

APROXIMAÇÕES DO INTERVALO DE CONFIANÇA EXATO PARA VARIÂNCIA DE UMA POPULAÇÃO NORMAL

Daniel Furtado FERREIRA¹
Denismar Alves NOGUEIRA¹

- RESUMO: A variância de uma população é um importante parâmetro a ser estimado. São inúmeras as áreas que requerem estimativas precisas da variância. A estimação por intervalo para variância tem por finalidade expressar a precisão de estimativas desse parâmetro. O objetivo desse trabalho é o de apresentar um intervalo de confiança para a variância de uma população normal resultante da transformação de Wilson e Hilferty (1931) e da aproximação assintótica de qui-quadrado de Bishop, Fienberg e Holland (1975) e avaliar suas *performances* por meio de simulação Monte Carlo utilizando 2.000 iterações. O intervalo de confiança baseado na aproximação de Wilson e Hilferty (1931) mostrou-se com acurácia semelhante ao do intervalo exato para coeficiente de confiança variando de 10% a 99,99%, devendo ser recomendado para $\nu > 4$; sendo $\nu = n - 1$ os graus de liberdade da amostra. O intervalo baseado na aproximação de Bishop, Fienberg e Holland (1975) pode ser recomendado somente para $\nu > 30$, sendo, no entanto, o intervalo anterior considerado melhor; o intervalo normal assintótico apresentou robustez para populações não normais e pode ser recomendado para estimar intervalos de confiança das variâncias dessas populações na ausência de um método exato ou mais eficiente; os intervalos de confiança propostos não necessitam da estimação de quantis da distribuição de qui-quadrado e apresentam resultados similares ao do intervalo exato, uma vez que para grandes valores de ν , situação essa em que os métodos numéricos são instáveis para obter quantis da distribuição qui-quadrado, estes métodos possuem maiores acurácia; o valor didático dessas aproximações deve ser considerado para suas utilizações e recomendações.
- PALAVRAS-CHAVE: Aproximação assintótica; qui-quadrado; Monte Carlo; simulação.

1 Introdução

A variância de variáveis aleatórias de uma população é um importante parâmetro a ser estimado em diferentes situações experimentais. São inúmeras as áreas que requerem estimativas precisas da variância para estabelecerem estratégias e para nortear as escolhas de métodos e técnicas adequadas para o destino futuro de ações de uma instituição de pesquisa ou de uma empresa. Estimativas de variâncias são reconhecidamente importantes, por exemplo, no melhoramento de plantas e de animais e no controle da qualidade de processos de produção. A estimação por intervalo para variância tem por finalidade expressar a precisão de estimativas desse parâmetro, que não são contempladas na estimação pontual.

¹ Departamento de Ciências Exatas, Universidade Federal de Lavras - UFLA, Lavras, MG, CEP 37200-000. E-mail: danielff@ufla.br. Bolsista CNPq.

Os processos de estimação intervalar de variâncias são largamente difundidos na literatura (Maatta e Casella, 1987), sobretudo nas situações em que a população amostrada é normal. Para a situação de normalidade, o intervalo de confiança baseado na distribuição de qui-quadrado (χ^2) com $\nu = n - 1$ graus de liberdade, sendo n o tamanho da amostra, pode ser utilizado. Para grandes valores de ν , no entanto, os algoritmos numéricos utilizados para obtenção dos quantis da distribuição de qui-quadrado, necessários para estimar os intervalos, apresentam reconhecida dificuldade de operação. Como é apontado por Benton e Krishnamoorthy (2003, p.258), algoritmos conhecidos na literatura e publicados na *Applied Statistic* em algumas situações, em que os graus de liberdade são elevados, não conseguem calcular o valor da função de distribuição de qui-quadrado por apresentarem “overflow”, principalmente na distribuição de qui-quadrado não-central. Os pacotes computacionais que utilizam essas rotinas podem apresentar problemas de “overflow” em consequência desse fato.

Para a obtenção de um critério para testar a homogeneidade de variâncias, Samiuddin (1976) apresenta uma aproximação normal de qui-quadrado utilizando a transformação de Wilson e Hilferty (1931). Essa aproximação é bastante acurada, até mesmo para pequenos valores de ν , como, por exemplo $\nu = 1$, desde que determinados valores de α não próximos de 0 ou de 1 sejam usados. Nesse caso, o valor de α é definido pelo seguinte evento probabilístico da distribuição de qui-quadrado: $P(\chi^2 > \chi_\alpha^2)$ (α é a probabilidade do erro tipo I). Para ν maiores a acurácia se estende para valores de α próximos de zero ou de um, que são as regiões limitantes em relação a precisão para pequenos valores de ν . Uma segunda aproximação assintótica de qui-quadrado é proposta por Bishop, Fienberg e Holland (1975). Essa última aproximação assintótica, considera uma expansão em série de Taylor de segunda ordem e é apresentada nas equações (3.6) e (3.7). Esses autores apontam uma melhor acurácia dessa aproximação.

Como para grandes amostras, os algoritmos numéricos para obtenção dos quantis de qui-quadrado são mais instáveis (Press et al., 1992) e, portanto, constituem situações inapropriadas para o emprego de algoritmos numéricos, pretende-se nesse trabalho propor alternativas baseadas apenas nos quantis da distribuição normal padrão. Sendo assim, essas situações de grandes amostras são circunstâncias ideais de trabalho das aproximações apresentadas, uma vez que são aproximações assintóticas.

Do exposto objetivou-se por intermédio desse trabalho apresentar um intervalo de confiança para a variância de uma população normal resultante da transformação de Wilson e Hilferty (1931) e da aproximação assintótica de qui-quadrado de Bishop, Fienberg e Holland (1975), bem como avaliar suas *performances* por meio de simulação Monte Carlo, comparando-os ao intervalo de confiança exato.

