

CURVA ROC

Como fazer e interpretar no SPSS

Paulo R. Margotto

Professor do Curso de Medicina da Escola Superior de Ciências da Saúde
(ESCS)/SES/DF

www.paulomargotto.com.br

pmargotto@gmail.com

Muitos exames dos nossos pacientes tem valores medidos numa escala numérica e assim a sensibilidade e a especificidade dependem de onde se coloca o ponto de corte (*cut off*) entre os resultados positivos e negativos.

Uma forma mais eficiente de demonstrar a relação normalmente antagônica entre a sensibilidade e a especificidade dos exames que apresentam resultados contínuos são as Curvas de Características de Operação do Receptor (Curvas ROC-*Receiver Operating Characteristic*). A Curva ROC é uma ferramenta poderosa para medir e especificar problemas no desempenho do diagnóstico em medicina por permitir estudar a variação da sensibilidade e especificidade para diferentes valores de corte.

As Curvas ROC foram desenvolvidas no campo das comunicações como uma forma de demonstrar as relações entre sinal-ruído. Interpretando o sinal como os verdadeiros positivos (sensibilidade) e o ruído, os falsos positivos (1-especificidade), podemos entender como este conceito pode ser aplicado em medicina.

A Curva ROC é um gráfico de sensibilidade (ou taxa de verdadeiros positivos) versus taxa de falsos positivos (Figura 1)

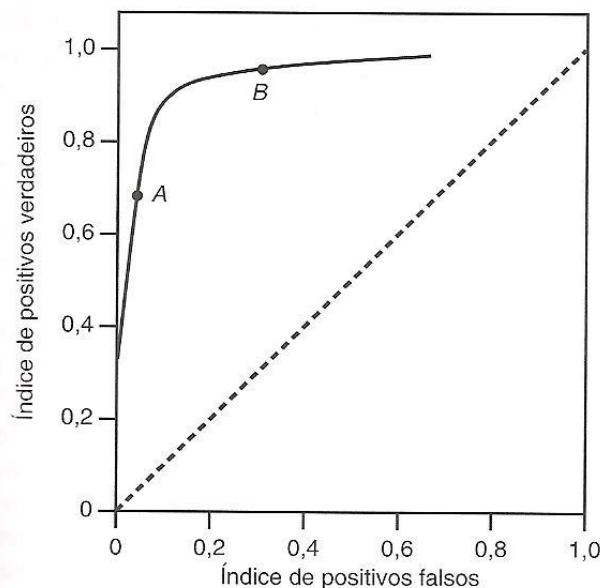


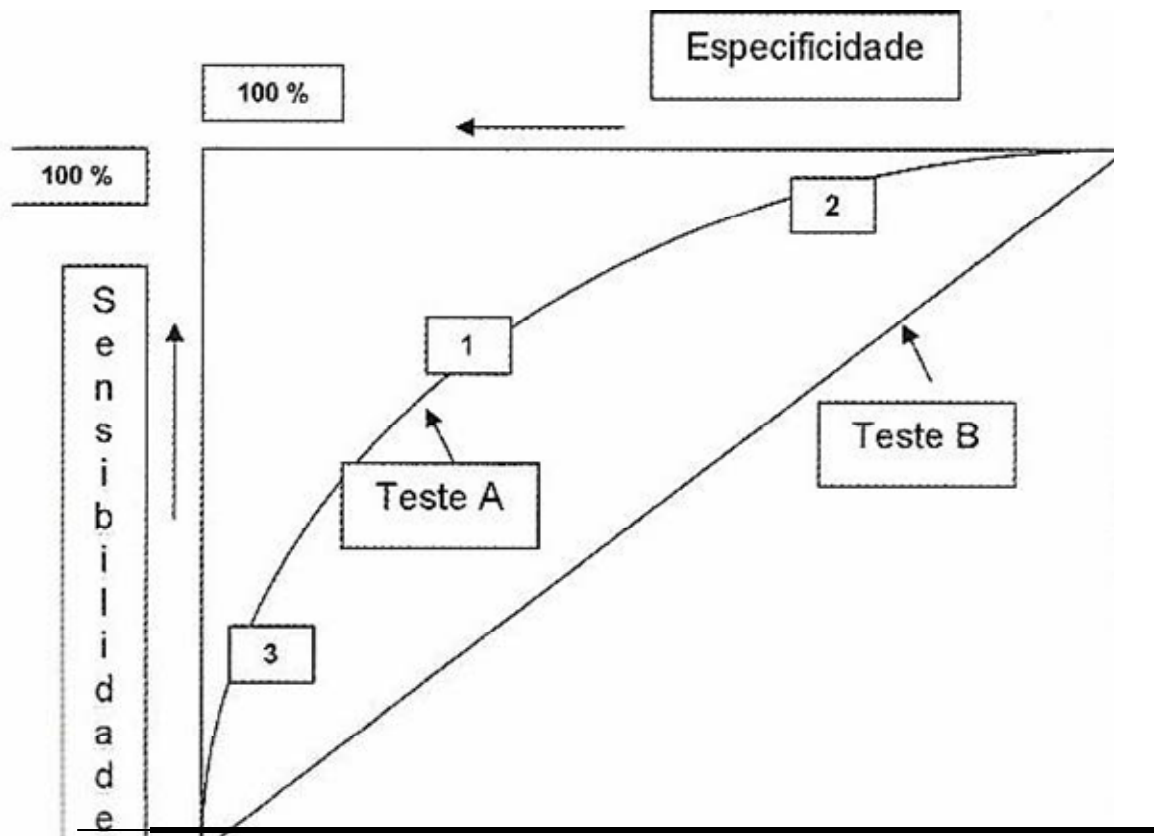
Fig.1. Curva de Características de Operação do Receptor (Curva ROC)

