

## A PESQUISA AVALIATIVA: SUGESTÕES ALTERNATIVAS PARA O TRATAMENTO SIMPLIFICADO DE DADOS QUANTITATIVOS EM ESCALA NOMINAL DICOTÔMICA NAS PESQUISAS PEDAGÓGICAS NO AMBIENTE ESCOLAR

Adroaldo C. A. Gaya<sup>1</sup>, Arlete Brasiliense e Anelise Gaya<sup>2</sup>

**RESUMO** - O objetivo deste ensaio é sugerir um conjunto de alternativas de fácil utilização para o tratamento de dados quantitativos na pesquisa avaliativa em investigações pedagógica no ambiente escolar. Define-se operacionalmente a pesquisa avaliativa aplicada às práticas pedagógicas como um conjunto de procedimentos sistemáticos de coleta, análise e interpretação de dados fidedignos e válidos para a avaliação da eficácia e tomada de decisões sobre os programas de intervenção pedagógica no ambiente escolar. Apresentam-se dois desenhos metodológicos: (1) desenhos de grupo único com pré e pós teste e, (2) desenhos com grupos de intervenção e controle não equivalentes. Para análise dos dados propõe-se medidas de dimensão de efeito e medidas de probabilidade baseado no valor de “p”. Nos estudos de grupo único as medidas de dimensão do efeitos são as medidas de prevalência, de ocorrência e a razão de chance. Para o cálculo de “p” utiliza-se o teste de McNemar. Nos estudos de dois grupos não equivalentes sugere-se inicialmente proceder-se ao ajustamento dos grupos por técnica de pareamento e, posteriormente, valer-se das medidas de tamanho do efeito: as medidas de prevalência, de incidência, razão de chance e acrescenta-se a medida de razão relativa de chance. Para o valor de “p” sugere-se o teste do Qui-quadrado com correção de Yates. Supõe-se que este ensaio possa constituir-se numa ferramenta útil aos estudantes de licenciatura na elaboração de seus trabalhos de conclusão de curso e dos professores-pesquisadores que valem-se do método científico para avaliar o impacto de sua prática pedagógica. São sugestões de fácil aplicação, que podem inclusive dispensar a necessidade de softwares estatísticos, sem, no entanto, descuidar da relevância do rigor inerente ao conhecimento científico.

**Palavras-chave:** Pesquisa avaliativa. Intervenção. Escola.

**ABSTRACT** -. The purpose of this essay is to suggest a set of user-friendly alternative for the treatment of quantitative data on evaluative research in educational research in the school environment. It is defined operationally evaluative research applied to teaching practices as a set of systematic collection procedures, analysis and interpretation of reliable and valid data for assessing the effectiveness and decision-making on educational intervention programs in the school environment. They present two methodological designs: (1) unique set of drawings with pre and post test and (2) designs with intervention groups and non-equivalent control. Data analysis proposes to effect size measures and probability measures based on the value of "p". The only group studies the effects of the measures dimension is the prevalence measures of occurrence and the odds ratio. To calculate "p" is used McNemar test. In studies of two non-equivalent groups is suggested initially proceed to the adjustment of the matching technique of groups and then make use of the effect size measures: measures the prevalence, incidence, and adds odds ratio the relative ratio measurement chance. For the value of "p" suggests the chi-square test with Yates correction. It is assumed that this test can constitute a useful tool for undergraduate students in preparing their work for completion of course and teacher-researchers who draw on the scientific method to assess the impact of their teaching. Suggestions are easy to apply, which may even obviate the need for statistical software, without, however, neglecting the importance of rigor inherent to scientific knowledge.

**Keywords:** Evaluative research. Intervention. School.



Revista  
Ciência e Conhecimento  
Volume 9 – Nº 2 – 2015.



1 – Universidade Federal do Rio Grande do Sul – UFRGS. Projeto Esporte Brasil – PROESP-Br. Porto Alegre, RS, Brasil.

2 – Secretaria Municipal de Educação de Alvorada, RS, Brasil. Assessora Pedagógica do Projeto Esporte Brasil – PROESP-Br.

**E-mail para contato:**  
Adroaldo C. A. Gaya  
acgaya@ufrgs.esef.br

Recebido em: 20/06/2015.  
Revisado em: 28/07/2015.  
Aceito em: 20/08/2015.

**Área:**  
Metodologias e estratégias de ensino aprendizagemens.

## INTRODUÇÃO

*Dois momentos em que a universidade bate à porta da escola como um pedinte faminto: o estágio dos nossos alunos da graduação e a realização de uma pesquisa sobre a escola, a sala de aula ou a prática pedagógica de um professor (TELLES, 2002, p. 92).*

Os cursos de licenciatura nas suas diversas áreas do saber se deparam com a crescente necessidade de produzir conhecimento científico sobre as práticas pedagógicas no ambiente escolar. Entretanto, principalmente nas licenciaturas intimamente ligadas às ciências biológicas e da natureza, os modelos quantitativos hegemônicos da produção do conhecimento científico migram diretamente dos laboratórios para o ambiente escolar levando consigo exigências metodológicas que causam dificuldades operacionais aos professores-pesquisadores. Não obstante, salienta-se que não se trata de questionar a elegância teórica, a qualidade e o rigor desses métodos quantitativos clássicos. O que está em causa não é sua idoneidade científica, mas sua possibilidade de adaptar-se ao ambiente escolar. Os modelos experimentais clássicos exigem um conjunto de pressupostos que na maioria das vezes inviabilizam sua aplicação nas pesquisas pedagógicas. Por exemplo: a exigência da aleatoriedade da amostra, a distribuição probabilística, a delimitação da dimensão da amostra e os pressupostos da estatística inferencial. Os requisitos de Fischer, como são conhecidas essas exigências (cf. SALSBURG, 2009), não se ajustam facilmente ao ambiente escolar. O objetivo deste ensaio é sugerir um conjunto de alternativas de fácil utilização para o tratamento de dados quantitativos na pesquisa avaliativa em investigações pedagógica no ambiente escolar.

