

TESTES ASSINTÓTICOS: ESCORE DE RAO

Luiz Guilherme Pessoa Aguiar*

RESUMO

A história dos testes de hipóteses estatísticos é realmente extensa. No entanto, se precisarmos escolher alguns artigos que tiveram uma influência profunda na prática atual, os seguintes três artigos devem estar entre eles: Neyman e Pearson (1928), Wald (1943) e Rao (1948). Esses artigos propuseram três princípios básicos de teste: Razão de Verossimilhança (RV), Wald (W) e o teste de Escore de Rao (ER), respectivamente. Geralmente é levantado a questão de que o teste ER pode ser inferior aos outros procedimentos assintoticamente equivalentes, como os testes W e RV, em termos de poder, uma vez que não utiliza informações precisas da hipótese alternativa. No entanto, os resultados de Monte Carlo de Godfrey (1981) e Bera e McKenzie (1986) sugerem que aparentemente não há retrocesso no desempenho do teste ER em comparação com o teste RV. Para enfatizar seu papel fundamental na Inferência Estatística, Rao (2005) chamou os três testes já mencionados anteriormente, RV, W e o ER de "A Santíssima Trindade". Embora as representações analíticas e equivalências assintóticas dessas estatísticas baseadas em verossimilhança sejam totalmente discutidas e facilmente manipuladas por meio de expansões de Taylor, o significado geométrico real delas, com suas implicações, permanece um tanto vago e, em certa medida, não muito claro para a maioria. Neste artigo de revisão vamos explorar a criação, propriedades, interpretação geométrica, função poder em hipóteses de Pitman, analisar diferenças e semelhanças e comparar o poder do teste para "A Santíssima Trindade" e ver como podemos transformar o teste log-rank em termos de um teste ER.

Palavras-chave: Escore; Teste; Comparação.

FATO HISTÓRICO

"Nos primeiros anos da minha nomeação para o Instituto Indiano de Estatística, tive a oportunidade de interagir com o pessoal e os investigadores e discutir com eles novos problemas em estatística decorrentes do trabalho de consulta. Um dos estudiosos, S.J. Poti, me perguntou sobre como testar uma hipótese simples $H : \theta = \theta_0$ relativa a um único parâmetro θ quando há alguma informação prévia sobre a alternativa, como $\theta > \theta_0$. Sugeri um procedimento pelo qual a potência local no lado direito de θ_0 é maximizada, levando a um teste da forma $P(\theta)^T/P(\theta) > \lambda$ onde P^T/P é a pontuação de Fisher. O resultado foi publicado em Rao e Poti (1946). Eles também propuseram um teste geral da forma $(P^T/P)^2 > \lambda$ que provavelmente terá boa potência local em ambos os lados de θ_0 . Dois anos depois, eu estava trabalhando em um problema na Universidade de Cambridge, no Reino Unido, que envolvia testes de hipóteses simples e compostas relativas a múltiplos parâmetros quando havia informações de que as alternativas estavam próximas

* Bacharelado em Estatística na Universidade Federal do Ceará — Campus do Pici.
E-mail: gui_aguiar@alu.ufc.br

daquelas especificadas pela hipótese nula. Isso levou à combinação dos critérios de testes individuais baseados nas pontuações de Fisher $s_1(\theta), \dots, s_r(\theta)$ em um único critério. Seguindo os métodos utilizados na análise multivariada, cheguei à estatística Escore para uma hipótese simples que tem uma distribuição qui-quadrado assintótica com r graus de liberdade, que foi aprovadas por R.A. Fisher, meu orientador de tese quando eu trabalhava na Universidade de Cambridge durante os anos 1946 a 1948. O artigo (Rao, 1948), contendo a discussão geral dos testes Escores, é publicado nos anais da Cambridge Philosophical Society."

- Prof. Calyampudi Radhakrishna Rao

1 INTRODUÇÃO

A história dos testes de hipóteses estatísticos é realmente extensa. No entanto, se precisarmos escolher alguns artigos que tiveram uma influência profunda na prática atual, os seguintes três artigos devem estar entre eles: Neyman e Pearson (1928), Wald (1943) e Rao (1948). Esses artigos propuseram três princípios básicos de teste: Razão de Verossimilhança (RV), Wald (W) e o teste de Escore de Rao (ER), respectivamente.

Por muito tempo, os econométricos utilizaram principalmente o teste RV e alguma versão do princípio de Wald. O teste de Escore era raramente utilizado explicitamente, embora alguns dos testes econométricos anteriores pudessem ter uma interpretação de teste de Escore. No entanto, nas décadas de 1970 e 1980, houveram atividades intensas na aplicação do princípio de Escore a uma variedade de problemas de testes econométricos e no estudo das propriedades dos testes resultantes. Atualmente, os testes ER são os itens mais comuns no conjunto de ferramentas de testes dos econométricos.

2 DESENVOLVIMENTO

Nesta seção falaremos sobre as propriedades do Teste Escore, sua interpretação geométrica e as funções poder.

2.1 Teste Escore E Suas Propriedades

Suponha que existam n observações independentes, representadas por y_1, y_2, \dots, y_n , com uma função de densidade de probabilidade $f(y; \theta)$, onde θ é um vetor de parâmetros $p \times 1$ com $\theta \in \Theta \subseteq \mathbb{R}^p$. Θ representa o espaço de parâmetros. Assume-se que $f(y; \theta)$ satisfaz as condições de regularidade mencionadas em Rao (1973, p. 364) e Serfling (1980, p. 144). A função de log-verossimilhança, o vetor Escore e a matriz de informação são definidas, respectivamente:

$$l(\theta) = \sum_{i=1}^n \ln f(y_i; \theta),$$

$$d(\theta) = \frac{\partial l(\theta)}{\partial \theta},$$

$$I(\theta) = -E \left[\frac{\partial^2 l(\theta)}{\partial \theta \partial \theta^T} \right].$$

A hipótese a ser testada é $H_0 : h(\theta) = c$, onde $h(\theta)$ é um vetor função de θ com dimensões $r \times 1$, sendo $r \leq p$, e c é um vetor constante conhecido. Assume-se que $H \equiv H(\theta) = \frac{dh(\theta)}{d\theta}$ possui posto colunar completo, ou seja, posto $[H(\theta)] = r$. A estatística ER para testar H_0 pode ser expressa como:

$$ER = d(\tilde{\theta})^T I(\tilde{\theta})^{-1} d(\tilde{\theta}) \quad (1)$$

onde " \sim " indica que as quantidades foram avaliadas na estimativa de máxima verossimilhança restrita (EMVR) de θ , dizemos $\tilde{\theta}$.

A ideia por trás deste teste é que, se H_0 for verdadeira, é esperado que $d(\theta)$ seja próximo de zero, devido ao fato de que $d(\theta) = 0$, onde θ é o EMVR. Sob H_0 , a estatística ER é assintoticamente distribuída como χ_r^2 [Rao (1973, pp. 418-419), Serfling (1980, pp. 156-160) e Godfrey (1988, pp. 13-15)]. Econometristas costumam usar o teste acima. No entanto, ele é frequentemente chamado de teste do multiplicador de Lagrange (ML). A terminologia "teste do ML" originou-se dos artigos Aitchison e Silvey (1958) e Silvey (1959), nos quais uma interpretação ML de (1) foi fornecida.

Observe que $\tilde{\theta}$ pode ser obtido a partir da solução das equações

$$d(\tilde{\theta}) - H(\tilde{\theta})\tilde{\lambda} = 0,$$

$$H(\tilde{\theta}) = c$$

onde λ 's são os multiplicadores de Lagrange. Portanto, temos $d(\tilde{\theta}) - H(\tilde{\theta})\tilde{\lambda}$. Dado que $H(\tilde{\theta})$ tem posto completo $d(\tilde{\theta}) = 0$ que é equivalente a $\tilde{\lambda} = 0$, ou seja, os multiplicadores de Lagrange se anulam. Esses multiplicadores podem ser interpretados como o custo implícito de impor as restrições. Pode ser demonstrado

$$\tilde{\lambda} = \frac{\partial l(\tilde{\theta})}{\partial c}$$

onde, os multiplicadores fornecem a taxa de mudança do valor máximo alcançável em relação à mudança nas restrições. $H_0 : h(\theta) = c$ for verdadeiro e $l(\tilde{\theta})$ fornecer o valor ótimo, $\tilde{\lambda}$ deve estar próximo de zero. Dada essa interpretação "econômica" em termos de multiplicadores, não é surpreendente que os econometristas prefiram o termo ML em vez de ER. Em termos de multiplicadores de Lagrange, (1) pode ser expresso como

$$ML = \tilde{\lambda}^T H(\tilde{\theta})^T I(\tilde{\theta})^{-1} H(\tilde{\theta}) \tilde{\lambda} \quad (2)$$

nota-se que a partir de (1) e (2) que a forma do teste ER é muito mais fácil de calcular, o que explica sua popularidade entre os econometristas. No entanto, como explicado anteriormente, eles o chamam por um nome mais próximo da maneira de pensar dos economistas.