A linha diagonal pontilhada corresponde a um teste que é positivo ou negativo, aleatoriamente. A Curva ROC Permite evidenciar os valores para os quais existe maior otimização da sensibilidade em função da especificidade que corresponde ao

ponto em que se encontra mais próxima do canto superior esquerdo do diagrama, uma vez que o índice de positivos verdadeiros é 1 e o de falsos positivos é zero.

A medida que um critério para o teste positivo torna-se mais rigoroso, o ponto da curva correspondente a sensibilidade e a especificidade (ponto A) movimenta-se para baixo e para a esquerda (sensibilidade menor e especificidade maior). Se adotar um critério menos evidente, para identificar os positivos, o ponto da curva (ponto B) movimenta-se para cima e para a direita (sensibilidade maior, especificidade menor).

A Curva ROC permite comparar dois ou mais exames diagnósticos e constitui uma das maiores virtudes das curvas ROC. No exemplo (Figura 2) a seguir, o Teste A tem melhor acurácia que o teste B (teste inválido: os seus resultados não são melhores do que os da chance). O ponto 1 confere maior valor de sensibilidade e especificidade; o ponto 2 confere maior sensibilidade, porém menor especificidade e já o ponto 3, maior especificidade, porém, menor sensibilidade.



Falso- positivo (1-especificidade)

Figura 2. Curva ROC

O exemplo a seguir (Figura 3) de Maulik et al que compararam 4 índices correspondentes nos traçados Doppler da artéria umbilical, no que se refere a uma precisão para prever a evolução perinatal desfavorável. Podemos observar que 3 índices são praticamente semelhantes, mas um deles, o IP (índice de pulsatilidade) era muito menos preciso.

