Revista da Estatística UFOP, Vol III(3), 2014, ISSN 2237-8111
Edição especial: 59ª Reunião Anual da Regional Brasileira da Sociedade Internacional de Biometria - RBRAS
Trabalhos apresentados na versão poster.

Estudo comparativo das proporções de evasão nos cursos técnicos na modalidade subsequente e concomitante do IFNMG - Campus Montes Claros, sob uma visão estatística

Maria Fátima Ferreira Almeida [1]
Gerson Rodrigues dos Santos[2]
Jaudir Aguiar Almeida Júnior [3]
Magaly Stefânia Almeida[4]

1 Introdução

Um teste de hipótese pode ser definido como um procedimento de decisão para não rejeitar ou rejeitar uma hipótese estatística, mencionado por CECON et al.(2012) como uma suposição quanto ao valor de um parâmetro populacional ou uma afirmação quanto à natureza da população, com base em informações da amostra.

A necessidade de comparar duas proporções binomiais ocorre com muita frequência nas pesquisas científicas. O método mais comumente tratado pela maioria dos livros de estatística é conhecido como método de Wald (Pan, 2002 apud FERREIRA, 2009), o qual se baseia na aproximação normal assintótica das proporções amostrais observadas. Devido a simplicidade e a familiaridade dos modelos normais , o método de Wald se familiarizou, porém, CARARI et al.(2010) advertem que, para pequenas amostras a qualidade dos intervalos é prejudicada.

Intervalos de confiança exatos para Δ são propostos por Chan e Zhang (1999) e por Agresti e Min (2001), ou seja, intervalos de confiança que não apresentem resultados aberrantes e que sejam conservativos, no entanto de difícil aplicação. Agresti e Coull (1998) propuseram o método que acrescenta 4 pseudo-observações (add-4) com metade sucesso e metade fracasso. Neste caso, duas pseudo-observações são acrescentada a cada amostra, sendo a metade considerada como sucesso, sua generalização para duas amostras independentes foi feita por Agresti e Caffo (2000) citados por FERREIRA et al., 2010.

Este trabalho apresenta uma comparação entre proporções binomiais independentes , cuja variável testada é a proporção de evasão entre as duas populações de estudantes de cursos técnicos na modalidade concomitante e na modalidade subsequente.

Pretende-se verificar se existe diferença significativa entre a proporção de evasão escolar na modalidade subsequente em relação a modalidade concomitante. Para tal, será considerado um nível de significância de 5%.

---

[1] DEMAT– IFNMG/ Montes Claros. e-mail: Professora de Estatística. E-mail: fatima.almeida@ifnmg.edu.br
[2] DET – UFV.e-mail: Professor de Estatística e Geoestatística E-mail: gerson.santos@ufv.br
[3]DCE- IFNMG. Montes Claros. Aluno no curso de Ciência da Computação. E-mail: aguiarjunior40@gmail.com
[4]FACIT- Fundação Educacional de Montes Claros. e-mail: Aluna do curso de Engenharia e Licenciatura em Química . E-mail: magalystefania@yahoo.com.br
Agradecimentos: Ao IFNMG, Campus Montes Claros, pelo apoio prestado.
.

2 Material e método

2.1.Estimação da diferença de duas proporções binomiais

O interesse é estimar Δ por intervalo com um coeficiente de confiança ɣ = 1-α. A distribuição conjunta dos possíveis resultados amostrais (y_1, y_2) que estão contidos no látice retangular de pontos {0, 1,2,...n_1}x{01,...,n_2} deve ser obtida. No entanto os valores amostrais ocorrem com probabilidade desconhecida, que é função de p_1, p_2, n_1 e n_2. Essa distribuição é dada pelo produto das funções de probabilidade das binomiais das duas populações, uma vez que são variáveis independentes. Essa função de probabilidade tem pouco valor para o processo de estimação direta, uma vez que p_1 e p_2 são desconhecidos. A aproximação pela normal apresenta-se como uma alternativa mais simples. O intervalo gerado a apartir desta aproximação é conhecido como intervalo de Wald. Sejam $\hat{p}_1 = \frac{y_1}{n_1}$ e $\hat{p}_2 = \frac{y_2}{n_2}$ os estimadores de máxima verossimilhança de p_1 e p_2, respectivamente, em que y_1 e y_2 referem-se ao número de sucessos do evento de interesse observados nas amostras de tamanho n_1 e n_2. O intervalo de Wald é dado por:

$$IC_{1-\alpha}(\Delta): \hat{p}_1 - \hat{p}_2 \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

em que $\Delta = \hat{p}_1 - \hat{p}_2$ e $Z_{\alpha/2}$ é o quantil superior $\alpha/2$ da distribuição normal padrão.