2 Métodos

2.1 Intervalos de confiança para variância de uma normal

Seja uma amostra aleatória X_1, X_2, \dots, X_n de uma população normal com média μ e variância σ^2 , $N(\mu, \sigma^2)$, desconhecidos. A distribuição de amostragem da quantidade

pivô (1) é conhecida por distribuição de qui-quadrado (Mood, Graybill e Boes, 1974) com parâmetro $\nu = n - 1$.

$$\chi^2_\nu = \frac{(n-1)S^2}{\sigma^2} = \frac{\nu S^2}{\sigma^2}, \quad (1)$$

em que: $\nu = n - 1$ são os graus de liberdade e S^2 é o estimador pontual não viesado de σ^2 , dado por:

$$S^2 = \frac{1}{n-1} \left[\sum_{i=1}^n X_i^2 - \frac{\left(\sum_{i=1}^n X_i \right)^2}{n} \right] \quad (2)$$

A distribuição de χ^2 é dada por:

$$f(\chi^2) = \frac{1}{2\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\chi^2}{2}\right)^{\frac{\nu}{2}-1} e^{-\frac{\chi^2}{2}}, \text{ para } \chi^2 > 0 \quad (3)$$

em que $\Gamma(\bullet)$ é função gama do argumento.

Os quantis da distribuição qui-quadrado, necessários para estimar intervalos de confiança para σ^2 , podem ser obtidos da inversão da função de distribuição de probabilidade correspondente. Esse processo pode ser realizado por intermédio de métodos numéricos de integração e de obtenção dos zeros de uma função. Esses métodos se tornam menos eficientes quando ν cresce.

Sejam os quantis superiores $(1-\alpha/2) \times 100\%$ e $\alpha/2 \times 100\%$, respectivamente, da distribuição de qui-quadrado definidos pelas seguintes afirmativas probabilísticas:

$$P\left(\chi^2_\nu > \chi^2_{1-\alpha/2, \nu}\right) = 1 - \alpha/2 \text{ e } P\left(\chi^2_\nu > \chi^2_{\alpha/2, \nu}\right) = \alpha/2.$$

Logo, é possível obter o já conhecido intervalo exato de confiança para σ^2 de uma normal por:

$$P\left(\chi^2_{\alpha/2, \nu} \leq \chi^2_\nu \leq \chi^2_{1-\alpha/2, \nu}\right) = 1 - \alpha \quad (4)$$

Substituindo o resultado de (1) em (4) e resolvendo o sistema de equações para σ^2 tem-se:

$$P\left(\chi^2_{\alpha/2, \nu} \leq \frac{\nu S^2}{\sigma^2} \leq \chi^2_{1-\alpha/2, \nu}\right) = P\left(\frac{\nu S^2}{\chi^2_{\alpha/2, \nu}} \leq \sigma^2 \leq \frac{\nu S^2}{\chi^2_{1-\alpha/2, \nu}}\right) = 1 - \alpha$$

Dessa afirmativa probabilística constrói-se o intervalo de confiança para σ^2 de uma normal (Mood, Graybill e Boes, 1974), o qual é dado por:

$$IC_{1-\alpha}(\sigma^2): \left[\frac{\nu S^2}{\chi_{\alpha/2, \nu}^2}; \frac{\nu S^2}{\chi_{1-\alpha/2, \nu}^2} \right] \quad (5)$$

Para grandes amostras é possível obter um outro intervalo de confiança, utilizando argumentos diferentes. Seja para isso o estimador de máxima verossimilhança de σ^2 dado por $\hat{\sigma}^2 = (n-1)S^2/n$. O intervalo de confiança pode ser obtido pela aproximação normal assintótica de um estimador de máxima verossimilhança. Para isso, considere a variância de S^2 de uma população qualquer, que é dada por:

$$V(S^2) = \frac{1}{n} \left(\mu_4 - \frac{n-3}{n-1} \sigma^4 \right) \quad (6)$$

Para a distribuição normal, o momento centrado de ordem 4 é dado por $\mu_4 = 3\sigma^4$, em que de uma forma geral defini-se $\mu_4 = \int_{-\infty}^{+\infty} (x-\mu)^4 f(x) dx$, sendo $f(x)$ a função de densidade de probabilidade do modelo considerado (Mood, Graybill e Boes, 1974). Substituindo esse resultado na expressão (6), obtém-se a variância de S^2 de uma normal:

$$V(S^2) = \frac{2\sigma^4}{\nu} \quad (7)$$

Assim, utilizando a propriedade de normalidade assintótica de um estimador de máxima verossimilhança, porém fazendo a sua substituição pelo estimador não viesado S^2 , do qual foi obtida a variância (7), obtém-se:

$$Z = \frac{S^2 - \sigma^2}{\sigma^2 \sqrt{\frac{2}{\nu}}} \quad (8)$$

que tem distribuição assintoticamente normal padrão, $N(0, 1)$.

Com base no quantil superior $Z_{\alpha/2}$ da distribuição normal padrão a seguinte afirmativa probabilística é válida para n suficientemente grande:

$$P \left(-Z_{\alpha/2} \leq \frac{S^2 - \sigma^2}{\sigma^2 \sqrt{\frac{2}{\nu}}} \leq Z_{\alpha/2} \right) \approx 1 - \alpha \quad (9)$$

De (9) pode-se explicitar σ^2 nas inequações e o resultado obtido é o intervalo normal assintótico para σ^2 de uma população normal. Esse intervalo é dado por:

$$IC_{1-\alpha}(\sigma^2): \left[\frac{S^2}{1 + Z_{\alpha/2} \sqrt{\frac{2}{\nu}}}; \frac{S^2}{1 - Z_{\alpha/2} \sqrt{\frac{2}{\nu}}} \right] \quad (10)$$

Se ν for pequeno, esse intervalo pode levar a resultados negativos para o limite superior do intervalo de confiança em função da escolha de α . Essa situação é, no entanto, inapropriada para se utilizar uma aproximação assintótica, que exige valores grandes de ν (n suficientemente grande).

