

Escolhendo o ponto de corte mais adequado para variáveis contínuas

Choosing the most appropriate cut-point for continuous variables

FRANCISCO TUSTUMI, TCBC-SP¹⁻³ 

Genericamente, as variáveis podem ser classificadas como categóricas ou contínuas. As variáveis categóricas são um número finito de categorias, como sexo e estágio oncológico (I-IV). Variáveis contínuas são variáveis numéricas com infinitos valores possíveis entre quaisquer dois valores.

Vários biomarcadores usados rotineiramente na prática clínica são contínuos, como contagem de glóbulos vermelhos e brancos, expressão de Ki-67, índice de massa corporal (IMC), nível de antígeno carcinoembrionário (CEA), captação de SUVmax, entre outros.

Os pontos de corte podem ser usados em variáveis contínuas para “discretizar” um biomarcador em diferentes categorias, fornecendo referências pelas quais os indivíduos serão classificados em grupos. A grande vantagem da aplicação de pontos de corte é que o uso de limiares nos parâmetros facilita a tomada de decisão. Na prática clínica, os médicos precisam saber em que grupo seus pacientes estão localizados para estabelecer o diagnóstico, tratamento ou prognóstico adequado. “Esse valor do Ki-67 é um sinal de preocupação ou não?”; “Esse IMC exige ou não procedimento bariátrico?”; “Esse valor de SUVmax é positivo ou negativo?”; “O que devo fazer com este biomarcador?”.

Vários métodos têm sido propostos para encontrar o melhor ponto de corte para cada estudo. Os autores e leitores de artigos científicos subestimam o valor dos métodos para decidir sobre qualquer ponto de corte. Os artigos geralmente não detalham qual método (ou por que um método específico) foi escolhido para selecionar um determinado ponto de corte para variáveis contínuas. Na verdade, alterar o ponto de corte pode afetar drasticamente as conclusões dos estudos.

Dependendo do ponto de corte escolhido, o p valor pode variar e uma hipótese nula pode ser aceita ou rejeitada. Nesse sentido, o pesquisador poderia facilmente trazer evidências de associações positivas e nulas entre duas variáveis apenas alterando o conjunto de pontos de corte! Nas Figuras 1 e 2, investigou-se o valor da relação do SUVmax no PET-CT com a sobrevivência no câncer de esôfago. Ao usar um ponto de corte de 13,25 para o SUVmax (Figura 1), o p valor do log-rank foi de 0,699 e demonstraríamos que não há diferença na taxa de sobrevivência entre SUVmax baixo (<13,25) e alto ($\geq 13,25$). No entanto, se optarmos por alterar o ponto de corte para 20 (Figura 2), evidenciaríamos uma diferença entre as curvas de sobrevivência, com p valor 0,045.

Provavelmente, a maneira mais simples de determinar o ponto de corte é pelo valor mediano. O valor mediano dicotomiza o biomarcador, garantindo um tamanho amostral igual para ambos os grupos.

Em contraste, os métodos baseados em resultados permitem encontrar um ponto de corte “ótimo”. Utilizando métodos estatísticos, os pesquisadores podem escolher o ponto de corte de acordo com a separação ideal entre os grupos em relação a algum desfecho.

A primeira coisa que os pesquisadores precisam considerar antes de escolherem o ponto de corte é estabelecer claramente o que estão investigando. A escolha do ponto de corte deve ser baseada em como a variável contínua independente “y” se relaciona com a variável dependente “x”: como resultado binário ou tempo até o evento.

Para estudos diagnósticos, o recurso mais utilizado para a variável contínua são as curvas Receiver Operating Characteristic (ROC), avaliando um desfecho