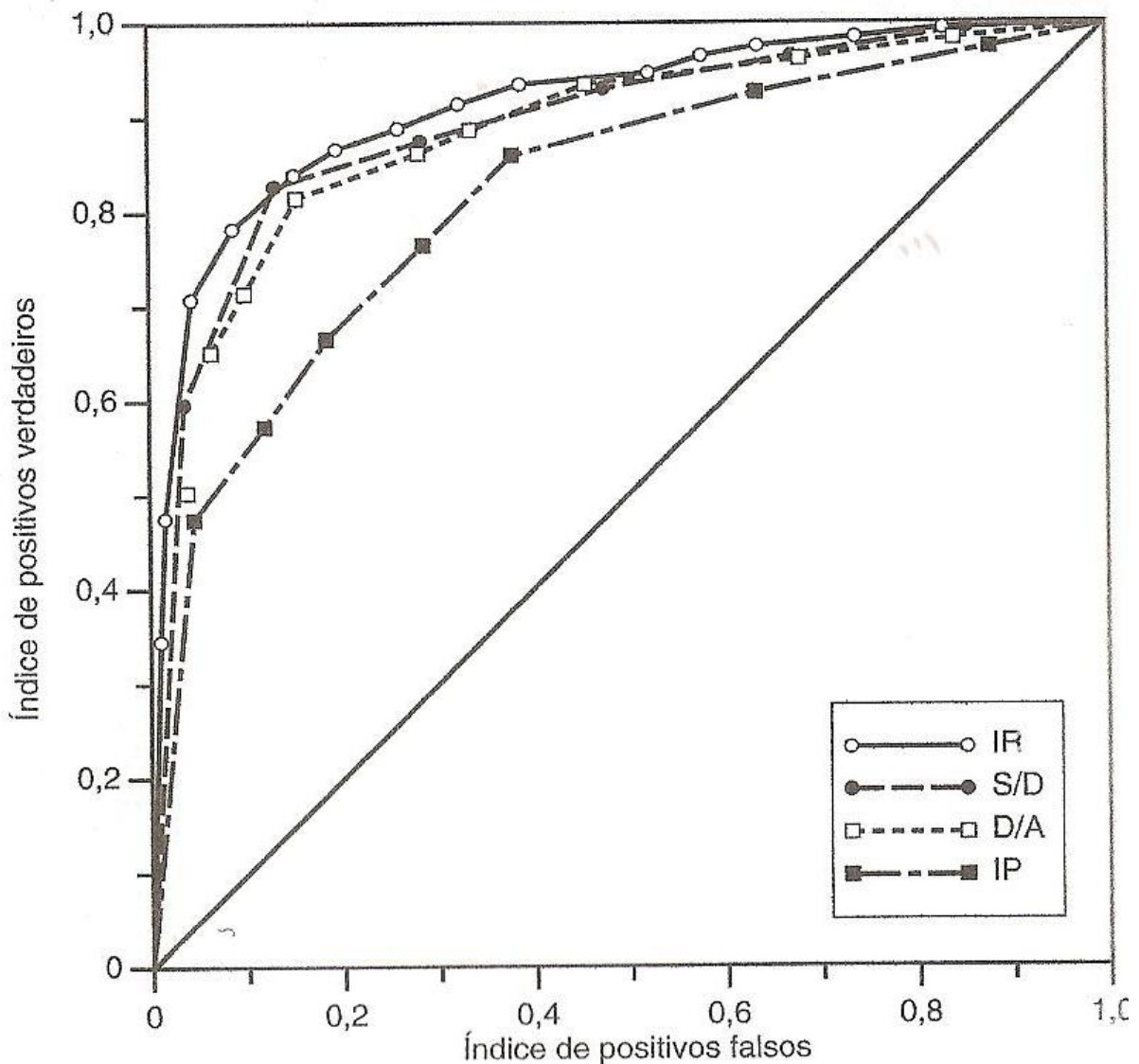


Fig. 3. Curvas de características de operação do receptor par os índices Doppler da artéria umbilical. Os pontos de dados são os valores medidos dos índices. IR: índice de resistência; D/A: relação diastólica à média; IP: índice de pulsatilidade. Maulik et al.

A Curva ROC discrimina entre dois estados, onde cada ponto da curva representa um compromisso diferente entre a sensibilidade e o falso positivo que pode ser definido pela adoção de um diferente valor do ponto de corte de anormalidade. Assim, podemos ter diferentes correspondências na curva. Um critério restrito (paciente positivo quando a evidência da doença é muito forte) é aquele que traduz uma pequena fração de falsos positivos e também uma pequena fração de verdadeiros positivos (canto inferior esquerdo da Curva ROC). O uso de critérios menos restritos, conduzem a maiores frações de ambos os tipos (pontos colocados no canto direito da curva (Figura 10)).

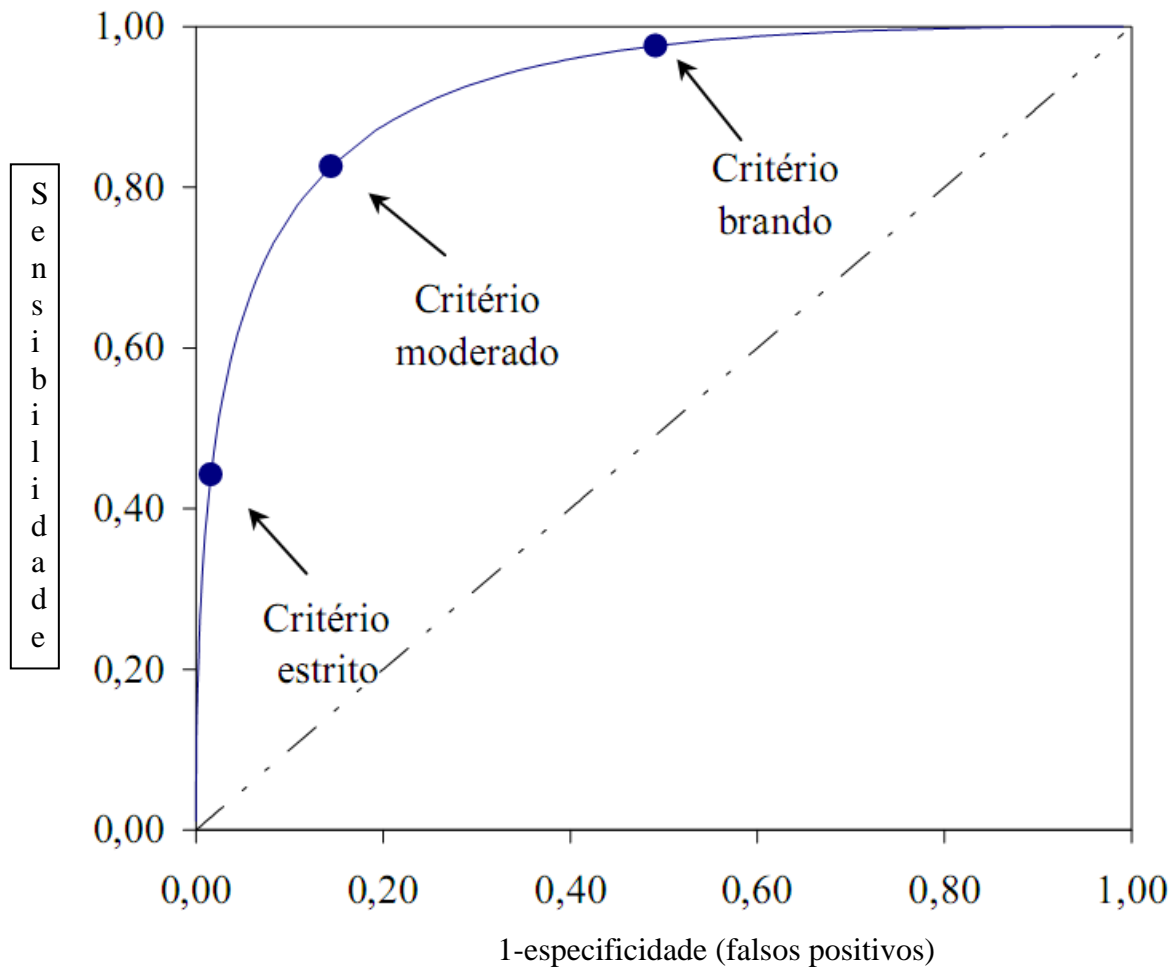


Fig.10. Curva ROC para uma dada capacidade de discriminação, com a variação do ponto de corte (critério de decisão)

