

*Subsídios Metodológicos para o Professor Pesquisador em
Ensino de Ciências*

*Pesquisa em Ensino:
Métodos Qualitativos e Quantitativos*



*Marco A. Moreira
Paulo R. S. Rosa*

*Porto Alegre, Brasil
2009/2016*

Ficha de apresentação

Tipo de publicação: Recopilação de trabalhos publicados ou apresentados em congressos sobre o tema *Métodos Qualitativos e Quantitativos* a fim de subsidiar metodologicamente o professor investigador, em particular da área de ensino de ciências.

Autores: *Marco A. Moreira* – Instituto de Física, UFRGS, Brasil
Paulo R. Rosa – Instituto de Física, UFMS, Brasil

Data e local: 2009 (1ª edição), 2016 (2ª edição revisada) Porto Alegre, Brasil.

Outras publicações da mesma série

- **Subsídios Teóricos:** *Comportamentalismo, Construtivismo e Humanismo.*
- **Subsídios Teóricos:** *A Teoria da Aprendizagem Significativa.*
- **Subsídios Epistemológicos:** *Epistemologias do Século XX.*
- **Subsídios Metodológicos:** *Pesquisa em Ensino: Aspectos Metodológicos.*
- **Subsídios Didáticos:** *Mapas conceituais, Diagramas V e Organizadores Prévios.*

Sumário

Apresentação.....4

Capítulo 1

Pesquisa em ensino: métodos qualitativos5

Capítulo 2

Pesquisa em ensino: métodos quantitativos.....32

Apresentação

Este material de apoio está constituído por dois textos sobre metodologias de pesquisa em ensino, particularmente em ensino de ciências. Foram escritos para serem usados independentemente um do outro. Consequentemente, podem apresentar certas superposições.

O primeiro deles começa com uma breve comparação entre os paradigmas quantitativo e qualitativo, depois concentra-se no enfoque qualitativo e ao final aborda o tema da triangulação metodológica. O segundo tenta resgatar a metodologia quantitativa aplicada à pesquisa.

Cabe chamar atenção que não são mais do que textos de apoio, ou textos introdutórios, que buscam oferecer subsídios metodológicos para professores que queiram fazer pesquisa em ensino a fim de melhorar sua prática docente.

Porto Alegre, 2016

Marco Antonio Moreira
Paulo Ricardo S. Rosa

Capítulo 1

Pesquisa em Educação em Ciências: Métodos Qualitativos¹

M. A. Moreira

Resumo

Após uma breve comparação entre os paradigmas quantitativo e qualitativo, o texto está dedicado apenas ao enfoque qualitativo descrevendo, com mapas conceituais, a etnografia, o estudo de casos e a investigação-ação. Ao final, é abordada a questão da triangulação metodológica.

Introdução

Neste texto, a pesquisa em educação em ciências está entendida como a produção de conhecimentos resultante da busca de respostas a perguntas sobre ensino, aprendizagem, currículo e contexto educativo em ciências, assim como sobre o professorado de ciências e sua formação permanente, dentro de um quadro epistemológico, teórico e metodológico consistente e coerente. Porém, o mesmo se ocupará só do domínio metodológico dessa pesquisa e, nesse domínio, o foco será particularmente a *metodologia qualitativa*.

A metodologia da pesquisa em educação em ciências é a mesma da pesquisa em educação e esta foi dominada, ao longo do século XX, por dois paradigmas clássicos: um inspirado na metodologia das ciências naturais, enfatizando observações empíricas quantificáveis e adequadas para tratamentos estatísticos; o outro derivado da área humanística com ênfase em informações holísticas e qualitativas e em abordagens interpretativas.

O filósofo alemão Wilhelm Dilthey argumentava (apud Husén, 1988) já em 1890 que as humanidades tinham sua própria lógica de pesquisa e que a diferença entre as ciências naturais e as humanidades era que estas buscavam compreender enquanto que as primeiras procuravam explicar (op. cit., p. 17). Essa distinção parece hoje muito simplificada, mas serve para mostrar que o debate é antigo.

A pesquisa em educação começa (segundo Landsheere, 1988), por volta de 1900, sob o nome de "pedagogia experimental", com pesquisadores como Meumann na Alemanha, Binet na França, Thorndike nos Estados Unidos e Claparède na Suíça, pouco tempo depois da "psicologia experimental" iniciada por Wundt em Leipzig, por volta de 1880, e fortemente influenciada por ela. De acordo com esse mesmo autor (op. cit. p. 11), nas três primeiras décadas do século passado, a pesquisa educativa teve uma acentuada ênfase quantitativa, dirigida ao estudo da eficácia no ensino, particularmente nos Estados Unidos.

¹ Programa Internacional de Doctorado en Enseñanza de las Ciencias. Universidad de Burgos, Espanha; Universidade Federal do Rio Grande do Sul, Brasil. *Texto de Apoio n° 14*. Publicado em *Actas del PIDEAC*, Vol. 4:25-55, 2002.

Posteriormente, nos anos 30 a 50, a crise econômica e a guerra levaram a uma grande redução na atividade de pesquisa em educação, em especial na Europa. Porém, nessa mesma época aparecem como campo de interesse dos pesquisadores os estudos de natureza sociológica questionando a escola como mecanismo de reprodução de distinções sociais e práticas discriminatórias (ibid., p. 13). Nas décadas de 60 e 70 outra vez houve um período de muito apoio financeiro à pesquisa educativa, particularmente aquela voltada ao desenvolvimento curricular em ciências e matemática. Foi também uma época de predomínio da abordagem quantitativa, porém a reação a essa "tradição positivista" começava a ser cada vez mais forte no contexto da pesquisa educativa em nível internacional. A tal ponto que nos anos 80 e 90 houve um claro predomínio da abordagem qualitativa na pesquisa em educação em geral e em ciências em particular.

Esta pequena e pouco rigorosa resenha histórica foi feita somente para reforçar a asserção de que os dois paradigmas clássicos – o quantitativo e o qualitativo – dominaram a pesquisa educativa no século XX, com uma certa alternância.

Desde o ponto de vista epistemológico, em todo esse tempo surgiram teses de incompatibilidade paradigmática kuhneana (Smith, 1983; Smith e Heshusius, 1986; Marshal, 1986), de compatibilidade prática, funcional, pragmática (e.g., Shulman, 1981; Miles e Huberman, 1984), de conciliação e triangulação metodológicas (e.g., Eisner, 1981; Firestone, 1987) ou integradoras como a de Keeves (1988) e a de Bericat (1998). Essas colocações estão discutidas em outro texto anterior e complementar a este (Moreira, 2000).

Neste texto ficaremos na perspectiva integradora, mas este tema será deixado para o final. Por enquanto continuaremos na distinção entre os dois paradigmas. Mesmo que esta distinção e inclusive a idéia de paradigma possam ser objeto de críticas por parte de pesquisadores (e.g., Walker e Evers, 1988) continuaremos nela, por algum tempo, por razões didáticas. Por essas razões, no item seguinte distinguiremos os dois paradigmas clássicos como se constituíssem uma dicotomia. Uma vez estabelecida essa distinção, o texto abordará sucessivamente metodologias qualitativas como a etnografia, o estudo de casos, a pesquisa ação e outras. Na conclusão abandonaremos a visão dicotômica e defenderemos uma postura integradora.

Os dois paradigmas clássicos

Na tabela 1 se estabelece uma comparação dicotômica entre os paradigmas quantitativo e qualitativo em termos de pressupostos, objetivos, métodos, papel do pesquisador e retórica de apresentação do conhecimento produzido. Por ser auto-explicativa tal tabela, não será comentada.

A pesquisa qualitativa (interpretativa)

Na figura 1 se apresenta um mapa conceitual para a abordagem qualitativa na pesquisa educativa. Já que os mapas conceituais não são auto-explicativos, faremos um breve comentário. No topo aparece o conceito de *pesquisa qualitativa* ao qual estão associados atributos como *interpretativa*, *holística*, *naturalista*, *participativa*, *interacionista simbólica*, *construtivista*, *etnográfica*, *fenomenológica* e *antropológica*.

O interesse central dessa pesquisa está em uma *interpretação dos significados* atribuídos pelos sujeitos a suas *ações* em uma *realidade socialmente construída*, através de

observação participativa, isto é, o pesquisador fica *imerso* no fenômeno de interesse. Os *dados* obtidos por meio dessa participação ativa são de *natureza qualitativa* e analisados correspondentemente. As *hipóteses* são *geradas* durante o processo investigativo. O pesquisador busca *universais concretos* alcançados através do estudo profundo de *casos particulares* e da comparação desse caso com outros estudados também com grande profundidade. Através de uma *narrativa detalhada*, o pesquisador busca *credibilidade* para seus modelos interpretativos.

