

Avaliando a relação entre as variáveis categóricas status final do curso e demográficas.

Oberdan Costa

*TRS (Technology, Networks and Society), Universidade Fernando Pessoa

10 de Julho de 2020

PhD CI SiTEGI 2020 - UFP

Porto - Portugal

Breve contextualização

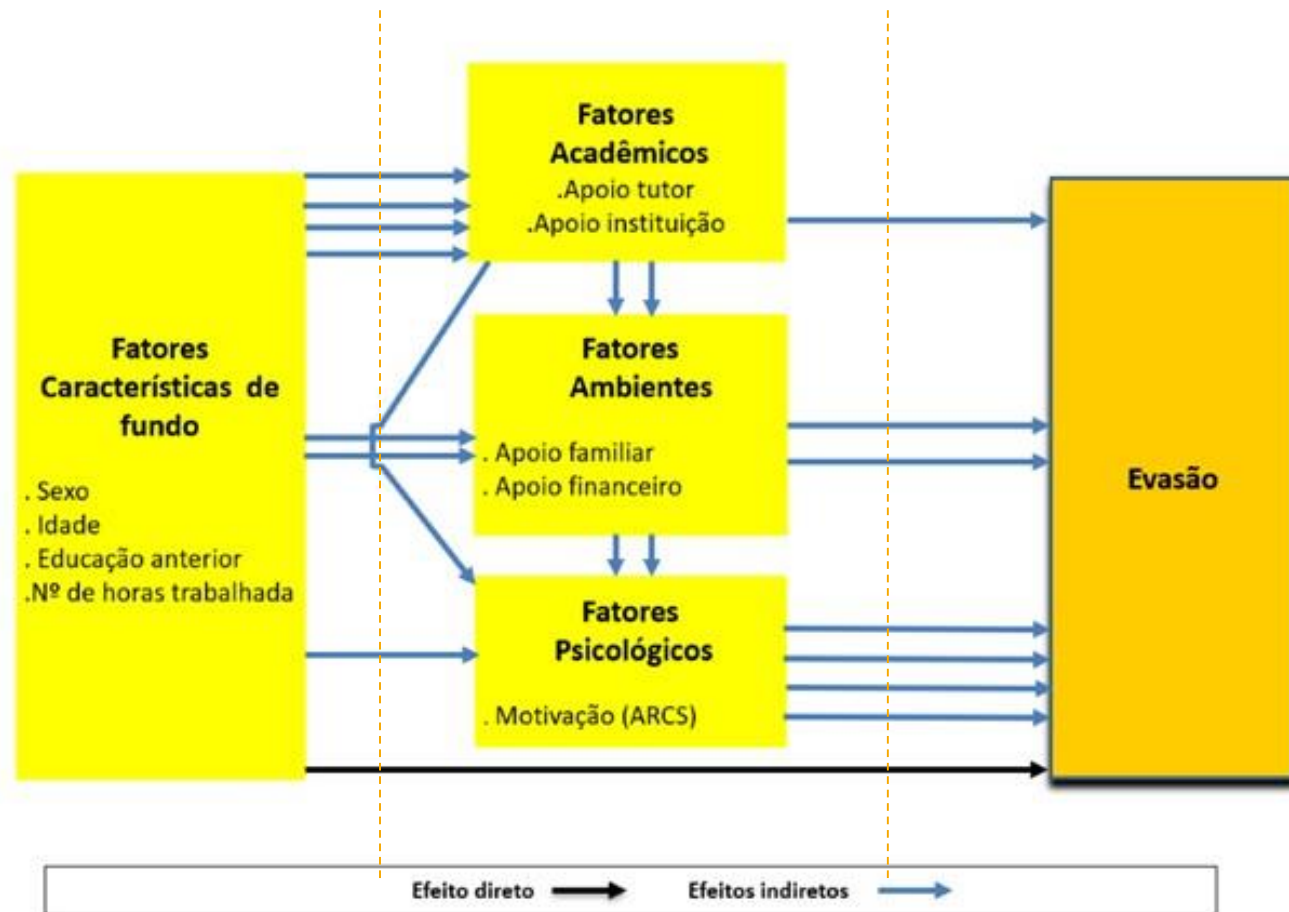
- O presente trabalho trás um recorte da tese, intitulada:

DROPOUT NO ENSINO À DISTÂNCIA E A SUSTENTABILIDADE DAS IES: um modelo de referência para um sistema integrado de alertas para evasão.