A partir de (1), fica claro que a estatística ER é essencialmente uma medida de distância entre as hipóteses nula e alternativa, e pode ser interpretada como uma norma de Mahalanobis [consulte Del Pino (1990)]. Isso leva ao desenvolvimento de formas alternativas do teste ER, utilizando diferentes tipos de medidas de distância [consulte Ullah (1989)]. Observe que a estatística ER não depende explicitamente da hipótese alternativa, pois utiliza a inclinação de $l(\theta)$ em $\theta = \tilde{\theta}$. Pode haver diferentes funções de verossimi-

lhança (com a mesma inclinação e possivelmente a mesma curvatura em H_0) que levarão à mesma estatística ER. Esse fenômeno é chamado de propriedade de invariância do teste ER. A estatística ER para testar a normalidade sugerida em Bera e Jargue (1981) e Jargue e Bera (1987), com a família de distribuições de Pearson como alternativa, permanece inalterada sob alternativas de Gram-Charlier (tipo A) [consulte Bera (1982a, p. 98)]. Estatísticas para testar homocedasticidade são invariantes para diferentes formas de alternativas, como heterocedasticidades multiplicativas e aditivas, como observado por Breusch e Pagan (1979) e Godfrey e Wickens (1982). Testar independência serial contra processos autoregressivos de ordem $AR(p)$ ou processos de média móvel de ordem $MA(p)$ leva à mesma estatística de teste, veja, por exemplo, Breusch (1978) e Godfrey (1978a). Pesaran (1979) constatou que o teste RS não diferencia entre defasagens distribuídas polinomial e racional.

Geralmente é levantado a questão de que o teste ER pode ser inferior aos outros procedimentos assintoticamente equivalentes, como os testes W e RV, em termos de poder, uma vez que não utiliza informações precisas da hipótese alternativa. No entanto, os resultados de Monte Carlo de Godfrey (1981) e Bera e McKenzie (1986) sugerem que aparentemente não há retrocesso no desempenho do teste ER em comparação com o teste RV. Utilizando modelos de alternativa localmente equivalentes (ALE), Godfrey e Wickens (1982) fornecem algumas bases teóricas por trás da propriedade de invariância do teste ER. Suas análises também destacam o fato de que uma escolha adequada da alternativa pode reduzir significativamente a complexidade computacional de um teste. A propriedade de invariância também alerta os pesquisadores contra aceitar automaticamente a hipótese alternativa quando a nula é rejeitada, ou seja, se rejeitarmos a nula, isso não implica que a hipótese alternativa do teste seja correta.

2.2 Interpretação Geométrica

Para enfatizar seu papel fundamental na Inferência Estatística, Rao (2005) chamou os três testes já mencionados anteriormente, RV, W e o ER de "A Santíssima Trindade". Embora as representações analíticas e equivalências assintóticas dessas estatísticas baseadas em verossimilhança sejam totalmente discutidas e facilmente manipuladas por meio de expansões de Taylor, o significado geométrico real delas, com suas implicações, permanece um tanto vago e, em certa medida, não muito claro para a maioria. Tendemos a aprendê-las separadamente, reconhecendo que são estatísticas razoáveis, cada uma medindo a "distância" entre a hipótese nula e a evidência da amostra em sua própria escala apropriada, mas sem compreender totalmente suas conexões profundas. Além da "Santíssima Trindade", uma quarta estatística de teste baseada na verossimilhança foi introduzida recentemente por Terrell (2002): o teste do Gradiente (G). Que é assintoticamente equivalente à "Santíssima Trindade".

A "Santíssima Trindade" (Rao, 2005) pode ser escrita como:

$$d = 2\{l(\hat{\theta}) - l(\theta_0)\} \quad (3)$$

$$w = \frac{(\hat{\theta} - \theta_0)^2}{\mathcal{J}(\hat{\theta})^{-1}} \quad (4)$$

$$s = \frac{u(\theta_0)^2}{\mathcal{I}(\theta_0)^{-1}} \quad (5)$$

considerando o caso de um parâmetro escalar e um modelo estatístico regular com log-verossimilhança $l(\theta)$, onde $\theta \in \Theta$, subconjunto de \mathbb{R} , o estimador de máxima verossimilhança $\hat{\theta} = \arg \max l(\theta)$, existe e é único. θ_0 é o verdadeiro valor do parâmetro. A derivada da log-verossimilhança $l(\theta)' = u(\theta)$ é a função escore, $\mathcal{I}(\theta)$ e $\mathcal{J}(\theta) = -l(\theta)''$ são as informações esperada e observada. Como o foco é a interpretação geométrica de estatísticas assintoticamente equivalentes, a distinção entre Informação observada e esperada é uma questão secundária, pois são assintoticamente equivalentes e até mesmo idênticas em casos importantes, como na família exponencial natural, portanto, será negligenciada.

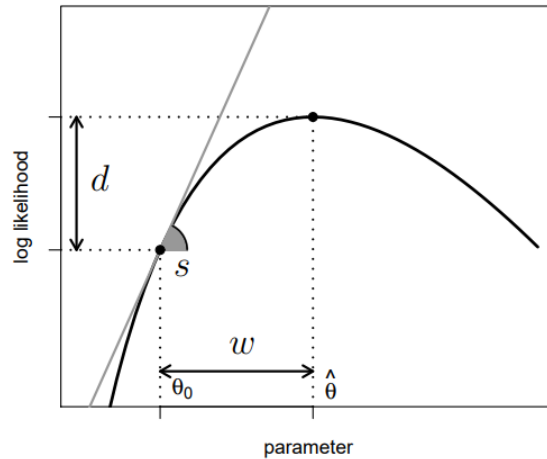
Como mencionado anteriormente, a nova estatística do Gradiente (Terrell, 2002), é definida como:

$$g = u(\theta_0)(\hat{\theta} - \theta_0) \quad (6)$$

assim como para a Santíssima Trindade, a distribuição nula em uma amostra grande da estatística do Gradiente é uma qui-quadrado com 1 grau de liberdade. É importante ressaltar que nenhum teste, incluindo a estatística do Gradiente, é uniformemente mais poderoso em todas as configurações: portanto, faz sentido aprender cada um deles, e apreciar a interpretação geométrica pode ajudar a atingir esse objetivo.

Como é relatado na maioria dos livros didáticos, cada teste utiliza sua própria escala: d opera na escala do logaritmo da verossimilhança, w na escala do parâmetro e s na escala da primeira derivada. Para explicá-los melhor, alguns livros didáticos também apresentam uma exibição gráfica clássica, como a mostrada na Figura 1, a Razão de Verossimilhança é relatada na escala y, Wald na escala x e o Score na escala da primeira derivada. As escalas diferentes não favorecem a compreensão das conexões subjacentes.

Figura 1 — Comparação geométrica da Santíssima Trindade.



Fonte: Muggeo, Vito & Lovison, Gianfranco. (2014).

Embora essa representação seja absolutamente correta, é proposto que a apresentação e discussão das conexões entre essas três estatísticas de teste podem ser proveitosamente aprimoradas usando uma ilustração gráfica para todas as estatísticas de teste na mesma escala.