Tabela 1. Um paralelo entre o paradigma quantitativo e qualitativo na pesquisa educativa (M.A. Moreira, 2000)

	PARADIGMA QUANTITATIVO REALISTA/RACIONALISTA	PARADIGMA QUALITATIVO IDEALISTA/NATURALISTA
<i>Pressupostos</i>	Realidade objetiva, independente de crenças, com existência própria. O pesquisar não afeta ao que se está pesquisando. Os instrumentos são uma maneira de alcançar medições precisas de objetos e eventos com existência própria; instrumentos válidos são os que produzem representações exatas da realidade. Se o pesquisador deixa de estudar algo, esse algo continuará existindo e permanecerá ligado a outras coisas da mesma maneira. Dualismo sujeito-objeto. Verdade é uma questão de correspondência com a realidade (Smith, 83).	Realidade socialmente construída; não há realidade independente dos esforços mentais de criar e moldar; o que existe depende da mente humana. O que se pesquisa não é independente do processo de pesquisa. Os instrumentos não têm lugar independentemente de aquilo que têm que medir, são extensões dos pesquisadores na sua tentativa de construir ou de dar forma à realidade. A realidade não tem existência prévia à pesquisa e deixará de existir se a pesquisa for abandonada. Não há dualismo sujeito-objeto. Verdade é questão de concordância em um contexto (Smith, 83).
<i>Objetivos</i>	Procuram explicar causas de mudanças em fatos sociais, principalmente através de medição objetiva e análise quantitativa (Firestone, 87). Focam comportamentos de grupos ou indivíduos (Eisner, 81). Buscam a predição e controle de eventos, algoritmos, verdades, universais abstratos aos que se chega através de generalizações estatísticas de amostras para populações (Erickson, 86).	Procuram a compreensão do fenômeno social segundo a perspectiva dos atores através de participação em suas vidas (Firestone, 87). Focam significados e experiências; ações em vez de comportamentos (Eisner, 81). Procuram a explicação interpretativa; heurísticas em vez de algoritmos; universais concretos alcançados através do estudo detalhado de um caso e da comparação com outros estudados com igual detalhe (Erickson, 86).
<i>Métodos</i>	Tomam emprestado o modelo das ciências físicas para pesquisar o mundo social e humano. Ocupam-se de desenhos experimentais, quase-experimentais e correlacionais; testes de hipóteses; instrumentos válidos e fidedignos; testes de significância; amostragem; inferência estatística; generalização. Seguem um modelo hipotético-dedutivo.	Usam técnicas etnográficas, estudos de caso, antropologia educativa. Ocupam-se de observação participativa; significados individuais e contextuais; interpretação; desenvolvimento de hipóteses; indicadores de baixa inferência; casos, grupos ou indivíduos específicos; particularização. Podem fazer uso de estatística descritiva. São bem mais indutivos.
<i>Papel do pesquisador</i>	Distante para evitar viés (Firestone, 87); objetivo. Limita-se ao que é. Quantifica registros de eventos. Faz uso de meios científicos. Busca fiabilidade e validade.	Imerso no fenômeno de interesse (Firestone, 87); participante. Faz anotações, ouve, observa, registra, documenta, busca significados, interpreta. Procura credibilidade.
<i>Retórica</i>	Padronizada, estatística, objetiva. Extenso uso de tabelas, gráficos, coeficientes. Procura neutralizar a personalidade do pesquisador. Fria, científica, buscando convencer o leitor de que a análise feita é neutra impessoal (Firestone, 87).	Persuasiva, descritiva, detalhada. Extenso uso de transcrições, vinhetas, documentos, exemplos, comentários interpretativos. Usa a linguagem cotidiana com suficiente detalhe para evidenciar que são válidas as interpretações dos significados tidos pelos atores (Erickson, 86).

A pesquisa qualitativa é chamada também *naturalista* porque não envolve manipulação de variáveis, nem tratamento experimental (é o estudo do fenômeno em seu acontecer natural); *fenomenológica* porque enfatiza os aspectos subjetivos do comportamento humano, o mundo do sujeito, suas experiências cotidianas, suas interações sociais e os significados que dá a essas experiências e interações; *interacionista simbólica* porque toma como pressuposto que a experiência humana é mediada pela interpretação, a qual não se dá de forma autônoma, mas na medida em que o indivíduo interage com outro, é por meio de interações sociais como vão sendo construídas as *interpretações*, os significados, a visão de realidade do sujeito (André, 1998, pp. 17-18).

Erickson (1986, p. 119), um pesquisador muito conhecido em educação, prefere o termo *pesquisa interpretativa* para se referir a toda uma família de abordagens de pesquisa participativa observacional, em lugar de pesquisa qualitativa, por ser mais abrangente, por evitar a idéia de que seja essencialmente não quantitativa e por apontar ao interesse central dessa pesquisa que é o significado humano em um contexto social e sua elucidação e exposição pelo pesquisador. Para ele (op. cit., p. 121), a pesquisa interpretativa envolve: a) intensa e ampla participação no contexto pesquisado, b) cuidadosos registros do que ocorre nesse contexto juntamente com outras fontes de evidência (e.g., anotações, documentos, exemplos de coisas feitas pelos sujeitos, gravações em áudio ou em vídeo) e c) análise reflexiva de todos esses registros e evidências assim como descrição detalhada (i.e., utilizando a narrativa e transcrições literais de verbalizações dos sujeitos).

Para Erickson (op. cit., p. 129), a tarefa da pesquisa interpretativa é descobrir maneiras específicas através das quais formas locais e não locais de organização social e cultural se relacionam com atividades de pessoas específicas em suas eleições e ações sociais conjuntas. Para a pesquisa na sala de aula, isso significa descobrir como as escolhas e ações de todos os atores constituem um currículo prescrito — um ambiente de aprendizagem. Professores e alunos juntos e interagindo adquirem, compartilham e criam significados não só através dos sistemas linguístico e matemático, mas também por meio de outros sistemas como a ideologia política, os pressupostos das subculturas étnicas e sociais a respeito do papel de mulheres e homens, das relações adequadas entre adultos e crianças, etc., isto é, por aculturação.

Uma vez apresentadas algumas características gerais da pesquisa qualitativa, passaremos a focar, sucessivamente, três metodologias principais dentro dessa abordagem: a *etnografia*, o *estudo de casos* e a *pesquisa-ação*.

A etnografia

A etnografia é uma metodologia, uma ferramenta para estudar e compreender uma cultura, a maneira de vida de um grupo de pessoas, isto é, suas idéias, crenças, valores e pressupostos, seus comportamentos e as coisas que fazem (Ogbu et al., 1988, p. 48). Em outras palavras, *a etnografia é uma tentativa de descrever uma cultura* (André, 1998, p. 19). A pesquisa etnográfica consta essencialmente de uma descrição de eventos que ocorrem no cotidiano da vida de um grupo com especial atenção às estruturas sociais e condutas de indivíduos a respeito do seu status de pertença ou membresia ao grupo, e uma interpretação do que significa todo isso para a cultura do grupo (Taft, 1988, p. 71).

