Estudo do poder de teste de hipóteses para a variância populacional

Siomara Cristina Broch [1,2,3]
Daniel Furtado Ferreira [1,3,4]

1 Introdução

Um dos principais objetivos da Estatística é a obtenção de conclusões válidas para toda a população com base em observações de amostras retiradas desta população. Esses processos de tomada de decisões são denominados de testes de hipóteses ou testes de significância e fazem parte da inferência estatística.

Toda inferência estatística está sujeita a erros, principalmente a dois tipos. O erro tipo I, denominado de nível de significância (α) do teste, é quantificado pela probabilidade da decisão de rejeitar a hipótese nula quando na realidade ela é verdadeira. É também conhecido como uma falsa decisão positiva, em que é declarado um efeito quando ele não existe [1]. Este erro está relacionado com o nível de confiança da inferência $1-\alpha$, que é a probabilidade de aceitar a hipótese nula e ela realmente ser verdadeira, em que é tomada uma decisão correta. O erro tipo II, simbolizado por β, refere-se à probabilidade da decisão de aceitar a hipótese nula quando na realidade ela é falsa. Este erro é conceituado também como uma falsa decisão negativa, em que não se declara um efeito que realmente existe [1]. Ele está relacionado com o poder do teste, dado por $1-\beta$, que é a capacidade da inferência identificar as diferenças reais entre os valores.

O pesquisador deve procurar um equilíbrio entre as probabilidades desses erros para que o teste de significância tenha um poder e confiança aceitáveis no estudo. Diversos autores ([6]; [4]; [5]; dentre outros) afirmam que um modo de reduzir a probabilidade de ambos os tipos de erros é aumentar o tamanho da amostra *n*. Este estudo visa discutir este fato considerando o poder do teste de hipóteses do parâmetro variância populacional em função do tamanho da amostra e do nível de significância.

2 Material e métodos

Para análise e discussão de como se dá a redução dos erros tipo I e tipo II e seu controle, é tomado como um parâmetro de teste a variância populacional $\theta = \sigma^2$ e amostras aleatórias de

[1]DEX - UFLA.
[2]Professora do IFFarroupilha/RS e-mail: *siomarabroch@jc.iffarroupilha.edu.br*
[3]Agradecimento ao CNPq pelo apoio financeiro.
[4]Bolsista do CNPq.

tamanho n extraídas de populações normais ($N(\mu,\sigma^2)$). Para cada amostra calcula-se a variância $S^2 = (1/(n-1))\sum_{i=1}^{n}(X_i-\bar{X})^2$. A estatística amostral $\chi^2_\nu = (n-1)S^2/\sigma^2$ tem distribuição exata qui-quadrado com $\nu = n-1$ graus de liberdade. Como a distribuição do qui-quadrado não é tão robusta e pode levar a resultados errôneos se a população não tem uma distribuição normal, é necessário inicialmente verificar a normalidade dos dados amostrais.

A distribuição qui-quadrado não é simétrica, sendo diferente para cada número de graus de liberdade ν e vai se tornando menos assimétrica tendendo para uma distribuição normal a medida que ν aumenta. Em testes de hipóteses, como a distribuição amostral da variância é exata para $\sigma^2 = \sigma_0^2$, no cálculo do poder que é determinado sendo H_1 verdadeira, o valor de σ^2 é desconhecido. Logo, tem-se apenas um valor aproximado da distribuição qui-quadrado escalado, dada por $(n-1)S^2/\sigma_0^2 \sim \sigma^2\chi^2_{\alpha,n-1}$, dependente de σ^2. O poder para o teste unilateral à esquerda e à direita é estimado, respectivamente, sendo S^2 uma estimativa de σ^2, pelas probabilidades

$$Poder_E = (1-\beta)(\sigma^2) = P\left(\chi^2 < \frac{\chi^2_{1-\alpha,n-1}\sigma_0^2}{S^2}\right), \tag{1}$$

$$Poder_D = (1-\beta)(\sigma^2) = 1 - P\left(\chi^2 < \frac{\chi^2_{\alpha,n-1}\sigma_0^2}{S^2}\right). \tag{2}$$

Segundo Ferreira [3] e Costa Neto [2], para obter boas aproximações dos valores dos quantis da distribuição qui-quadrado, dado uma probabilidade α, simbolizada por $\chi^2_{\alpha,\nu}$ como o quantil superior α da distribuição com $\nu = n-1$ graus de liberdade, pode-se usar o estimador

$$\widehat{\chi}^2_{\alpha,\nu} = \nu\left(1 - \frac{2}{9\nu} + Z_\alpha\sqrt{\frac{2}{9\nu}}\right)^3, \tag{3}$$

que supõe a variável $(S^2/\sigma^2)^{1/3} \sim N(1-(2/9\nu), 2/9\nu)$, em que Z_α é o valor da variável normal padrão que corresponde em probabilidade ao quantil $100\alpha\%$ e equivale ao $\chi^2_{\alpha,\nu}$ desejado.