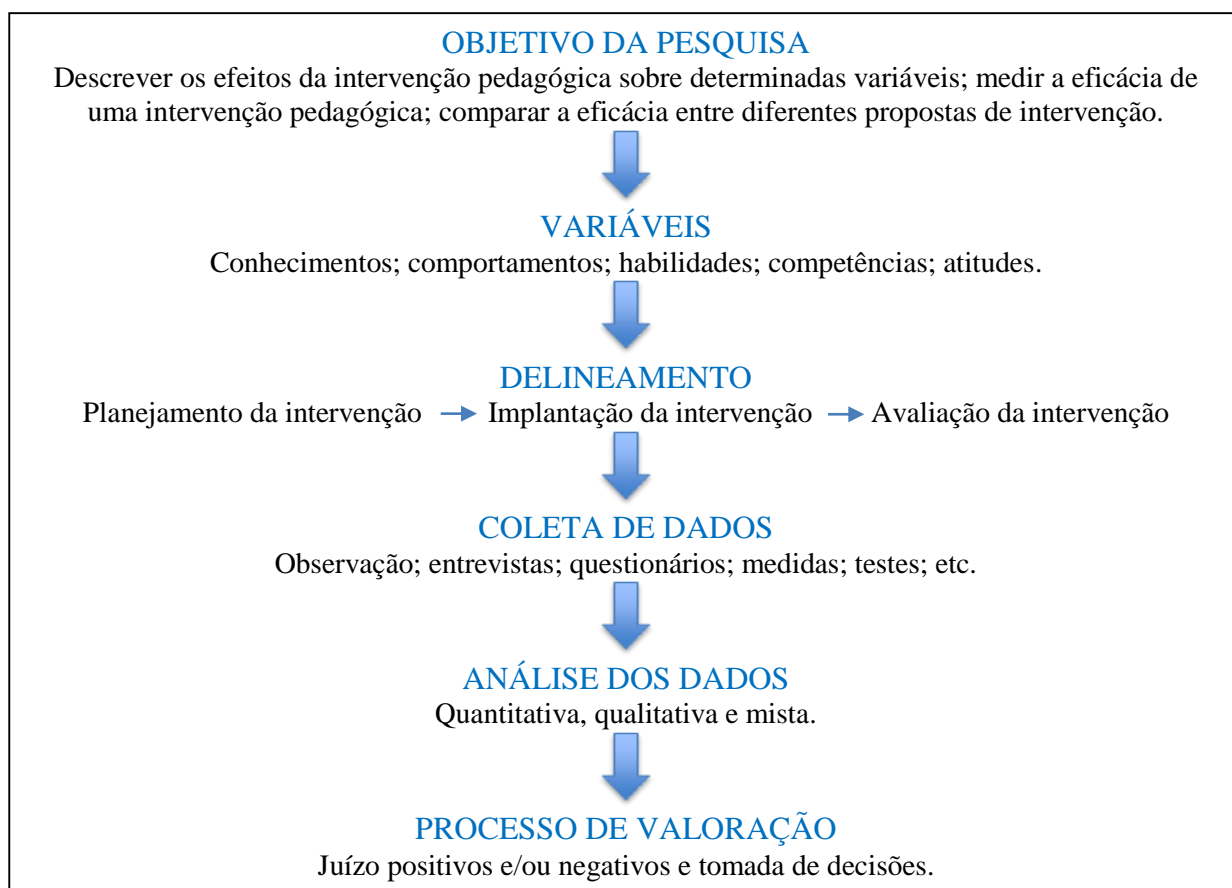
Define-se operacionalmente a pesquisa avaliativa aplicada às práticas pedagógicas como um conjunto de procedimentos sistemáticos de coleta, análise e interpretação de dados fidedignos e válidos para a avaliação da eficácia e tomada de decisões sobre os programas de intervenção pedagógica no ambiente escolar.

A pesquisa avaliativa, em relação a pesquisa experimental tradicional, é mais focada na busca de efetivos resultados práticos. Como referem Seltz, Wrightsman e Cook (1987), seus resultados não se destinam apenas a incrementar o corpo de conhecimentos ou a desenvolver teorias, são utilizados imediatamente para a tomada de decisões. Há duas classes de pesquisa avaliativa: a pesquisa somativa que avalia o produto final da intervenção e; a pesquisa formativa que avalia o processo de intervenção (CONTANDRIOPOULOS, 2006). Este ensaio restringe-se aos delineamentos da pesquisa avaliativa somativa com abordagem quantitativa com medidas nominais dicotômicas (do tipo sucesso/insucesso).

## 1. Estratégias operacionais das pesquisas avaliativas somativas com abordagem quantitativa

Os projetos de pesquisa avaliativa, como se pode observar no quadro 1, não diferem substancialmente dos projetos tradicionais de investigação científica, senão essencialmente por um processo de valoração subjetiva que, como se verá adiante, baseada num conjunto de valores manifestos pelo professor-pesquisador servem de referências “*ad hoc*”<sup>1</sup> para a definição dos procedimentos de avaliação e de tomadas de decisão (DE LA ORDEM, 1985)<sup>2</sup>. Em outras palavras, a pesquisa avaliativa **consiste em aplicar um julgamento de valor a uma intervenção, através de um dispositivo capaz de fornecer informações cientificamente válidas e pedagogicamente legítimas, permitindo aos diferentes atores envolvidos se posicionarem e construírem um julgamento capaz de ser traduzido em ação** (CONTANDRIOPOULOS, 2006).

**Quadro 1.** As possíveis fases de um projeto de pesquisa avaliativa somativa.



**Fonte:** Adaptado pelos autores de ARNAL, J.; Del RINCÓN, D.; LATORRE, A. (1992).

<sup>1</sup> “Ad hoc”: referente a argumento, proposta ou hipótese formulada com o único objetivo de legitimar ou defender uma teoria, e não em decorrência de uma compreensão objetiva e isenta da realidade.

<sup>2</sup> DE LA ORDEM, A (coord). Dicionário de Ciencias de la Educación. Investigación Educativa. Madrid: Anaya, 1985.

Este ensaio discorre sobre as propostas de estratégias operacionais das pesquisas avaliativas somativas com abordagem quantitativa e medidas nominais dicotômicas para a avaliação da eficácia de programas de intervenção pedagógica a partir de dois delineamentos: (1) com pré e pós teste de grupo único e; (2) com grupo de intervenção e grupo controle não equivalente. Para todos os modelos sugere-se o tratamento dos dados a partir de análises estatísticas descritivas. Em outras palavras, sem a pretensão de inferências por modelação estatística<sup>3</sup>.