2.2.1 ER Na Escala Da Log-verossimilhança E Escore

Escala da Log-verossimilhança:

Sob as condições regulares assumidas, $l(\theta)$ é suave até a segunda ordem, e é possível aproximá-lo por meio de polinômios de segunda ordem. Como é bem conhecido, existem duas aproximações quadráticas disponíveis. A primeira é baseada na expansão de Taylor em $\theta = \hat{\theta}$:

$$\mathcal{P}_w(\theta) = \ell(\hat{\theta}) + (\theta - \hat{\theta})\ell(\hat{\theta})' + \frac{1}{2}(\theta - \hat{\theta})^2\ell(\hat{\theta})'', \quad (7)$$

onde $\ell(\hat{\theta})' = 0$; e a segunda aproximação baseia-se em uma expansão de Taylor em $\theta = \theta_0$:

$$\mathcal{P}_s(\theta) = \ell(\theta_0) + (\theta - \theta_0)\ell(\theta_0)' + \frac{1}{2}(\theta - \theta_0)^2\ell(\theta_0)'' \quad (8)$$

Tradicionalmente, os testes de W e ER são discutidos com referência às fórmulas usuais (4) e (5), que são empregadas para justificar sua fundamentação e, ao mesmo tempo, para ilustrar suas diferentes escalas. O fato de que eles são baseados nas duas aproximações quadráticas mencionadas anteriormente é apenas mencionado para explicar como os numeradores $(\hat{\theta} - \theta_0)^2$ e $u(\theta_0)^2$ são ponderados pela curvatura do logaritmo da verossimilhança. "A estatística do teste da Razão de Verossimilhança é um múltiplo da diferença do logaritmo da verossimilhança; a estatística do teste de Wald é um múltiplo da diferença ao quadrado de $(\hat{\theta} - \theta_0)$; e a estatística do teste de Escore é um múltiplo da inclinação ao quadrado em θ_0 " escrevem Boos e Stefanski (2013). Portanto, não é comum ver as relações reais de w e s com o log-verossimilhança em si. Ao introduzir os testes de Wald e Escore, seria útil apresentá-los usando a mesma abordagem usada para a Razão de Verossimilhança, ou seja, como maneiras alternativas de comparar os mesmos dois valores de plausibilidade, mas empregando as aproximações quadráticas (7) e (8), em vez do log-verossimilhança em si.

Ao subtrair e dobrar os valores de plausibilidade em θ_0 e $\hat{\theta}$ fornecidos pela aproximação (7) resulta em:

$$2 \left\{ \mathcal{P}_w(\hat{\theta}) - \mathcal{P}_w(\theta_0) \right\} = w$$

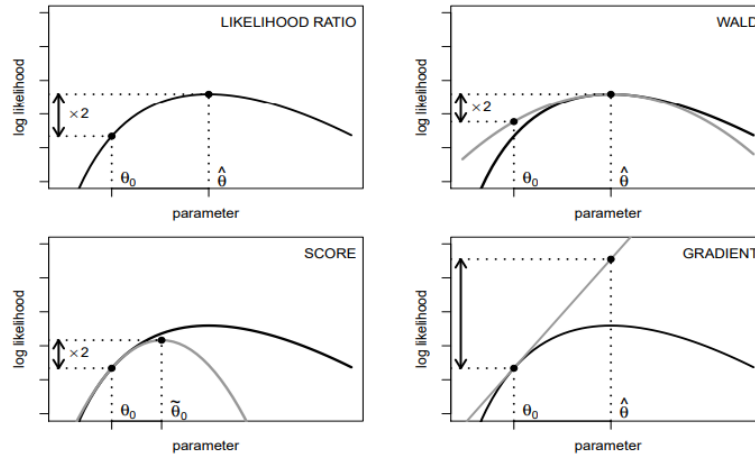
Da mesma forma, s em (5) pode ser obtido subtraindo e dobrando os valores de plausibilidade no máximo e sob H_0 de acordo com a aproximação (8). Observe que a plausibilidade máxima de (8) é alcançada em $\tilde{\theta}_0 = \theta_0 - \ell(\theta_0)' / \ell(\theta_0)''$. O leitor pode obtê-lo igualando $\mathcal{P}_s(\theta)'$ a zero, ou reconhecendo que $\tilde{\theta}_0$ é simplesmente a atualização do passo do algoritmo de Newton para maximizar $\ell(\theta)$ a partir de θ_0 . Portanto, temos:

$$2 \left\{ \mathcal{P}_s(\tilde{\theta}_0) - \mathcal{P}_s(\theta_0) \right\} = s$$

A estatística do Gradiente também pode ser expressa como uma diferença de plausibilidade em θ_0 e $\hat{\theta}$ em uma aproximação apropriada do log-verossimilhança, usando uma expansão de Taylor de primeira ordem para $\ell(\theta)$ em θ_0 , $\mathcal{P}_g(\theta) = \ell(\theta_0) + (\theta - \theta_0)\ell(\theta_0)'$. Aqui não precisamos dobrar a diferença para obter:

$$\left\{ \mathcal{P}_g(\hat{\theta}) - \mathcal{P}_g(\theta_0) \right\} = g$$

Figura 2 — Comparação geométrica na escala da Log-verossimilhança.



Fonte: Muggeo, Vito & Lovison, Gianfranco. (2014).

A Figura 2 apresenta as representações gráficas das quatro estatísticas de teste na escala comum do logaritmo da verossimilhança. Claro, para inferências sobre o parâmetro de localização de modelos gaussianos, as duas aproximações \mathcal{P}_w e \mathcal{P}_s coincidem com o logaritmo da verossimilhança, assim $d = w = s$, também. A estatística do G reduz para o mesmo valor. De fato, as quatro estatísticas de teste são idênticas neste caso especial.

Comparando as quatro estatísticas de teste na escala do logaritmo da verossimilhança. Em cada gráfico, a verossimilhança é ilustrada (linha preta) junto com a aproximação relevante subjacente à estatística do teste: no painel Wald, a linha cinza é $\mathcal{P}_w(\theta)$, no painel Score, é $\mathcal{P}_s(\theta)$, e no painel Gradient, é $\mathcal{P}_g(\theta)$.

As setas no lado esquerdo quantificam a estatística de teste observada correspondente, quanto mais longa a seta, maior a evidência contra H_0 . Observe que, para a Razão de Verossimilhança, Wald e Score, os comprimentos das setas precisam ser dobrados para obter os valores reais comparáveis com os da estatística do Gradiente.

Escala Escore:

As quatro estatísticas de teste também têm uma representação comum na escala Escore. Conforme mostrado abaixo, na Figura 3, a área cinza do "triângulo" \mathcal{A} em cada gráfico é a metade do valor da estatística correspondente.

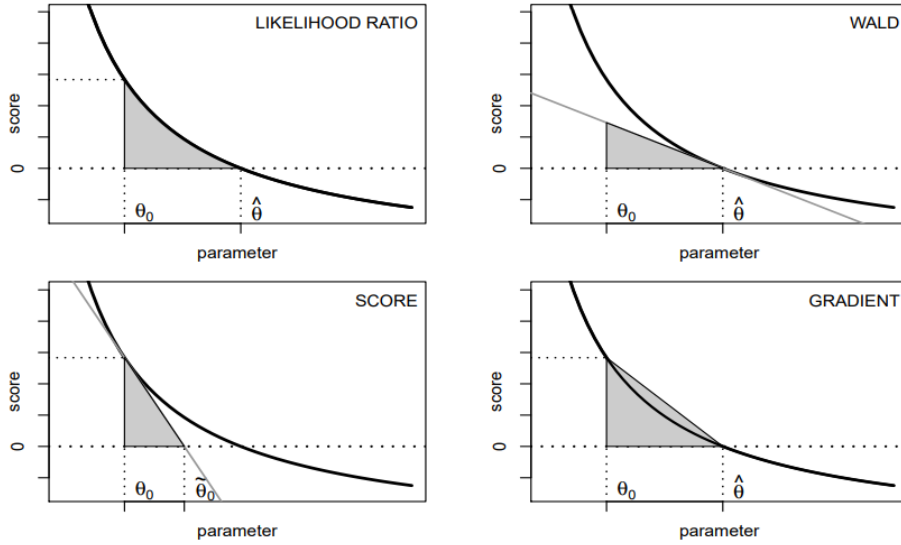
- A estatística RV, mais uma vez, é simples. Como $l(\theta)' = u(\theta)$ até uma constante, a área do triângulo curvilíneo \mathcal{A} é

$$\mathcal{A} = \int_{\theta_0}^{\hat{\theta}} u(\theta) d\theta = \left\{ \ell(\hat{\theta}) - \ell(\theta_0) \right\} = \frac{1}{2}d.$$

Em cada triângulo restante exibido, seja b a base, h a altura, e lembre-se de que $u(\theta)'$ fornece a inclinação em cada $\theta \in \Theta$.

- Para a estatística W, $b = (\hat{\theta} - \theta_0)$ a partir de relações trigonométricas básicas $h/b =$

Figura 3 — Comparação geométrica na escala de Escore.



Fonte: Muggeo, Vito & Lovison, Gianfranco. (2014).

$-u(\hat{\theta})'$, portanto $h = (\hat{\theta} - \theta_0)\{-u(\hat{\theta})'\}$. Com \mathcal{A} sendo:

$$\mathcal{A} = \frac{bh}{2} = \frac{1}{2} (\hat{\theta} - \theta_0)^2 \{-u(\hat{\theta})'\} = \frac{1}{2} (\hat{\theta} - \theta_0)^2 / \{-u(\hat{\theta})'\}^{-1} = \frac{1}{2} w$$

onde, $\{-u(\hat{\theta})'\} = \mathcal{J}(\hat{\theta})$.