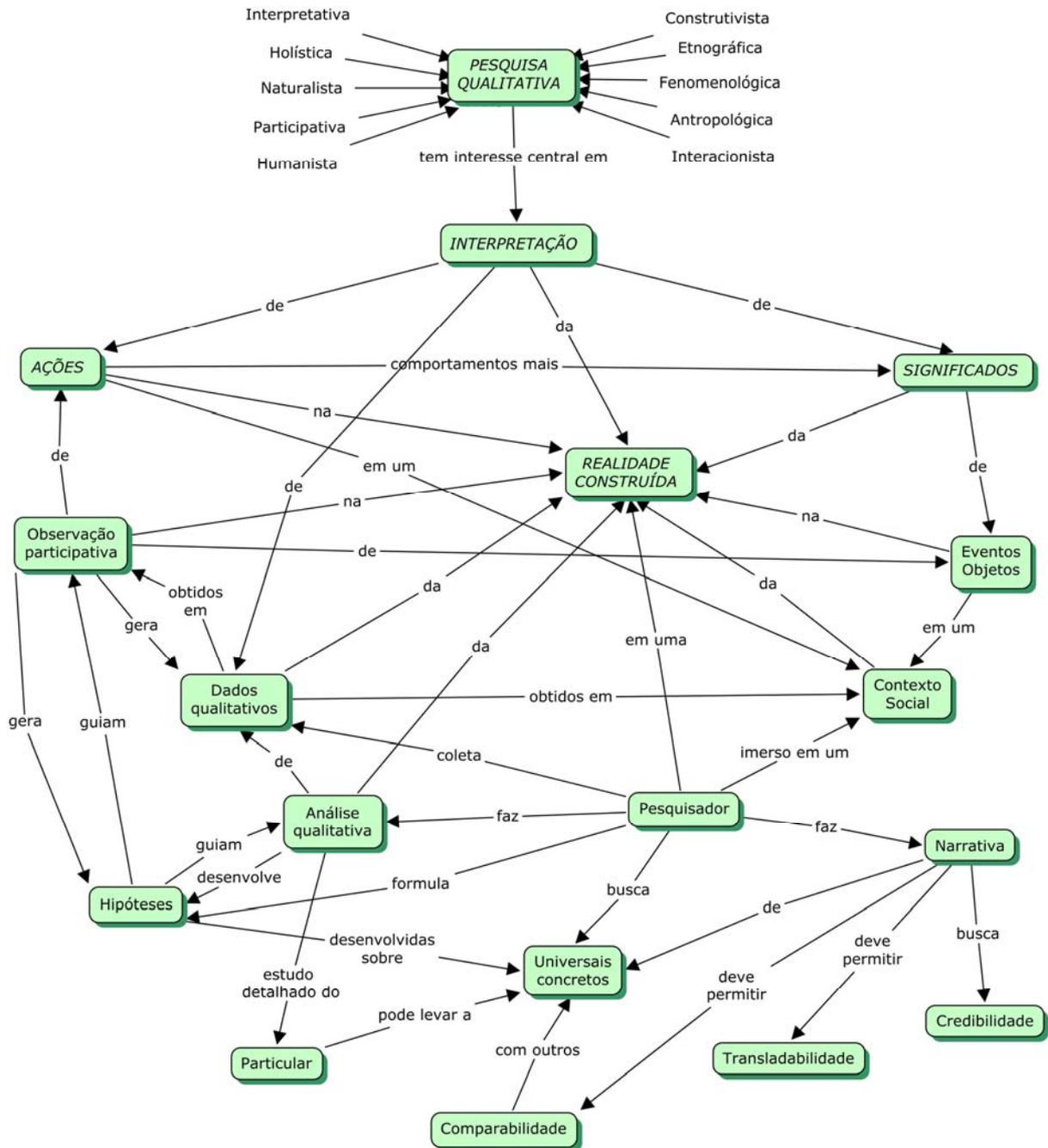


Figura 1. Um mapa conceitual para a pesquisa educacional qualitativa (M.A. Moreira, 2000)

Na etnografia, o pesquisador participa, o quanto é possível, da vida normal do grupo pesquisado, da cultura pesquisada. A pesquisa é conduzida no cenário natural dos eventos, no contexto no qual ocorrem os acontecimentos, através de observação participativa. Para chegar a uma *compreensão descritiva contextualizada* da cultura, o pesquisador tem que se meter em tal cultura, aprender a “*linguagem nativa*”, como disse o célebre antropólogo Malinowski, *interagir* com os membros dessa cultura, desenvolver uma *compreensão empática* da vida das pessoas tal como elas a percebem, assim como uma *perspectiva holística* do grupo. Tudo isso, é lógico, implica um longo “tempo de residência” nessa cultura. Isto é, o pesquisador deve permanecer “imerso” na cultura pesquisada durante um período de tempo “suficientemente grande” para, de acordo com Malinowski, contextualizar os dados em um “account” holístico e coerente e descrever “a vida tal como é vivida” (Ogbu et al., 1988, p. 50).

O pesquisador etnográfico tem, por conseguinte, um duplo papel: participante e observador. Por um lado, ele tem que se envolver com o grupo, “aculturar-se” nele. Por outro lado, deve ser capaz de observar, interpretar, discernir, desenvolver uma perspectiva holística. Sendo ao mesmo tempo observadores e participantes, os pesquisadores etnográficos não são desprendidos do fenômeno de interesse; eles influenciam seus dados e são influenciados por estes em todas as etapas de observação, interpretação e descrição (Taft, 1988, p. 72). A grande vantagem de ser observador participante parece ser ao mesmo tempo a principal dificuldade que deve enfrentar o pesquisador participante. Ao mesmo tempo que tenta “pertencer” à cultura pesquisada, ele ou ela deve também ser capaz de “mirá-la desde fora”, interpretá-la, descrevê-la.

A observação participativa é a principal técnica de pesquisa etnográfica. Porém, as entrevistas são também muito utilizadas. Os dados gerados por essas duas técnicas são frequentemente complementados por outros como documentos, narrativas, histórias de vida, artefatos, diagramas, produzidos no grupo pesquisado. No geral, o pesquisador etnográfico busca recolher toda a informação possível, não só através de observação participativa e entrevistas, para interpretá-la indutivamente e construir uma realidade social que é sua compreensão descritiva contextualizada da cultura pesquisada.

A metodologia etnográfica é qualitativa e holística, fazendo uso da intuição, empatia e outras habilidades do pesquisador para interpretar descritivamente uma cultura. Seu interesse está em descobrir (no sentido de construir uma descrição compreensiva contextualizada) e não em verificar. Porém, isso não implica não ter nenhuma hipótese ou teoria inicial. O pesquisador etnográfico não começa um trabalho de campo “sem ter nada na cabeça”. Isso não existe. Ele ou ela sempre terão conhecimentos teóricos prévios que de alguma maneira vão orientar seus passos iniciais, mas não devem ter hipóteses e teorias que serão verificadas ou rejeitadas no estudo. Isto é, o pesquisador etnográfico não deve ter idéias pré-concebidas, tal como recomendou Malinowski (apud Taft, 1988, p. 74).

As hipóteses são formuladas recursivamente durante o processo, durante o desenvolvimento da pesquisa. Gradativamente pode emergir uma base teórica para a compreensão dos processos grupais. Essa base teórica é conhecida como *teoria fundamentada*, isto é, fundamentada no próprio processo de pesquisa (ibid.), ou fundamentada nos dados. Tal teoria, que foi desenvolvida de maneira indutiva, provavelmente gerará hipóteses úteis para guiar, inicialmente, novas observações participativas. Porém, não se está falando aqui do indutivismo científico ingênuo tão criticado epistemologicamente, nem de hipóteses que serão “comprovadas” em estudos “mais rigorosos”. (O tema da credibilidade dos “resultados” da pesquisa qualitativa será discutido mais adiante em outro item).

A etnografia é uma metodologia de pesquisa em antropologia que chegou à pesquisa em educação não faz muito, nos anos 60 do século passado. Uma grande diversidade de etnografias educativas tem sido desenvolvida desde essa época, porém o conceito de cultura permaneceu como constructo unificador. Três orientações principais podem ser identificadas (Ogbu et al., 1988, p. 50-51) a partir de distintos níveis de análise e diferentes ênfases em suas definições de cultura: *etnografia holística* (também conhecida como etnografia tradicional, velha etnografia ou macro-etnografia), a *etno-ciência* (também chamada nova etnografia ou antropologia cognitiva) e a *micro-etnografia* (etnografia da comunicação).

A *etnografia holística*, da qual falamos até agora, tenta descrever a cultura, ou o grupo, como um todo enquanto que a etno-ciência e a micro-etnografia focam unidades muito menores como palavras, indivíduos ou cenas (ibid.).

A *micro-etnografia* é uma etnografia focada, isto é, uma etnografia que se ocupa de olhar repetidas vezes e de analisar detalhadamente registros audiovisuais de interações humanas em cenas-chave, em situações-chave de interação social, acompanhadas de observação participativa do contexto mais amplo no qual ocorrem tais cenas (op. cit., p. 51). É uma etnografia da comunicação, focando sujeitos individuais e seu discurso em certos cenários.

A *etno-ciência* se afasta da etnografia holística tradicional ao definir cultura primariamente em termos de cognições das pessoas. Seus pressupostos básicos são que o conteúdo dos dados culturais consta de regras, códigos e um ordenamento ideativo da sociedade que está organizado em diversos domínios culturais de conhecimento.

As experiências são codificadas em “lexemas” ou palavras; portanto, a linguagem é a principal fonte de dados culturais e as técnicas de estudo da linguagem podem ser aplicadas ao estudo da cultura ideativa ou cognição. Por conseguinte, há menos ênfase na observação participativa e mais ênfase na coleta de vocabulários sobre eventos particulares, assim como nos esquemas classificatórios (op.cit., p. 52).

Relacionada com a etno-ciência está a *etno-metodologia* que, segundo André (1998, p. 18), não é exatamente uma metodologia, mas um campo de pesquisa: é o estudo de como os indivíduos compreendem e estruturam seu cotidiano, isto é, é a tentativa de descobrir “os métodos” que as pessoas utilizam em seu dia-a-dia para entender e construir a realidade que as envolve. Em consequência, seus principais focos de interesse são os conhecimentos tácitos, as formas de compreensão do sentido comum, as práticas cotidianas e as atividades rotineiras que moldam as condutas dos atores sociais (ibid.).

Independente desses aparentemente diferentes tipos de etnografia, podemos caracterizá-la de maneira geral como a tentativa de descrição de uma cultura. A principal preocupação na etnografia se refere ao significado que têm as ações e eventos para as pessoas ou grupos estudados (op. cit., p. 19). A etnografia é um esquema de pesquisa desenvolvido pelos antropólogos para estudar uma cultura e uma sociedade. Etimologicamente, etnografia significa “descrição cultural” (ibid., p. 27).