- Problema

Em que medida o modelo de referência proposto para um sistema integrado de alertas contribui para prever a evasão de estudantes nas instituições de ensino superior a distância.

Breve contextualização



. Fatores preditores da evasão de estudantes anteriores a admissão da graduação (Costa e Gouveia ,2019).

. Fatores preditores da evasão de estudantes posteriores a admissão da graduação

Avaliando a relação entre as variáveis categóricas status final do curso e demográficas.

Oberdan Costa



Introdução

Há na literatura diversos trabalhos que oferecem uma visão geral da pesquisa relacionada a associação entre as variáveis demográficas e a variável resposta evasão no ensino superior a distância. Entre eles:

- . Os estudos de (XENOS, PIERRAKEAS, PINTELAS, 2002; WALTER, 2006) apontam que as características de estudantes adulto antes da sua admissão no Ensino Superior estão associadas a evasão.
- . Um estudo por meio da estatística descritiva, qui-quadrado e análise multivariada de variância (MANOVA) de (Park e Choi,2009) revela as características individuais idade, o gênero e o nível educacional não tiveram um efeito significativo e direto na decisão de evasão dos estudantes.

- **Objetivo**

Investigar a associação entre as variáveis categóricas Status Final de Curso (SFC) e demográficas (sexo, idade, grau de escolaridade, situação de emprego e tempo de duração do curso) e entre suas categorias.

- **Hipóteses**

- H1: A associação entre as variáveis categóricas SFC e sexo não se dá de forma aleatória.
- H2: A associação entre as variáveis categóricas SFC e idade não se dá de forma aleatória.
- H3: A associação entre as variáveis categóricas SFC e grau de escolaridade não se dá de forma aleatória.
- H4: A associação entre as variáveis categóricas SFC e Horas de Trabalho Semanais (HTS) não se dá de forma aleatória.
- H5: A associação entre as variáveis categóricas SFC e Tempo de Duração do Curso (TDC) não se dá de forma aleatória

Metodologia

- O trabalho emprega análise de correspondência simples (Anacor) - teste do χ^2 (mostra a existência de associação estatisticamente significativa entre variáveis categóricas) e Resíduos Padronizados Ajustados (revela as relações de dependências entre cada par de categoria).
- Dados da amostra: 228 estudantes de programas acadêmicos de licenciatura e bacharelado de 42, 48 e 60 meses da modalidade de ensino a distância finalizados em dezembro de 2014 e 2015, mantidos por uma instituição privada da região sul do Brasil
- Software SPSS v.25

Quadro geral características da amostra e distribuições de frequência

Variável categórica	Categoria	Freq.	Porc.
Status Final de Curso (SFC)	0= "Não evadiu/Concluiu"	143	62,7
	1= "Não concluiu/Evadiu"	85	37,3
Variável categórica	Categoria	Freq.	Porc.
Sexo	1 ="masculino"	92	40,4
	2 ="feminino"	136	59,6
Idade	1= 19 ou menos	2	,9
	2=20 a 24 anos	31	13,6
	3=25 a 29 anos	49	21,5
	4=30 a 34 anos	69	30,3
	5=35 a 39 anos	43	18,9
	6=40 a 44 anos	29	12,7
	7= 45 a 49 anos	4	1,8
	8=50 a 54 anos	1	,4
	9 =55 ou mais	,0	,0
Educação	1= Diploma do ensino médio	150	65,8
	2= EJA	45	19,7
	3= PROEJA	,0	,0
	4= Licenciado ou de dois anos	21	9,7
	5= Diploma de bacharel	12	5,3
Horas de trabalho semanais	1= Menos de 20 horas	3	1,3
	2= 20 a 29 horas	2	,9
	3= 30 a 39 horas	8	3,5
	4 = 40 a 49 horas	199	87,3
	5 = 50 ou mais horas	16	7
Tempo de duração do curso	1= 42 meses	17	7,5
	2= 48 meses	196	86,0
	3= 60 meses	14	6,1

Apresentação de resultados

- Teste χ^2 para verificação de associação entre as variáveis categóricas
- Análise de resíduos padronizado ajustado.