- Para a estatística ER, $h = u(\theta_0)$ a partir de relações trigonométricas básicas $h/b = -u(\theta_0)'$, portanto $b = u(\theta_0)/\{-u(\theta_0)'\}$. Com \mathcal{A} sendo:

$$\mathcal{A} = \frac{bh}{2} = \frac{1}{2} u(\theta_0)^2 / \{-u'(\theta_0)\} = \frac{1}{2} s,$$

onde, $\{-u(\theta_0)'\} = \mathcal{J}(\theta_0)$.

- Para a estatística G, é visto que $b = (\hat{\theta} - \theta_0)$, $h = u(\theta_0)$. Com \mathcal{A} sendo:

$$\mathcal{A} = \frac{bh}{2} = \frac{1}{2} u(\theta_0) (\hat{\theta} - \theta_0) = \frac{1}{2} g,$$

Para inferências sobre o parâmetro de localização de modelos normais, o escore é linear, e portanto, os quatro triângulos coincidem.

2.3 Funções Poder: Hipóteses Locais De Pitman

Supomos que temos um conjunto de n observações $x = (x_1, \dots, x_n)^T$ cuja função de verossimilhança depende de um parâmetro desconhecido $\theta \in \Theta$. Deseja-se testar a hipótese $H_0 : \theta_1 = \theta_{10}$ contra $H_1 : \theta_1 \neq \theta_{10}$, onde $\theta_1 = (\theta_{(1)}, \dots, \theta_{(m)})^T$ e o componente restante de θ , $\theta_2 = (\theta_{(m+1)}, \dots, \theta_{(p)})^T$ constitui um vetor não especificado de parâmetros de incômodo.

Procedimentos de teste exatos e ótimos como na situação acima existem apenas em uma pequena parcela de casos, e por isso a atenção é dada aos critérios de teste, como o baseado na RV, que são assintoticamente ótimos. Para distinguir entre procedimentos que têm propriedades assintóticas idênticas, é necessário examinar mais de perto o comportamento em amostras pequenas. Considerar a distribuição não nula sob uma série de alternativas de Pitman convergindo para H_0 na taxa $O(n^{-1/2})$, de uma estatística sugerida por Rao (1948), nos permite comparações com resultados correspondentes derivados por Hayakawa (1975) para a RV e a estatística de Wald. Nenhuma estatística é uniformemente superior às demais, mas essa análise nos permite identificar regiões onde uma estatística pode superar suas concorrentes.

Sendo x é uma amostra aleatória de uma distribuição com função de densidade de probabilidade $f(\cdot, |\theta)$, de modo que a função de log-verossimilhança de x é dada por:

$$l(x; \theta) = \sum_{i=1}^n \ln f(x_i | \theta) \quad (9)$$

Foram utilizadas as seguintes suposições e notações:

- A função $l(x; \theta)$ é regular em relação às suas derivadas em θ incluindo até as de quarta ordem;
- Todos os estimadores de máxima verossimilhança relevantes são únicos, pelo menos para n não muito grande;
- O valor verdadeiro do parâmetro será simplesmente escrito como θ ;
- A estimativa de máxima verossimilhança condicional de θ_2 sob H_0 será escrita como $\tilde{\theta}_2$;
- funções avaliadas no ponto $\theta_1 = \theta_{10}$, $\theta_2 = \tilde{\theta}_2$ serão diferenciadas pela adição (\sim), enquanto qualquer função avaliada no ponto $\theta_1 = \theta_{10}$ terá um subscrito zero.

Foram introduzidas as seguintes derivadas do logaritmo da verossimilhança:

$$y_i = n^{-\frac{1}{2}} \frac{\partial l(x; \theta)}{\partial \theta_i}, \quad y_{ij} = n^{-1} \frac{\partial^2 l(x; \theta)}{\partial \theta_i \partial \theta_j}, \quad y_{ijk} = n^{-3/2} \frac{\partial^3 l(x; \theta)}{\partial \theta_i \partial \theta_j \partial \theta_k}$$

com as matrizes:

$$y = (y_1, \dots, y_p)^T, \quad Y = ((y_{ij})), \quad Y_{\dots} = ((y_{ijk}))$$

Juntamente com os cumulantes correspondentes:

$$\kappa_{ij} = E(y_{ij}), \quad \kappa_{i,j} = E(y_i y_j), \\ n^{-\frac{1}{2}} \kappa_{ijk} = E(y_{ijk}), \quad n^{-\frac{1}{2}} \kappa_{i,jk} = E(y_i y_{jk}), \quad n^{-\frac{1}{2}} \kappa_{i,j,k} = E(y_i y_j y_k)$$

com as matrizes:

$$K = ((\kappa_{i,j})), \quad K_{\dots} = ((\kappa_{ijk})), \quad K_{\dots} = ((\kappa_{i,jk})), \quad K_{\dots} = ((\kappa_{i,j,k})).$$

Os κ' s conforme definidos aqui, são todos $O(1)$. Eles estão relacionados pelas seguintes identidades:

$$\kappa_{i,j} + \kappa_{ij} = 0, \quad \kappa_{ijk} + \kappa_{i,jk} + \kappa_{j,ki} + \kappa_{k,ij} + \kappa_{i,j,k} = 0, \quad \kappa_{ijk} + \kappa_{i,jk} = \frac{\partial}{\partial \theta_i} \kappa_{j,k}.$$

Para a convolução de quantidades com três sufixos, foi adotado uma notação essencialmente igual a de Hayakawa (1975),

$$Y_{...} * a * b * c = \sum_{i,j,k=1}^p y_{ijk} a_i b_j c_k, \quad Y_{...} * Y = \left(\sum_{i,j=1}^p y_{ijk} y_{ij} \right).$$

e a partição $\theta = \theta^T = (\theta_1, \theta_2)^T$ induz partições correspondentes em

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad Y = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix}, \quad K = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}$$

Rao (1948) introduziu a estatística de teste ER

$$S_R = \tilde{y}^T \tilde{K}^{-1} \tilde{y},$$

que possui assintoticamente, sob H_0 a distribuição χ^2 com m graus de liberdade. Para obter uma representação mais precisa de sua distribuição, uma expansão em série de S_R em termos de y e Y é derivada até a ordem $O(n^{-1/2})$. Utilizando uma expansão do tipo A de Edgeworth [(Malva, et al. (2008)] da densidade conjunta de y e Y é possível exibir a distribuição de S_R correta para esta ordem. As técnicas envolvidas são aquelas consideradas em maior detalhe por Peers (1971) e Hayakawa (1975). Após uma álgebra bastante pesada chegaram a representação da função de densidade de probabilidade de S_R no entorno de H_0 na forma

$$g(S_R) = f(S_R; m, \lambda_0) + n^{-1/2} \sum_{j=0}^3 c_{Rj} f(S_R; m + 2j, \lambda_0) + O(n^{-1}),$$

onde $f(\chi^2; \nu, \lambda_0)$ denota a função de densidade de probabilidade de uma variável χ^2 não central com ν graus de liberdade e parâmetro de não centralidade λ_0 . Nesse contexto foi encontrado

$$\begin{aligned} c_{R0} &= \frac{1}{6} \{ (K_{.....} - K_{...}) * \eta * \eta * \eta + 3(K_{1..} + K_{1,..}) * \varepsilon * \eta * \eta - 3(K_{...} + 2K_{,..}) * J * \eta \}_0, \\ c_{R1} &= \frac{1}{6} \{ (K_{...} - 2K_{,..}) * \eta * \eta * \eta - 3(K_{1..} + K_{1,..}) * \varepsilon * \eta * \eta - 3K_{,...} * B * \eta \\ &\quad + 3(K_{...} 2K_{,..}) * J * \eta \}_0, \\ c_{R2} &= \frac{1}{2} (K_{.....} * B * \eta)_0, \quad c_{R3} = \frac{1}{6} (K_{...} * \eta * \eta * \eta)_0 \end{aligned}$$

em que

$$J = \begin{bmatrix} 0 & 0 \\ 0 & K_{22}^{-1} \end{bmatrix}, \quad \eta = \begin{bmatrix} 1 \\ -K_{22}^{-1} & K_{21} \end{bmatrix} \varepsilon, \quad B = K^{-1} - J$$

e o parâmetro de não centralidade é dado por

$$\lambda_0 = (\eta^T K \eta)_0$$

3 APLICAÇÕES

Nesta seção falaremos sobre as diferenças e semelhanças entre a "Santíssima Trindade" e comparando o poder dos três testes.