2.2 Aproximações normais do intervalo de confiança exato

A distribuição de qui-quadrado apresentada na expressão (3) possui média ν e variância 2ν . Sabe-se que o coeficiente de assimetria dessa distribuição é dado por $\sqrt{\beta_1} = 2\sqrt{2}/\sqrt{\nu}$ e o coeficiente de curtose é $\beta_2 = 3 + 12/\nu$. Para grandes valores de ν , $\sqrt{\beta_1}$ tende a zero e β_2 tende a 3, que são os valores dos coeficientes de assimetria e de curtose da distribuição normal. Assim, poderia se pensar em uma aproximação normal padrão para a quantidade pivô $\hat{Z} = (\chi^2 - \nu)/\sqrt{2\nu}$. Fazendo a substituição de χ^2 por $\nu S^2/\sigma^2$ e Z pelo quantil superior $Z_{\alpha/2}$ é possível utilizando uma afirmativa probabilística semelhante a empregada em (9) obter o intervalo de confiança para a variância de uma normal. Não é nada surpreendente observar que o resultado que pode ser facilmente obtido, resulta no intervalo de confiança (10). Essa aproximação de qui-quadrado pela normal, no entanto, tem convergência bastante lenta. Assim, espera-se que o intervalo (10) tenha acurácia somente para grandes valores de n .

Todavia, Samiuddin (1976) apresenta uma aproximação de qui-quadrado (χ^2) por meio de uma normal padrão (Z) utilizando a transformação de Wilson e Hilferty (1931), dada por:

$$\hat{\chi}_\nu^2 = \nu \left[Z \sqrt{\frac{2}{9\nu}} + 1 - \frac{2}{9\nu} \right]^3 \quad (11)$$

A inversão dessa função, necessária para a derivação do intervalo almejado, foi obtida pelos autores e é dada por:

$$\hat{Z} = \frac{\sqrt[3]{\frac{\hat{\chi}_\nu^2}{\nu} - 1 + \frac{2}{9\nu}}}{\sqrt{\frac{2}{9\nu}}} \quad (12)$$

Tomando por base a afirmativa probabilística (9), utilizando o resultado da expressão (12), substituindo, ainda, $\hat{\chi}_\nu^2$ por $\nu S^2/\sigma^2$, obtém-se o intervalo de confiança explicitando σ^2 na expressão resultante.

$$P \left(-Z_{\alpha/2} \leq \frac{\sqrt[3]{\frac{\nu S^2/\sigma^2}{\nu} - 1 + \frac{2}{9\nu}}}{\sqrt{\frac{2}{9\nu}}} \leq Z_{\alpha/2} \right) \approx 1 - \alpha \quad (13)$$

Assim, o resultado obtido é:

$$P\left(\frac{S^2}{\left(1-\frac{2}{9\nu}+Z_{\alpha/2}\sqrt{\frac{2}{9\nu}}\right)^3}\leq\sigma^2\leq\frac{S^2}{\left(1-\frac{2}{9\nu}-Z_{\alpha/2}\sqrt{\frac{2}{9\nu}}\right)^3}\right)\approx 1-\alpha \quad (14)$$

Da afirmativa probabilística apresentada na expressão (14) verifica-se que o almejado intervalo de confiança aproximado é dado por:

$$IC_{1-\alpha}(\sigma^2): \left[\frac{S^2}{\left(1-\frac{2}{9\nu}+Z_{\alpha/2}\sqrt{\frac{2}{9\nu}}\right)^3}; \frac{S^2}{\left(1-\frac{2}{9\nu}-Z_{\alpha/2}\sqrt{\frac{2}{9\nu}}\right)^3} \right] \quad (15)$$

Uma outra aproximação assintótica da distribuição de qui-quadrado é apresentada por Bishop, Fienberg e Holland (1975). Essa aproximação é dada por:

$$\hat{\chi}_\nu^2 = \nu + Z\sqrt{2\nu} + \frac{7}{12}(Z^2 - 1) \quad (16)$$

cuja função inversa é dada por:

$$\hat{Z} = \frac{6\left[\sqrt{\frac{7}{3}\left(\hat{\chi}_\nu^2 + \frac{7}{12}\right) - \frac{\nu}{3}} + \sqrt{2\nu}\right]}{7} \quad (17)$$

Da afirmativa probabilística seguinte pode-se derivar facilmente o intervalo de confiança almejado.

$$P(-Z_{\alpha/2} \leq \hat{Z} \leq Z_{\alpha/2}) \approx 1-\alpha \quad (18)$$

Assim,

$$IC_{1-\alpha}(\sigma^2): \left[\frac{S^2}{1+\frac{7}{12\nu}(Z_{\alpha/2}^2-1)+Z_{\alpha/2}\sqrt{\frac{2}{\nu}}}; \frac{S^2}{1+\frac{7}{12\nu}(Z_{\alpha/2}^2-1)-Z_{\alpha/2}\sqrt{\frac{2}{\nu}}} \right] \quad (19)$$

A acurácia desses intervalos apresentados em (15) e (19) foi checada por meio de simulação Monte Carlo, detalhada na seção 4, tanto para situações ideais de normalidade, quanto em situações em que a população amostrada não é normal, para avaliar a robustez. A grande vantagem dos intervalos propostos, (15) e (19), é a necessidade de obter somente quantis da distribuição normal padrão, que evidentemente não depende dos graus de liberdade e, portanto, não implica na necessidade de obter os quantis da distribuição de qui-quadrado.