Teste χ^2 para avaliação de associação entre as variáveis categóricas

01. Par variáveis categóricas (SFC * sexo)

Testes qui-quadrado^c

	Valor	gl	Significância Assintótica (Bilateral)	Sig. exata (2 lados)	Sig. exata (1 lado)	Probabilidade de ponto
Qui-quadrado de Pearson	,007 ^a	1	,934	1,000	,523	
N de Casos Válidos	228					

a. 0 células (0,0%) esperavam uma contagem menor que 5. A contagem mínima esperada é 34,30.

02. Par variáveis categóricas (SFC * idade)

Testes qui-quadrado

	Valor	gl	Significância Assintótica (Bilateral)
Qui-quadrado de Pearson	10,453 ^a	7	,164
N de Casos Válidos	228		

a. 6 células (37,5%) esperavam uma contagem menor que 5. A contagem mínima esperada é ,37.

03. Par variáveis categóricas (SFC * grau de escolaridade)

Testes qui-quadrado

	Valor	gl	Significância Assintótica (Bilateral)
Qui-quadrado de Pearson	2,997 ^a	3	,392
N de Casos Válidos	228		

a. 1 células (12,5%) esperavam uma contagem menor que 5. A contagem mínima esperada é 4,47.

Figura 5. Resultado do teste χ^2 para verificação de associação entre o SFC e grau de escolaridade.

04. Par variáveis categóricas (SFC * horas de trabalho semanais)

Testes qui-quadrado

	Valor	gl	Significância Assintótica (Bilateral)
Qui-quadrado de Pearson	2,037 ^a	4	,729
N de Casos Válidos	228		

a. 5 células (50,0%) esperavam uma contagem menor que 5. A contagem mínima esperada é ,75.

Teste χ^2 para verificação de associação entre as variáveis categóricas

05. Par variáveis categóricas (SFC * tempo de duração do curso)

Testes qui-quadrado

	Valor	gl	Significância Assintótica (Bilateral)
Qui-quadrado de Pearson	4,631 ^a	2	,099
N de Casos Válidos	228		

a. 0 células (0,0%) esperavam uma contagem menor que 5. A contagem mínima esperada é 5,22.

Análise de resíduos padronizado ajustado - Tabulação cruzada



Tabulação cruzada SFC * idade

		idade							Total	
		19 ou menos	20 a 24 anos	25 a 29 anos	30 a 34 anos	35 a 39 anos	40 a 44 anos	45 a 49 anos		50 a 54 anos
2	(a)	2	16	29	50	26	15	4	1	143
	(b)	1,3	19,4	30,7	43,3	27,0	18,2	2,5	,6	143,0
	(c)	,7	-3,4	-1,7	6,7	-1,0	-3,2	1,5	,4	
	(d)	,7	-,8	-,3	1,0	-,2	-,7	,9	,5	
	(e)	1,1	-1,4	-,6	2,0	-,3	-1,3	1,6	,8	
1	(a)	0	15	20	19	17	14	0	0	85
	(b)	,7	11,6	18,3	25,7	16,0	10,8	1,5	,4	85,0
	(c)	-,7	3,4	1,7	-6,7	1,0	3,2	-1,5	-,4	
	(d)	-,9	1,0	,4	-1,3	,2	1,0	-1,2	-,6	
	(e)	-1,1	1,4	,6	-2,0	,3	1,3	-1,6	-,8	
Total	(f)	2	31	49	69	43	29	4	1	228
	(g)	2,0	31,0	49,0	69,0	43,0	29,0	4,0	1,0	228,0

(1) Status Final de Curso; (2) concluiu; (3) evadiu; (a) Contagem; (b) Contagem esperada; (c) resíduo; (d) resíduo padronizado; (e) resíduo ajustado; (f) contagem TT; (g) contagem esperada TT