1 - Universidade de São Paulo, Gastroenterologia - São Paulo - SP - Brasil 2 - Hospital Israelita Albert Einstein, Cirurgia - São Paulo - SP - Brasil 3 - Centro Universitário Lusiada, Departamento de Medicina Baseada em Evidências - Santos - SP - Brasil

binário: “O paciente tem a doença ou não?”; “O teste é positivo ou negativo?”. Este gráfico (Figura 3) mostra a relação entre a sensibilidade e 1 especificidade, ou a relação entre verdadeiros e falsos positivos. Cada ponto no gráfico mostra a relação de sensibilidade e especificidade, de acordo com um determinado ponto de corte. A sensibilidade e especificidade em relação ao ponto de corte de probabilidade são demonstradas na Figura 4. Em 1950, Youden¹ relatou pela primeira vez uma das estratégias mais utilizadas para identificação dos pontos de corte das curvas ROC. O índice de Youden (J) maximiza as taxas de verdadeiro positivo e verdadeiro negativo. A equação é $J = \text{Sensibilidade} + \text{Especificidade} - 1$, que é aplicada a cada ponto da curva ROC. O máximo J é frequentemente usado como ponto de corte. Graficamente, o ponto máximo J representa a distância vertical máxima entre a linha de 45 graus e o ponto na curva ROC.

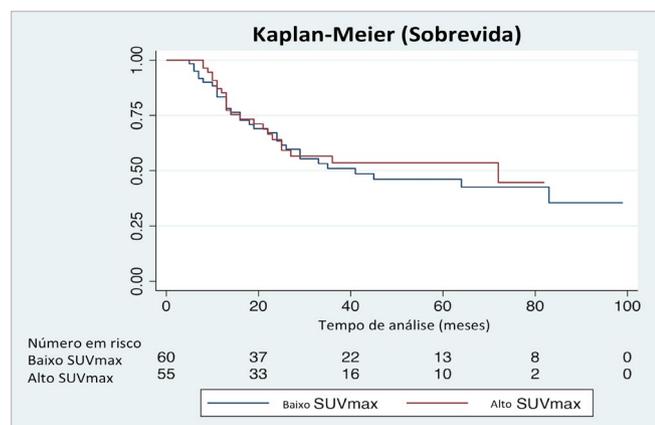


Figura 1. Análise de sobrevida comparando grupos de SUVmax baixo e alto para câncer de esôfago. Foi determinado um ponto de corte de 13,25. A análise sugeriu que o SUVmax não era uma variável prognóstica (p -valor para log-rank: 0,699). Gráfico desenhado no software estatístico Stata 16.0 (StataCorp LLC, Texas, EUA).

No entanto, qualquer investigação diagnóstica é baseada nos parâmetros de desempenho diagnóstico, como especificidade, sensibilidade, razão de verossimilhança e valores preditivos. A escolha do ponto de corte ideal deve ser baseada no que se esperaria para os parâmetros de desempenho do teste diagnóstico para uma doença específica. Uma doença de alto risco ou de rápido crescimento exigiria testes com alta sensibilidade e alto valor preditivo negativo (diagnosticando a maioria dos pacientes mesmo à custa de uma alta percentagem de falsos positivos). O objetivo do teste é “afastar” uma determinada doença. O ponto de corte seria escolhido como

conveniente de acordo com a doença e localizado mais à direita da curva ROC. Por exemplo, qualquer biomarcador esperado para triagem de câncer colorretal deve ter alta sensibilidade e alto valor preditivo negativo.

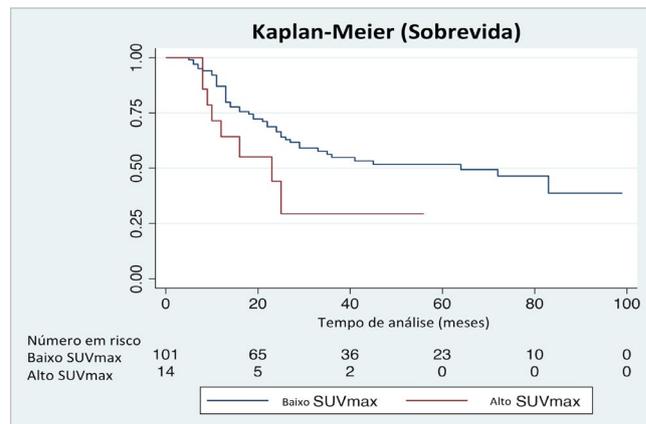


Figura 2. Análise de sobrevida comparando grupos de SUVmax baixo e alto para câncer de esôfago. Um ponto de corte de 20 foi determinado. A análise sugeriu que um SUVmax alto estava associado a um pior prognóstico (p -valor para log-rank: 0,045). Gráfico desenhado no software estatístico Stata 16.0 (StataCorp LLC, Texas, EUA).