2.3 Simulação Monte Carlo

A princípio é interessante que se destaque o valor didático das aproximações apresentadas para o intervalo de confiança de uma variância de uma população normal. Os aspectos que podem ser apontados como vantajosos para serem explorados nesse contexto são: as derivações teóricas da expansão assintótica de Wilson e Hilferty (1931) e de Bishop, Fienberg e Holland (1975); a utilização de intervalos de confiança que dependem apenas do quantil da distribuição normal com o intuito de ilustrar a utilização de uma aproximação assintótica; e finalmente a vantagem de se apresentar aproximações assintóticas dos quantis de qui-quadrado, utilizando apenas quantis da normal padrão. Para avaliar a acurácia dos intervalos (15) e (19) foram realizadas simulações Monte Carlo. Os resultados de a probabilidade de cobertura (Mood, Graybill e Boes, 1974) e do comprimento médio do intervalo foram obtidos para os intervalos (5), (10), (15) e (19). Foram adotados os valores dos coeficientes de confiança nominais $(1-\alpha)$ iguais a 10%, 50%, 95%, 99% e 99,99%, tamanhos amostrais $n = 2$ (1) 50 (de 2 a 50 variando de 1 em 1) e amostras de diferentes populações. As distribuições de probabilidade consideradas foram a normal, $N(10, 100)$, sem perda de generalidade, exponencial com média $1/\lambda = 10$ e variância $1/\lambda^2 = 100$ e distribuição gama com parâmetros $\alpha = 2$ e $\beta = 1/2$ e 15, cujas variâncias são iguais a $\alpha^2\beta$, ou seja, iguais a 2 e 60 respectivamente. As densidades de probabilidade consideradas para a geração de as populações estão apresentadas em (20) (a) normal, (b) exponencial e (c) gama. O teorema da probabilidade integral foi utilizado para a geração de amostras de tamanho n de cada população (Mood, Graybill e Boes, 1974). As simulações foram todas realizadas no Proc IML do programa SAS (1998).

$$\left\{ \begin{array}{ll} f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} & x \in]-\infty; +\infty[\quad (a) \\ f(x) = \lambda e^{-\lambda x} & x \geq 0 \quad (b) \\ f(x) = \frac{1}{\alpha\Gamma(\beta)} \left(\frac{x}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x}{\alpha}\right)} & x \geq 0 \quad (c) \end{array} \right. \quad (20)$$

Foram realizadas 2.000 simulações de cada configuração (α , n e população), sendo que em cada caso, para avaliar a probabilidade de cobertura, foi obtida nas 2.000 simulações a proporção de vezes que o intervalo considerado continha o verdadeiro valor populacional e, também, computou-se a média dos comprimentos dos 2.000 intervalos para cada um dos métodos utilizados.

3 Resultados e discussões

Nas Figuras de 1 a 5 são apresentas as probabilidades de cobertura dos intervalos de confiança estudados, em função do tamanho da amostra e do método empregado para as diferentes populações. As Figuras de 1 a 5 diferem basicamente na alteração do valor

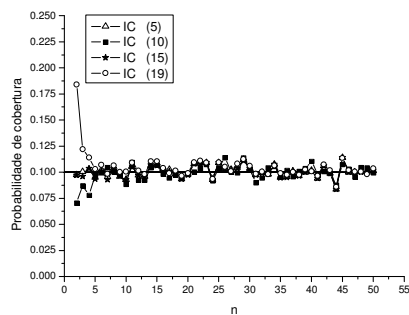
nominal de cobertura utilizado. É importante enfatizar que as escalas das ordenadas em cada figura diferem na amplitude de variação. Como o intervalo de confiança exato foi usado na situação de normalidade, as oscilações encontradas nas Figuras de 1 a 5 (a) em torno do valor nominal adotado para a confiança foram atribuídas ao erro de Monte Carlo nas 2.000 amostras consideradas. Assim, na Figura 1 (a) todos os métodos apresentaram praticamente valores iguais de probabilidade de cobertura e iguais ao valor nominal, considerando $1 - \alpha = 0,10$ e população normal, exceto para graus de liberdade $\nu \leq 3$ com os métodos (10) e (19). Para distribuições assimétricas como a exponencial (Figura 1 (b)) e a gama (Figura 1 (c)) a probabilidade de cobertura decresceu com o aumento de n , e os quatro métodos apresentaram resultados semelhantes. Nesse caso, é possível perceber exceções para essa última afirmativa. O intervalo normal assintótico (10) apresentou menor cobertura que os demais para ν pequeno, menor do que 4 e a aproximação qui-quadrado assintótica (19) valores de probabilidade de cobertura superiores às dos demais para $\nu < 3$. Com a distribuição gama com parâmetros $\alpha = 2$ e $\beta = 15$, Figura 1 (d), que é uma distribuição que se aproxima da normal, verifica-se que os quatro métodos tenderam a se igualar e a apresentarem valores de probabilidade de cobertura semelhantes ao valor do coeficiente de confiança nominal, principalmente para n grande. Com n pequeno o intervalo (19) apresentou probabilidade de cobertura muito superior ao valor nominal adotado.

A concordância entre os métodos (5) e (15) foi muito elevada para o coeficiente de confiança de 10%. O método (19) também apresentou concordância com esses dois intervalos, principalmente para $\nu \geq 5$. Com o aumento do coeficiente de confiança para 50% (Figura 2), resultados semelhantes aos anteriores ocorreram. No entanto, quando a distribuição amostrada foi a normal, Figura 2 (a), com valores pequenos de ν , $\nu < 3$, o intervalo normal assintótico (10) apresentou cobertura superior ao valor nominal e diferente dos métodos (5) e (15), cujas coberturas foram idênticas entre si e iguais ao valor do coeficiente de confiança nominal. O intervalo (19) apresentou cobertura igual a zero para $\nu = 1$ e superior ao valor nominal de confiança para $\nu \leq 7$. A razão para a probabilidade de cobertura ter sido igual a zero com $\nu = 1$ para (19) é devida ao problema de o limite superior desse intervalo ser negativo com o aumento de $1 - \alpha$ e redução de ν . O preço para a maior probabilidade de cobertura do intervalo (10) foi um maior comprimento médio. Nas distribuições assimétricas (Figura 2 (b) e (c)) observou-se uma redução do valor da cobertura em relação ao valor nominal à medida que n aumentou. Para $n = 2$ a 8, os intervalos (10) e (19) apresentaram cobertura, em geral, superior ao valor nominal, exceto para $\nu = 1$ com (19).