Em educação, rigorosamente falando, o que se faz são estudos etnográficos, isto é, uma adaptação da etnografia à educação, uma vez que o fenômeno de interesse da pesquisa educativa é, em última análise, o processo educativo, não uma cultura ou um grupo social em si mesmos. Tais estudos etnográficos incluem, por exemplo, uma sala de aulas em particular, um pequeno grupo em uma sala de aulas ou em uma escola, cenas ou diálogos na sala de aulas, relações escola-comunidade, etc.

A modo de conclusão deste item a respeito da etnografia se apresenta na figura 2 um mapa conceitual para a etnografia. No topo aparece como conceito mais abrangente o próprio conceito de *etnografia* que pode ser *holística* (a etnografia tradicional ou “velha” etnografia), *micro-etnografia* (a etnografia da comunicação) ou *etno-ciência* (a antropologia cognitiva ou a “nova” etnografia). Porém, a etnografia é sempre uma tentativa de descrever uma *cultura* (ou uma micro-cultura) que é caracterizada principalmente por *significados*, construídos e partilhados pelo grupo social, isto é, pelo “ponto de vista nativo” (idéias, crenças, valores, pressupostos), segundo Ogbu et al. (1988, p. 50).

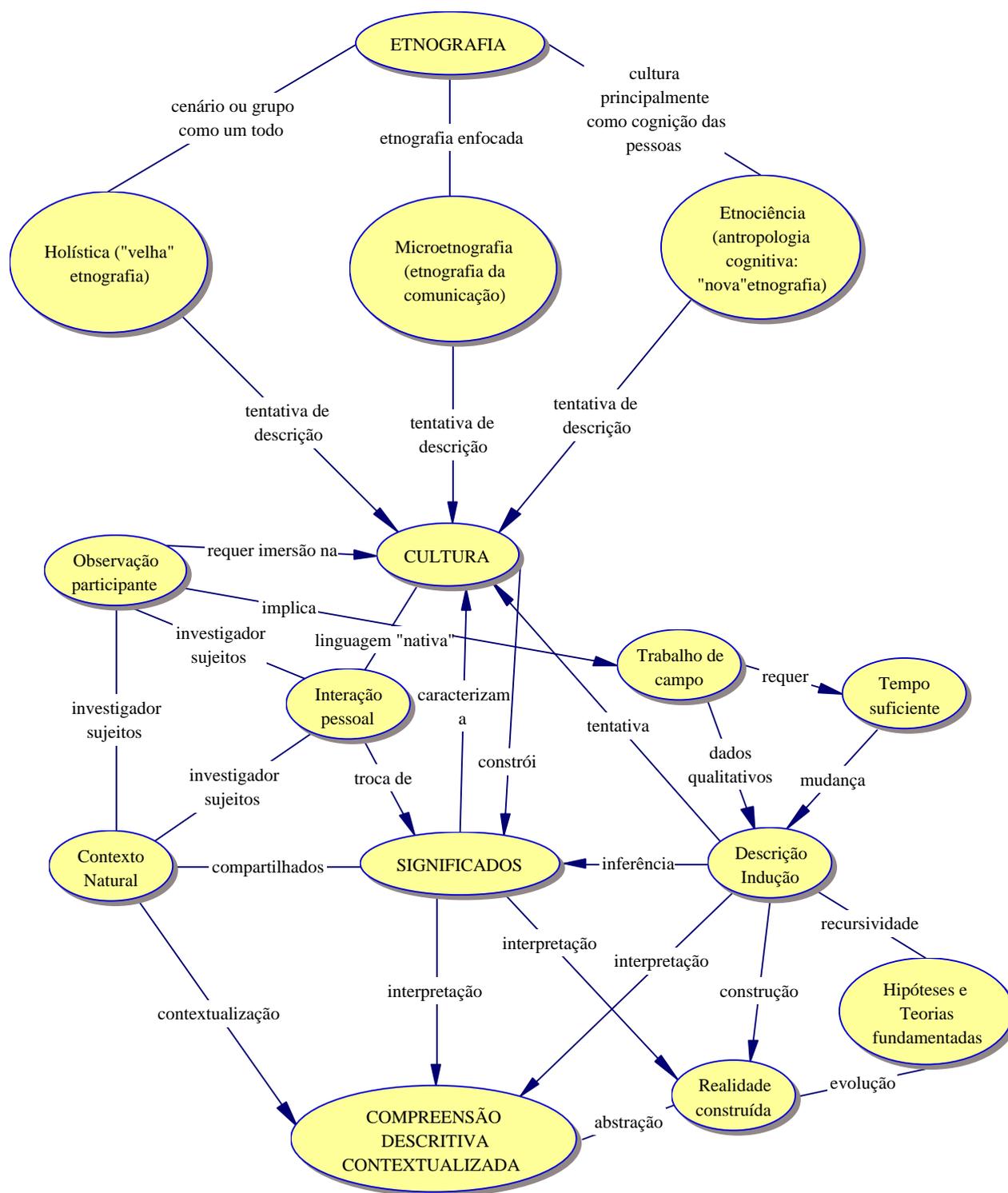


Figura nº 2. Um mapa conceitual para a etnografia (Moreira, 2002).

A descrição de uma cultura requer *observação participante*, *trabalho de campo* (durante um *tempo suficiente*) e *interação pessoal* (inter-subjetividade, empatia) em um *contexto natural*. A etnografia busca *descrição* e utiliza a *indução* para chegar a uma

realidade construída; as *hipóteses* são desenvolvidas ao longo do processo e as teorias emergem dos dados, isto é, são *teorias fundamentadas* (nesse sentido, os métodos qualitativos são indutivos). O resultado de todo o processo é uma *compreensão descritiva contextualizada*, de um grupo social, de umas cenas, de um discurso, de umas cognições ou, em termos mais abrangentes e originais, de uma cultura.

O estudo de casos

De acordo com Sturman (1988, p. 61), estudo de caso é um termo genérico para a pesquisa de um indivíduo, um grupo ou um fenômeno. Enquanto que as técnicas usadas nessa pesquisa podem variar e incluir tanto enfoques qualitativos como quantitativos, a característica que mais distingue o estudo de caso é a crença de que os sistemas humanos desenvolvem uma completude e integração, isto é, não são simplesmente um conjunto de partes ou de traços. Por conseguinte, o estudo de caso encaixa em uma tradição holística de pesquisa segundo a qual as características de uma parte são determinadas grandemente pelo todo ao qual pertence. A compreensão das partes requer a compreensão de suas inter-relações no todo. É uma visão sistêmica que pressupõe que os elementos de um evento educativo, por exemplo, são interdependentes e inseparáveis e uma mudança em um elemento implica uma mudança no resto.

Portanto, fazer uma pesquisa do tipo estudo de caso, isto é, para entender um caso, para compreender e descobrir como as coisas ocorrem e por que ocorrem, para talvez prever algo a partir de um único exemplo ou para obter indicadores que possam ser usados em outros estudos (talvez quantitativos) é necessário uma profunda análise das interdependências das partes e dos padrões que emergem. O que se requer é um estudo de padrões, não de variáveis isoladas (ibid.). Para tudo isso, as técnicas de pesquisa qualitativa são frequentemente as mais adequadas.

O estudo de casos se utiliza há muito tempo em campos como o direito, a medicina, a psicologia e a administração. Porém, na pesquisa educativa, o uso dessa metodologia é mais recente e dentro de uma concepção bastante restrita, ou seja, o estudo descritivo de uma unidade que pode ser uma escola, um professor, um aluno ou uma aula (André, 1998, p. 21). Pode também, por exemplo, ser o estudo de um currículo ou de um sistema escolar.

Qual seria a diferença entre estudo de caso e etnografia? Segundo André (ibid.), se pode fazer um estudo de caso etnográfico, isto é, aplicar a abordagem etnográfica ao estudo de um caso. Porém, nem todos os tipos de estudos de caso encaixam na perspectiva etnográfica, nem todos os estudos etnográficos são estudos de caso.

Quais seriam então os distintos tipos de estudos de caso?

Em primeiro lugar é necessário distinguir entre estudo de caso e conceitos afins. De acordo com Serrano (1998, p. 85), o *trabalho com casos* denota os procedimentos corretivos, remediais, de desenvolvimento ou ajuste que seguem ao diagnóstico das causas de desajuste; o *método de casos* é uma estratégia didática na qual os elementos principais do estudo de casos são apresentados aos estudantes com propósitos ilustrativos, sem necessidade de oferecer uma visão completa dos fatos (o propósito é mais o de estabelecer um marco de discussão e debate); a *história de casos* é a busca do passado de uma pessoa, grupo ou instituição; o *estudo de casos*, por sua vez, pode ser definido como uma descrição intensiva, holística e uma análise profunda de uma entidade singular, um fenômeno ou unidade social.