## **1.1 Delineamentos para pesquisa avaliativa somativa com pré e pós teste de grupo único**

### **2.1.1 Avaliação Somativa através de Medidas de Dimensão do Efeito com escalas nominais dicotômicas**

Este é o modelo mais simples e mais fácil de se realizar no ambiente escolar. É um delineamento do tipo pré-experimental (Cf. CAMPBELL e STANLEY, 2001). O professor pode realizar a pesquisa com sua própria turma. O desenho corresponde: pré-teste → intervenção → pós-teste. Salienta-se que há evidentes limitações no que se refere a validação externa dos resultados (indução dos resultados para além dos sujeitos avaliados). Não obstante, ao utilizar-se medidas de pré e pós teste com o mesmo grupo de sujeitos (onde cada sujeito é seu próprio controle), minimizam-se algumas principais dificuldades no que se refere ao controle das variáveis entre os sujeitos (intrasujeitos), estratégia que permite admitirmos uma boa validade interna.

Sugere-se, neste ensaio, que as variáveis de desfecho (variáveis dependentes) sejam medidas em escala nominal dicotômica. As variáveis medidas em escala nominal dicotômica permitem que o professor-pesquisador possa afirmar a eficácia de seu projeto de intervenção pedagógica (medidas de dimensão do efeito) diretamente pela ocorrência dos alunos que obtiveram sucesso (ou insucesso). Tem-se a convicção de que para o professor-pesquisador é mais relevante identificar os alunos que beneficiaram-se (ou não) de sua intervenção pedagógica do que comparar as médias e desvios padrão do grupo de alunos antes a após intervenção. Médias e Desvios Padrão reduzem os resultados do coletivo de estudantes a um valor único que tem reduzida utilidade na compreensão do fenômeno pedagógico. Outra vantagem prática é que as variáveis medidas em escala nominal dicotômica permitem que o

---

<sup>3</sup> Talvez seja importante esclarecer uma confusão conceitual recorrente na comunidade científica. Trata-se da asserção de que as pesquisas que limitam-se a estatística descritiva não possam generalizar seus resultados para além do grupo investigado. É preciso estarmos atentos para dois conceitos distintos: generalizações estatísticas e, generalizações analíticas (Cf. YIN, 2015). No primeiro caso trata-se de inferências sobre uma população com base nos dados empíricos coletados de uma amostra. Trata-se de generalizar resultados. No segundo caso não se trata de generalizar resultados empíricos (dados) mas sim de generalizar hipóteses ou teorias que resultam da valoração do pesquisador sobre os resultados da pesquisa.

professor-pesquisador possa criar critérios a priori “*ad hoc*” (pontos de corte). Ou seja, permite que antes mesmo da realização do trabalho de campo, ainda durante a fase de planejamento ele possa anunciar os critérios que vai adotar para avaliar o sucesso ou insucesso de sua intervenção (por exemplo: quantos alunos devem ser beneficiados para considerar o sucesso (ou insucesso) da sua intervenção?). Tais critérios funcionam como parâmetros que subsidiam a subjetiva valoração do projeto a partir das expectativas dos resultados definidas operacionalmente a priori pelo professor pesquisador, procedimentos que ficam inviabilizados quando o julgamento da eficácia intervenção pedagógica ocorre exclusivamente pelo valor de “p”.

Sugere-se os seguintes indicadores<sup>4</sup> estatísticos para a descrição e avaliação do impacto de uma intervenção pedagógica com variáveis dependentes em escala nominal dicotômica através das medidas de dimensão do efeito; (1) Medidas de Prevalência (PP); (2) Medidas de Incidência (PI); (3) Medidas de Chance (PC) e; (4) Processo de Valoração (PV). Vamos discorrer sobre esses indicadores através do exemplo apresentado no quadro 2.

**Quadro 2.** Exemplo de pesquisa avaliativa somativa com medidas de tamanho de efeito em escala nominal dicotômica de grupo único.

Numa escola com 400 alunos, identificam-se 80 deles com distorção idade/ano escolar<sup>5</sup>. O serviço de orientação pedagógica propôs uma intervenção para a correção de fluxo escolar. Após a intervenção (com grupos de reforço escolar, laboratórios de aprendizagem, etc., com os alunos em distorção idade/ano escolar) o número absoluto de alunos com distorção idade/ano escolar (80) reduziu para 50 alunos. Portanto 30 alunos corrigiram seu fluxo escolar.

Os professores-pesquisadores com a intenção de socializar os resultados de sua experiência pretendem publicá-los em forma de artigo científico. Quais os indicadores estatísticos mais adequados para descrever os dados e medir a dimensão do efeito da intervenção pedagógica? Sugere-se quatro indicadores:

- 1) **Medida de prevalência (MP):** expressa a medida em que, num determinado momento, ocorre um evento de interesse em relação ao total de sujeitos de uma população. (MERCHÁN-HAMANN, TAUIL & COSTA 2000). É um valor descritivo. Em nosso exemplo: (1) a MP antes da intervenção pedagógica =  $(80 / 400) \times 100 = 20$ . Interpreta-se que 20% dos estudantes estão em distorção idade/ano escolar. (2) Supõe-se que após a intervenção pedagógica o número absoluto de alunos nesta condição (80) reduziu para 50 alunos. A MP após a intervenção passa a ser de  $(50 / 400) \times 100 = 12\%$ .
- 2) **Medida de incidência (MI):** Descreve a ocorrência de novos casos. Em nosso exemplo representa em percentagem os 30 alunos que corrigiram o fluxo escolar em relação a todos os 400 alunos da escola =  $(30 / 400) \times 100 = 7,5$ . Interpreta-se que 7,5% do total de (400) alunos corrigiram o fluxo escolar. Todavia, este resultado é pouco esclarecedor para a interpretação da medida da dimensão do efeito da intervenção pedagógica neste caso. Deve-se perceber que o universo de alunos com distorção idade/ano escolar é 80 (e não 400). Desses 80 alunos que participaram da intervenção pedagógica 30 corrigiram o fluxo escolar e 50 permaneceram com distorção idade/ano escolar. Sendo assim, a melhor alternativa é medir a razão de chance.

<sup>4</sup> Pode-se utilizar outros indicadores similares: a razão de prevalência (RR). A razão de Incidência (RI) e a razão de chance ou: “odds ratio (RC). Nestes casos não se multiplica os resultados da divisão por 100. Em nosso exemplo a RP antes da intervenção pedagógica seria = 0,2; a RP após intervenção = 0,07 e; a RC = 0,6.

<sup>5</sup> Se define a distorção idade/ano escolar quando há uma defasagem superior a dois anos entre a idade dos alunos e o ano que está cursando. As medidas para minimizar ou eliminar estas diferenças é definido como correção de fluxo escolar.