3 Resultados e discussões

Na Figura 1 tem-se o comportamento da variância amostral determinada a partir de diferentes tamanhos de amostras, o que comprova que à medida que n cresce a variância amostral tende a variar menos e se igualar à variância populacional. Na Figura 2 observa-se que o poder para os testes bilaterais tem um ponto de mínimo quando a variância amostral S^2, estimativa da variância populacional σ^2, é igual ao valor em teste $\sigma_0^2 = 36$ e, neste caso, a probabilidade de erro tipo I é igual a α. Acima ou abaixo desse valor H_0 é falsa, ocasionando diferenças em relação ao valor em teste e, portanto, o poder do teste aumenta.

As Figuras 3 e 4 representam o poder para testes unilaterais com $\alpha = 5\%$ e $\alpha = 1\%$, para as hipóteses $H_0 : \sigma^2 \geq 36$ *vs* $H_1 : \sigma^2 < 36$ e $H_0 : \sigma^2 \leq 36$ *vs* $H_1 : \sigma^2 > 36$, respectivamente. Para variâncias amostrais que se afastam do valor em teste no sentido de tornar verdadeira H_1, o poder do teste aumenta atingindo valores próximos a 1. Este aumento ocorre mais rapidamente quanto maior for o nível de significância e maior o tamanho da amostra. Para um mesmo

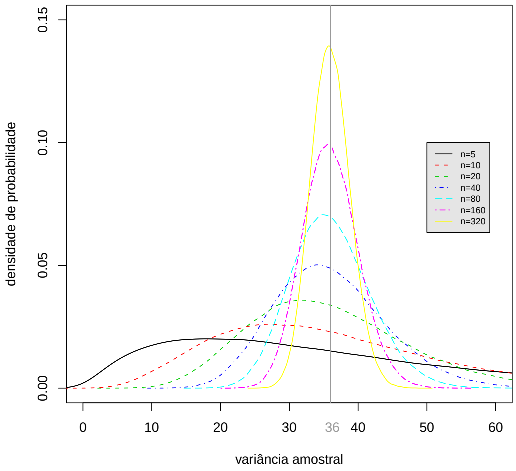

Figura 1: Distribuição amostral das variâncias obtidas de amostras de uma população normal com $\mu = 25$ e $\sigma = 6$ para diferentes tamanhos de amostras (n).

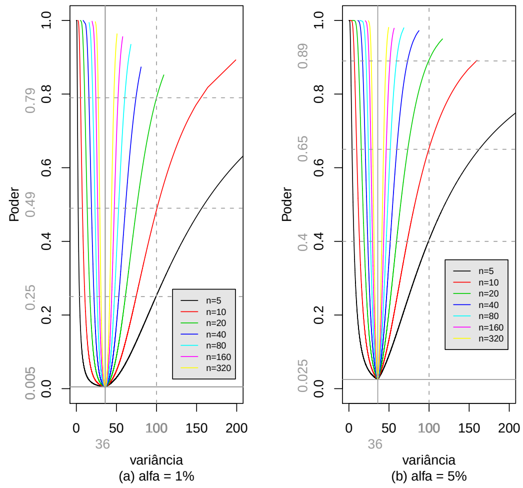

Figura 2: Curvas de poder de teste bilateral para diferentes tamanhos de amostras (n) e níveis de significância $\alpha = 1\%$ e $\alpha = 5\%$.

tamanho amostral o poder do teste é maior para $\alpha = 5\%$. Observa-se que em testes unilaterais a probabilidade de erro tipo I é de, no máximo, α. No caso a) quando a variância amostral é maior do que 36, indica que o valor real do parâmetro é distinto do valor testado, porém coerente com H_0 e, portanto, nesta situação ao não se rejeitar H_0 estará sendo cometido um erro tipo I menor do que α. Neste caso o poder do teste diminui, a probabilidade de erro tipo II aumenta enquanto que a probabilidade de erro tipo I diminui. Só quando a variância amostral é igual ao valor em teste o poder do teste é de α para qualquer tamanho de amostra.