Antes de falar de tipos de estudos de casos, é interessante também destacar seus traços essenciais. Segundo essa mesma autora (op. cit., p. 91), as *propriedades essenciais* de um estudo de casos qualitativo são a *particularização* (se centram em uma situação, evento, programa ou fenômeno particular), a *descrição* (o produto final é uma descrição rica e densa do objeto de estudo), a *heurística* (iluminam a compreensão do leitor respeito ao objeto de estudo) e a *indução* (baseiam-se no raciocínio indutivo; as teorias, os conceitos ou as hipóteses surgem de um exame dos dados fundados no contexto mesmo).

Quanto aos *tipos* de estudos de caso, Serrano (ibid., p. 97) argumenta que podem classificar-se pela natureza do informe final, independentemente de sua orientação disciplinar ou área de interesse, em *descritivos*, *interpretativos* e *avaliativos*.

Estudos de caso descritivos (ibid.) se caracterizam por um informe detalhado de um fenômeno objeto de estudo sem fundamentação teórica prévia; são inteiramente descritivos, não se guiam por generalizações estabelecidas ou hipotéticas, nem desejam formular hipóteses ou teorias.

Estudos de caso interpretativos (op. cit., p. 98) contêm descrições ricas e densas; porém, os dados descritos são utilizados para desenvolver categorias conceituais ou para ilustrar, defender ou desafiar pressupostos teóricos difundidos antes do estudo. O pesquisador deve reunir tanta informação sobre o objeto de estudo quanto seja possível, com a pretensão de interpretar ou teorizar sobre o fenômeno.

Estudos de caso avaliativos (ibid.) implicam descrição, explicação e juízo; sobretudo, este tipo de estudo de casos examina a informação para emitir um juízo; a emissão de juízos é o ato final e essencial da avaliação.

Não obstante, a mesma autora coloca que mesmo que se possa estabelecer esta classificação e mesmo que alguns estudos de casos possam ser puramente descritivos, em educação a maioria dos estudos de casos é uma combinação de descrição e avaliação ou de descrição e interpretação.

O estudo de caso interpretativo nos remete outra vez ao tema da *teoria fundamentada* referido na etnografia. Este tipo de estudo de casos parece ser uma metodologia ideal para fundamentar uma teoria, isto é, para induzir uma teoria a partir de dados descritivos muito ricos. Porém, não se trata de uma teoria formal no sentido usado nas ciências naturais, também não do "método indutivista". São mais bem categorias, hipóteses compreensivas.

Outra classificação de tipos de estudos de caso é fornecida por Stenhouse (1985, apud Sturman, 1988, p. 63):

Estudo de caso etnográfico, do qual já falamos, que envolve o estudo profundo de uma entidade singular geralmente através de observação participante e entrevistas.

Estudo de caso pesquisa-ação, no qual o foco está em gerar uma mudança no caso em estudo.

Estudo de caso avaliativo que envolve avaliação de programas e no qual, muitas vezes, um trabalho de campo mais condensado substitui a abordagem etnográfica mais demorada.

Estudo de caso educativo que está desenhado para melhorar a compreensão da ação educativa.

Como se pode perceber nas classificações de Serrano e Stenhouse, é difícil separar o estudo de casos de outros tipos de pesquisa qualitativa como a etnografia e a pesquisa-ação. Podemos caracterizar bem o estudo de casos argumentando que sua preocupação central é a compreensão de uma instância singular, o que significa que o objeto estudado é caracterizado como único, como uma representação singular da realidade que é multidimensional e historicamente localizada (André, 1998, p. 21). Porém, como vimos, um estudo de caso pode ser feito através de uma etnografia ou de uma pesquisa-ação, por exemplo.

As características essenciais dos estudos de casos, assim como seus distintos tipos estão mapeadas conceitualmente na figura nº 3. Os dois conceitos chaves são *estudo de casos* e *instância singular*. Na parte superior do mapa, acima do conceito de estudo de casos estão suas propriedades essenciais (*indução*, *particularização*, *heurística* e *descrição*); na parte inferior, embaixo do conceito de instância singular aparecem instâncias de tal conceito. No eixo central do mapa estão os diferentes tipos de estudo de caso identificados por Serrano (1998) e Stenhouse (1985). Os conectores tentam explicitar as relações entre os conceitos e as flechas sugerem certas convergências.

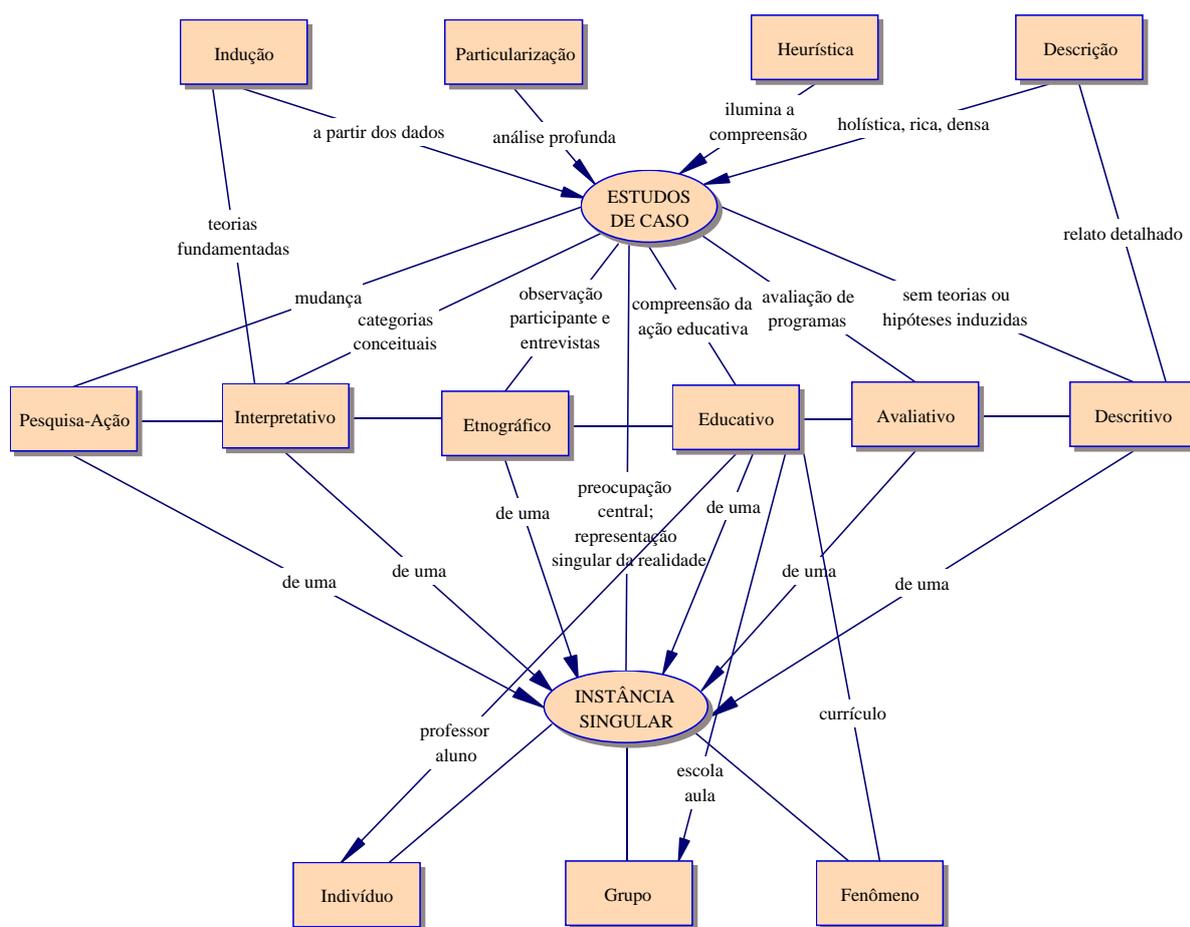


Figura 3. Um mapa conceitual para estudos de caso (Moreira, 2002).

A pesquisa-ação

O objetivo fundamental da pesquisa-ação consiste em melhorar a prática em vez de gerar conhecimentos. A produção e utilização do conhecimento se subordinam a este objetivo e estão condicionadas por ele (Elliott, 1993, p. 67). A melhora na prática consiste em implantar aqueles valores que constituem seus fins, por exemplo, a educação no ensino (ibid.). Porém, o conceito de educação como fim do ensino transcende a conhecida distinção entre processo e produto. A melhora da prática supõe levar em conta ao mesmo tempo os resultados e os processos.