### Continuação do quadro 2.

- 3) Razão de chance (RC): é o coeficiente entre a chance de que algo ocorra e a chance de que algo não ocorra. No exemplo, entre os 80 alunos com distorção idade/série, após a intervenção pedagógica 30 deles regularizaram seu fluxo (o sucesso ocorreu) e 50 não corrigiram (o sucesso não ocorreu). Com tal, a RC é  $= (30 / 50) = 0,60$ . Pode-se concluir do resultado que a chance dos estudantes com distorção idade/ano escolar regularizaram seu fluxo escolar a partir da intervenção pedagógica é de 60% ( $0,60 \times 100$ ). Em outras palavras, a intervenção pedagógica obteve uma eficácia de 60% (medido pela RC). Ou ainda, podemos inferir que a razão de chance de cada estudante com distorção idade/ano escolar regularizar seu ciclo, ao participar deste programa de intervenção pedagógica, é de 6 alunos em 10, ou 3 em cinco.
- 4) Processo de valoração: é subjetivo. Se aproxima ao que os estatísticos denominam como probabilidade pessoal (SALSBURG, 2009). Neste exemplo sobre a distorção idade/ano escolar que critérios o pedagogo pode assumir para validar sua proposta de intervenção? O que é um bom projeto? Aquele que corrige o fluxo escolar 20%, 50%, 60% ou 80%? Ressalta-se que se trata do processo de valoração, que é subjetivo e deve ser explicitado durante a fase de planejamento da pesquisa pelo professor-pesquisador coerente com suas concepções teóricas e experiências profissionais<sup>6</sup>. Outra possibilidade para a avaliação seria criar um conjunto de critérios qualitativos em escala ordinal: Exemplo: <20% ineficaz; 21 a 40% Pouco eficaz; 41 a 60% Razoavelmente eficaz; 61 a 80% Eficaz e; >81% Muito Eficaz (baseado no modelo probabilístico pessoal de Patrick Suppes. (Cf. SALSBURG, 2009, p. 247-48). Todavia, na opinião dos autores deste ensaio, a melhor avaliação decorre dos critérios de valoração pessoal do pesquisador tendo como referência o contexto onde atua e realiza sua tarefa pedagógica. A probabilidade pessoal “ad hoc”.

### 2.1.2 Avaliação Somativa através de medidas de probabilidade estatística em escala nominal dicotômica de grupo único - Teste de McNemar

As avaliações de impacto por medidas de probabilidade estatística valem-se do valor de “p” (probabilidade) oriundo de um teste estatístico de significância da discrepância entre a comparação de resultados observados (medidos na pesquisa) e previstos (dados provenientes de uma distribuição normal teórica) (SALSBURG, 2009; THOMAS, NELSON e SILVERMAN, 2012; BARROS et al. 2012). O Teste de McNemar expressa a proporção das mudanças pré e pós intervenção. É um modelo de qui-quadrado para amostras repetidas (onde cada indivíduo é o seu próprio controle), medidas em escala nominal (ou ordinal). Assim como o  $X^2$ , o teste de McNemar é um teste de adequação de ajuste (SALSBERG, 2009), calcula um valor de  $X^2$  que é comparado aos valores críticos associados a um determinado nível de significância (no anexo 1 consta um corte da tabela de valores críticos de  $X^2$ ). O quadro 3 exemplifica a utilização do teste de McNemar.

<sup>6</sup> Evidentemente esta escolha também depende do tema que estamos tratando (por isso a denominação “ad hoc”). Talvez, caso estivéssemos avaliando um projeto para adolescentes dependentes de drogas, 20% de proporção de ocorrência de sucesso, embora seja uma percentagem relativamente baixa, poderia ser considerada pertinente.

**Quadro 3.** Exemplo de pesquisa avaliativa somativa com medidas de probabilidade estatística em grupo único -Teste de McNemar.

O professor de educação física, observando a alta prevalência de excesso de peso entre seus alunos planeja um projeto de intervenção em suas aulas. Promove duas estratégias: aumentar os níveis de atividade física em suas aulas e incrementar informações teóricas sobre hábitos de vida fisicamente ativo e hábitos da alimentação saudável. Sua intervenção será durante um semestre letivo, todavia o professor pretende publicar um artigo na Revista Nova Escola, sobre o resultado ou os efeitos de sua proposta pedagógica. Planeja um estudo com seus alunos (grupo único) em que vai medir o perfil nutricional, antes e após o programa de intervenção através do índice de Massa Corporal (IMC - variável de desfecho ou dependente). Utiliza um modelo de avaliação com medidas em escala nominal e dicotômica cujos critérios são: (1) alunos com excesso de peso (EP) e; (2) alunos eutróficos ou com peso “normal” (EU). Dos 35 alunos que compõe sua turma nas medidas de IMC antes da intervenção (pré-teste), 14 estavam com EP e 21 EU. Ao final do semestre o professor volta a medir seus alunos (pós-teste) e obtém os seguintes resultados: (1) entre os 14 alunos com EP no início 12 tornaram-se EU (12 melhoraram (+)) e dois mantiveram-se como EP (2 mantiveram (+-)); (2) Entre os 21 alunos EU 2 tornaram-se EP (2 pioraram (-)) e 19 permaneceram EU (19 mantiveram-se (+)).

Como avaliar através da probabilidade estatística (teste de McNemar) se seu programa de intervenção teve ou não efeito significativo sobre a incidência de EU na sua turma de estudantes?

Para visualizar a lógica do Teste de McNemar vamos constituir uma tabela de contingência da seguinte forma.

Pré-teste	Pós-teste		n
	EP (-)	EU (+)	
EP (-)	(A) 2 (--)	(C) 2 (+-)	4
EU (+)	(B) 12 (-+)	(D) 19 (++)	31
<b>n</b>	<b>4</b>	<b>31</b>	<b>35</b>

Na célula A o número de alunos que permaneceram EP [2 (--)]

Na célula B o número de alunos que migraram de EP para EU [12 (-+)]

Na célula C o número de alunos que migraram de EU para EP [2 (+-)]

Na célula D o número de alunos que permaneceram EU [(19++)]

A equação para cálculo do  $X^2$  utiliza os valores discrepantes na tabela de contingência (+ -) e (- +). Então tem-se a seguinte equação:

$$X^2 = [(A - D) - 1]^2 / A + D$$

$$X^2 = [12 - 2] - 1^2 / 12 + 2 = 81 / 14 = 5,78$$

Obteve-se um  $X^2 = 5,78$ . O próximo passo é comparar o valor do  $X^2$  aos valores críticos fornecidos por uma tabela de distribuição de  $X^2$ <sup>(7)</sup> (Anexo 1). O valor crítico fornecido pela tabela {para 1 grau de liberdade [número de colunas da tabela de contingência -1 multiplicado pelo número de linhas -1 = (2-1) (2-1) = 1] e nível de significância de 0,05} é 3,84. Como o valor calculado (5,78) é maior que o valor crítico (3,84) conclui-se que houve diferença estatisticamente significativa entre os resultados do pré e pós teste. Dessa forma, pode-se inferir que a proposta pedagógica do professor para de minimizar a ocorrência de excesso de peso foi estatisticamente eficaz a um nível de confiança de 95%.