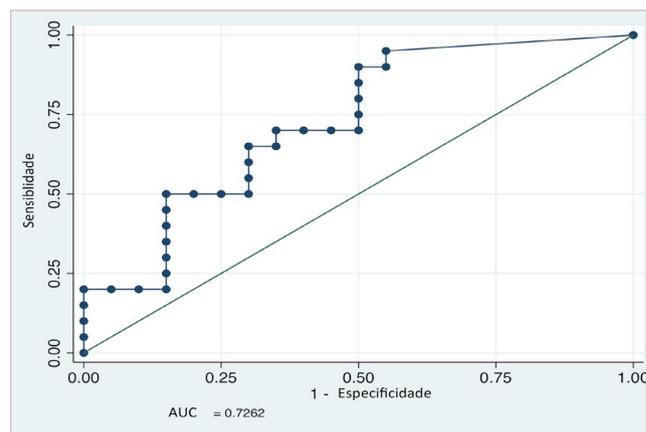


Figura 3. Curva ROC. Gráfico desenhado no software estatístico Stata 16.0 (StataCorp LLC, Texas, EUA).

Por outro lado, uma doença crônica ou uma doença que demandaria um tratamento de alto risco deve ter alta especificidade e um teste de alto valor preditivo positivo. O objetivo do teste seria “ratificar” uma suspeita diagnóstica. O ponto de corte do biomarcador de escolha seria escolhido conforme a conveniência de acordo com a doença e colocado mais à esquerda da curva ROC.

Os estudos prognósticos geralmente avaliam os resultados de tempo até o evento, como as taxas de sobrevida. Esse tipo de avaliação traz uma nova variável:

o tempo de acompanhamento. Se uma curva ROC for usada para resultados de tempo até o evento, a duração heterogênea do acompanhamento devido às observações censuradas entre os pacientes é interpretada erroneamente como acompanhamento homogêneo. Por exemplo, na análise de sobrevida, os pacientes que morreram em um acompanhamento de 5 anos seriam avaliados como os mesmos pacientes que morreram em um acompanhamento de 1 ano, o que inadequado.

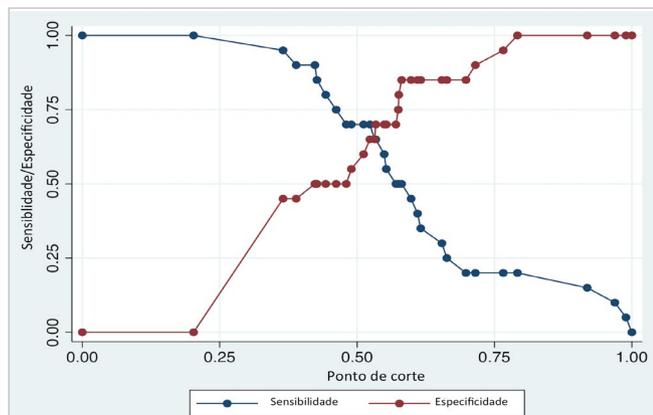


Figura 4. Relação da sensibilidade e especificidade com o ponto de corte de probabilidade. Gráfico desenhado no software estatístico Stata 16.0 (StataCorp LLC, Texas, EUA).

Consequentemente, os pesquisadores frequentemente escolhem seu ponto de corte de acordo com a abordagem do p valor mínimo². Os pesquisadores avaliam todos os possíveis pontos de corte e selecionam aquele que produz o menor p valor entre os grupos na análise de sobrevivência. No entanto, esse padrão de avaliação é propenso ao “viés de busca do pônei”, também conhecido como “viés de dragagem de dados”. Quando a análise de dados é repetida várias vezes até que os dados possam ser apresentados como estatisticamente significativos, o risco de falsos positivos aumenta (você encontrará o “p” que estava procurando!). Além disso, ao escolher o p valor mínimo, existe o risco de superestimação da medida do tamanho do efeito.

