AVALIAÇÃO DO ERRO TIPO 1 NA APLICAÇÃO DE UM TESTE DE HIPÓTESES PARA A MÉDIA

Cleber Giugioli Carrasco
Luciano Amaral da Silva

Resumo: Neste artigo avaliamos a probabilidade de se cometer o erro tipo I, ao utilizarmos um teste de hipóteses para a média, através de um estudo de simulação conhecido como método de Monte Carlo. Essa avaliação é feita verificando a influência do tamanho amostral e do nível de significância á, na avaliação do erro tipo I. Todo esse procedimento é implementado computacionalmente no *software free R*. Exemplos numéricos ilustram a metodologia adotada.

Palavras-chave: Erro tipo I, método de Monte Carlo, nível de significância, teste de hipóteses.

1 Introdução

Uma hipótese é uma pressuposição a respeito de certa característica de uma determinada população, ou ainda, uma pressuposição sobre os parâmetros dessa população ou de um modelo. Uma vez formulada a hipótese, desejamos testá-la a fim de verificar se a rejeitamos ou não. O procedimento que irá verificar se tal hipótese deve ser rejeitada ou não é denominado teste de hipóteses (T. H.). Essa hipótese será testada através de resultados amostrais. Dessa forma, o objetivo de um teste de hipóteses é concluir, através de uma estatística obtida de uma amostra ($\hat{\theta}$), se rejeitamos ou não a hipótese nula H_0. Isto pode ser feito construindo uma região crítica (*RC*), também conhecida como região de rejeição (BUSSAB; MORETTIN, 1987). Se a estatística obtida da amostra pertencer a RC, temos evidência para rejeitarmos H_0. Caso contrário, não há evidências suficientes para rejeitarmos essa hipótese. Esta região crítica pode ser construída de maneira que P (Î *RC* | H_0 é verdadeira) = α . Isto é, fixado $\alpha = 5\%$ por exemplo, esperamos rejeitar H_0 quando H_0 for verdadeira, isto é, cometermos o erro tipo 1 em 5% das vezes que aplicarmos o teste de hipóteses.

A hipótese a ser testada em um teste de hipóteses é chamada hipótese nula (H_0). Geralmente, a hipótese nula é uma afirmação do tipo "não há diferença", ou seja, do tipo igualdade (MOORE; MCCABE, 1999). A hipótese complementar à hipótese nula é denominada hipótese alternativa (H_A). O teste de hipóteses é planejado para avaliar a intensidade da evidência contra a hipótese nula. As hipóteses se referem a alguma população ou algum modelo, e não a um resultado particular. Dessa forma, devemos formular H_0 e H_A em termos de parâmetros populacionais. Em particular, nem sempre é claro se H_A deve ser unilateral ou bilateral.

Cleber Giugioli Carrasco é Mestre em Estatística (UFSCar) e professor assistente da UEG. E-mail: cleber.carrasco@ueg.br

Luciano Amaral da Silva é graduando em Matemática (UEG Bolsista PBIC/UEG).

Em geral, as hipóteses de um teste são dadas por:

$H_0 : \theta \ = \theta_0$ **(1)**

$H_A : \theta\ ^{1}\ \theta_0$

As hipóteses alternativas também podem ser do tipo unilateral dadas por:

$H_A : \theta < \theta_0$ ou ?$H_A : \theta > \theta_0$ **(2)**

É necessário, entretanto, tomar cuidado na tomada de decisão sobre a rejeição ou não da hipótese nula, uma vez que estamos sujeitos a cometer erros. Devido a esse problema, as regras de decisão são construídas seguindo critérios que nos permitam reduzir erros na tomada de uma decisão.

Em geral, podemos cometer dois tipos de erros ao aplicar um teste de hipóteses:

Erro tipo I: Ocorre quando rejeitamos H_0, e H_0 é de fato verdadeira. **(3)**

Erro tipo II: Ocorre quando não rejeitamos H_0, e H_0 é de fato falsa.