Para aplicar o teste de McNemar há “softwares” livremente disponíveis na internet, como o QuickCalcs ([www.graphpad.co/scientific-software](http://www.graphpad.co/scientific-software))<sup>8</sup>.

<sup>7</sup> Tabelas de valor critic de  $X^2$  pode ser facilmente encontrada em qualquer livro básico de estatística ou na Internet.

<sup>8</sup> QuickCalcs ([www.graphpad.co/scientific-software](http://www.graphpad.co/scientific-software)). Acesso em 10 de maio de 2015.

## **2.2 Delineamentos para pesquisa avaliativa somativa com grupo de intervenção e grupo controle**

Neste modelo de pesquisa avaliativa somativa com abordagem quantitativa o professor-pesquisador dispõe de, pelo menos, duas turmas de alunos para planejar sua pesquisa. O professor defini aquela que será objeto de sua intervenção pedagógica e, compara os resultados com a outra. É um delineamento do tipo quase-experimental (Cf. CAMPBELL & STANLEY, 2001). O desenho corresponde a intervenção pedagógica em apenas um grupo e a comparação dos resultados entre os dois grupos no pré e pós teste.

Todavia, deve-se considerar que as duas turmas que participam da experiência não são organizadas por procedimentos aleatórios, portanto não se pode assegurar a priori que as turmas sejam equivalentes. Que os alunos ou parte deles não detenham algumas características relevantes que possam intervir (variáveis intervenientes) sobre os resultados. Sendo assim, para minimizar tais inconvenientes sugere-se que as medidas de pré-teste sejam utilizadas para proceder-se ao ajustamento dos grupos por técnica de pareamento (GAYA, 2008) como será demonstrado a seguir. Ou seja, o pareamento é uma alternativa para diminuir as prováveis diferenças naturais existentes entre os indivíduos, o professor-pesquisador organiza os alunos em pares, em que cada aluno do grupo que será submetido a intervenção pedagógica tenha seu próprio controle entre os alunos do grupo que não será submetido a intervenção pedagógica. As duplas devem ser o mais semelhante possível quanto às características que possam influenciar os resultados.

### **2.2.1 Ajustamento por pareamento**

Sugere-se discorrer sobre o ajustamento por pareamento através do exemplo que segue no quadro 4.



**Quadro 4.** Exemplo de pesquisa com ajustamento de grupos não equivalentes por pareamento.

EXEMPLO: Professor de língua estrangeira pretende investigar os efeitos de um novo programa de ensino sobre a aprendizagem da língua inglesa em adolescente da sua escola. O professor pretende aplicar na turma “A” o novo programa de ensino e, na turma “B” o ensino tradicional. Ao final do semestre vai comparar os resultados de proficiência em língua inglesa dos alunos da turma “A” com os alunos da turma “B”.

Considerando, que a composição das turmas não obedece a um critério aleatório, o professor supõe que as turmas não sejam equivalentes em algumas das variáveis intervenientes relevantes. Ele resolve então utilizar os escores de proficiência dos alunos no pré-teste para formar duplas homogêneas. No quadro abaixo estão as notas dos alunos (de zero a 10) e os critérios de avaliação: Não proficientes (NP) notas inferiores a 6 e Proficientes (P) notas iguais ou superiores a 6).

**Notas dos alunos e avaliação das turmas “A” e “B” no pré-teste (dados fictícios)**

Turma A:			Turma A:			Turma B:			Turma B:		
Notas dos alunos no Pré-teste			Notas dos alunos no Pré-teste			Notas dos alunos no Pré-teste			Notas dos alunos no Pré-teste		
1	7,5	(P)	11	3,5	(NP)	1	4,0	(NP)	11	4,5	(NP)
2	2,5	(NP)	12	6,0	(P)	2	1,5	(NP)	12	4,0	(NP)
3	4,0	(NP)	13	2,0	(NP)	3	2,0	(NP)	13	5,0	(NP)
4	6,5	(P)	14	1,5	(NP)	4	5,5	(NP)	14	3,0	(NP)
5	5,5	(NP)	15	4,5	(NP)	5	6,0	(P)	15	1,5	(NP)
6	5,0	(NP)	16	3,0	(NP)	6	6,5	(P)	16	4,5	(NP)
7	3,0	(NP)	17	4,5	(NP)	7	1,0	(NP)	17	6,5	(P)
8	3,5	(NP)	18	5,5	(NP)	8	3,0	(NP)	18	2,5	(NP)
9	5,0	(NP)	19	2,0	(NP)	9	3,5	(NP)	19	5,0	(NP)
10	0,5	(NP)	20	5,0	(NP)	10	0,0		20	8,0	(P)

**Continuação do quadro 4.**

Considerando as notas o professor pareou os seus alunos das turmas “A” e “B” da seguinte forma:

**Pares de alunos da Turma “A” e “B” conforme as notas do pré-teste**

Pares de alunos	Alunos da Turma “A”		Alunos da Turma “B”	
1	10	0,5	10	0,0
2	14	1,5	2	1,5
3	13	2,0	3	2,0
4	19	2,0	15	1,5
5	2	2,5	18	2,5
6	7	3,0	8	3,0
7	16	3,0	14	3,0
8	8	3,5	9	3,5
9	11	3,5	1	4,0
10	2	4,0	12	4,0
11	15	4,5	11	4,5
12	17	4,5	16	4,5
13	6	5,0	13	5,0
14	9	5,0	19	5,0
15	20	5,0	-	
16	5	5,5	4	5,5
17	19	5,5	-	
18	12	6,0	5	6,0
19	4	6,5	17	6,5
20	1	7,5	20	8,0

Este pareamento manifesto na tabela acima, sugere que os grupos são equivalentes quanto aos resultados do pré-teste. Todavia, como se pode observar, não foi possível proceder ao pareamento de dois alunos. Os alunos 19 e 20 da turma “A”, não tiveram colegas com notas semelhantes na turma “B”. Esta é uma limitação do método, e desta forma o professor vai selecionar para sua experiência os resultados de 18 alunos (do total de 20 em cada turma) para avaliar o impacto de sua intervenção pedagógica. Assim, através do ajustamento por pareamento pretende-se que estejam satisfeitas as exigências de que os grupos sejam equivalentes antes da intervenção pedagógica e, como tal, se possa adotar técnicas estatísticas para comparação entre grupos. (Atenção: evidentemente, isto não significa na prática que o professor deva retirar os alunos 19 e 20 de suas aulas. Apenas, seus resultados não serão computados no cálculo dos indicadores estatísticos dos grupos).