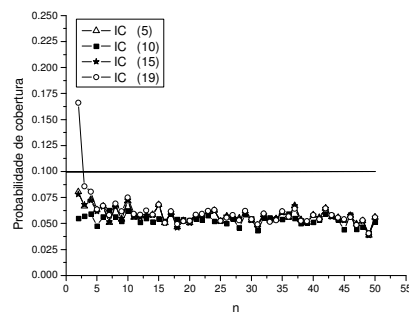
No caso de uma gama tendendo a simétrica (Figura 2 (d)) os métodos apresentaram cobertura próximas ao valor nominal, com um pequeno desvio a fim de subestimar o real valor, para valores de ν superiores a 7. Novamente os intervalos (10) e (19) para ν pequenos, apresentaram probabilidades de cobertura superiores ao valor nominal.

Para grandes valores do coeficiente nominal, quais sejam, 95%, 99% e 99,99% (Figuras 3, 4 e 5) as probabilidades de cobertura no caso normal, Figuras 3, 4 e 5 (a), dos intervalos (5), (10), (15) e (19) tenderam a ser iguais para grandes valores de n e também a serem iguais ao valor nominal. Para valores pequenos de ν as aproximações (10), (15) e (19) apresentam problemas com relação ao limite de confiança superior, resultando em valores negativos, e, portanto, apresentando cobertura nula. Os valores de ν para os quais

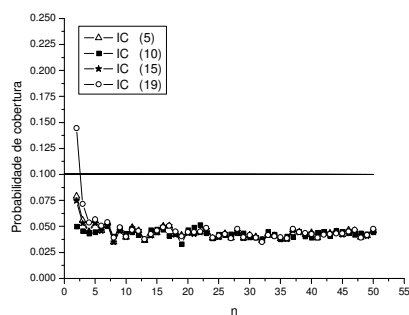
esse fato ocorre são muito menores para o intervalo (15) em relação aos intervalos (10) e (19).



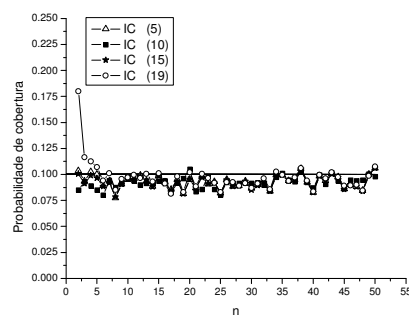
(a) Normal



(b) Exponencial



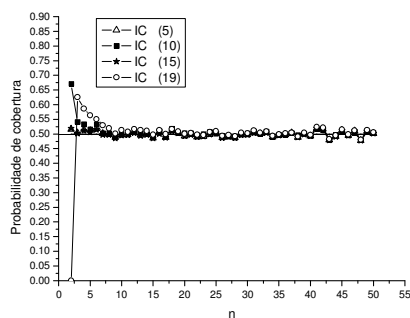
(c) Gama com $\beta = 1/2$



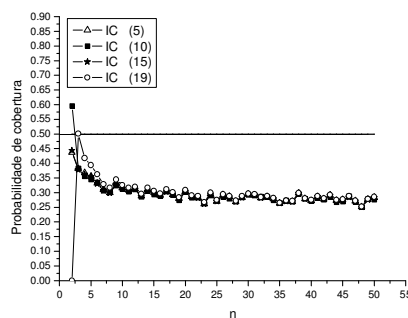
(d) Gama com $\beta = 15$

FIGURA 1 - Probabilidade de cobertura dos intervalos de confiança (5), (10), (15) e (19) em função do tamanho da amostra considerando diferentes populações (a) normal, (b) exponencial com $\lambda = 0,1$, (c) gama com parâmetros $\alpha = 2$ e $\beta = 1/2$ e (d) gama com parâmetros $\alpha = 2$ e $\beta = 15$, considerando valor nominal do coeficiente de confiança igual a 10%.

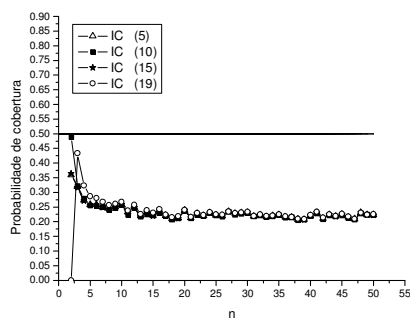
O intervalo (19) apresenta problemas para o limite superior do intervalo, para menores valores de ν em relação àqueles do intervalo (10). Com o aumento de $1-\alpha$ há um aumento dos valores de ν para que as aproximações resultem em intervalos válidos. Assim, para $1-\alpha$ igual a 95%, o valor de ν mínimo para o método proposto (15) é igual a 2, para 99% é igual a 2 e para 99,99% é igual a 4. A partir desses valores a probabilidade de cobertura praticamente se iguala ao do intervalo exato. Para o intervalo (10) os valores de ν necessários para resultar intervalos válidos são iguais a 7 para 95%, 14 para 99% e 31 para 99,99%. E para o intervalo de confiança (19) os valores de ν necessários para resultar intervalos válidos são iguais a 4 para 95%, 4 para 99% e 5 para 99,99%.



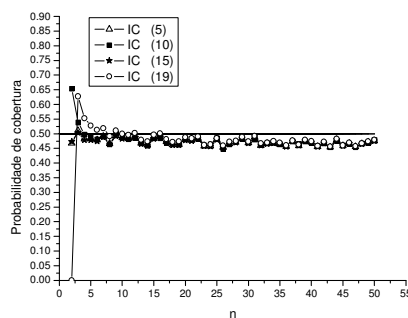
(a) Normal



(b) Exponencial



(c) Gama com $\beta = \frac{1}{2}$



(d) Gama com $\beta = 15$

FIGURA 2 - Probabilidade de cobertura dos intervalos de confiança relativos às equações (5), (10), (15) e (19) em função do tamanho da amostra considerando diferentes populações (a) normal, (b) exponencial com $\lambda = 0,1$, (c) gama com parâmetros $\alpha = 2$ e $\beta = \frac{1}{2}$ e (d) gama com parâmetros $\alpha = 2$ e $\beta = 15$, considerando valor nominal do coeficiente de confiança igual a 50%.