As probabilidades de cometermos os erros tipo I e II são dadas respectivamente por:

α = P(erro tipo I) = P(rejeitar H_0 | H_0 é verdadeira) **(4)**

β = P(erro tipo II) = P(não rejeitar H_0 | H_0 é falsa)

A probabilidade α de cometermos o erro tipo I é denominada nível de significância. Ou seja, α é a probabilidade do teste rejeitar H_0, quando de fato H_0 é verdadeira. Em geral α é fixado em 10%, 5% ou 1%.

Os possíveis resultados da aplicação de um teste de hipóteses e suas respectivas probabilidades de ocorrência condicionadas à realidade são apresentados na Tabela 1 retirada de Costa Neto (1977).

Tabela 1: Resultados de um T.H. e suas probabilidades condicionadas à realidade

		Realidade	
		H_0 verdadeira	H_0 falsa
Decisão	Aceitar H_0	decisão correta $(1 - \alpha)$	Erro tipo II (β)
	Rejeitar H_0	Erro tipo I (α)	decisão correta $(1 - \beta)$

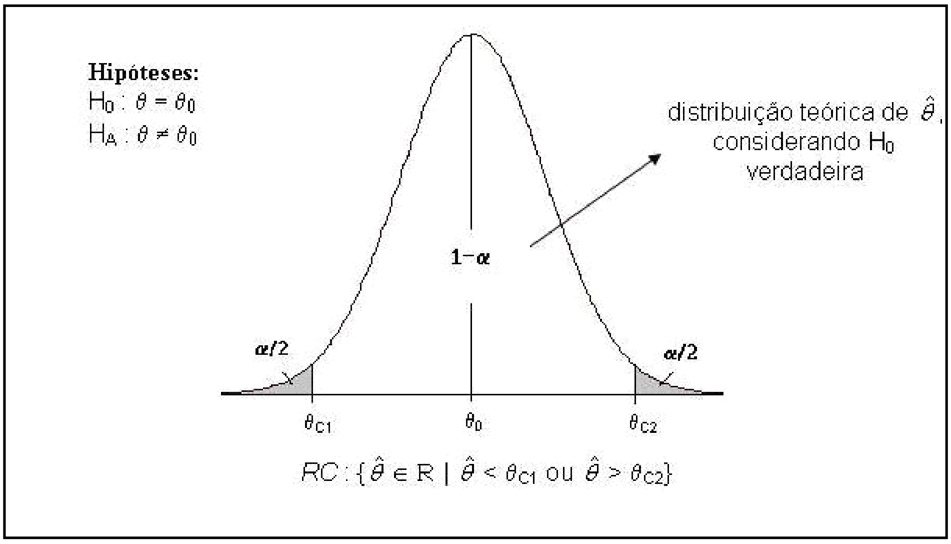
Em geral, ao realizar um teste de hipóteses controlamos o erro do tipo I, fixando a sua probabilidade de ocorrência á. Entretanto, a determinação de α, ou seja, do erro tipo II não é fixada, e o seu cálculo é feito atribuindo alguns valores, escolhidos dentro do caso alternativo.

Segundo Bussab e Morettin (1987), o objetivo do teste de hipóteses é dizer, através de uma estatística obtida de uma amostra, se a hipótese H_0 é ou não refutável, isto é, conseguido através de uma região crítica (*RC*) ou de rejeição. Esta região é construída de modo que P (Î *RC* | H0 é verdadeira) = α. Se o valor da estatística observada pertencer a RC, rejeitamos H_0. Caso contrário, não a rejeitamos.