Após a sugestão de ajustamento dos grupos por pareamento, para avaliação de intervenções pedagógicas em desenhos de grupos intervenção e controle com variáveis dicotômicas em escala nominal sugere-se: (1) testes de medidas de efeito e; (2) o teste do Qui-

quadrado ( $X^2$ ) com correção de Yates quando os sujeitos da pesquisa forem menos de 20 ou quando numa das células da tabela de contingência houver menos de 5 alunos.

Para discorrer sobre estas estratégias de avaliação do impacto da intervenção pedagógica, no quadro 5 dá-se seguimento ao exemplo anterior.

**Quadro 5.** Notas e avaliações dos alunos das turmas “A” e “B” no Pós-teste (dados fictícios).

Supõe-se que na tabela seguinte estão os resultados das duas turmas de língua inglesa após a intervenção pedagógica com o método inovador no grupo “A” e o modelo tradicional de ensino no grupo “B”.

**Notas e avaliação dos alunos das turmas “A” e “B” no pós-teste (dados fictícios)**

Turma A: Notas dos alunos no Pós-teste			Turma A: Notas dos alunos no Pós-teste			Turma B: Notas dos alunos no Pós-teste			Turma B: Notas dos alunos no Pós-teste		
1	7,5	(P)	11	5,5	(NP)	1	6,0	(NP)	11	4,0	(NP)
2	7,5	(P)	12	8,5	(P)	2	6,5	(P)	12	4,0	(NP)
3	6,0	(P)	13	6,0	(P)	3	5,0	(NP)	13	5,0	(NP)
4	8,5	(P)	14	3,5	(NP)	4	5,5	(NP)	14	5,0	(NP)
5	7,5	(P)	15	6,5	(P)	5	7,0	(P)	15	3,5	(NP)
6	7,0	(P)	16	5,0	(NP)	6	6,5	(P)	16	4,5	(NP)
7	6,0	(P)	17	6,5	(P)	7	5,0	(NP)	17	6,0	(P)
8	5,5	(NP)	18	6,5	(P)	8	7,5	(P)	18	3,5	(NP)
9	7,0	(P)	19	6,0	(NP)	9	4,0	(NP)	19	6,0	(P)
10	4,5	(NP)	20	6,0	(P)	10	6,5	(P)	20	8,0	(P)

### 2.2.2 Avaliação Somativa através de Medidas de Dimensão do Efeito com escalas nominais dicotômicas

Para que se possa interpretar com clareza a medida da dimensão do efeito sugere-se os seguintes indicadores: (a) medidas de prevalência e incidência e (b) medidas de eficácia (razão de chance e razão de chance relativa) e; (c) processo de valoração. Retorna-se ao exemplo:

**Quadro 6.** Avaliação do impacto por medidas de dimensão do efeito.**Medidas de prevalência**

**A prevalência de alunos proficientes em inglês antes das intervenções pedagógicas (pré-teste)** na turma “A” (primeira e segunda linhas da primeira tabela do quadro 4) é  $= 3/18 = 0,16$ . Interpreta-se que 16% ( $0,16 \times 100$ ) dos alunos são proficientes antes das aulas.

O mesmo resultado de prevalência é encontrado (no pré-teste) com os alunos da turma “B” (terceira e quarta linhas da primeira tabela do quadro 4)  $3/18 = 0,16$  ou 16%.

**A prevalência de alunos proficientes em inglês após as intervenções pedagógicas (pós-teste)** com o método inovador (Turma “A”) é  $= 14/18 = 0,77$ . Interpreta-se que 77% dos alunos estão proficientes em inglês após as aulas com o método inovador.

**A Prevalência de alunos proficientes em inglês após as aulas (pós-teste) com o método tradicional** (Turma “B”) é  $= 8/18 = 0,44$ . Interpreta-se que 44% dos alunos estão proficientes em inglês após as aulas tradicionais.

**Medidas de incidência**

Ao observar as medidas de prevalência, se subtrairmos do valor de prevalência no pós-teste o valor de prevalência do pré-teste obtém-se: (1) na Turma “A” ( $77\% - 16\% = 61\%$ ). Interpreta-se que houve uma proporção de incidência (novos casos de alunos proficientes) de 61% com o método inovador. Na Turma “B” ( $44\% - 16\% = 28\%$ ). Interpreta-se que houve uma proporção de incidência de novos alunos com proficiência em inglês de 28% com o método tradicional.

**A Razão de prevalência**

É a razão entre os valores de incidência nas duas turmas, considerando o total de alunos em cada turma (18). A razão de prevalência é  $= 61/28 = 2,17$ . Interpreta-se que os alunos da turma “A” com o método inovador têm 2,17 vezes mais chance de serem proficientes em inglês do que os alunos da Turma “B” com o método tradicional.

**Medidas de razão de chance (eficácia)**

Considere-se que as medidas de razão de chance constituem-se pela razão entre a ocorrência de alunos proficiente e não proficiente em cada análise. Nas duas turmas os resultados são semelhantes. São 3 alunos proficientes e 15 não proficientes. Portanto a chance dos alunos serem proficientes antes das intervenções pedagógicas é  $= 3/15 = 0,20$ . Interpreta-se que 0,20 é a chance dos alunos serem proficientes antes das práticas pedagógicas.

Na Turma “A”, com o método inovador a razão de chance é  $= 14/4 = 3,5$ . Interpreta-se que a chance de um aluno tornar-se proficiente em inglês com o método inovador aumenta em 3,5 vezes

Na Turma “B” com o método tradicional a razão de chance é  $= 8/10 = 0,80$ . Interpreta-se que a chance de um aluno tornar-se proficiente em inglês com o método tradicional aumenta em 0,80 vezes.