A Tabela 1 foi construída retirando-se das Figuras 2, 3 e 4 alguns valores aproximados do poder dos testes. O teste unilateral em que a H_1 tem um sentido determinado é mais poderoso do que o teste bilateral, para o mesmo tamanho de amostra e mesma probabilidade de erro tipo I [6]. Além disso, para ambos os testes, para o mesmo α e valor fixo de S^2, quanto maior o tamanho da amostra maior o poder. Para um mesmo tamanho de amostra, para cada valor de S^2, quanto maior o nível de significância do teste maior o poder.

Tabela 1: Poder aproximado dos testes para uma variância amostral $S^2 = 100$ em função de do tamanho da amostra e nível de significância adotado.

Tamanho da amostra	Nível de significância			
	Teste unilateral à direita		Teste bilateral	
	1%	5%	1%	5%
5	0,31	0,49	0,25	0,49
10	0,55	0,73	0,49	0,65
20	0,84	0,93	0,79	0,89

Assim, as formas de controlar a probabilidade do erro tipo II são através do controle da

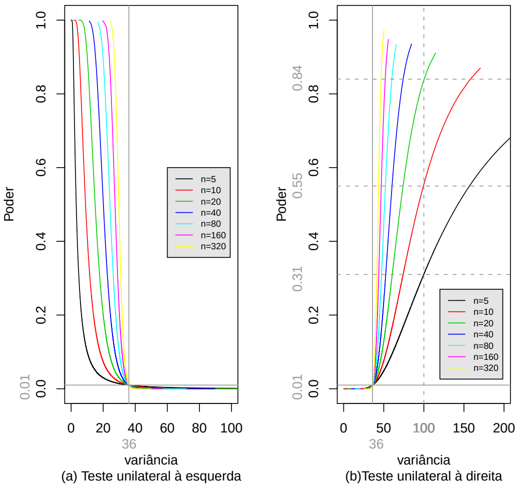

Figura 3: Curvas de poder de testes unilaterais para diferentes tamanhos de amostras (n) e nível de significância $\alpha = 1\%$.

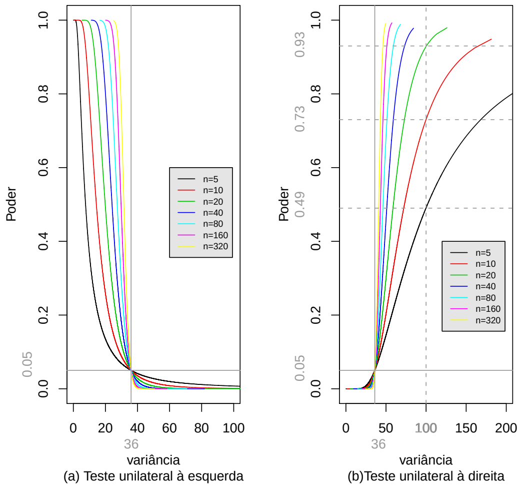

Figura 4: Curvas de poder de testes unilaterais para diferentes tamanhos de amostras (n) e nível de significância $\alpha = 5\%$.

probabilidade do erro tipo I ou pelo aumento do tamanho da amostra. Como o pesquisador normalmente conhece e fixa num valor aceitável a probabilidade α a alternativa para diminuir a probabilidade do erro tipo II é o correto dimensionamento do tamanho da amostra.

3.1 Dimensionamento da amostra

Como o erro tipo II depende da hipótese alternativa e ela é sempre um intervalo de valores, existem diversos valores possíveis para σ^2 e para cada um o seu β_{σ^2} correspondente. Costa Neto [2] sugere estabelecer tecnicamente até que ponto uma divergência entre a realidade e H_0 pode ser tolerada e assim determinar o tamanho de amostra mínimo n sob o controle de α e β. Supondo as hipóteses $H_0 : \sigma^2 \leq \sigma_0^2$ *vs* $H_1 : \sigma^2 > \sigma_0^2$, sendo $\chi_c^2 = \nu S^2/\sigma_0^2$ a estatística de teste, a rejeição de H_0 ocorre se $P\left(\chi_c^2 > \chi_{\alpha,\nu}^2\right) = \alpha$. O poder do teste é dado pela expressão (2) e usando a aproximação (3), após algumas simplificações, resulta em

