[label="Standard Error"],,
  lower[label="95% CI Lower limit"] upper[label="95% CI Upper limit"];
quit;

%mend;

/***** Here is the example-1 *****/
proc import out=mysasdata
  datafile="/home/luizacaest0/TCC/Segmento/dados s.xlsx"
  dbms=xlsx replace;
  sheet='Planilha1';

```

```
run;

data a;set mysasdata;
  if histologia = . or CINTILOGRAFIA = . then delete;
run;

proc print;
run;

proc format;
  value resp 0 = 'Cicatrização'
            1 = 'Atividade';
run;

proc freq data=a;
  tables CINTILOGRAFIA*histologia/nocol norow agree;
  format CINTILOGRAFIA resp. histologia resp.;
run;

proc freq data=a;
  tables colonoscopia*histologia/nocol norow agree;
  format colonoscopia resp. histologia resp.;
run;

proc freq data=a;
  tables cintilografia*colonoscopia/nocol norow agree;
  format cintilografia resp. colonoscopia resp.;
run;

data b; set a;
  cluster=id;
  subject=segmento;
  trt2=1;
  resp2=histologia;
  trt1=2;
  resp1=CINTILOGRAFIA;
  output;
run;
```

```
proc sort data=b;by cluster id;
run;

data response (keep=cluster subject trt2 resp2 trt1 resp1);
  set b;
  if Resp1 = 1 then Resp1 = 2; else Resp1 = 1;
  if Resp2 = 1 then Resp2 = 2; else Resp2 = 1;
run;

%ClusterKappa(dsin=response, clusV=cluster, unitV=subject,
  Resp1V=Resp1, Resp2V=Resp2);
```

Programa SAS para computação da acurácia a nível de segmento.

```
proc import out=mysasdata
datafile="/home/luizacaest0/TCC/Segmento/dados s.xlsx"
dbms=xlsx replace;
sheet='Planilha1';
run;

data sorted; set mysasdata;
ptid=id;
artery=segmento;
invasive=histologia;
mdct=CINTILOGRAFIA;
output;
run;

proc print;
run;

proc sort data=sorted;by ptid;
run;

proc freq;
tables histologia*CINTILOGRAFIA/nopct;
run;

data segm_tbl2a (keep=ptid TPi segmEvnts ) /* Sensitivity data set */
  segm_tbl2b (keep=ptid NotTPi NotsegmEvnts) /* Specificity data set */;
retain TPi segmEvnts NotTPi NotsegmEvnts ;
set sorted; /* We assume that data set has been sorted already */
by ptid;
if first.ptid then
do;
TPi =0; /* No of TPi */
segmEvnts =0; /* No. of segment (Ni) with Event */
NotTPi =0; /* No of TNi */
NotsegmEvnts=0; /* No. of segment (Ni) without Event*/
end;
segmEvnts =sum(segmEvnts,(invasive=1) );
```

```
TPi =sum(TPi, (invasive=1 & MDct=1));
NotsegmEvnts=sum(NotsegmEvnts, (invasive=0));
NotTPi =sum(NotTPi, (invasive=0 & MDct=0));
if last.ptid then
do;
if TPi >=0 & segmEvnts >=0 then output segm_tbl2a; * Sentitivity *;
if NotTPi >= 0 & NotsegmEvnts >=0 then output segm_tbl2b; * Specificity *;
end;
label
ptid = "Patient ID"
TPi = "No of TPi"
segmEvnts = "No. of segment (Ni) with Event"
NotTPi = "No of TNi"
NotsegmEvnts = "No. of segment (Ni) without Event";
run;

/* Cálculo Acurácia e IC 95 % */
data b;merge segm_tbl2a segm_tbl2b;
TPi=TPi+NotTPi;
Ni=segmEvnts + NotsegmEvnts;
Sei=TPi/Ni;
data c;set b;
SE_numerador = Ni*Sei;
run;

proc print;
var SE_numerador Ni;
sum SE_numerador Ni;
run;

data d;set c;
SE = 0.8359375;
N_barra = 4;
run;

data e;set d;
Numerador = (Ni/N_barra)**2 * (Sei - SE)**2;
proc print;
var Numerador;
sum Numerador;
```

```
run;
```

```
data f;set e;  
variancia_ACC = 1.57617/(32*31);  
run;
```

```
data g;set f;  
if _N_ = 1;  
li = SE - 1.96*Sqrt (variancia_ACC);  
ls = SE + 1.96*SQRT(variancia_ACC);  
run;
```

```
proc print;  
var li SE ls variancia_ACC;  
run;
```

Programa SAS para computação da curva roc a nível de segmento.

```
proc import out=mysasdata
datafile="/home/luizacaest0/TCC/Segmento/dados s.xlsx"
dbms=xlsx replace;
sheet='Planilha1';
run;

data sorted; set mysasdata;
ptid=id;
artery=segmento;
invasive=histologia;
mdct=CINTILOGRAFIA;
output;
run;

proc print;
run;

proc sort data=sorted;by ptid;
run;

proc freq;
tables histologia*CINTILOGRAFIA/nopct;
run;

data segm_tbl2a (keep=ptid TPi segmEvnts ) /* Sensitivity data set */
  segm_tbl2b (keep=ptid NotTPi NotsegmEvnts) /* Specificity data set */;
retain TPi segmEvnts NotTPi NotsegmEvnts ;
set sorted; /* We assume that data set has been sorted already */
by ptid;
if first.ptid then
do;
TPi =0; /* No of TPi */
segmEvnts =0; /* No. of segment (Ni) with Event */
NotTPi =0; /* No of TNi */
NotsegmEvnts=0; /* No. of segment (Ni) without Event*/
end;
segmEvnts =sum(segmEvnts,(invasive=1) );
```

```
TPi =sum(TPi, (invasive=1 & MDct=1));
NotsegmEvnts=sum(NotsegmEvnts, (invasive=0));
NotTPi =sum(NotTPi, (invasive=0 & MDct=0));
if last.ptid then
do;
if TPi >=0 & segmEvnts >=0 then output segm_tbl2a; * Sentitivity *;
if NotTPi >= 0 & NotsegmEvnts >=0 then output segm_tbl2b; * Specificity *;
end;
label
ptid = "Patient ID"
TPi = "No of TPi"
segmEvnts = "No. of segment (Ni) with Event"
NotTPi = "No of TNi"
NotsegmEvnts = "No. of segment (Ni) without Event";
run;

/* Cálculo Acurácia e IC 95 % */
data b;merge segm_tbl2a segm_tbl2b;
TPi=TPi+NotTPi;
Ni=segmEvnts + NotsegmEvnts;
Sei=TPi/Ni;
data c;set b;
SE_numerador = Ni*Sei;
run;

proc print;
var SE_numerador Ni;
sum SE_numerador Ni;
run;

data d;set c;
SE = 0.8359375;
N_barra = 4;
run;

data e;set d;
Numerador = (Ni/N_barra)**2 * (Sei - SE)**2;
proc print;
var Numerador;
sum Numerador;
```

```
run;
```

```
data f;set e;  
variancia_ACC = 1.57617/(32*31);  
run;
```

```
data g;set f;  
if _N_ = 1;  
li = SE - 1.96*Sqrt (variancia_ACC);  
ls = SE + 1.96*SQRT(variancia_ACC);  
run;
```

```
proc print;  
var li SE ls variancia_ACC;  
run;
```