O valor do ponto de corte é definido com um valor que pode ser selecionado arbitrariamente pelo pesquisador entre os valores possíveis para a variável de decisão, acima do qual o paciente é classificado como positivo (teste positivo, paciente com a doença) e abaixo do qual é classificado como negativo (teste de diagnóstico negativo, ausência de doença).

Para cada ponto de corte são calculados valores de sensibilidade e especificidade, que podem então serem dispostos no gráfico. Um classificador perfeito corresponderia a uma linha horizontal no topo do gráfico, porém esta dificilmente será alcançada. Na prática, curvas consideradas boas estarão entre a linha diagonal e a linha perfeita, onde quanto maior a distância da linha diagonal, melhor o sistema. A linha diagonal indica uma classificação aleatória, ou seja, um sistema que aleatoriamente seleciona saídas como positivas ou negativas, como jogar uma moeda para cima e esperar cara ou coroa. Finalmente, a partir de uma curva ROC, devemos poder selecionar o melhor limiar de corte para obtermos o melhor desempenho possível.

Para a determinação se duas ou mais Curvas ROC são significativamente diferentes, a avaliação é feita através da determinação da área sob a curva, usando uma modificação do teste da soma de ordens de Wilcoxon para esta comparação. Assim é

possível quantificar a exatidão de um teste diagnóstico (proporcional à área sob a curva), além da possibilidade de comparar testes diagnósticos.

Como realizar a Curva ROC pelo SPSS

Vamos usar o exemplo disponível no programa **SPSS 18: hivassay.sav**

Um laboratório farmacêutico está desenvolvendo um teste rápido para detecção de infecção pelo HIV. O atraso na obtenção de resultados de testes tradicional reduz a sua eficácia, porque muitos pacientes não retornam para saber os resultados. O desafio é desenvolver um teste que forneça resultados em 10 a 15 minutos e é tão preciso quanto os testes tradicionais. Os resultados do ensaio são oito tons de vermelho, com tons mais profundos, indicando maior probabilidade de infecção. O teste é rápido, mas é preciso? Para ajudar a responder esta pergunta, um ensaio laboratorial foi realizado em 2.000 amostras de sangue, metade dos quais estavam infectadas com o HIV e metade, não. Os resultados estão em *hivassay.sav*. Vamos usar a Curva ROC para determinar em que tom de vermelho o médico deve assumir que o paciente é HIV-positivo.

Em *Analyse* escolha *ROC Curve* (Figura 4) e será aberta a caixa de diálogo *Roc Curve* (Figura 5)