O esquema abaixo, juntamente com a Figura 1, apresenta uma formulação de um teste de hipóteses bilateral:

1. Determinar as hipóteses a serem testadas H_0 (hipótese nula) e H_A (alternativa);
2. Escolher qual estatística (estimador) será usada para julgar H_0;
3. Fixar a probabilidade α de cometer o erro tipo I, ou seja, a P(rejeitar H_0 | H_0 é verdadeira) e usar este valor para construir a *RC* (região crítica);
4. Usar as informações fornecidas pela amostra para encontrar o valor da estatística () que definirá a decisão;
5. Se o valor da estatística observada na amostra pertencer à *RC*, rejeitar H_0, caso contrário não rejeitar H_0.

Figura 1 : Distribuição teórica de $\hat{\theta}$



A Conclusão do teste dependerá do valor da estatística , se Î *RC* rejeita-se H_0, caso contrário, não teremos evidência suficiente para rejeitarmos H_0.

Neste trabalho, propomos avaliar a probabilidade de se cometer o erro do tipo 1, através do efeito do tamanho amostral e da probabilidade nominal de α, ou nível de significância do teste, através de um estudo de simulação conhecido como método de Monte Carlo, que consiste em repetir o mesmo procedimento diversas vezes. Para verificar o efeito do tamanho amostral no cálculo do erro tipo 1, utilizaremos amostras de tamanho variado. De forma análoga, utilizaremos alguns valores para α. Toda implementação computacional será feita utilizando o software *free R* (VENABLES; SMITH, 2006).

Material e Métodos

Consideremos inicialmente o teste de hipóteses para a média com variância conhecida e fixamos a probabilidade do erro tipo 1, ou seja, o nível de significância á. A partir disso, geramos uma amostra de tamanho n de um modelo normal com média ì = $ì_0$ e aplicamos o teste bilateral e/ou unilateral. Dessa forma, definimos α = 0 se o teste não rejeitar H_0 e $\alpha = 1$ se o teste rejeitar H_0.

Utilizando o método de Monte Carlo, repetimos o processo descrito acima *r* vezes, e verificamos em cada repetição quantas vezes o teste rejeitou a hipótese nula de ì = $ì_0$. Assim temos que,

$$\begin{cases} \delta_i = 0 \text{ se o teste não rejeitar } H_0 \\ \delta_i = 1 \text{ se o teste rejeitar } H_0 \end{cases} \qquad i = 1, 2, \ldots, r \qquad (5)$$

Dessa forma, podemos estimar o valor de α (a probabilidade do erro tipo 1), ou seja, P(rejeitar H_0 | H_0 é verdadeira) da seguinte maneira,

$$\alpha^* = \frac{\sum_{i=1}^{r} \delta_i}{r}, \qquad i = 1, 2, \ldots, r. \qquad (6)$$

Esperamos a priori que o valor estimado de á esteja próximo do valor nominal de á. Se a probabilidade do erro tipo 1 estimada (α^*) for maior que α, então o procedimento é dito ser conservativo, caso contrário, se a probabilidade α^* for menor que a probabilidade nominal de α, então o procedimento é dito anti-conservativo, ou simplesmente não conservativo.

Para avaliar o efeito do tamanho amostral e do nível de significância na probabilidade de se cometer o erro tipo 1 num teste de hipóteses para a média com variância conhecida, através do método de Monte Carlo, tomaremos amostras de tamanhos distintas e fixaremos o nível de significância do teste em 10%, 5% e 1%. Dessa forma, teremos várias combinações de amostras, e para cada caso será verificado o valor estimado de α, ou seja, α^*.

A metodologia adotada foi implementada no *software R* (VENABLES; SMITH, 2006) em que foram geradas amostras de uma distribuição normal com parâmetros ì = 5 e variância $ó^2 = 1$, e os tamanhos amostrais fixamos em *n* = 5, 10, 30, 50, 20 e 1000, os níveis de significância dos testes em α = 0,10, 0,05 e 0,01. Ao todo foram 18 combinações de amostras para cada tipo de hipótese alternativa.