3.1 Diferenças E Semelhanças Entre Os Testes

Um pesquisador gostaria de saber se certo modelo (com quatro variáveis preditoras) se ajusta significativamente melhor que um modelo apenas com duas variáveis preditoras e para isso ele usou três testes que são comumente usados para testar esse tipo de questão, que são: o teste da razão de verossimilhança (RV), o teste de Wald e o teste do escore (também conhecido como multiplicador de Lagrange), embora todos os três testes abordem a mesma questão básica, eles são ligeiramente diferentes.

Características dos testes:

- Esses testes podem ser descritos como testes de diferenças entre modelos aninhados, porque pode-se dizer que um dos modelos está aninhado dentro do outro;
- A hipótese nula para todos os três testes é que o modelo menor é o modelo “verdadeiro”;
- Uma estatística de teste grande indica que a hipótese nula é falsa.

3.1.1 Verossimilhança

Os três testes usam a verossimilhança dos modelos a serem comparados para avaliar seu ajuste. A verossimilhança é a probabilidade dos dados dada as estimativas dos parâmetros. O objetivo de um modelo é encontrar valores para os parâmetros (coeficientes) que maximizam o valor da função de verossimilhança, ou seja, encontrar o conjunto de estimativas de parâmetros que tornam os dados mais prováveis. Muitos procedimentos utilizam o logaritmo da probabilidade, em vez da probabilidade em si, porque é mais fácil de trabalhar. A log-verossimilhança (ou seja, o logaritmo da verossimilhança) será sempre negativa, valores mais próximos de zero indicam um modelo com melhor ajuste. Esses testes são muito gerais e podem ser aplicados a qualquer modelo com função de verossimilhança.

3.1.2 Detalhando Os Três Testes

Teste da razão de verossimilhança:

O teste RV é realizado estimando dois modelos e comparando o ajuste de um

modelo com o ajuste do outro. A remoção de variáveis preditoras de um modelo quase sempre fará com que o modelo se ajuste pior (ou seja, um modelo terá uma probabilidade logarítmica mais baixa), mas é necessário testar se a diferença observada no ajuste do modelo é estatisticamente significativa ou não. O teste RV faz isso comparando as probabilidades logarítmicas dos dois modelos. Se essa diferença for estatisticamente significativa, então diz-se que o modelo menos restritivo (aquele com mais variáveis) ajusta os dados significativamente melhor do que o modelo mais restritivo (aquele com menos variáveis). A estatística de teste resultante segue uma distribuição qui-quadrado, com graus de liberdade iguais ao número de parâmetros que são restringidos.

Usando exemplo em estudo, foi executado um modelo completo e o restrito e foi avaliado a diferença de ajuste usando o teste RV. O modelo completo teve uma log-verossimilhança de $-82,42$ e o restrito $-102,45$, resultando em uma estatística de teste RV de $36,05$ que seguindo uma distribuição qui-quadrado com dois graus de liberdade tem um valor p associado de $p < 0,001$, indicando que o modelo com todos os quatro preditores se ajusta significativamente melhor do que o modelo com apenas dois preditores.

Teste de Wald:

O teste de Wald aproxima-se do teste RV, mas com a vantagem de requerer apenas a estimação de um modelo. O teste de Wald funciona testando a hipótese nula de que um conjunto de parâmetros é igual a algum valor. No exemplo em estudo testado, a hipótese nula é que os dois coeficientes de interesse sejam simultaneamente iguais a zero. Se o teste falhar em rejeitar a hipótese nula, isso sugere que a remoção das variáveis do modelo não prejudicará substancialmente o ajuste desse modelo, uma vez que um preditor com um coeficiente que é muito pequeno em relação ao seu erro padrão geralmente não está ajudando pouco a prever a variável dependente. O teste funciona testando a que distância os parâmetros estimados estão de zero (ou qualquer outro valor sob a hipótese nula) em erros padrão, semelhantes aos testes de hipótese normalmente impressos na saída de regressão. A diferença é que o teste de Wald pode ser usado para testar vários parâmetros simultaneamente, enquanto os testes normalmente impressos na saída de regressão testam apenas um parâmetro por vez.

Usando exemplo em estudo, após a execução do modelo de regressão logística, pode-se utilizar o teste de Wald, no qual tivemos o resultado que ao igualar simultaneamente as variáveis preditoras de interesse a zero, obtivemos pelo teste de Wald um valor qui-quadrado gerado de $27,53$ com dois graus de liberdade, resultando em um valor p de $p < 0,0001$ portanto podemos rejeitar a hipótese nula, indicando que os coeficientes não são simultaneamente iguais a zero. Como a inclusão de preditores estatisticamente significativos deveria levar a uma melhor previsão (isto é, a um melhor ajuste do modelo).

Teste Escore:

Assim como acontece com o teste de Wald, o teste Escore requer a estimativa de apenas um único modelo. A diferença é que com o teste Escore o modelo estimado não inclui o(s) parâmetro(s) de interesse. Isso significa que, no exemplo em estudo, podemos usar o teste Escore para testar se adicionar as duas variáveis preditoras ao modelo resultará em uma melhoria significativa no ajuste do modelo, após executar um modelo com apenas duas variáveis preditoras. A estatística de teste é calculada com base na inclinação da

função de verossimilhança nos valores observados das variáveis do modelo duas variáveis preditoras. Esta inclinação estimada, ou “pontuação”, é a razão pela qual o teste Escore é às vezes chamado de teste de pontuação. As pontuações são então utilizadas para estimar a melhoria no ajuste do modelo se variáveis adicionais forem incluídas no modelo. A estatística de teste é a mudança esperada na estatística qui-quadrado do modelo se uma variável ou conjunto de variáveis for adicionado ao modelo. Como testa a melhoria do ajuste do modelo se variáveis atualmente omitidas forem adicionadas ao modelo, o teste Escore às vezes também é chamado de teste para variáveis omitidas. Às vezes também são chamados de índices de modificação, particularmente na literatura de modelagem de equações estruturais.

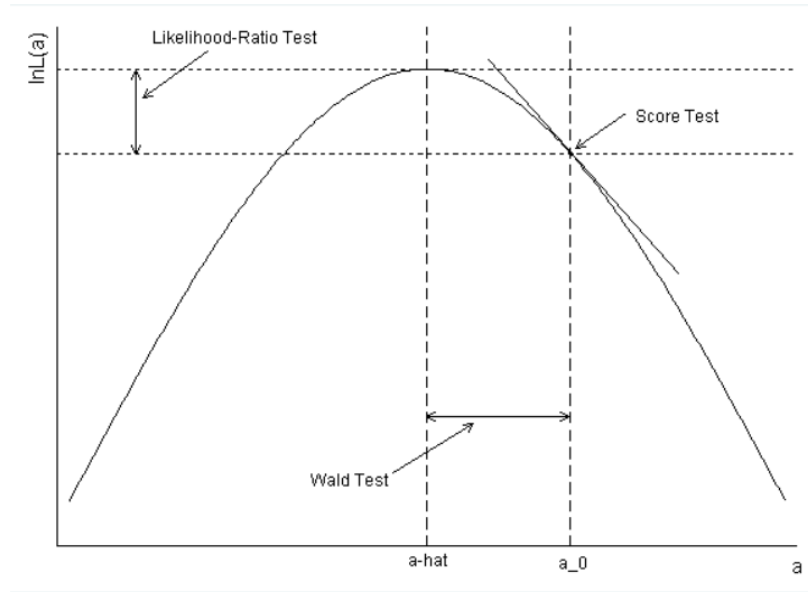
Usando exemplo em estudo, podemos observar os resultados do teste Escore. Ao contrário dos dois testes anteriores, que são usados principalmente para avaliar a mudança no ajuste do modelo quando mais de uma variável é adicionada ao modelo, o teste Escore pode ser usado para testar a mudança esperada no ajuste do modelo se um ou mais parâmetros que são atualmente limitados podem ser estimados livremente. No exemplo em estudo, isso significa testar se adicionar as duas variáveis preditoras ao modelo melhoraria significativamente o ajuste do modelo. A estatística de teste Escore para adicionar as duas variáveis preditoras ao modelo é 31,51, seguindo uma distribuição qui-quadrado, com graus de liberdade iguais ao número de variáveis adicionadas ao modelo, portanto, no exemplo em estudo, duas. O valor p está abaixo do limite estipulado de 0,05, sugerindo que a inclusão das variáveis duas variáveis preditoras no modelo criaria uma melhoria estatisticamente significativa no ajuste do modelo. Esta conclusão é consistente com os resultados dos testes RV e Wald.