Nas distribuições assimétricas (Figuras 3, 4 e 5 (b) e (c)) para valores de ν maiores que os apontados anteriormente à probabilidade de cobertura do intervalo normal assintótico foi superior aos demais, mas inferior ao valor nominal. O intervalo proposto (15) e o intervalo exato, acima dos limites de ν retro mencionados, apresentaram probabilidades de cobertura extremamente similares.

A aproximação (19) apresentou probabilidade de cobertura superior a do intervalo exato (5) e da aproximação qui-quadrado (15). Para ν grande, em geral superior a 15, as probabilidades de cobertura dos intervalos (5), (15) e (19) se igualam. Para a distribuição aproximadamente simétrica (Figura 3, 4 e 5 (d)), também, com valores de ν acima dos limites mencionados, os métodos tenderam a se igualar e, também, a apresentarem valores iguais ao valor nominal.

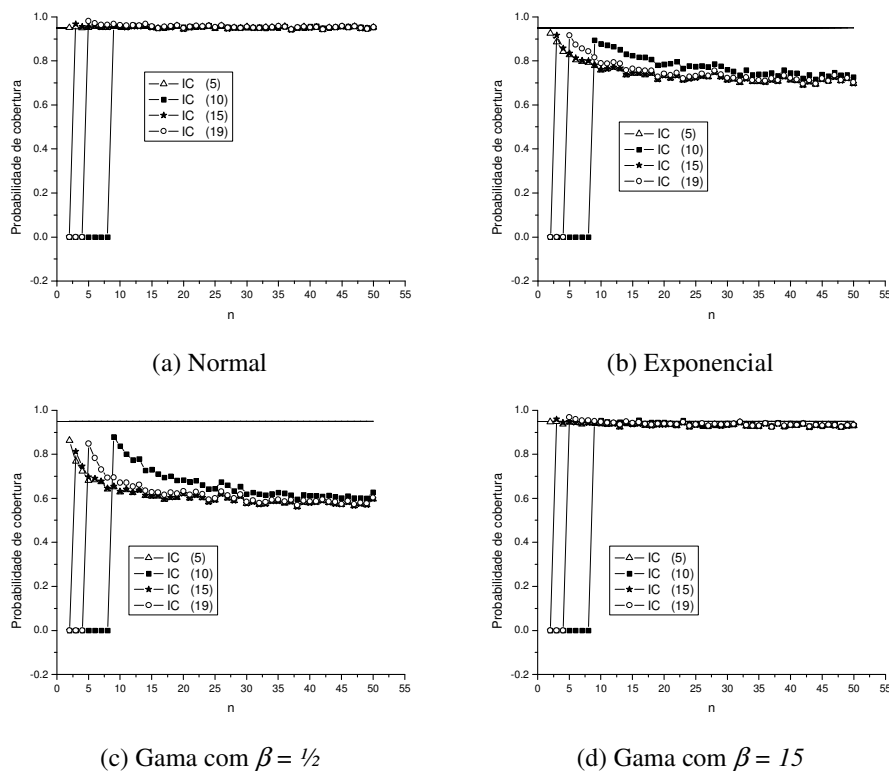


FIGURA 3 - Probabilidade de cobertura dos intervalos de confiança relativos às equações (5), (10), (15) e (19) em função do tamanho da amostra considerando diferentes populações (a) normal, (b) exponencial com $\lambda = 0,1$, (c) gama com parâmetros $\alpha = 2$ e $\beta = 1/2$ e (d) gama com parâmetros $\alpha = 2$ e $\beta = 15$, considerando valor nominal do coeficiente de confiança igual a 95%.

É curioso notar que as perdas na robustez dos intervalos ocorrem com mais intensidade para valores de coeficiente de cobertura nominal menores. Chama a atenção também o fato de o intervalo proposto (15) apresentar resultados praticamente idênticos ao do intervalo exato para a probabilidade de cobertura. A exceção ocorre principalmente quando o coeficiente de confiança aumenta excessivamente e os graus de liberdade são extremamente pequenos, quais sejam, $\nu \leq 4$. Da mesma forma, verifica-se que as probabilidades de cobertura de (19) e de (5) se aproximam somente para ν maiores do que aquelas observadas quando se comparam (5) e (15).

Finalmente, na Figura 6 apresentam-se duas situações contrastantes para a aplicação dos métodos: uma considerando a distribuição gama com maior assimetria à direita (Figura 6 (b) e (d)) e outra, com a distribuição normal (Figura 6 (a) e (c)). Também os gráficos foram subdivididos em duas regiões para um detalhamento maior, uma vez que as escalas tinham uma grande amplitude de variação para n inferior a 15 em relação a

valores maiores de n . Nesse caso os comprimentos médios dos intervalos são apresentados.