Para o teste bilateral, as hipóteses dos testes foram definidas por testar:

$$H_0 : \mu = 5 \qquad (7)$$
$$H_A : \mu \neq 5$$

No caso dos testes unilaterais à esquerda e à direita, as hipóteses alternativas foram definidas respectivamente por:

$$H_A : \mu < 5 \quad \text{e} \quad H_A : \mu > 5 \qquad (8)$$

Para o cálculo estimado do erro tipo 1 no estudo de Monte Carlo utilizou-se *r* = 1.000, ou seja, repetiu-se o processo 1000 vezes e verificou-se em cada uma se o teste rejeitou a hipótese nula ou não.

Resultados e Discussão

As Tabelas 2, 3 e 4 apresentam os resultados de simulação para os testes bilateral, unilateral à esquerda e unilateral à direita respectivamente, para todos os casos. Nessas tabelas, observamos que as probabilidades de se cometer o erro tipo 1 estão próximas do nível de significância α fixado, tanto para o teste bilateral como para os testes unilaterais. Com relação ao tamanho amostral, não observamos um padrão quando diminuimos ou aumentamos a amostra, ou seja, não observamos efeito na estimativa do erro tipo 1 com relação ao tamanho da amostra.

Tabela 2: Resultado da simulação para o teste bilateral

α	Tamanho da amostra (n)					
	5	**10**	**30**	**50**	**200**	**1000**
0,10	0,108	0,116	0,102	0,095	0,088	0,105
0,05	0,057	0,062	0,049	0,046	0,038	0,058
0,01	0,013	0,014	0,011	0,006	0,005	0,009

Tabela 3: Resultado da simulação para o teste unilateral à esquerda

α	Tamanho da amostra (n)					
	5	**10**	**30**	**50**	**200**	**1000**
0,10	0,098	0,089	0,110	0,096	0,126	0,106
0,05	0,044	0,050	0,061	0,058	0,060	0,061
0,01	0,007	0,007	0,011	0,010	0,011	0,015

Tabela 4: Resultado da simulação para o teste unilateral à direita

α	Tamanho da amostra (n)					
	5	**10**	**30**	**50**	**200**	**1000**
0,10	0,108	0,116	0,102	0,095	0,088	0,105
0,05	0,057	0,062	0,049	0,046	0,038	0,058
0,01	0,013	0,014	0,011	0,006	0,005	0,009

Vale ressaltar ainda que para amostras geradas com distintos valores para variância ó², os resultados encontrados permanecem o mesmo, devido à área da região crítica ser sempre fixada em á%.

Conclusão

Neste trabalho, através de um estudo de Monte Carlo, verificou-se que a probabilidade de cometer o erro tipo I num teste de hipóteses para média com variância conhecida está muito próxima da probabilidade nominal de a, ou seja, quando fixamos o nível de significância do teste, a probabilidade de se cometer um erro do tipo I está sob controle, mesmo para tamanhos amostrais variados.

Referências

BUSSAB, W. O.; MORETTIN, P. A. *Estatística básica*. 4. ed. São Paulo: Atual, 1987.

COSTA NETO, P. L. O. *Estatística*. São Paulo: Edgar Blücher, 1977.

MOORE, D. S.; McCABE, G. P. *Introdução à prática da estatística*. 3. ed. New York: LTC, 1999.

VENABLES, W.N.; SMITH, D. M. *An Introduction to R*: notes on R: A Programing Environment for Data Analysis and Grafics, Version 2.4.1, 2006. Disponível em <http://cran-r.c3sl.ufpr.br/.> Acesso em: 15 jun. 2008.

Artigo aprovado para publicação em maio de 2009.

Abstract: In the present paper we can evaluate the probability of making the type I error, when we make use of hypothesis testing on average, through the simulation study known as Monte Carlo method. This evaluation is done to verify the influence of the sample size and the level of significance á in the evaluation of the type I error. All this process is implemented by the computer with the *free software R*. Numerical examples illustrate the proposed methodology.

Keywords: Type I error, Monte Carlo simulation, level of significance, hypothesis testing.