Este intervalo, embora de fácil aplicação, tem performance muito precária para pequenas amostras e moderadas para $\hat{p}_1$ e $\hat{p}_2$ afastados de ½.

O método proposto por Agresti e Coull (1998) considera que duas pseudo-observações devem ser acrescentadas a cada amostra, sendo a metade considerada como sucesso. O estimador modificado de p_i, com i =1, 2, é dado por:

$$\hat{p}_i = \frac{y_{i+1}}{n_i + 2}\text{ , i = 1,2 ; cuja variância é estimada por: } \hat{V}(\hat{p}_i) = \hat{V}(\hat{p}_i) = \frac{\hat{p}_i(1-\hat{p}_i)}{n_i + 2} \qquad (1)$$

Assim o intervalo de confiança utilizando a aproximação normal com generalização para duas populações independentes feita por Agresti e Caffo(2000) é dado por:

$$IC_{1-\alpha}(\Delta): \hat{p}_1 - \hat{p}_2 \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1+2} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2+2}} \qquad (2)$$

Pela comparação dos dois primeiros momentos dos estimador da variância da diferença das duas proporções com aqueles da distribuição qui-quadrado escalonado, FERREIRA et al., (2010) apresenta o resultado para $\text{Var}(\hat{V}(\hat{p}_i))$ e os graus de liberdade v para aplicação do teste t, dados por:

$$v = \frac{2\left[\hat{V}(\hat{p}_1)+\hat{V}(\hat{p}_2)\right]^2}{\hat{V}ar[\hat{V}(\hat{p}_1)]+\hat{V}ar[\hat{V}(\hat{p}_2)]} \tag{3}$$

O intervalo de Confiança T avaliado por Pan(2002) que se baseia na distribuição t de Student é dado, portanto, por:

$$C_{1-\alpha}(\Delta): \hat{p}_1 - \hat{p}_2 \pm t_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1+2}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2+2}} \tag{4}$$

Em que $t_{\alpha/2}$ é o quantil superior $\alpha/2$ da distribuição t de student com v graus de liberdade dado por (3). Deste modo, para testar a hipótese nula:

$H_0 : \hat{p}_1 - \hat{p}_2 - \Delta = 0$ versus $H_1 : \Delta \neq 0$

2.2. O Teste de Agresti e Caffo (2000) modificado por Pan (2002)

Para a aplicação deste teste é necessário obter os graus de liberdade dado por (3) e em seguida aplicar o teste, dado por :

$$t_{\alpha/2} = \frac{\hat{p}_1 - \hat{p}_2 - \Delta}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{(n_1+2)}+\frac{\hat{p}_2(1-\hat{p}_2)}{+(n_2+2)}}} \tag{5}$$

O calculo da variância $\hat{V}ar(\hat{v}(p_i))$ é feito fórmula (6).

$$\begin{aligned}\hat{V}ar(\hat{v}(\hat{p}_i)) &= (\hat{p}_i - \hat{p}_i\ )/(n_i+2)^3 \\ &+\left[\hat{p}_i + (6n_i+5)\hat{p}_i^{\,2} + 4(n_i+1)(n_i-1)\hat{p}_i^{\,3} - 2(n_i+1)(2n_i+1)\hat{p}_i^{\,4}\right]/(n_i+2)^5 \\ &-2\left[\hat{p}_i + (2n_i+5)\hat{p}_i^{\,2} - 2(n_i+1)(2n_i+1)\hat{p}_i^{\,3}\right]/(n_i+2)^4\end{aligned} \tag{6}$$

A fórmula (6) é utilizada para o cálculo de graus de liberdade (v) dado pela fórmula (3) para obter valor crítico na tabela t de Student.

O p-valor é obtido por: P(t > $t_{\alpha/2}$/v). (7)

2.3.Caracterização dos dados da pesquisa

Os dados analisados referem-se ao número de estudantes evadidos anualmente nos cursos técnicos do Instituto Federal de Educação do Norte de Minas Gerais, campus Montes Claros, discriminada a modalidade de ensino em Concomitante e Subsequente.

A motivação para aplicar um teste estatístico reside no fato de que o número de alunos evadidos anualmente na modalidade concomitante é inferior ao subsequente, mas isso não garante que estatisticamente as proporções sejam também inferiores em relação a modalidade subsequente, daí, utilizou o teste de hipóteses de Agresti e Caffo (2000) modificado por Pan (2002) mencionado por FERREIRA (2010), que garante uma análise mais coerente para casos de pequena amostra.