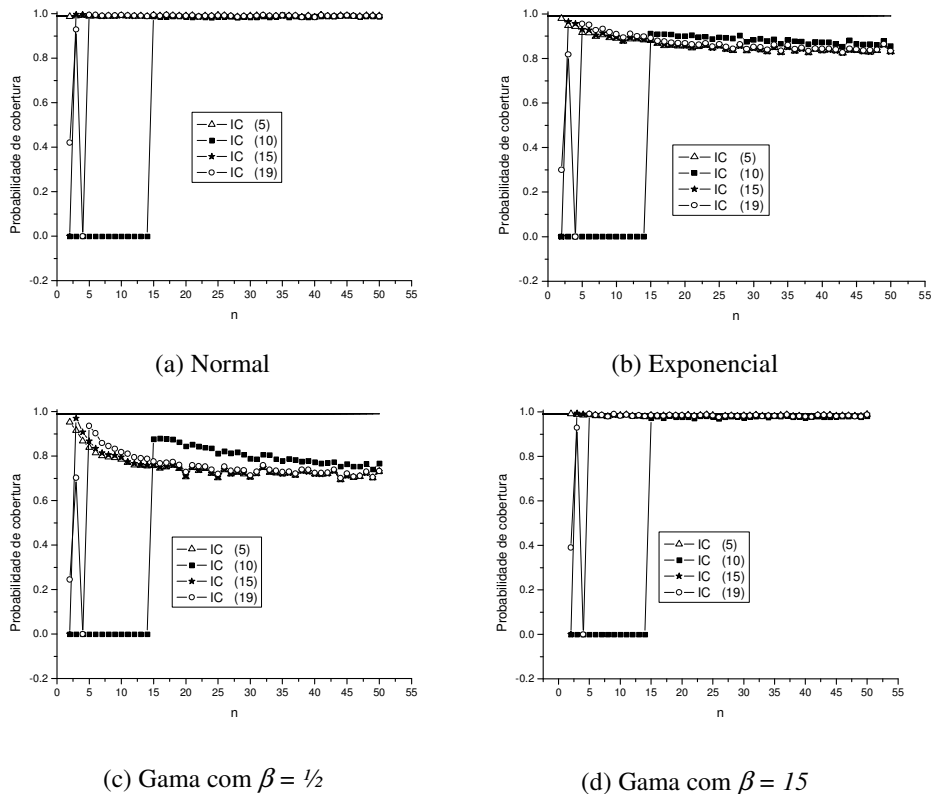
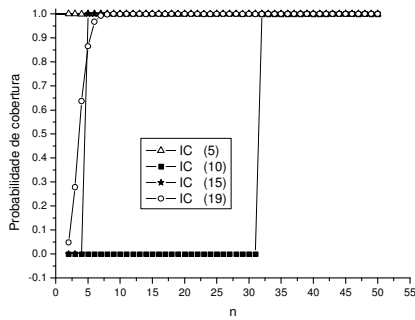


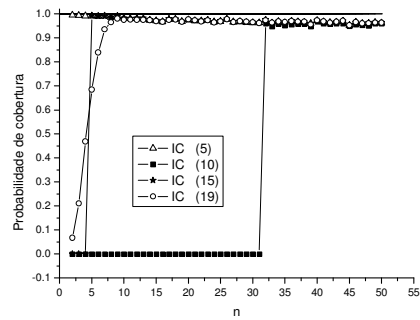
FIGURA 4 - Probabilidade de cobertura dos intervalos de confiança relativos às equações (5), (10), (15) e (19) em função do tamanho da amostra considerando diferentes populações (a) normal, (b) exponencial com $\lambda = 0,1$, (c) gama com parâmetros $\alpha = 2$ e $\beta = 1/2$ e (d) gama com parâmetros $\alpha = 2$ e $\beta = 15$, considerando valor nominal do coeficiente de confiança igual a 99%.

De maneira geral, pode-se observar que os intervalos (5) e (15) apresentaram praticamente os mesmos comprimentos médios, cujos valores são inferiores àqueles obtidos para o intervalo (10). O intervalo (19) apresentou comprimentos médios muito próximos daqueles valores de (5) e (15), porém foram quase sempre superiores.

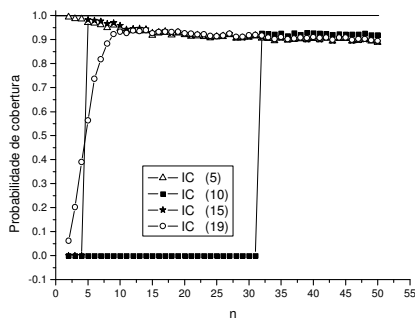
Os quatro intervalos tendem a apresentar valores semelhantes somente para n próximo a 50. Já os intervalos (5), (15) e (19) tenderam a se igualar em comprimento médio para n próximos a 30.



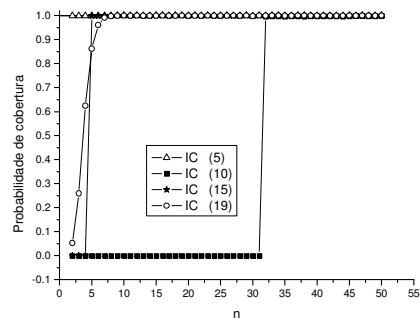
(a) Normal



(b) Exponencial



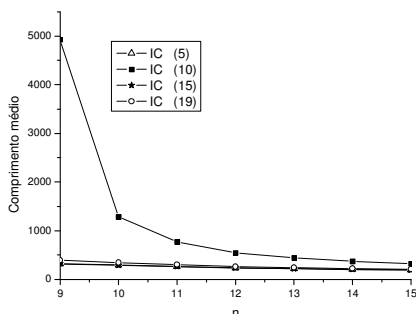
(c) Gama com $\beta = 1/2$



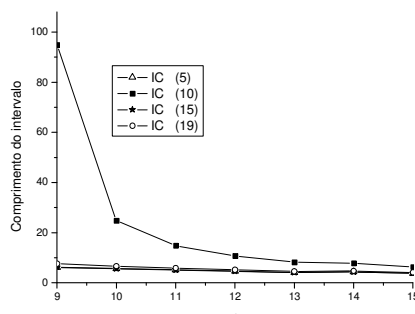
(d) Gama com $\beta = 15$

FIGURA 5 - Probabilidade de cobertura dos intervalos de confiança relativos às equações (5), (10), (15) e (19) em função do tamanho da amostra considerando diferentes populações (a) normal, (b) exponencial com $\lambda = 0,1$, (c) gama com parâmetros $\alpha = 2$ e $\beta = 1/2$ e (d) gama com parâmetros $\alpha = 2$ e $\beta = 15$, considerando valor nominal do coeficiente de confiança igual a 99,99%.

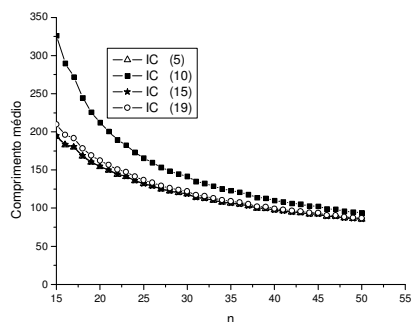
Nas distribuições reconhecidamente não normais, apesar de o intervalo de confiança normal assintótico (10) apresentar valores de comprimento médio superior aos demais, a probabilidade de cobertura desse intervalo é maior, porém ainda inferior ao valor nominal, mostrando-se, portanto, não conservador. Nas situações de não-normalidade esse intervalo pode ser recomendado por se tratar de uma alternativa mais robusta. Para o caso normal, no entanto, a probabilidade de cobertura desse intervalo é maior, em geral, do que a dos seus concorrentes e maior do que o valor nominal e, ainda, apresenta comprimento médio superior ao de seus competidores. Assim, não deve ser recomendado para situações normais, exceto para n igual ou superior a 50.