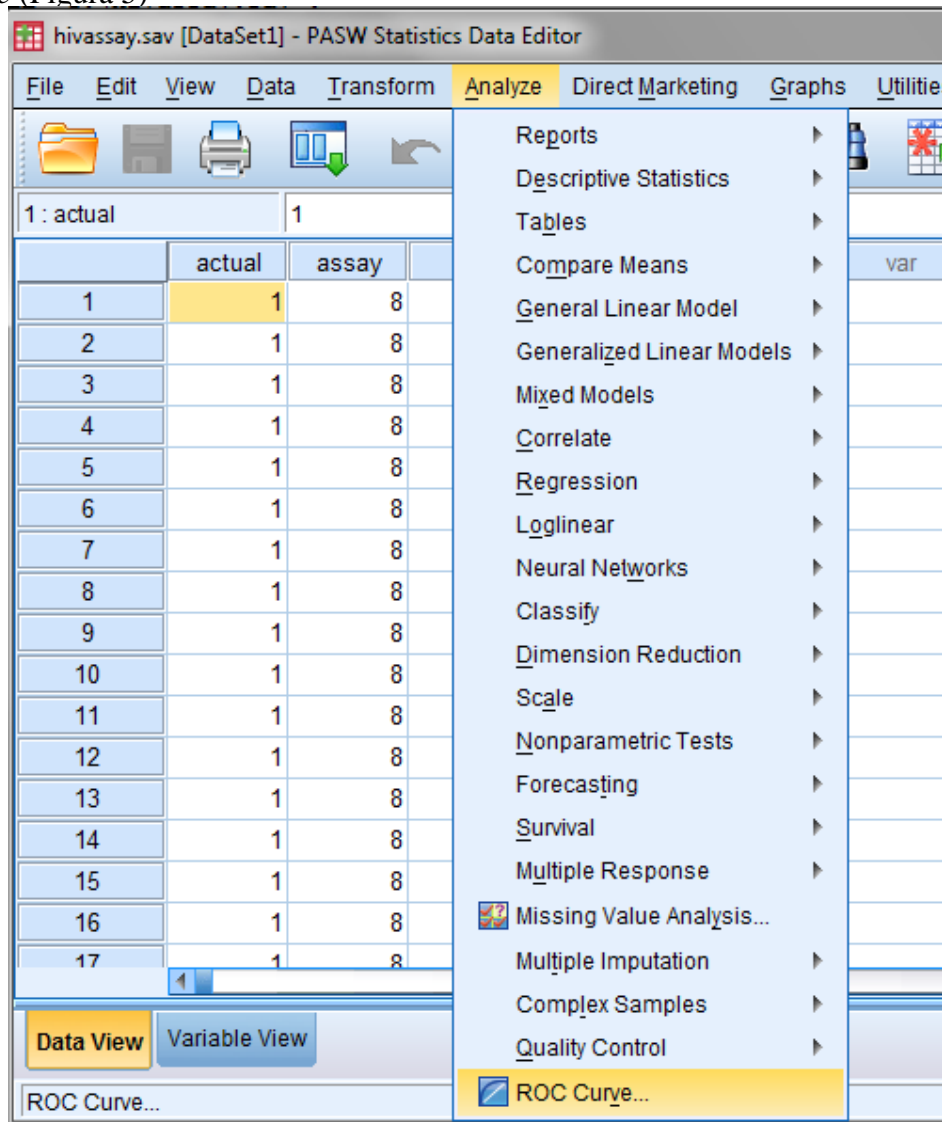


Fig.4. *Analyse>ROC Curve*

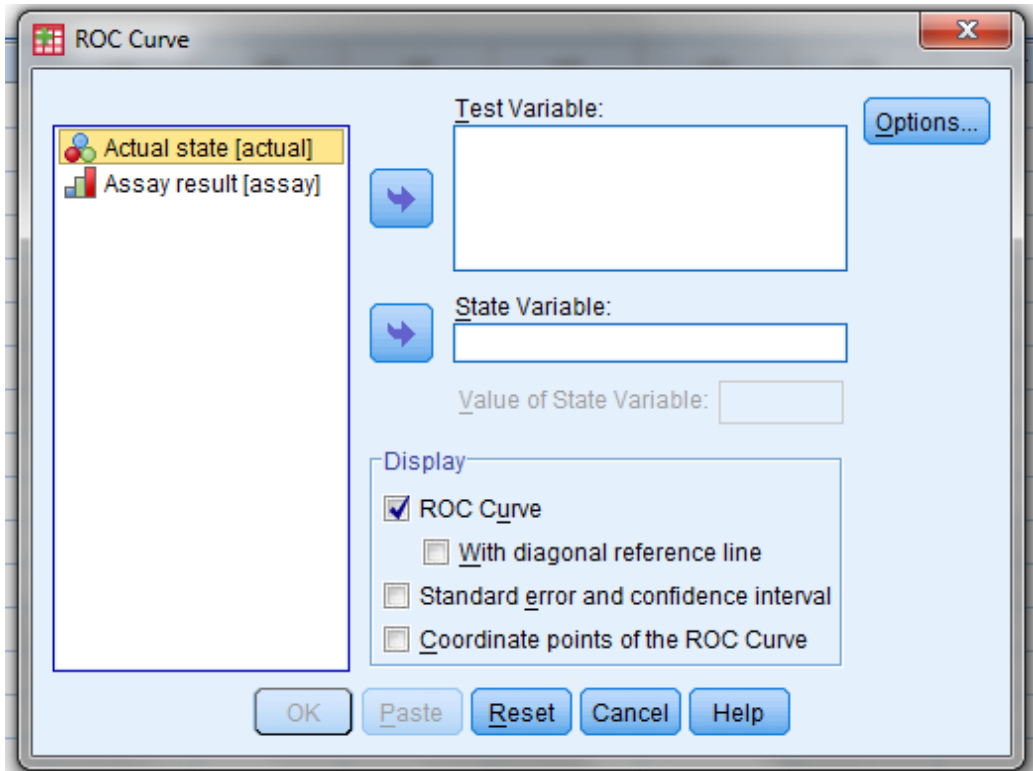


Fig. 5. Caixa de diálogo *ROC Curve* (B)

Na caixa de diálogo *Roc Curve*, selecione *assay result* como *Test variabe* e em *State variable*, selecione *actual satate* e o tipo 1, como seu valor positivo. No *Display*, selecione *With diagonal reference line*, *Standard error and confidence interval* e *Coordinate points of the ROC* (Figura 6)