3.1.3 Comparando Os Três Testes

Conforme falado anteriormente, todos os três testes abordam a mesma questão básica, restringir os parâmetros a zero, ou seja, deixar de fora essas variáveis preditoras. A diferença entre os testes é como eles fazem essa possível redução. Como vimos, para realizar um teste RV, é necessário estimar ambos os modelos que se deseja comparar. A vantagem dos teste W e teste ER é que eles se aproximam do teste RV, mas exigem que apenas um modelo seja estimado. Tanto o teste de Wald quanto o teste Escore são assintoticamente equivalentes ao teste de RV, ou seja, à medida que o tamanho da amostra se torna infinitamente grande, os valores das estatísticas dos testes se tornarão cada vez mais próximos da estatística do teste de RV. Em amostras finitas, os três tendem a gerar estatísticas de teste um tanto diferentes, mas geralmente chegarão à mesma conclusão. O teste de Wald é comumente usado para realizar testes de múltiplos graus de liberdade em conjuntos de variáveis fictícias usadas para modelar variáveis preditoras categóricas em regressão. A vantagem do teste Escore é que ele pode ser usado para procurar variáveis omitidas quando o número de variáveis candidatas é grande.

Uma maneira de entender melhor como os três testes estão relacionados e como são diferentes é observar uma representação gráfica do que eles estão testando. A figura acima ilustra o que cada um dos três testes faz. Ao longo do eixo x (denominado “a”) estão os valores possíveis do parâmetro a , no exemplo em estudo, este seria o coeficiente de regressão para variáveis preditoras. Ao longo do eixo y estão os valores do log de verossimilhança correspondentes aos valores de a . O teste RV compara as probabilidades logarítmicas de um modelo com valores do parâmetro a restrito a algum valor (em nosso

Figura 4 — Comparação geométrica dos testes RV, Wald e Escore.



Fonte: Fox (1997, p. 570);

exemplo em estudo é zero) para um modelo onde a é estimado livremente. Ele faz isso comparando a altura das probabilidades dos dois modelos para ver se a diferença é estatisticamente significativa, lembre-se, valores mais altos de probabilidade indicam melhor ajuste. Na figura acima, isto corresponde à distância vertical entre as duas linhas pontilhadas. Em contraste, o teste de Wald compara a estimativa do parâmetro \hat{a} com a_0 ; a_0 é o valor de a sob a hipótese nula, que geralmente afirma que $a = 0$. Se \hat{a} for significativamente diferente de a_0 , isso sugere que estimar livremente a (usando \hat{a}) melhora significativamente o ajuste do modelo. Na figura, isso é mostrado como a distância entre a_0 e \hat{a} no eixo x (destacado pelas linhas sólidas). Finalmente, o teste Escore analisa a inclinação do log de verossimilhança quando a é restringido (em nosso exemplo em estudo, a zero). Ou seja, analisa a rapidez com que a probabilidade está mudando no valor hipotético (nulo) de a . Na figura acima, isso é mostrado como a linha tangente em a_0 .

3.2 Comparação De Poder Do Teste

Para comparar o poder do teste em tabelas de contingência foram feitas algumas modificações nas estatísticas ER, W e RV para serem equivalentes ao conhecido teste de Pearson para tabelas de contingência. Temos então:

$$RS = \frac{1}{n} \sum_{i=1}^2 \sum_{j=1}^c \frac{n_{ij}^2}{p_{ij}^0} - n, \quad (10)$$

$$W = n^2 \sum_{i=1}^2 \sum_{j=1}^c \frac{(p_{ij}^0)^2}{n_{ij}} - n, \quad (11)$$

$$LR = 2 \sum_{i=1}^2 \sum_{j=1}^c n_{ij} \ln \left(\frac{\hat{p}_{ij}}{p_{ij}^0} \right). \quad (12)$$

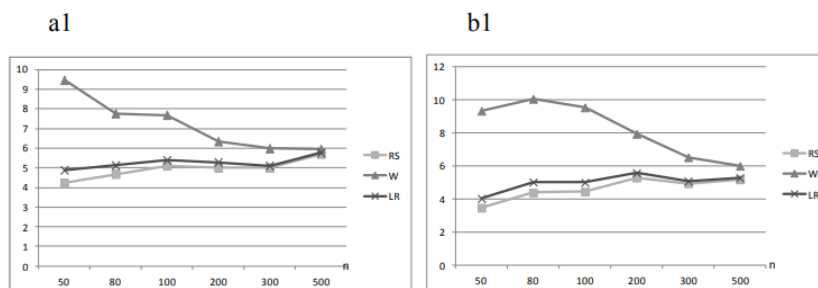
Tabela 1 — Probabilidade para diferentes casos de Tabelas de contingência ($2 \times c$)

a1	0.12	0.3	0.18	b1	0.12	0.24	0.06	0.18
	0.08	0.2	0.12		0.08	0.16	0.04	0.12
a2	0.1	0.3	0.2	b2	0.1	0.24	0.06	0.2
	0.1	0.2	0.1		0.1	0.16	0.04	0.1
a3	0.08	0.3	0.22	b3	0.08	0.24	0.06	0.22
	0.12	0.2	0.08		0.12	0.16	0.04	0.08
a4	0.06	0.3	0.24	b4	0.06	0.24	0.06	0.24
	0.14	0.2	0.06		0.14	0.16	0.04	0.06
a5	0.04	0.3	0.26	b5	0.04	0.24	0.06	0.26
	0.16	0.2	0.04		0.16	0.16	0.04	0.04
c1	0.18	0.2	0.06	0.1	0.06			
	0.08	0.12	0.04	0.06	0.1			
c2	0.18	0.12	0.08	0.06	0.1	0.06		
	0.08	0.06	0.06	0.04	0.06	0.1		

Por ser difícil comparar as estatísticas descritas anteriormente de forma analítica, para verificar suas propriedades foi realizado alguns estudos de simulação, tomando como critério o poder do teste. Para diversos conjuntos de probabilidades em tabelas de contingência (Tabela 1) foi gerado 5,000 amostras para $n = 50, 80, 100, 200, 300$ e 500 , utilizando o software R com a função `rmultinom()` do pacote `stats` (R Core Equipe (2013)).

Para cada amostra foram calculados os valores dos testes ER, W e RV através das fórmulas (9), (10) e (11), e tomando 0,05 como nível de significância foi estabelecido o poder desses testes. Como a função de poder é a probabilidade de rejeição de uma hipótese falsa, o poder é calculado como uma porcentagem de rejeições. A Figura 5 contém os gráficos das funções que representam o tamanho do teste (nível de significância). Nos casos a1 e b1 a hipótese de independência não é rejeitada.

Figura 5 — Tamanho (em porcentagens) dos testes ER, W e RV.

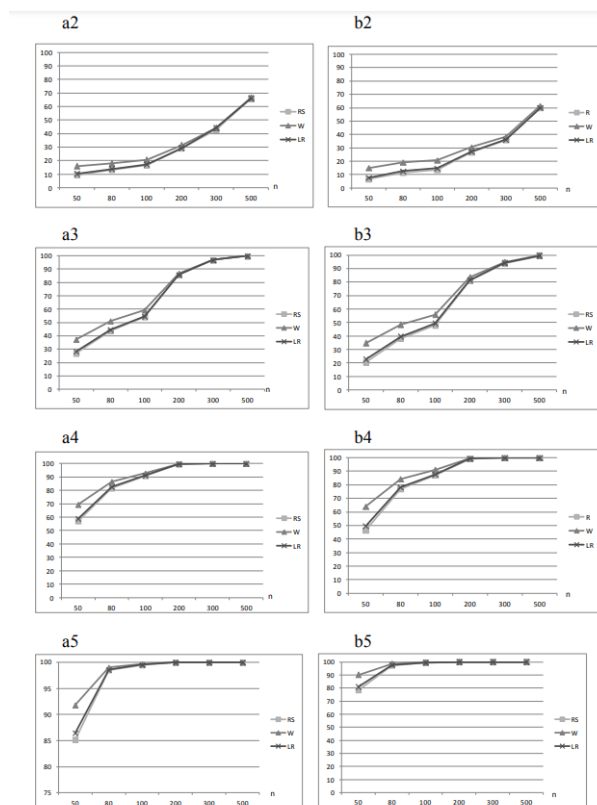


Fonte: Dobek, Anita & Moliński, Krzysztof & Skotarczak, Ewa. (2015).;

A Figura 6 contém os gráficos das funções que representam o poder do teste. Em todos estes casos a hipótese deve ser rejeitada. Uma conclusão geral que pode ser

tirada dos resultados é que para tamanhos de amostra variando de 50 a 300, o teste de W demonstra superioridade aos outros testes em termos de poder. Nos casos *a1* e *b1*, à medida que *n* aumenta, o tamanho do teste (nível de significância) aproxima-se do valor assumido. Para pequenos *n* W excede significativamente o valor de 0,05. As outras duas estatísticas comportam-se de forma semelhante com valores próximos de 0,05.