Segundo Kemmis e McTaggart (1988; apud Kemmis, 1988, p. 174), a pesquisa-ação é definida como uma forma de pesquisa *coletiva* auto-reflexiva empreendida por participantes de situações sociais para melhorar a produtividade, racionalidade e justiça de suas próprias práticas sociais ou educativas, assim como sua compreensão em relação a tais práticas e às situações em que ocorrem. Os participantes podem ser professores, alunos, diretores, pais e outros membros da comunidade, isto é, qualquer grupo que partilha uma preocupação, um objetivo. É uma *pesquisa colaborativa*; porém, é importante enfatizar que esta ação colaborativa depende de que cada indivíduo examine criticamente suas próprias ações (ibid.).

Na educação, quando se pretende melhorar a prática, têm que se considerar conjuntamente os processos e os produtos. Este tipo de reflexão simultânea sobre a relação entre processos e produtos constitui, segundo Elliott (op. cit., p. 68), uma característica fundamental da pesquisa-ação. É uma *prática reflexiva* que aspira a melhorar a concreção dos valores do processo, muito diferente do raciocínio técnico que versa sobre os meios para conseguir um fim. Para Elliot (op. cit., p. 69), é ao mesmo tempo ético e filosófico. Na medida em que a reflexão trata da escolha de um curso de ação em um determinado conjunto de circunstâncias para levar à prática os próprios valores, reveste-se de caráter ético. Mas, como a eleição ética supõe a interpretação dos valores que hão de se traduzir à prática – a reflexão sobre os meios não se pode separar da reflexão sobre os fins – a reflexão ética tem uma dimensão filosófica (ibid.).

Na pesquisa-ação, os professores são incentivados a questionar suas próprias idéias e teorias educativas, suas próprias práticas e seus próprios contextos como objetos de análise e crítica (Kemmis, 1988, p. 174). A partir de uma reflexão cuidadosa, os professores podem desvelar idéias ou suposições teóricas que resultam injustificadas e os deixam perdidos em sua tarefa docente; por exemplo, tem-se suposições muito rígidas a respeito da natureza de habilidades inatas dos estudantes (ibid.).

Analogamente, os docentes, através da reflexão crítica, podem concluir que práticas antigas moldadas por hábito e tradição são inúteis ou irrelevantes nos tempos atuais; por exemplo, práticas disciplinares que funcionavam antes, hoje já não são aceitáveis ou são contraproducentes (ibid.). Quanto ao contexto, eles podem chegar à conclusão de que sua estrutura é inadequada e obstaculiza o alcance de metas educativas; por exemplo, a estrutura física da aula pode dificultar o trabalho em grupos, a interação pessoal, o ensino centrado no aluno.

Não obstante, não se pode esquecer que a pesquisa-ação é uma pesquisa coletiva, colaborativa. A reflexão pessoal é importante, mas a verdadeira mudança vem da auto-reflexão coletiva. Que os participantes sintam a necessidade de iniciar mudanças, de inovar, é condição necessária antecedente da pesquisa-ação, mas não suficiente.

O processo de pesquisa-ação, segundo Kemmis e McTaggart (1988) e Elliot (1993, p. 88), se caracteriza por uma espiral de ciclos de reconhecimento (descoberta de fatos): reconhecimento de uma situação que se quer mudar; planificação geral da ação objetivando a mudança; *desenvolvimento, implementação e avaliação* dessa ação; reflexão à luz da evidência recolhida na implementação; revisão do plano geral; planificação de nova ação; implementação, avaliação, reflexão, revisão do plano; planificação e implementação de uma terceira ação...

Naturalmente, este caráter cíclico não significa um processo linear, automático, mecânico. Tal como se disse no começo deste item, a pesquisa-ação, através dessa espiral de ciclos, tem por objetivo a *melhoria* das práticas e da compreensão de situações, e o *envolvimento* tanto quanto possível de todos os afetados intimamente pelas ações em todas as fases do processo investigativo. A pesquisa-ação é um processo colaborativo, auto-reflexivo, no qual o envolvimento direto dos professores e outros implicados, na coleta de dados, análise crítica, reflexão, cria imediatamente um sentido de responsabilidade quanto à melhora da prática (Kemmis, 1988, p. 174).

A pesquisa-ação unifica processos considerados com frequência independentes; por exemplo: o ensino, o desenvolvimento do currículo, a avaliação, a pesquisa-educativa e o desenvolvimento profissional (Eliott, 1993, p. 72). O ensino, por exemplo, no contexto da pesquisa-ação se concebe como uma forma de pesquisa dirigida a compreender como traduzir os valores educativos a formas concretas de prática.

O desenvolvimento do currículo não é um processo que antecede o ensino; o desenvolvimento de programas curriculares se produz através da prática reflexiva do ensino (ibid.). A pesquisa-ação não reforça a postura dos professores enquanto conjunto de indivíduos que operam de forma independente e autônoma, que não partilham suas reflexões com os demais.

De maneira geral, pode-se dizer que a pesquisa-ação sempre implica um plano de ação baseado em objetivos de mudança (melhora), a implementação e controle desse plano através de fases de ação, assim como a descrição concomitante do processo cíclico resultante. Porém, Kemmis e McTaggart (1988, apud Kemmis, 1988) identificam várias características básicas da pesquisa-ação que ajudam a distingui-la de outros tipos de pesquisa qualitativa. Segundo eles, a pesquisa-ação:

- é uma abordagem para melhorar a educação através de mudanças e para aprender desde as consequências das mudanças;
- se desenvolve através de uma espiral auto-reflexiva de ciclos de planificação, ação, observação sistemática, reflexão, replanificação, nova ação, observação e reflexão;
- é participativa, as pessoas trabalham para melhorar suas próprias práticas;
- é colaborativa, cria grupos auto-críticos que participam e colaboram em todas as fases do processo investigativo;
- envolve os participantes em um processo de teorização sobre suas práticas, questionando circunstâncias, ações e consequências dessas práticas;
- requer que as pessoas ponham em xeque suas idéias e suposições com relação às instituições;
- é aberta com relação ao que conta como evidência, ou dados, mas sempre implica manter e analisar registros das consequências das ações implementadas;
- permite que os participantes, ao mesmo tempo, mantenham registros de suas próprias mudanças pessoais e analisem criticamente as consequências dessas mudanças;

- começa pequena; normalmente com pequenas mudanças que um pequeno grupo, ou talvez uma só pessoa, possa tentar, mas se desloca, gradativamente, rumo a mudanças mais extensivas;
- requer que os participantes analisem criticamente as situações (salas de aulas escolas, sistemas educativos) nos quais trabalham;
- é um processo político porque envolve mudanças nas ações e interações que constituem e estruturam práticas sociais; tais mudanças tipicamente afetam as expectativas e interesses de outros além dos participantes imediatos nessas ações e interações.

Tal como se fez nas seções anteriores, esta finaliza com um mapa conceitual com relação ao tema abordado. A figura 4 apresenta um mapa conceitual para pesquisa-ação. O conceito central é *mudança*: a pesquisa-ação tem como meta melhorar a *prática* através da mudança. É também central a tríade (*re*) *planejamento* ↔ *ação* ↔ (*auto*) *reflexão* que caracteriza o *processo cíclico* da pesquisa-ação. Por outro lado, a pesquisa-ação é um processo *participativo, coletivo, colaborativo, político, auto-reflexivo, autocrítico, auto-avaliativo* que requer o envolvimento dos *participantes* em todas as fases e em todos os aspectos característicos desse processo.