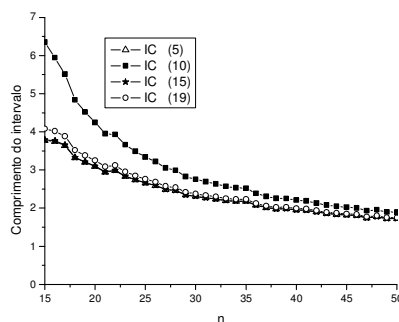
(a) Normal para $9 < n < 15$



(b) Gama com $\beta = 1/2$ para $9 < n < 15$



(c) Normal para $n \geq 15$



(d) Gama com $\beta = 1/2$ para $n \geq 15$

FIGURA 6 - Comprimento médio dos intervalos de confiança relativos às equações (5), (10), (15) e (19) em função do tamanho da amostra considerando diferentes populações (a) e (c) normal e (b) e (d) gama com parâmetros $\alpha = 2$ e $\beta = 1/2$, considerando valor nominal do coeficiente de confiança igual a 95%.

Conclusões

O intervalo de confiança para a variância de uma população normal resultante da transformação de Wilson e Hilferty (1931) foi derivado e apresentado na equação (15). A segunda aproximação assintótica de qui-quadrado de Bishop, Fienberg e Holland (1975) permitiu derivar o intervalo de confiança (19). A avaliação das suas performances por meio de simulação Monte Carlo permitiu que se obtivessem as seguintes conclusões:

- o intervalo de confiança (15) mostrou-se com acurácia semelhante ao do intervalo exato para coeficiente de confiança variando de 10% a 99,99%, devendo ser recomendado para $\nu > 4$;

- o intervalo (19) pode ser recomendado somente para $\nu > 30$, sendo, no entanto, o método (15) considerado melhor;
- o intervalo normal assintótico (10) apresentou robustez para populações não normais e pode ser recomendado para estimar variâncias dessas populações na ausência de um método exato ou mais eficiente para a situação em questão;
- os intervalos de confiança propostos não necessitam da estimação de quantis da distribuição de qui-quadrado e apresentam resultados similares ao do intervalo exato, uma vez que, para grandes valores de ν (que são situações em que os métodos numéricos são instáveis para obter quantis da distribuição qui-quadrado) estes métodos possuem maior acurácia;
- o valor didático dessas aproximações deve ser considerado para suas utilizações e recomendações.

FERREIRA, D. F.; NOGUEIRA, D. A. An asymptotic normal approach for the variance exact confidence interval of a normal population. *Rev. Mat. Estat.*, São Paulo, v.23, n.1, p.31-46, 2005.

- *ABSTRACT: The variance of a population is an important parameter to be estimated. There are many areas that require accurate estimates of the variance. The interval estimate has the purpose to express the precision of estimates of such parameter. This work aimed at presenting two confidence intervals for the normal variance that result from the transformation of Wilson and Hilferty (1931) and from the asymptotic chi-square approach of Bishop, Fienberg and Holland (1975). It also aims at evaluating their performances by means of Monte Carlo simulation using 2.000 iterations. The confidence interval based on Wilson and Hilferty (1931) approach showed basically the same accuracy as that of the exact confidence interval for confidence coefficient varying from 10% to 99.99% and should be recommended for $\nu > 4$, where $\nu = n - 1$ is the degrees of freedom of the sample. The interval based on the approach of Bishop, Fienberg and Holland (1975) can only be recommended for $\nu > 30$; however, the previous interval is considered to be better. The normal asymptotic interval presented robustness for non normal populations and it can be recommended to estimate variances for those populations in the absence of one or more efficient method for the situation in question. The proposed confidence intervals do not need the estimate of chi-square quantiles and present similar results to those of the exact interval. For large values of ν , these methods possess larger accuracy and although, in this situation, numerical methods are unstable for obtaining chi-square quantiles; the didactic value of those approaches should be considered for their uses and recommendations.*
- *KEYWORDS: Asymptotic approach; chi-square; Monte Carlo simulation.*

Referências

BENTON, D.; KRISHNAMOORTHY, K. Computing discrete mixtures of continuous distributions: noncentral chisquare, noncentral t and the distribution of the square of the sample multiple correlation coefficient. *Comp. Stat. Data An.*, Amsterdam, v.43, p.249-267, 2003.

BISHOP, Y. M. M.; FIENBERG, S. E.; HOLLAND, P. W. *Discrete multivariate analysis: theory and practice*. Massachusetts: The Massachusetts Institute of Technology, 1975. 557p.

MAATTA, J. M.; CASELLA, G. Conditional properties of interval estimators of the normal variance. *Ann. Stat.*, Hayward, v.4, 1372-1388,1987.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. *Introduction to the theory of statistics*. Singapore: International Editions, 1974, 564p.

PRESS, W. H. et al. *Numerical recipes in Fortran: the art of scientific computing*. Cambridge: Cambridge University Press. 1992. 994p.

SAMIUDDIN, M. Bayesian test of homogeneity of variance. *J. Am. Stat. Assoc.*, New York, v.71, n.354, p.515-517, 1976.

SAS INSTITUTE. *SAS: functions users guide*, vs. 6., 12. ed. Cary, 1998.

WILSON, E. B.; HILFERTY, M. M. The distribution of chi-square. *Proc. Acad. Sci.*, Washigton, v.17, p.684-688, 1931.

Recebido em 02.07.2003.

Aprovado após revisão em 26.04.2005.