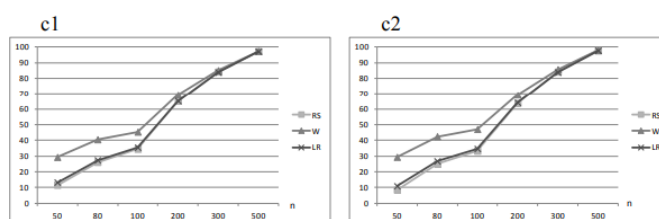
Figura 6 — Poder (em porcentagens) dos testes ER, W e RV.



Fonte: Dobek, Anita & Moliński, Krzysztof & Skotarczak, Ewa. (2015).;

Os demais gráficos da Figura 6 ilustram uma situação onde nas tabelas de contingência estão se afastando da independência, com aumento da distância euclidiana entre *a1* (*b1*) e *a2* – *a5* (*b2* – *b5*). As curvas de potência são funções crescentes até o valor máximo (100%).

Figura 7 — Poder (em porcentagens) dos testes ER, W e RV.



Fonte: Dobek, Anita & Moliński, Krzysztof & Skotarczak, Ewa. (2015).;

Na Figura 7 superioridade do teste de W também é observada para tabelas de contingência maiores (*c1*, *c2*). Simulações semelhantes foram realizadas para tabelas

2×2 e os resultados são os mesmos das tabelas maiores. Nos estudos de simulação também foi alterado as probabilidades marginais e novamente os resultados foram os mesmos. Com base nestes estudos de simulação podemos afirmar que a estatística de W tem o maior poder no que diz respeito ao problema de testar a independência em tabelas de contingência ($2 \times c$). Recorde-se, no entanto, que a sua validade está sujeita a uma limitação significativa que não se aplica às restantes estatísticas, nomeadamente a exigência de um número diferente de zero de casos em cada célula.

4 EXTENSÕES

O teste ER é muito útil na prática estatística, pois fornece uma maneira simples e unificada de construir testes para algumas hipóteses simples e compostas em muitos problemas estatísticos, especialmente quando os cálculos baseados em funções de verossimilhança são difíceis, mas as funções de verossimilhança e suas derivadas sob as hipóteses nulas são mais fáceis de avaliar. Por exemplo, recentemente, Stiger, Barnhart e Williamson (1999) utilizaram o teste ER para avaliar a suposição de proporcionalidade em um modelo de chances proporcionais ajustado com a metodologia de equações de estimação generalizadas. Eles apontaram que o teste de pontuação requer apenas o ajuste do modelo de chances proporcionais, enquanto a aplicação do teste de Wald exige o ajuste de modelos de regressão logística separados para cada dicotomização da variável de resposta.

O objetivo dessa extensão é apontar outro fato interessante sobre o teste ER, alguns testes populares na prática estatística, derivados, são na verdade testes ER sob alguns modelos estatísticos. O teste log-rank em análise de sobrevivência será usado como exemplo para ilustrar esse ponto. Há uma vantagem em formular esse teste como teste ER, os resultados assintóticos derivados para o teste ER em um modelo estatístico geral pode ser aplicado para avaliar as propriedades assintóticas desse teste.

Na maioria dos ensaios clínicos de fase III, os pacientes são randomizados para dois ou mais grupos de tratamento. Por exemplo, o primeiro grupo pode consistir em pacientes tratados com um novo medicamento, e o segundo grupo, aqueles tratados com um medicamento padrão. As experiências de sobrevivência dos pacientes em diferentes grupos são então comparadas. Neste artigo, consideramos apenas o caso em que existem apenas dois grupos. Seja $h_1(t)$ a taxa de risco para pacientes no primeiro grupo e $h_2(t)$ para a segunda população. As hipóteses que estamos interessados em testar são:

$$H_0 : h_1(t) = h_2(t) \quad H_1 : h_1(t) \neq h_2(t).$$

Suponha que, para testar as hipóteses acima, randomizamos n_1 pacientes para o primeiro grupo e n_2 para o segundo grupo de tratamento. O teste de log-rank, proposto por Mantel e Haenszel (1959), é o teste estatístico mais popular na literatura oncológica para testar as hipóteses acima.

Assume-se que, na amostra combinada, houve D tempos distintos, $t_{(1)} < t_{(2)} < \dots < t_{(D)}$, onde pelo menos uma falha foi observada. Para $i = 1, \dots, D$, seja d_{1i} o número de falhas observadas na amostra do estudo no tempo $t_{(i)}$ e n_i o número de sujeitos na amostra do estudo em risco no tempo $t_{(i)}$. Então, sob H_0 , d_{1i} pode ser assumido distribuído como hipergeométrica com expectativa $e_{1i} = n_i d_i / n_i$, onde d_i é o número total de falhas observadas no tempo $t_{(i)}$ e n_i é o número total de sujeitos em risco no tempo $t_{(i)}$. Portanto, a soma das diferenças entre d_{1i} e e_{1i} ao longo de $i = 1, 2, \dots, D$

pode ser usada como uma estatística para medir quão provável é que a hipótese nula seja verdadeira.

Seja ν_{1i} a variância de d_{1i} calculada a partir da distribuição hipergeométrica e assumamos que todas as diferenças são independentes. Então, a variância para a soma das diferenças pode ser escrita como uma soma de ν_{1i} ao longo de $i = 1, 2, \dots, D$. Portanto, uma estatística de teste para as hipóteses ditas anteriormente pode ser definida como:

$$W_L = \frac{\left[\sum_{i=1}^D (d_{1i} - e_{1i}) \right]^2}{\sum_{i=1}^D \nu_{1i}},$$

A distribuição de W_L é próxima à de uma distribuição χ^2 com um grau de liberdade. Portanto, rejeitaríamos H_0 no nível α se $W_L > C_\alpha$. Este teste possui vários nomes, incluindo teste de Mantel-Haenszel, teste de Mantel-Cox e teste de Peto-Mantel-Haenszel, mas é mais conhecido na análise de sobrevivência e na literatura oncológica como teste de log-rank, após Peto et al. cunharem esse nome em seu famoso artigo sobre o design e análise de ensaios clínicos com desfechos de sobrevivência (Peto et al., (1977)).

Este teste também foi apresentado como um teste de escore. Defina $Z = 1$ se um paciente for atribuído ao primeiro grupo e $Z = 0$ se ao segundo grupo. Seja $h(t, z)$ a função de risco dado $Z = z$ e assumamos um modelo de regressão de riscos proporcionais, temos:

$$h(t, z) = h_0(t) \exp(\beta z),$$

onde $h_0(t)$ é uma função de risco desconhecida. Então, as hipóteses anteriores em que estamos interessados são equivalentes a:

$$H_0 : \beta = 0 \quad \text{vs.} \quad H_1 : \beta \neq 0.$$

Assumamos que os dados que temos estão na seguinte forma:

$$\{X_i = \min(T_i, C_i), \delta_i = I(T_i \leq C_i), Z_i\}, \quad i = 1, \dots, N$$

onde N é o tamanho total da amostra de pacientes randomizados para dois grupos. Assumamos também que:

- $Y_i = (T_i, C_i, Z_i), i = 1, 2, \dots, N$ são independentes e identicamente distribuídos como $Y = (T, C, Z)$;
- T e C são condicionalmente independentes dado Z .