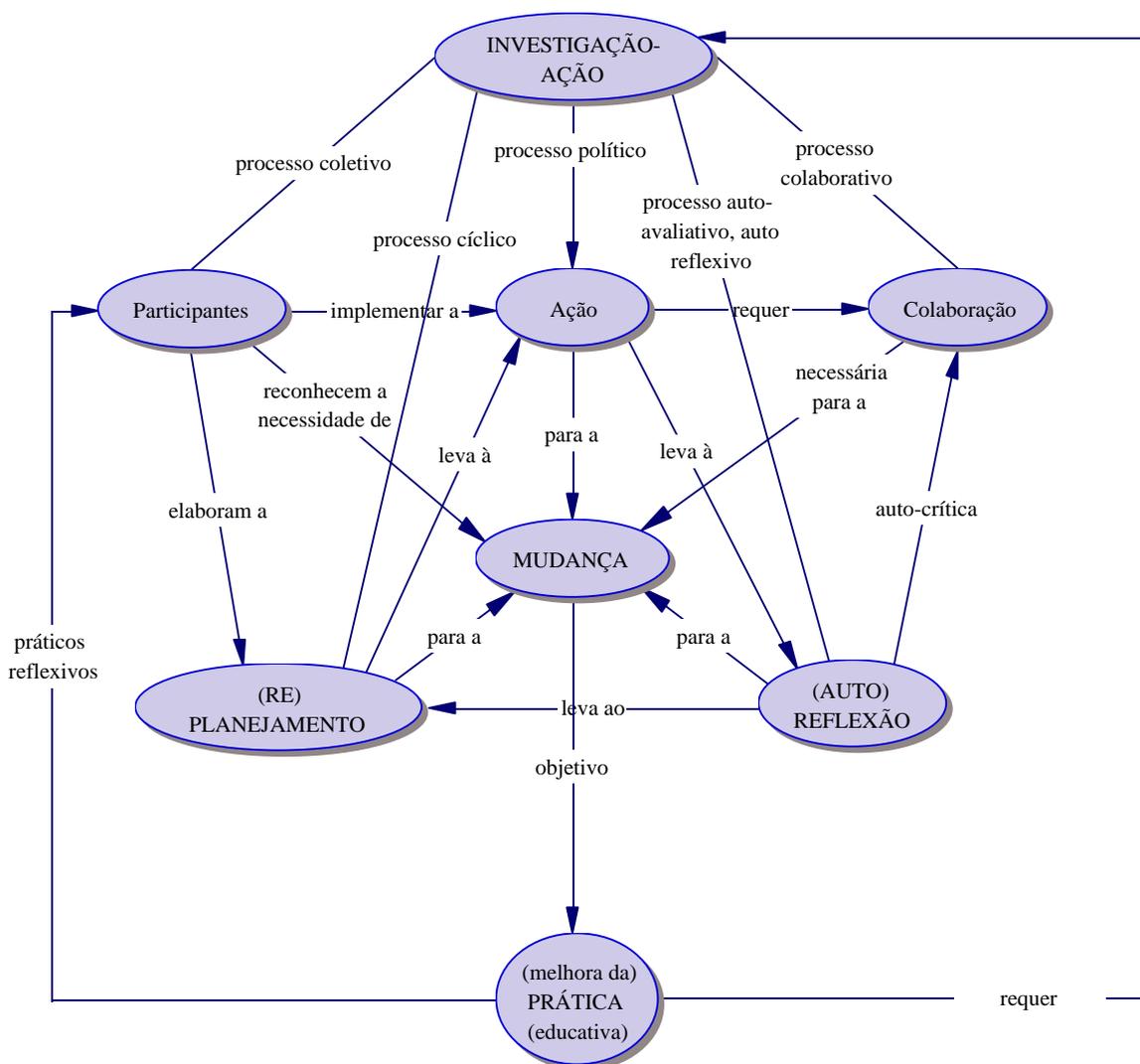


Figura 4. Um mapa conceitual para pesquisa-ação (Moreira, 2002).

Outros tipos de pesquisa qualitativa

Começamos este texto comparando as abordagens qualitativa e quantitativa em termos bastante dicotômicos; depois focamos características gerais da pesquisa qualitativa e passamos logo a descrever a *etnografia*, o *estudo de casos* e a *pesquisa-ação* que consideramos as *três metodologias principais* dessa abordagem. Porém, há outras, das quais apresentaremos algumas, sucintamente, a seguir.

Uma vez concluídas essas breves apresentações voltaremos a temas gerais, no sentido de que se aplicam a várias metodologias qualitativas, focando tópicos como a apresentação e a análise de dados, a fiabilidade e validade dos estudos qualitativos, possibilidades de generalização, triangulação e outros. No final, retomaremos a questão dos paradigmas a partir de um olhar integrador.

A fenomenografia

A fenomenografia é o estudo empírico dos diferentes modos através dos quais as pessoas vivenciam, percebem, apreendem, compreendem, ou conceituam vários fenômenos no, e aspectos do, mundo em seu entorno. As palavras vivência, percepção, compreensão ou conceituação são usadas de maneira intercambiável. Porém, isso não significa que não há diferenças em seus significados, mas sim que o número limitado de maneiras, através das quais certo fenômeno é interpretado pelas pessoas pode ser identificado, por exemplo, independente de se estão embebidas na experiência imediata do fenômeno ou em uma reflexão sobre o mesmo fenômeno (Marton, 1988, p. 95-97).

As diferentes experiências, compreensões, percepções, etc., são caracterizadas em termos de "categorias de descrição" logicamente relacionadas e hierarquizadas. Tais categorias representam diferentes capacidades de lidar com (ou entender) um fenômeno. Como algumas maneiras de experienciar um fenômeno são mais eficientes que outras com relação a algum critério, é possível estabelecer uma hierarquia de categorias de descrição (ibid.). A fenomenografia se aplica, por exemplo, ao estudo da resolução de problemas ou em pesquisas que procuram diferenças críticas nos significados atribuídos a certos fenômenos, conceitos ou princípios-chave em certo campo de conhecimentos.

A hermenêutica

A hermenêutica foi originalmente definida como a arte, ou a ciência, da interpretação, em particular da Bíblia. Porém, contemporaneamente é definida como a teoria e a prática da interpretação e compreensão em diferentes tipos de contextos humanos (religiosos, seculares e cotidianos). Isto é, a hermenêutica não se refere só à exegese e interpretação de textos, mas considera a compreensão e a interpretação como uma marca definitiva da existência humana e da vida social (Ödman e Kerdeman, 1988, p. 185).

Assim como a hermenêutica, a fenomenologia também se ocupa da estrutura da compreensão; porém, a fenomenologia constrói a compreensão primeiramente em termos de construtos e funções cognitivos, enquanto que para a hermenêutica a compreensão não é somente uma função cognitiva, é também a condição ontológica da existência humana (op. cit., p. 186).

Assim como a teoria crítica, a hermenêutica mantém que compreensão e significado são constituintes da vida social; não obstante, a hermenêutica mantém que uma vez que o sujeito está sempre envolvido em algum processo de compreensão, é impossível captar em qualquer forma final ou definitiva os significados embebidos em uma tradição. A hermenêutica, portanto, evita envolver-se na tentativa de fundamentar a compreensão em um marco teórico ou em um método e se concentra em interpretar culturas desde dentro de certas situações e contextos (ibid.).

Na pesquisa educativa, a hermenêutica pode aprofundar a compreensão do fenômeno educativo focando os significados que subjazem a estratégias e práticas educativas específicas. Por exemplo, como devem ser interpretadas certas práticas administrativas? Quais são os significados subjacentes? Analogamente, a hermenêutica pode aprofundar a compreensão dos significados e propósitos subjacentes a um currículo.

A pesquisa participativa

A pesquisa participativa é descrita por Hall (1988, p. 198), de uma maneira geral, como um processo que combina três atividades: pesquisa, educação e ação. É uma ação social em favor dos dominados, explorados, pobres, excluídos. A preocupação por poder e democracia e suas interações é central na pesquisa participativa. É também crítica a atenção a gênero, raça, etnia, orientação sexual, habilidades físicas e mentais, e outros fatores sociais (ibid.). A pesquisa participativa está desenhada para contribuir nos processos de mudança de poder ou democratização em uma variedade de contextos. Na pesquisa participativa não há receitas nem ortodoxias metodológicas: as questões e os métodos de trabalho devem fluir dos sujeitos envolvidos e de seu contexto.

(Não se deve confundir pesquisa participativa com observação participante que é uma mistura de várias técnicas, um estilo ou estratégia de pesquisa, na qual, como já vimos, o principal instrumento é o próprio pesquisador que deve ficar imerso no cenário para ouvir, ver e começar a experimentar a vida como os sujeitos a vivem; Ball, 1988, p. 310).

História oral

Como uma técnica de pesquisa, a história oral vai além do que o entrevistado responde às perguntas do entrevistador. Todos os matizes de seu testemunho são significativos: a não-resposta, o silêncio, a vacilação, tudo conta como evidência. As fontes de história oral são, portanto, mais que uma questão de evidência falada e mais do que simplesmente registrar fatos (Hyams, 1988, p. 91).

Como uma abordagem para a pesquisa, a história oral permite dar voz aos "sem voz", aos anônimos. É como construir a história de baixo para cima. Mesmo que sofra restrições entre os historiadores (por exemplo, só os sobreviventes são entrevistados; não se pode generalizar), esta abordagem é usada em outros campos, como o da educação. Por exemplo, as atitudes dos professores com relação ao sistema educativo ou às reformas educativas podem ser obtidas de maneira mais expansiva, mais aberta. As reminiscências dos professores podem permitir uma melhor compreensão da implementação das políticas educativas. A história oral tem potencial para explicar interpretações não-oficiais de problemas educativos.

O caráter único do testemunho individual é considerado importante para construir uma história. Porém, como foi dito antes, é preciso levar em conta muito mais que a informação factual provida. Trata-se de procurar informação que não está nos registros escritos, nas biografias.

Fidedignidade, generalização e validade

A *fidedignidade* se refere ao grau em que se podem replicar as medidas e os estudos. Em uma abordagem quantitativa, a *fidedignidade* das medições e dos instrumentos é um requisito básico, porém, em uma abordagem qualitativa, tal conceito não tem sentido ou deve ter outro significado, pois, em grande medida, o pesquisador é o principal instrumento ou, em outras palavras, o instrumento é uma extensão do pesquisador.