$$1-\beta = P\left[Z > \frac{\sqrt[3]{\delta}\left(1-\frac{2}{9\nu}+Z_\alpha\sqrt{\frac{2}{9\nu}}\right)-1+\frac{2}{9\nu}}{\sqrt{\frac{2}{9\nu}}}\right],$$

em que $\delta = \sigma_0^2/\sigma^2$ reflete a diferença entre o parâmetro e o valor hipotético, que deve ser fixado pelo pesquisador. A parte depois da desigualdade é o quantil superior $Z_{1-\beta}$ da normal padrão e, o tamanho da amostra é a solução para ν desta equação que, rearranjada resulta em $\left(\frac{Z_{1-\beta}-Z_\alpha\sqrt[3]{\delta}}{\sqrt[3]{\delta}-1}\right)^2 = \frac{(\nu-\frac{2}{9})^2}{18\nu}$. Resolvendo e isolando $n = \nu+1$ chega-se a

$$n = \frac{138}{135}+\frac{A}{9}+\frac{\sqrt{A^2+4A}}{9}, \quad com \quad A = \left(\frac{Z_{1-\beta}-Z_\alpha\sqrt[3]{\delta}}{\sqrt[3]{\delta}-1}\right)^2. \tag{4}$$

Para o teste de hipóteses unilateral à esquerda $H_0 : \sigma^2 \geq \sigma_0^2$ *vs* $H_1 : \sigma^2 < \sigma_0^2$, de modo semelhante obtém-se a expressão

$$n = \frac{3}{135} + \frac{B}{9} + \frac{\sqrt{B^2 + 4B}}{9}, \quad com \quad B = \left(\frac{Z_\beta - Z_{1-\alpha}\sqrt[3]{\delta}}{\sqrt[3]{\delta} - 1} \right)^2. \tag{5}$$

Para o teste de hipóteses bilateral $H_0 : \sigma^2 = \sigma_0^2$ *vs* $H_1 : \sigma^2 \neq \sigma_0^2$, chega-se a

$$1 - \beta = P\left[\frac{\sqrt[3]{\delta}\left(1 - \frac{2}{9\nu} + Z_{\frac{\alpha}{2}}\sqrt{\frac{2}{9\nu}}\right) - 1 + \frac{2}{9\nu}}{\sqrt{\frac{2}{9\nu}}} < Z < \frac{\sqrt[3]{\delta}\left(1 - \frac{2}{9\nu} + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{2}{9\nu}}\right) - 1 + \frac{2}{9\nu}}{\sqrt{\frac{2}{9\nu}}} \right]. \tag{6}$$

A parte à direita da desigualdade é o quantil inferior $Z_{1-\frac{\beta}{2}}$ da normal padrão enquanto que a parte à esquerda é o quantil superior $Z_{\frac{\beta}{2}}$. Como a distribuição do qui-quadrado tem quantis assimétricos, não se pode mensurar quanto do poder do teste está na desigualdade da direita e quanto está na da esquerda chegando a uma expressão explícita para n. Dessa forma a equação só poderá, pelo menos a princípio, ser resolvido por métodos numéricos.

4 Conclusões

Neste trabalho discutiu-se como controlar as probabilidades dos erros tipo I e tipo II para que um teste de hipóteses da variância populacional tenha um poder e confiança pré-definidos, estabelecendo-se expressões para dimensionar o tamanho da amostra com base numa diferença entre o parâmetro e o valor hipotético fixada pelo pesquisador.

Referências

[1] BRETZ, F.; HOTHORN, T.; WESTFALL, P. **Multiple comparisons using R**. USA: Chapman & Hall/CRC, 2010.

[2] COSTA NETO, P.L.O. **Estatística**. 2ª ed. rev. ampl., São Paulo: Edgard Blucher, 2002.

[3] FERREIRA, D.F. **Estatística Básica**. 2. ed. rev., Lavras: Ed. Ufla, 2009.

[4] FONSECA, J.S.; MARTINS, G.A. **Curso de estatística**. São Paulo: Atlas, 1996.

[5] MAGALHÃES, M.N.; LIMA, A.C.P. **Noções de Probabilidade e Estatística**. 6. ed. rev., São Paulo: Editora da USP, 2005.

[6] SIEGEL, S. **Estatística Não-paramétrica para as Ciências do Comportamento**. Tradução de Alfredo Alves Farias. São Paulo: McGraw-Hill, 1975.