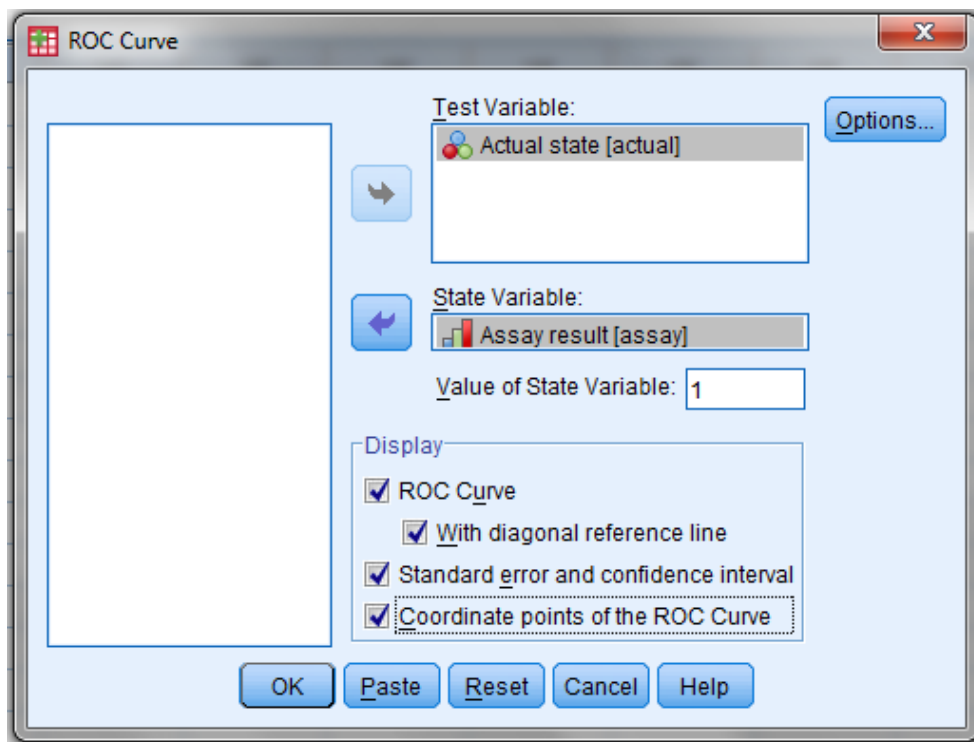
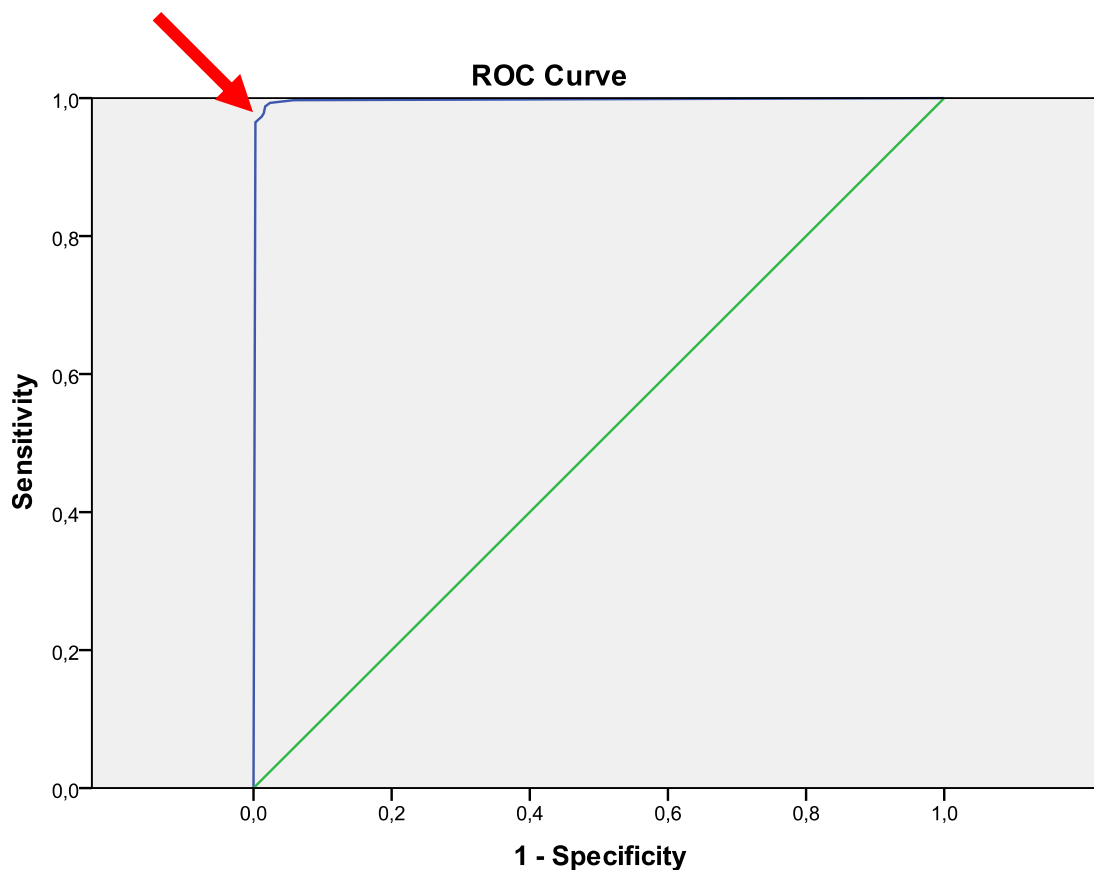


Fig.6. Caixa de diálogo *ROC Curve*

A seguir, clicar em [OK] e teremos a Curva ROC, que é um índice visual da precisão do ensaio (Figura 7)



Diagonal segments are produced by ties.