Conclui-se, com esses dados que os dois métodos de ensino tem bom nível de impacto sobre a aprendizagem do inglês. Todavia, a hipótese de trabalho do professor-pesquisador é que o método inovador é mais efetivo do que o método tradicional. Daí a relevância de medir a razão de chance relativa.

**Medidas de razão de chance relativa**

A razão de chance relativa é a razão entre a razão de chance entre os valores obtidos com o método inovador (Turma “A”) e o método tradicional (Turma “B”). Portanto:  $3,5/0,80 = 4,37$ . Interpreta-se que a chance de um aluno tornar-se proficiente pelo método inovador aumenta em 4,37 vezes em relação ao método tradicional.

**Processo de valoração**

O que o professor-pesquisador considera como um impacto satisfatório sobre a aprendizagem de inglês? Quais são os seus critérios para avaliação do impacto dos modelos de ensino? Supõe-se que ele adote os critérios de avaliação sugeridos por Cohem (1988) para estudos de associação: tamanho do efeito a até 0,10 pequeno efeito (Insatisfatório); 0,10 a 0,30 médio efeito (Satisfatório); superior a 0,30 grande efeito (Bom). Outra possibilidade é o uso do modelo probabilístico pessoal de Patrick Suppes. (Cf. SALSBERG, 2009, p. 247-48) onde: < 20% ineficaz; 21 a 40% Pouco eficaz; 41 a 60% Razoavelmente eficaz; 61 a 80% Eficaz e; > 81% Muito Eficaz. De qualquer forma, neste caso, os resultados da experiência sugerem: (1) que os dois modelos de ensino são eficazes. O modelo tradicional é Eficaz e o modelo inovador é Muito Eficaz; (2) Não obstante, confirma-se a hipótese do professor-pesquisador, o modelo inovador (turmas “A”) é bem mais eficaz que o modelo tradicional (Turma “B”). A chance de sucesso na aprendizagem em inglês com o modelo inovador é de 4,37 vezes em relação ao modelo tradicional.

### 2.2.3 Avaliação Somativa através de medidas de probabilidade estatística em escala nominal dicotômica de grupos não equivalentes - Teste do Qui-Quadrado com correção de Yates

O Qui-quadrado, representado por  $X^2$  é um teste de adequação de ajuste criados por Karl Pearson em 1900 (SALSBERG, 2009). Calcula um valor de  $X^2$  que é comparado aos valores críticos associados a um determinado nível de significância. O princípio básico deste(s) método(s) é comparar proporções, isto é, as possíveis divergências entre as frequências observadas e esperadas para um certo evento. Pode-se dizer que dois grupos se comportam de forma semelhante se as diferenças entre as frequências observadas e as esperadas em cada categoria forem muito pequenas, próximas a zero. O quadro 7 exemplifica a utilização do teste do  $X^2$  com correção de Yates<sup>9</sup>.

#### Quadro 7. Exemplo de pesquisa avaliativa somativa com medidas de probabilidade estatística em dois grupos – O Teste do $X^2$ com correção de Yates (Cf. CLEGG 1995)<sup>10</sup>.

Numa escola de educação básica o professor de educação física está preocupado com a atitude dos estudantes perante ao ensino de sua disciplina. Reconhece que a motivação entre os rapazes é superior a motivação entre as moças. Portanto, está interessado em investigar se o gênero influencia na opinião dos alunos sobre a educação física. Entrevista 350 alunos. São 200 rapazes e 150 moças. Os resultados são os seguintes: 140 rapazes afirmam gostar das aulas de educação física e 60 discordam. Entre as moças 60 dizem que gostam de participar das aulas e 90 discordam. Para calcular  $X^2$  há inúmeros programas de estatística disponíveis. Não obstante, neste ensaio, apresenta-se os procedimentos passo à passo de forma a permitir a compreensão lógica do modelo.

O primeiro passo para aplicação do  $X^2$  é criar uma tabela de contingência 2x2 a partir das frequências observadas.

Tabela de contingência com as frequências observadas

	Gostam de EF	Não gostam de EF	Total
Rapazes	135 (A)	65 (B)	200
Moças	60 (C)	90 (D)	150
Total	195	155	350

Os valores das células representam as frequência observadas.

O segundo passo é calcular as frequências esperadas: (1) multiplica-se o total da linha 1 pelo total da coluna 1 e divide-se pelo total (350) e tem-se a frequência esperada da célula (A):  $200 \times 195 / 350 = 111,5$ ; (2) multiplica-se o total da linha 1 com o total da coluna 2 e divide-se pelo total e tem-se a frequência esperada da célula (B):  $200 \times 155 / 350 = 88,5$ ; (3) multiplica-se o total da linha 2 com o total da coluna 1 e divide-se pelo total e tem-se a frequência esperada da célula (C):  $150 \times 195 / 350 = 83,5$ ; (4) multiplica-se o total da linha 2 com o total da coluna 2 e divide-se pelo total e tem-se a frequência esperada da célula (D):  $150 \times 155 / 350 = 66,4$ .

<sup>9</sup> Sempre que uma amostra for inferior a 20 ou, que numa das células da tabela de contingência do  $X^2$  ocorram valores inferiores a 5 recomenda-se a Correção de Yates. No exemplo do quadro 6, este procedimento está incorporado no terceiro passo representado pela subtração de 0,5 na frequência de cada célula.

<sup>10</sup> CLEGG, F. Estatística para Todos. Tradução de Catarina Horta. Lisboa: Gradiva, 1995.

**Continuação do quadro 6.**

Tabela de contingência com as frequências esperadas			
	<b>Gostam de EF</b>	<b>Não gostam de EF</b>	<b>Total</b>
Rapazes	111,5 (A)	88,5 (B)	<b>200</b>
Moças	83,5 (C)	66,5 (D)	<b>150</b>
<b>Total</b>	<b>195</b>	<b>155</b>	<b>350</b>

O terceiro passo é calcular a diferença entre as frequências esperadas e observadas (sempre diminuindo o menor valor do maior valor) para cada célula e ainda subtraindo 0,5 (Correção de Yates).

Célula (A) =  $135 - 111,5 - 0,5 = 23,1$ ; Célula (B) =  $88,5 - 65 - 0,5 = 20$ ; Célula (C)  $83,5 - 60 - 0,5 = 23$ ; Célula (D) =  $90 - 66,5 - 0,5 = 24$ .