Denote

$$\begin{aligned} N_i(t) &= I(X_i \leq t, \delta_i = 1), \\ Y_i(t) &= I(X_i \geq t), \pi(t) = E\{Y_1(t)\} = P\{X_1 \geq t\}, \\ S^{(j)}(\beta, t) &= N^{-1} \sum_{i=1}^n Z_i^j \exp(\beta Z_i) Y_i(t). \end{aligned}$$

Para τ satisfazendo $\pi(\tau) > 0$, defina a função de log-verossimilhança parcial para

β , proposto por Cox (1972):

$$l(\beta) = \sum_{i=1}^N \int_0^{\tau} [\beta Z_i - \log \{s^{(0)}(\beta, u)\}] dN_i(u).$$

A partir dessa função de log-verossimilhança parcial, podemos calcular a função de escore associada

$$U(\beta) = \frac{\partial l(\beta)}{\partial \beta} = \sum_{i=1}^N \int_0^{\tau} \left\{ Z_i - \frac{S^{(1)}(\beta, u)}{S^{(0)}(\beta, u)} \right\} dN_i(u).$$

e a função de informação observada

$$I(\beta) = \frac{\partial^2 l(\beta)}{\partial \beta^2} = \sum_{i=1}^N \int_0^{\tau} \left[\frac{S^{(2)}(\beta, u)}{S^{(0)}(\beta, u)} - \left\{ \frac{S^{(1)}(\beta, u)}{S^{(0)}(\beta, u)} \right\}^2 \right] dN_i(u).$$

Cox (1972) mostrou que o teste de escore, definido como rejeitar H_0 se

$$W_c = I^{-1}(0)U(0) > C_{\alpha},$$

é o mesmo que o teste de log-rank introduzido anteriormente. Dado que tanto o teste de log-rank pode ser expressos como teste ER, suas propriedades matemáticas podem ser discutidas rigorosamente com base na teoria assintótica geral desenvolvida para testes ER em um modelo estatístico geral.

5 CONSIDERAÇÕES FINAIS

A presente revisão proporcionou uma análise sobre a criação, propriedades, interpretação geométrica, função poder em hipóteses de Pitman, diferenças e semelhanças e comparar o poder do teste para "A Santíssima Trindade" e ver como podemos transformar o teste log-rank em termos de um teste Escore de Rao. Ao longo da revisão, foi possível constatar a relevância e a versatilidade dessa estatística, destacando seu papel fundamental em diversas áreas de pesquisa. O Escore de Rao é uma ferramenta estatística versátil, cujas aplicações são vastas e continua a desempenhar um papel fundamental no avanço do conhecimento estatístico e na análise de dados.

REFERÊNCIAS

- 1 Aitchison, J. and S. D. Silvey (1958), "Maximum-Likelihood Estimation of Parameters Subject to Restraints," *Annals of Mathematical Statistics* , 29, 813-828.
- 2 Bera, A. K. (1982a), *Aspects of Econometric Modelling* . Unpublished Ph.D. Dissertation, Australian National University.
- 3 Bera, A. K. and C. M. Jarque (1981), "An Efficient Large-Sample Test for Normality of Observations and Regression Residuals," *Working Papers in Economics and Econometrics* , 40, The Australian National University.

- 4 Bera, A. K. and C. R. McKenzie (1986), "Alternative Forms and Properties of the Score Test," *Journal of Applied Statistics* , 13, 13-25.
- 5 Bera, A.K. and Ullah, A., (1991). "Rao's Score Test in Econometrics," Discussion Paper 1991-43, Tilburg University, Center for Economic Research.
- 6 Boos, D. D. and Stefanski, L. A. (2013). *Essential Statistical Inference: Theory and Methods*. Springer.
- 7 Breusch, T. S. (1978), "Testing for Autocorrelation in Dynamic Linear Models," *Australian Economic Papers* , 17, 334-355.
- 8 Breusch, T. S. and A. R. Pagan (1979), "A Simple Test for Heteroskedasticity and Random Coefficient Variation," *Econometrica* , 47, 1287-1294.
- 9 Cox, D.R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc., B*, 34, 187–220.
- 10 Del Pino, G. E. (1990), "A Coordinate Free Approach to Score Tests," *Communications in Statistics, Theory and Methods* , 19, 155-167.
- 11 Dobek, Anita and Moliński, Krzysztof and Skotarczak, Ewa. (2015). Power comparison of Raos score test, the Wald test and the likelihood ratio test in (2xc) contingency tables. *Biometrical Letters*. 52. 10.1515/bile-2015-0009.
- 12 FAQ: How are the likelihood ratio, Wald, and Lagrange multiplier (score) tests different and/or similar? Disponível em: <<https://stats.oarc.ucla.edu/other/mult-pkg/faq/general/faqhow-are-the-likelihood-ratio-wald-and-lagrange-multiplier-score-tests-different-andor-similar/>>.
- 13 Fox, J. (1997) *Análise de regressão aplicada, modelos lineares e métodos relacionados*. Thousand Oaks, CA: Publicações Sage.
- 14 Godfrey, L. G. (1978a), "Testing Against General Autoregressive and Moving Average Error Models When the Regressors Include Lagged Dependent Variables," *Econometrica* , 46, 227-236.
- 15 Godfrey, L. G. (1981), "On the Invariance of the Lagrange Multiplier Test with Respect to Certain Changes in the Alternative Hypothesis," *Econometrica* , 49, 1443-1455.
- 16 Godfrey, L. G. (1988), *Misspecification Tests in Econometrics, The Lagrange Multiplier Principle and Other Approaches*, (New York: Cambridge University Press).
- 17 Harris, P. and Peers, H.W. (1980). The local power of the efficient scores test statistic, *Biometrika*, Volume 67, Issue 3, Pages 525–529, <https://doi.org/10.1093/biomet/67.3.525>.
- 18 Hayakawa, T. (1975). The likelihood ratio criterion for a composite hypothesis under a local alternative. *Biometrika* 62, 451-60.

- 19 Jarque, C. M. and A. K. Bera (1987), "Test for Normality of Observations and Regression Residuals," *International Statistical Review* , 55, 163-172.
- 20 Johnston, J. and DiNardo, J. (1997) *Métodos Econométricos Quarta Edição*. Nova York, NY: The McGraw-Hill Companies, Inc.
- 21 Malva, M., Mendonça, S. and Pestana, D. (2008). *Expansões de Edgeworth e Aproximações Pré-Assintóticas*.
- 22 Mantel, N.; Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute* comma; 22, 719–748.
- 23 Muggeo, Vito and Lovison, Gianfranco. (2014). The “Three Plus One” Likelihood-Based Test Statistics: Unified Geometrical and Graphical Interpretations. *The American Statistician*. 68. 10.1080/00031305.2014.955212.
- 24 Neyman, J. and E. S. Pearson (1928), On the Use and Interpretation of Certain Test Criteria for Purpose of Statistical Inference, *Biometrika*, 20, 175-240.
- 25 Peers, H. W. (1971). Likelihood ratio and associated test criteria. *Biometrika* 58, 577-87.
- 26 Pesaran, M. H. (1979), "Diagnostic Testing and Exact Maximum Likelihood Estimation of Dynamic Models," in E. G. Charatsis (ed.), *Proceedings of the Econometric Society Meeting* , 1979, Amsterdam: North Holland.
- 27 Peto, R., Pike, M.C., Armitage, P., Breslow, N.E., Cox, D.R., Howard, S.V., Mantel, N., McPherson, K., Peto, J., Smith, P.G. (1977). Design and analysis of randomized clinical trials requiring prolonged observation of each patient, II. *Br. J. Cancer*, 35, 1–39.
- 28 R Core Team (2013): *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.Rproject.org>.
- 29 Rao, C. R. (1948), Large Sample Tests of Statistical Hypotheses Concerning Several Parameters with Applications to Problems of Estimation, *Proceedings of the Cambridge Philosophical Society* , 44, 50-57.
- 30 Rao, C. R. (1973), *Linear Statistical Inference and Its Applications* , New York: John Wiley and Sons.
- 31 Rao, C. R. (2005). Score test: historical review and recent developments. In N. Balakrishnan, N. Kannan, and H. N. Nagaraja, editors, *Advances in Ranking and Selection, Multiple Comparisons, and Reliability*. Birkhäuser, Boston.
- 32 Rao, C. R. and Poti, S. J. (1946). On locally most powerful tests when alternatives are one sided.
- 33 Rao, C.R (2009). Teste de pontuação Rao. *Scholarpedia*, 4(10):8220., revisão #121946

- 34 Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, New York: John Wiley and Sons.
- 35 Silvey, S. D. (1959), "The Lagrange Multiplier Test," *Annals of Mathematical Statistics* , 30, 389-407.
- 36 Stiger, T.R., Barnhart, H.X. and Williamson, J.M. (1999) Testing proportionality in the proportional odds model fitted with GEE. *Stat. Med.* 18, 1419–1433.
- 37 Terrell, G. (2002). The gradient statistic. *Computing Science and Statistics*, 34, 206–215.
- 38 Tu, D. (2001). Rao's Score Tests in Survival Analysis: Examples and Bartlett Type Adjustments. <https://doi.org/10.1081/sta-100105682>.
- 39 Ullah, A. (1989), "Distance Measures and Their Applications in Economics," mimeo, Department of Economics, University of Western Ontario.
- 40 Wald, A. (1943), Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observation is Large, *Transactions of American Mathematical Society* , 54, 426-482.