Lausen³ propôs uma nova forma de determinação do ponto de corte, na qual o tempo de seguimento seria eventualmente incorporado para os resultados de tempo até o evento. A estatística de classificação máxima selecionada divide os pacientes em dois grupos por meio das estatísticas mais significativas entre si. O objetivo é descobrir o máximo das estatísticas padronizadas, de todos os pontos de corte possíveis, que possam fornecer a melhor

separação em dois grupos de sobrevida (Figura 5). O p valor condicional exato pode ser estimado com a simulação de Monte-Carlo. Os autores também² propuseram uma melhoria matemática, corrigindo o p-valor usando uma fórmula para minimizar o erro do tipo I. Outros autores também propuseram diferentes métodos para ajuste do nível α ^{4,5}. Bonferroni⁵ sugeriu uma correção simples, dividindo o nível α pelo número de pontos de corte possíveis.

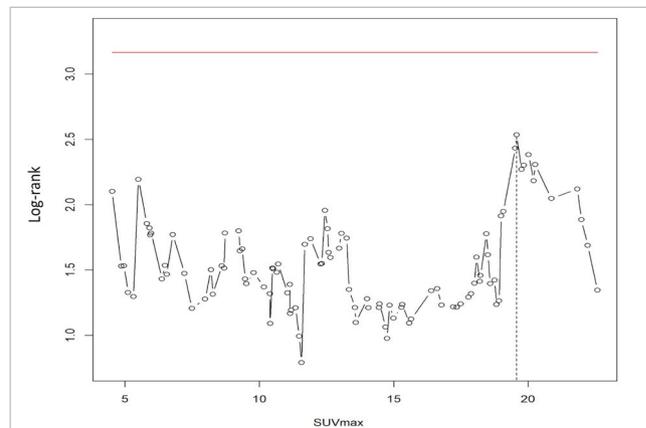


Figura 5. Estatística de classificação máxima selecionada, de acordo com Lausen³. Gráfico desenhado em R Core Team, software estatístico de 2016 (R Foundation for Statistical Computing, Viena, Áustria).

No entanto, é essencial notar que a aplicação de pontos de corte sempre impõe algumas limitações na sua interpretação. A utilização do limiar supõe a existência de um passo direto que separa “positivo” e “negativo”, “alto risco” e “baixo risco”, quando na verdade, variáveis contínuas representam um amplo espectro de possibilidades prognósticas, ou mesmo uma ampla gama de possibilidades de parâmetros de desempenho diagnóstico.

Independentemente do modelo de ponto de corte escolhido, os pesquisadores devem aplicar diferentes estratégias para demonstrar o valor real dos modelos de ponto de corte. Modelos de regressão devem ser usados para estudos de prognóstico, e a análise de sensibilidade também deve ser considerada. Estabelecer o ponto de corte ideal em diferentes conjuntos de dados independentes também pode ajudar a minimizar o risco de viés e o erro tipo I. Além disso, os leitores de manuscritos científicos devem sempre ter uma mentalidade crítica e a capacidade de perceber se o ponto de corte utilizado em determinados artigos foi razoável ou pode estar fornecendo informações enviesadas.

REFERÊNCIAS

1. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3(1):32-5. doi: 10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3.
2. Altman DG, Lausen B, Sauerbrei W, Schumacher M. Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst*. 1994;86(11):829-35. doi: 10.1093/jnci/86.11.829.
3. Lausen B, Schumacher M. Maximally selected rank statistics. *Biometrics*, 1992;48(1):73-85. doi: 10.2307/2532740.
4. Miller R, Siegmund D. Maximally selected chi-square statistics. *Biometrics*. 1982;38(4):1011-6. doi: 10.2307/2529881.
5. Mazumdar M, Glassman JR. Categorizing a prognostic variable: review of methods, code for easy implementation and applications to decision-making about cancer treatments. *Stat Med*. 2000;19(1):113-32. doi: 10.1002/(sici)1097-0258(20000115)19:1<113::aid-sim245>3.0.co;2-o.

Recebido em: 10/04/2022

Aceito para publicação em: 06/05/2022

Conflito de interesses: não.

Fonte de financiamento: nenhuma.

Endereço para correspondência:

Francisco Tustumi

E-mail: franciscotustumi@gmail.com