Analise de resíduos padronizado ajustado - Tabulação cruzada

+

Tabulação cruzada SFC * Horas de trabalho semanais

		Horas de trabalho semanais					Total	
		Menos de 20 horas	20 a 29 horas	30 a 39 horas	40 a 49 horas	50 a 59 horas		
1	2	(a)	2	1	4	128	8	143
		(b)	1,9	1,3	5,0	124,8	10,0	143,0
		(c)	,1	-,3	-1,0	3,2	-2,0	
		(d)	,1	-,2	-,5	,3	-,6	
		(e)	,1	-,4	-,8	1,3	-1,1	
	3	(a)	1	1	4	71	8	85
		(b)	1,1	,7	3,0	74,2	6,0	85,0
		(c)	-,1	,3	1,0	-3,2	2,0	
		(d)	-,1	,3	,6	-,4	,8	
		(e)	-,1	,4	,8	-1,3	1,1	
Total	(f)	3	2	8	199	16	228	
	(g)	3,0	2,0	8,0	199,0	16,0	228,0	

(1) Status Final de Curso; (2) concluiu; (3) evadiu; (a) Contagem; (b) Contagem esperada; (c) resíduo; (d) resíduo padronizado; (e) resíduo ajustado; (f) contagem TT; (g) contagem esperada TT

Analise de resíduos padronizado ajustado - Tabulação cruzada

Tabulação cruzada SFC * Tempo de duração do curso

		Tempo de duração do curso			Total	
		42 meses	48 meses	60 meses		
1	2	Contagem	13	120	10	143
		Contagem Esperada	9,4	124,8	8,8	143,0
		Resíduo	3,6	-4,8	1,2	
		Resíduos padronizados	1,2	-,4	,4	
		Resíduos ajustados	2,0	-2,0	,7	
	3	Contagem	2	79	4	85
		Contagem Esperada	5,6	74,2	5,2	85,0
		Resíduo	-3,6	4,8	-1,2	
		Resíduos padronizados	-1,5	,6	-,5	
		Resíduos ajustados	-2,0	2,0	-,7	
Total	Contagem	15	199	14	228	
	Contagem Esperada	15,0	199,0	14,0	228,0	

(1) Status Final de Curso; (2) concluiu; (3) evadiu



Discussão dos resultados

Teste χ^2 para verificação de associação entre as variáveis categóricas

RESULTADOS DA ASSOCIAÇÃO			χ^2_c VALOR CRITICO - Tabela de Distribuição		
Variáveis	χ^2 Valor Cal	gl.	χ^2_c	gl.	Nível de significância (α)
SFC e idade	10,453	7 graus	14,067	7 graus	5%

Regras de decisão Hipóteses

.Se X^2 calculado $\geq \chi^2_c$ tabelado: Rejeita-se H2

.Se X^2 calculado $< \chi^2_c$ tabelado: **Aceita-se H2, portanto não existe associação estatisticamente significativa**

Análise de Resíduo Padronizado Ajustado (RPA)

Variáveis	Categoria	Resíduo	RPA na célula	Tabela Distribuição Normal Padrão (TDNP)
SFC e idade	30 a 34 anos	6,7	2	-1.96 - 1.96

Como RPA ≥ 1.96 TDNP, **temos uma relação de dependência entre as categorias concluiu e 30 a 34 anos**

Discussão dos resultados

Teste χ^2 para verificação de associação entre as variáveis categóricas

RESULTADOS DA ASSOCIAÇÃO			χ^2_c VALOR CRITICO - Tabela de Distribuição		
Variáveis	χ^2 Valor Cal	gl.	χ^2_c	gl.	Nível de significância (α)
SFC e HTS	2,037	4 graus	9,488	4 graus	5%

Regras de decisão Hipóteses

.Se X^2 calculado $\geq \chi^2_c$ tabelado: Rejeita-se H4

.Se X^2 calculado $< \chi^2_c$ tabelado: **Aceita-se H4, portanto não existe associação estatisticamente significativa**

Análise de Resíduo Padronizado Ajustado (RPA)