Fig.7 Curva ROC

A Curva está difícil de se visualizar porque está muito próximo ao eixo vertical, ou seja, maior sensibilidade na detecção dos pacientes com HIV positivo e com baixo falso positivo.

O conhecimento da área sob a curva possibilita quantificar exatidão de um teste diagnóstico (proporcional à área sob a curva), além de possibilitar a comparação de testes diagnósticos. A área sob a curva ROC constitui um dos índices mais usados para sumarizar a qualidade da curva.

A área sob a curva ROC é uma medida do desempenho de um teste (índice de exatidão do teste). Um teste totalmente incapaz de discriminar indivíduos doentes e não doentes, teria uma área sob a curva de 0.5 (seria a hipótese nula). Acima de 0,70 é considerado desempenho satisfatório. Vejam o cálculo da área sob curva do presente exemplo (Tabela 1)

Area Under the Curve (Área sob a Curva)

Test Result Variable(s): Assay result

Area	Std. Error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,996	,001	,000	,994	,999

Tabela 1 Cálculo da área sob a curva

A área sob a curva representa a probabilidade de que o resultado do ensaio de um caso positivo escolhidos aleatoriamente irá exceder o resultado de um processo escolhido aleatoriamente negativo. No exemplo: 0.99 (IC a 95% de 0.994 a 0.99. A significância é inferior a 0,05, o que significa que a utilização do ensaio é melhor do que a chance.

A área sob a curva é um resumo estatístico útil para a determinação da acurácia do teste. A seguir, precisamos escolher um critério específico em que as amostras de sangue são classificados e estimar a sensibilidade e a especificidade do ensaio nesse critério. Veja as coordenadas da curva para comparar diferentes pontos de corte (Figura 8).

Coordinates of the Curve

Test Result Variable(s): Assay result

A variável resultado do teste (s)

Positive if Greater Than or Equal To ^a	Sensitivity	1 - Specificity
,00	1,000	1,000
1,50	,997	,058
2,50	,995	,040
3,50	,993	,024
4,50	,988	,017
5,50	,978	,015
6,50	,973	,012
7,50	,965	,003
9,00	,000	,000

Tabela 2. Coordenadas da curva ROC: A variável resultado do teste (s)

Esta tabela relata a sensibilidade e 1- especificidade para cada ponto de corte possível para a classificação positiva. A **sensibilidade** é a proporção de amostras de HIV-positivos com os resultados do teste maior que o corte e **1-especificidade** é a proporção de amostras de HIV-negativo com os resultados do teste maior que o de corte. No ponto de corte 0 é equivalente a assumir que todos são HIV-positivos. No ponto de corte 9 é equivalente a assumir que todos HIV-negativos. Vejam assim, que

ambos os extremos são insatisfatórios, o desafio é escolher um corte que devidamente equilibrar as necessidades de sensibilidade e especificidade.

Por exemplo, considere o ponto de corte 5.5. Usando este ponto de corte vamos ter uma sensibilidade de 0,978 e 1-especificidade de 0,015. Assim, aproximadamente 97,8% de todas as amostras do HIV-positivos seriam corretamente identificadas como tal, e 1,5% de todas as amostras do HIV-negativo poderiam ser incorretamente identificadas como positivas.

Se usarmos o ponto de corte 2,52, 99,5% de todas as amostras do HIV-positivos seriam corretamente identificados como tal e 4,0% de todas as amostras de HIV-negativo poderia ser incorretamente identificadas como positivas.

A escolha do ponto de corte será decidida pela necessidade de aumentar a sensibilidade ou a especificidade e vai depender do pesquisador. Devemos notar que os valores desta tabela representam as melhores orientações para os quais devemos considerar os pontos de cortes. Esta tabela não inclui as estimativas de erro, portanto, não há garantia da exatidão da sensibilidade ou especificidade para um dado ponto de corte na tabela.

O uso da Curva ROC possibilitou-nos avaliar a acurácia deste teste. A área sob a curva mostrou que a utilização do teste é melhor do que a chance, mas, além disto, as coordenadas da curva foi muito útil porque forneceu algumas orientações para determinar qual o melhor ponto de corte para a determinação os resultados do teste positivo e negativo.

Vejamos agora, exemplo de estudos com o uso da Curva ROC na neonatologia, comparando Índices Prognósticos CRIB (clinical risk index for babies), CRIB –II e SNAPPE-II (*score for neonatal acute physiology-perinatal extension*).

Parry G et al desenvolveram o CRIB-II para RN de 32 semanas após 1 hora de admissão na UTI sendo excluída a FiO_2 uma vez que esta não representa uma medida fisiológica verdadeira por ser determinada pelo cuidador. Assim, o CRIB II foi calculado a partir de 5 itens: sexo, peso ao nascer, idade gestacional, pior *base excess* e temperatura na admissão. A área sob a curva ROC para morte ou graves anormalidades cerebrais foi de 0,82, de 0,79 pra o CRIB, 0,80 para a idade gestacional e 0,77 para o peso ao nascer. Portanto, o CRIB II mostrou maior acurácia em relação ao CRIB, a idade gestacional e o peso ao nascer para a morte ou graves anormalidades cerebrais.