O quarto passo é aplicar a fórmula  $X^2 = \sum[(O - E)^2/E]$ . Elevar ao quadrado os valores de cada célula obtidas no terceiro passo e dividir o resultado pela frequência esperada para cada célula específica: Célula (A) =  $(23,1)^2 / 111,4 = 4,7$ ; Célula (B) =  $(20)^2 / 111,4 = 3,5$ ; Célula (C) =  $(23)^2 / 83,5 = 6,3$ ; Célula (D) =  $(24)^2 / 66,4 = 8,6$ .

O quinto passo é calcular o valor da soma dos quatro valores obtidos no passo 4:  $4,7 + 3,5 + 6,3 + 8,6 = 23,1$ .

O sexto passo é calcular os grau de liberdade (gl)  $gl = (r - 1)(c - 1)$ , sendo r= o número de linhas da tabela de contingência e c= ao número de colunas:  $gl = (2 - 1)(2 - 1) = 1$

Por fim, o último passo é comparar o valor do  $X^2$  com o valor encontrado na tabela de valores crítico de  $X^2$ , observando o valor de gl e o nível de significância (anexo 1).

Observa-se na tabela de valores críticos do  $X^2$  que o valor do  $X^2 = 23,1$ , para 1gl é bem superior aos valores críticos da tabela. Inclusive é superior ao valor crítico ao nível de significância de 0,001 (10,83). Portanto, conclui-se que há diferença estatisticamente significativa entre os rapazes e as moças. Os rapazes realmente gostam mais de fazer as aulas de educação física do que as moças. O resultado tem um nível de confiança  $(1 - 0,001) = 0,999$ .

**CONCLUSÃO**

Este ensaio é fruto da experiência de seus autores na docência de metodologia da pesquisa em cursos de formação de professores e da vivência em escolas de educação básica. Originou-se da constatação sobre as dificuldades encontradas pelos estagiários e professores no trato da investigação científica no âmbito das práticas pedagógicas. Repetidas vezes quando tais estagiários em seus trabalhos de conclusão de curso planejavam pesquisas de abordagem quantitativa no ambiente escolar deparavam-se com os modelos hegemônicos cujas exigências de rigor inviabilizavam sua aplicação. Os pressupostos de Fischer para as pesquisas probabilísticas constituem-se em obstáculos praticamente intransponíveis aos modelos de pesquisa que se poderia realizar na escola. Daí a necessidade de alternativas capazes de viabilizar pesquisas científicas que pudessem auxiliar os estagiários e professores em seu ambiente de trabalho a avaliar o impacto de suas práticas pedagógicas. A intenção dos autores neste ensaio foi a de sugerir alternativas que pudessem ser facilmente utilizadas sem descuidar das exigências de rigor inerentes ao conhecimento científico. Deu-se ênfase a

Pesquisa Avaliativa Somativa e a desenhos metodológicos simplificados, com medidas em escalas nominais dicotômicas. Dois métodos foram apresentados: (1) grupo único com pré e pós-teste e, (2) e dois grupos (intervenção e controle) não equivalentes. Para análise dos dados sugeriu-se: (1) medidas de dimensão do efeito: medidas de prevalência, incidência, razão de chance e razão relativa de chance e; (2) medidas do valor de “p”: Teste de McNemar e  $X^2$  com correção de Yates. Pretende-se com este estudo incentivar estudantes de cursos de licenciatura e professores a valerem-se do método científico como ferramenta para a avaliação do impacto de suas intervenções pedagógicas e, principalmente incentivar a publicação de trabalhos científicos no âmbito das pesquisas pedagógicas no ambiente escolar.

### REFERÊNCIAS

- ARNAL, J.; DEL RINCON, D, e LATORRE, A. Investigación educativa: fundamentos y metodología. Barcelona: Labor, 1992
- BARROS, M.; REIS, R; HALLAL, P.; FLORINDO, A.; FARIAS JÚNIOR, J. Análise de dados em saúde. 3. ed. Londrina: Midiograf, 2012.
- CLEGG, F. Estatística para todos. Lisboa: Gradiva, 1995.
- COHEM, J. Statistics power analysis for the behavioral sciences. 2. ed. New York: Academic Press, 1988.
- CONTANDRIOPOULO, A. P.; CHAMPAGNE, F.; DENIS, J.L. e PINEAULT, R. In. HARTZ, Z. M. A, Avaliação em Saúde: Dos modelos conceituais à prática na análise de implementação de programas. Rio de Janeiro: Osvaldo Cruz, 1987.
- DE LA ORDEM, A. Diccionario de ciências de la educación. Investigación educativa. Madrid: Anaya, 1985.
- GAYA, A. Ciências do Movimento Humano. Introdução à metodologia da pesquisa. Porto Alegre: Armed, 2008.
- KIDDER, L. H. (org). Métodos de Pesquisa nas relações sociais/ Seltz, Wrightman e Cook. 2.ed. São Paulo: USP, 1987.
- MERCHAN-HAMMAN, E.; TAUIL, P. L.; COSTA, M. P. Terminologia das medidas e indicadores em epidemiologia: subsídios para uma possível padronização da nomenclatura. Inf. Epidemiol. Sus v. 9, n. 4, Brasília dez. 2000.
- SALSBURG, D. Uma senhora toma chá... como a estatística revolucionou a ciência no século XX. Rio der Janeiro: Lorge Zahar, 2009.
- SELTZ, WRIGHTSMAN e COOK. Métodos de pesquisa nas relações sociais. São Paulo: EPU, 1987.
- TELLES, J. A. “É pesquisa? Ah, não quero, não, bem!”. Sobre pesquisa acadêmica e sua relação com a prática do professor de línguas. Linguagem e Ensino, v. 5, n. 2, p. 91-116, 2002.

THOMAS, J. R.; NELSON, J. K e SILVERMAN, S. J. Métodos de pesquisa em atividade física. Porto Alegre: ARTMED, 2012.

YIN, R. K. Estudo de Caso. Planejamento e métodos. 5. ed. Porto Alegre: Bookman, 2015.

**ANEXO.** Valores críticos de  $X^2$

gl	Nível de significância			
	0,10	0,05	0,01	0,001
1	2,706	3,841	6,635	10,83
2	4,605	5,991	9,210	13,82
3	6,251	7,815	11,34	16,27
4	7,779	9,488	13,28	18,47
5	9,236	11,07	15,09	20,52
6	10,64	12,59	16,81	22,46
7	12,02	14,07	18,48	24,32
8	13,36	15,51	20,09	26,12
9	14,68	16,92	21,67	27,88
10	15,99	18,31	23,21	29,59