Variáveis	Categoria	Resíduo	RPA na célula	Tabela Distribuição Normal Padrão (TDNP)
SFC e HTS	40 a 49 horas	3,2	1,3	-1.96 - 1.96

Como RPA < 1.96 TDNP, **não há relação de dependência entre as categorias concluiu e 40 a 49 horas**

Discussão dos resultados

Teste χ^2 para verificação de associação entre as variáveis categóricas

RESULTADOS DA ASSOCIAÇÃO			χ^2_c VALOR CRITICO - Tabela de Distribuição		
Variáveis	χ^2 Valor Cal	gl.	χ^2_c	gl.	Nível de significância (α)
SFC e TDC	4,631	2 graus	5,991	2 graus	5%

Regras de decisão Hipóteses

.Se X^2 calculado $\geq \chi^2_c$ tabelado: Rejeita-se H_0

.Se X^2 calculado $< \chi^2_c$ tabelado: **Aceita-se H_0 , portanto não existe associação estatisticamente significativa**

Análise de Resíduo Padronizado Ajustado (RPA)

Variáveis	Categoria	Resíduo	RPA na célula	Tabela Distribuição Normal Padrão (TDNP)
SFC e TDC	42 meses	3,6	2	-1.96 - 1.96

Como $RPA \geq 1.96$ TDNP, temos **uma relação de dependência entre as categorias concluiu e 42 meses**

Conclusão

- Verificou-se pelo teste do qui-quadrado que todos os pares de variáveis categóricas testados apresentaram resultado menores que o da tabela, ou seja, χ^2 (calculado) $< \chi_c^2$ (Valor crítico, grau de liberdade correspondente e nível de significância de 5%). Dessa forma considerando o teste χ^2 em todos os pares de variáveis categóricas aceitar-se a hipótese nula de que as duas variáveis categóricas se associam de forma aleatória, ou seja, **não existe associação estatisticamente significativa**, ao nível de significância de 5% entre a variável categórica SFC e cada uma das variáveis categorias demográficas.

Conclusão

Relação de dependências entre pares de categoria

Item(ns) destaque		Maior Frequência SFC		Valores		Dependência? (S/N) ¹
Variável	Categoria	Concluiu	Evadiu	Resíduo	RPA	
Sexo	masculino	58	34	0,3	0,1	Não
	feminino	85	51	0,3	0,1	Não
Idade	20 a 24 anos	16	15	3,4	1,4	Não
	30 a 34 anos	50	19	6,7	2,0	Sim
	40 a 44 anos	15	14	3,2	1,3	Não
Grau de escolaridade	diploma de ensino médio	98	52	3,9	1,1	Não
horas de trabalho semanais	40 a 49 horas	128	71	3,2	1,3	Não
Tempo de duração do curso	42 meses	13	2	3,6	2,0	Sim
	48 meses	120	79	4,8	-2,0	Não

¹ Sim ou Não(S/N)

Referenciais

- Costa, O. e Gouveia, L. (2019). Educação superior a distância: fatores preditores da evasão anteriores a admissão de estudantes. *Revista Educação Em Questão*, 57(51).
<https://doi.org/10.21680/1981-1802.2019v57n51ID15671>.
- FÁVERO, Luiz Paulo Lopes; BELFIORE, Patrícia Prado. Manual de análise de dados: estatística e modelagem multivariada com excel, SPSS e stata. [S.l: s.n.], 2017.
- PARK, Ji-Hye; CHOI, Hee Jun. Factors influencing adult learners' decision to drop out or persist in online learning. *Educational Technology & Society*, Taiwan, v. 12, n. 4, p. 207-217, 2009.
- WALTER, Amanda Moura. Variáveis preditoras de evasão em cursos a distância. 2006. 177f. Dissertação – Instituto de Psicologia, Universidade de Brasília, Brasília, 2006.
- XENOS, Michalis; PIERRAKEAS, Christos; PINTELAS, Panagiotis. A survey on student dropout rates and dropout causes concerning the students in the Course of Informatics of the Hellenic Open University. *Computers and Education*, Amesterdã, v. 39, n. 4, p. 361-377, 2002.