Gagliardi I et al, comparando CRIB, CRIB-II e SNAPE-II, relataram maior grau de discriminação para o CRIB e CRIB II em relação ao SNAPPE-II (área sob a Curva ROC de 0,90 e 0,91 versus 0,84- $p < 0,0004$). Vejamos a Curva ROC deste estudo (Fig.8).

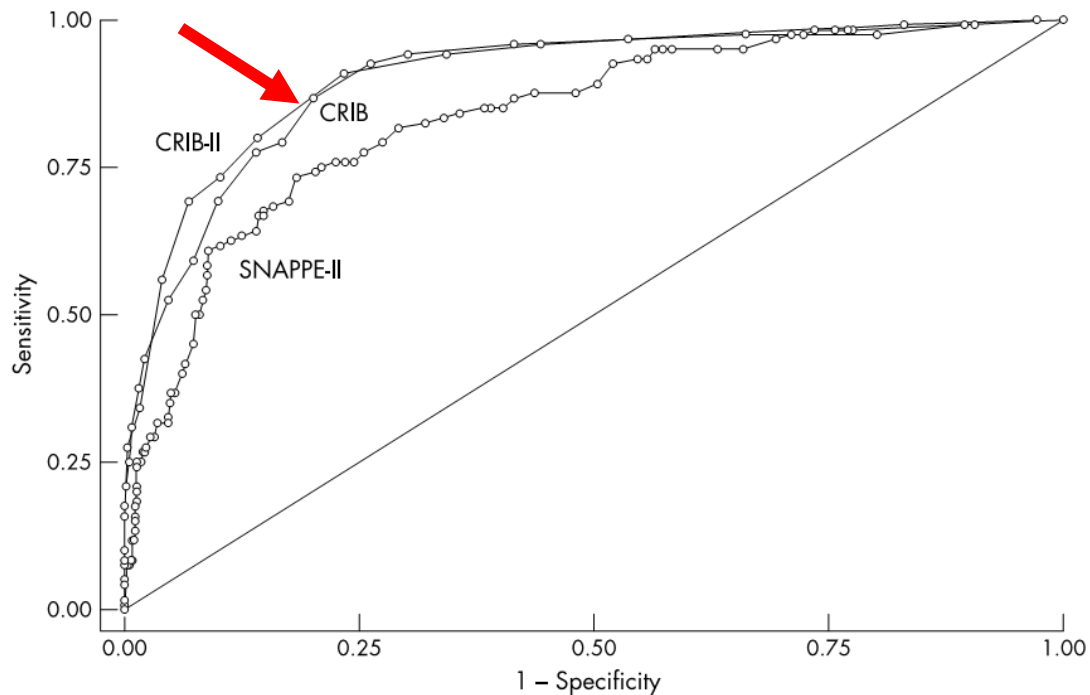


Fig. 8. Curva ROC para o CRIB, CRIB -II e SNAPPE-PE. A área sob a curva foi: CRIB, 0,903; CRIB-II, 0,907; SNAPPE-II, 0,837.

Observem que as Curva do CRIB e CRIB-II estão mais próximas do canto superior esquerdo do diagrama, o que confere maior área sob a curva e portanto, maior poder de discriminação em relação ao SNAPPE-II.

Referências:

- 1-Martinez EZ, Lozada-Neto F, Pereira BB. Cadernos Saúde Coletiva, Rio de Janeiro 2003;11:7-31 (disponível em [PDF] [A curva ROC para testes diagnósticos](#))
- 2- Braga ACS. Curvas ROC: Aspectos funcionais e aplicações. Tese de Doutorado, Universidade do Minho, dezembro de 2003, Portugal
- 3- Dawson B, Trapp RG. Métodos da medicina baseada em evidências. In. Dawson B, Trapp RG Bioestatística Básica e Clínica, McGrawHill, Rio de Janeiro, 3ª Edição, 2003, p.233-249
- 4- Maulik D, Yarlalagadda P, Youngblood JP, Ciston P. Comparative efficacy of umbilical arterial Doppler índices for predicting adverse perinatal outcome. Am J Obstet Gynecol 1991;164:1434-1440
- 5-Parry G, Tucker J et al. CRIB II: an update of the clinical risk index for babies score. Lancet 361:1789, 2003
- 6- Gagliardi L, Cavazza A, et al. [Assessing mortality risk in very low birthweight infants: a comparison of CRIB, CRIB-II, and SNAPPE-II.](#) Arch Dis Child Fetal Neonatal Ed. 89:F419, 2004
- 7-Curso de Epidemiologia Clínica-CLAP, 1988

Pesquisado, digitado, revisado por Paulo R. Margotto

www.paulomargotto.com.br

pmargotto@gmail.com.br

Brasília, 11 de junho de 2010