3 Resultados e discussão

Segue abaixo a descrição das hipóteses:

H_0: A proporção de alunos evadidos é igual estatísticamente nas duas modalidades de ensino, concomitante e subsequente

H_1 : A proporção de alunos evadidos diferem nas duas modalidades de ensino, concomitante e subsequente.

A Tabela 1 apresenta a aplicação do método de Agresti e Caffo(2000) modificado por Pan (2002) para testar as hipóteses acima descritas.

Tabela 1 – Resultado da Proporção de p_1: para o evento evasão na modalidade concomitante e p_2: para o evento evasão na modalidade subsequente a cada ano de 2010 a 2013 e total.

Ano	**Total de evadidos**		**Total de alunos**		**Probabilidade**	
	Concomitante	**Subsequente**	**Concomitante**	**Subsequente**	**Concomitante (p_1)**	**Subsequente (p_2)**
2010	11	91	30	184	0,38	0,49
2011	4	52	18	121	0,25	0,43
2012	18	25	118	94	0,16	0,27
2013	6	32	107	183	0,06	0,18
Total	39	200	273	582	0,15	0,34

Utilizando as fórmulas (1), (6), (3), (5) e (7) para aplicar o método modificado por Pan (2002) foi obtido os resultados apresentados na Tabela 2 , referente a cada ano e para o *Total* do período.

Tabela 2 – Valores de $V(pi)$ e $Var(V(pi))$, graus de liberdade v, valor $t_{\alpha/2}$ e *p -valores* anuais e total de acordo com o método de Agresti e Caffo (2000) modificado por Pan (2002).

Ano	**$V(p_1)$**	**$V(p_2)$**	**$Var(V(p_1))$**	**$Var(V(p_2))$**	**$t_{\alpha/2}$ calculado**	**v (*grau de liberdade*)**	***p-valor***
2010	0,00705	0,00134	0,00000072	1,083X 10^{-10}	-1,5934	194,33	0,056
2011	0,00938	0,00199	0,00000571	2,998 X10^{-09}	-1,6889	45,28	0,049
2012	0,00111	0,002057	0,00000004	4,683 X10^{-08}	-1,9918	243,13	0,024
2013	0,00055	0,00792	0,00000003	9,512 X10^{-09}	-3,1601	81,74	0,001
Total	0,00045	0,00387	0,00000000	1,106 X10^{-10}	-6,7479	453,7	2,3 x10-11

De acordo com os resultados constantes da Tabela 2, e considerando um erro α de 5%, tem-se o seguinte resultado para as hipóteses testadas:

No ano de 2010 Aceita-se a hipótese H_0, com 95% de confiança, de que as taxas de evasão dos cursos técnicos no IFNMG campus Montes Claros são iguais na modalidade concomitante e subsequente. Apartir do ano de 2011 até o ano de 2013, com 95% de confiança, não foi aceita a hipótese H_0 de que as taxas taxas de evasão dos cursos técnicos no IFNMG- Campus Montes Claros são iguais na modalidade concomitante e subsequente, ou seja, há evidências estatatísticas que podem ser constatadas na Tabela 1 e 2, de que neste período a taxa de evasão entre os alunos que cursam os cursos técnicos na modalidade subsequente são superiores aos que cursam na modalidade concomitante. A análise global do período de 2010 a 2013 evidencia por meio do p-valor (Tabela 2), que há diferença entre as taxas de evasão nas duas modalidades, prevalecendo a maior taxa de evasão na modalidade subsequente (Tabela 1).

4 Conclusão

De acordo com os resultados obtidos conclui-se que a metodologia mostrou eficiente para o teste de hipótese de proporções e deixou evidente a diferença existente entre as duas modalidades de ensino técnico (Concomitante e subsequente) quanto a evasão escolar, que irão contribuir para a tomada de decisão e gerenciamento de ações para corrigir o problema.

5 Referências

[1] CARARI,M.L.; LIMA, P. C.; FERREIRA, D. F.; CIRILLO, M. A. **Estimação de diferenças entre duas proporções binomiais via bootstrap.** Rev. Bras. Biom., São Paulo, v.28, n.3, p.112-134, 2010 .

[2] CECON,P.R.; SILVA, A.R.; NASCIMENTO, M.;FERREIRA,A. **Métodos estatísticos.** Viçosa: Editora UFV, 229p (Didática), 2012.

[3] FERREIRA, D. F. **Estatística básica.** 2ª ed. revisada**.** Lavras: Editora UFLA, 663p, 2009.

[4] PAN, W. Approximate confidence intervals for one proportion and two proportions.*Comput. Stat. Data Anal.*, Amsterdan, v.40, n.1, p.143-157, 2002.