Em relação aos estudos, a *fidedignidade* exige que um pesquisador que utilize os mesmos métodos que outro, chegue a idênticos resultados (Goetz e Lecompte, 1988, p. 214). Isso coloca um enorme problema nas pesquisas sobre o comportamento natural ou dos fenômenos únicos. O estabelecimento da *fidedignidade* de um delineamento interpretativo se complica ainda mais pela natureza dos dados e do processo de pesquisa, pelos usos da apresentação dos resultados e pela visão de mundo dos pesquisadores nesse campo (ibid.).

De acordo com esses autores (op. cit., p. 215), se comparados com os delineamentos experimentais de laboratórios, estritamente controlados, ou com os experimentos de campo, os delineamentos da pesquisa naturalística parecem resistir-se a toda tentativa de réplica. Por exemplo, os problemas de unicidade e idiossincrasia podem levar a afirmar que é impossível replicar um estudo etnográfico ou um estudo de caso. Além disso, sendo que o comportamento humano nunca é estático, nenhum estudo, independentemente de seus métodos e desenhos, pode ser replicado com exatidão (ibid.).

Porém, se pode contra-argumentar que tais unicidades e idiossincrasias não são tão extremas que não tenham nenhum grau de “representatividade”, alguns traços semelhantes a outras. Isto é, os grupos ou fenômenos sociais ou indivíduos pesquisados podem não ser tão únicos e idiossincráticos que não tenham nada a ver com outros grupos, fenômenos ou indivíduos. Ou se, de fato, o são, não são de interesse da pesquisa educativa. Para que serve uma etnografia de uma aula de ciências que não tem absolutamente nada a ver com outras aulas de ciências ou que não gere uma compreensão contextualizada com algum valor para elas? Para que serve estudar um professor de ciências tão único, que não tem nada a ver com outros professores de ciências?

Isso nos remete ao tema da *generalização*: devem-se procurar generalizações em estudos qualitativos, através de, por exemplo, estudos de “casos representativos”?

Segundo André (1998, p. 58), a generalização no sentido de leis que se aplicam universalmente não é um objetivo da pesquisa qualitativa. Porém, a idéia de generalização é bastante aceita nessa abordagem, no sentido de que os dados de um estudo podem ser úteis para compreender dados de outros estudos (ibid.). Por isso, a descrição densa é considerada vital quando se pretende fazer comparações ou transferências de uma situação a outra; a análise de similitudes e diferenças torna possível julgar em que medida as compreensões construídas em um estudo podem ser consideradas hipóteses sobre o que pode ou não ocorrer em outras situações. Esta posição é compartilhada por Ogbu et al. (1988, p. 53), que diz que a generalização se torna possível porque o conhecimento construído através da abordagem

qualitativa é profundo e contextualizado; as descrições detalhadas compreensivas permitem aos leitores fazer comparações e tomar decisões bem fundamentadas a respeito da generalização.

Posição semelhante é tomada por Taft (1988, p. 74), que argumenta que para generalizar de um caso individual a outros é necessário alcançar uma compreensão suficientemente detalhada sobre a significatividade dos eventos a respeito do contexto em que ocorrem para poder estender interpretações a outros contextos e grupos. Quando um pesquisador tenta compreender um grupo, ele ou ela é ajudado por conhecer outros grupos; as generalizações são feitas através da capacidade que tenha o pesquisador de mediar entre um grupo e outros. Portanto, a descrição etnográfica de uma escola, por exemplo, deriva seu valor em grande parte do fato de que o pesquisador – assim como os leitores – tem familiaridade com outras escolas, e com escolas em geral (ibid.).

A postura de Erickson (1986, p. 130), como vimos no começo deste texto, é que na pesquisa interpretativa o que se busca não é “universais abstratos” alcançados através de inferências estatísticas de amostras para populações, mas “universais concretos” aos quais se chega estudando um caso com muito detalhe e comparando-o com outros casos estudados com igual detalhe.

Não obstante, nenhuma generalização dessa natureza deve ser considerada final, mas apenas como hipótese de trabalho para outros estudos interpretativos ou como subsídio para questionários, entrevistas ou testes (Taft, 1988, p. 74).

O tema da generalização tem a ver com o da *validade externa* dos estudos, isto é, em que medida os construtos e universais concretos criados pelos pesquisadores são aplicáveis a mais de um grupo? Por outro lado, é necessário considerar também a questão da *validade interna*, ou seja, os pesquisadores estão interpretando o que acreditam interpretar? (Desde o ponto de vista quantitativo, a pergunta seria se os pesquisadores estão medindo o que acreditam medir).

A validade pode ser pensada como uma qualidade das conclusões e dos processos através dos quais são alcançadas, mas seu significado exato depende do critério de verdade que se está utilizando. Considerando que, na abordagem qualitativa, verdade é uma questão de concordância em contexto, o melhor significado de validade nessa abordagem parece ser o de *credibilidade*, como sugere Taft (op. cit., p. 73). A credibilidade depende do convencimento da comunidade de pesquisadores e leitores com relação às evidências apresentadas e aos processos utilizados. Sturman (1988, p. 65) propõe as seguintes estratégias para alcançar credibilidade:

- os procedimentos de coleta de dados devem ser explicados;
- os dados recolhidos devem ser apresentados e estar prontos para reanálise;
- instâncias negativas devem ser relatadas;
- vieses devem ser reconhecidos;
- análises de trabalhos de campo devem ser documentados;
- a relação entre asserção e evidência deve ser esclarecida;
- evidências primárias devem ser distinguidas das secundárias, assim como as descrições das interpretações;
- diários ou logs devem dar conta do que foi feito durante as distintas fases do estudo;
- técnicas devem ser desenhadas para "checar" a qualidade dos dados.

Erickson (1986, p. 140) se reporta ao mesmo tema, desde outra perspectiva, ao indicar cinco tipos de inadequações das evidências apresentadas:

1. *quantidade inadequada de evidências*; o pesquisador tem pouca evidência para garantir certas asserções-chave;
2. *diversidade inadequada de tipos de evidências*; o pesquisador não tem diferentes fontes de dados (por exemplo, observações, entrevistas, documentos); não busca triangulação de dados;
3. *interpretações incorretas das evidências*; o pesquisador não compreende bem certos aspectos-chave da complexidade da ação ou dos significados atribuídos pelos atores no contexto;
4. *inadequadas evidências desconfirmadoras*; o pesquisador não tem dados que possam desconfirmar uma asserção-chave; ou, mais importante, não apresenta evidência de que tenha sido conduzida uma busca deliberada de dados potencialmente desconfirmadores;
5. *análise inadequada de casos discrepantes*; o pesquisador não indaga as instâncias desconfirmadoras nem as compara com as confirmadoras para determinar quais aspectos dos casos desconfirmadores eram iguais ou diferentes dos aspectos análogos dos casos confirmadores.

O mesmo autor (op. cit., p. 145) recomenda que o informe de estudos etnográficos contenha nove elementos principais:

1. *asserções empíricas* (uma tarefa básica da análise de dados é gerar tais afirmações em grande medida através de indução);
2. *vinhetas narrativas analíticas* (uma vinheta narrativa é uma representação vívida de um evento no cotidiano do caso ou grupo pesquisado; a vinheta tenta persuadir o leitor de que as coisas no contexto eram como o autor diz que eram);
3. *citações das notas de campo* (podem-se citar diretamente as notas no informe, indicando a data em que foram tomadas; uma série de extratos de notas de campo podem servir como evidência de que o modo particular em que ocorreu um certo evento foi típico);
4. *citações de entrevistas* (as palavras dos entrevistados são um meio de transmitir aos leitores os pontos de vista dos sujeitos do estudo);
5. *informes sinópticos dos dados* (mapas, tabelas de frequências, diagramas);
6. *comentários interpretativos enquadrando uma certa descrição*;
7. *comentários interpretativos enquadrando a descrição geral* (a descrição geral tem como principal objetivo estabelecer a possibilidade de generalizar os padrões que foram ilustrados nas descrições particulares);
8. *discussão teórica* (comentários interpretativos com relação ao significado mais amplo dos padrões que emergiram dos dados);
9. *informe sobre a história natural da indagação no estudo* (isto é, uma discussão/descrição com relação a como certos conceitos-chave na análise evoluíram ou como padrões não esperados foram encontrados durante o trabalho de campo e na reflexão subsequente).

Os comentários interpretativos que enquadram as descrições particular e geral podem ser de três tipos: os que precedem e seguem uma descrição particular no texto, a discussão teórica que aponta à significatividade mais ampla dos padrões identificados nos eventos mencionados, e uma resenha das mudanças que ocorreram desde o ponto de vista do pesquisador durante o transcurso da indagação (op. cit., p. 152).

Para Erickson (ibid., p. 145), cada um desses nove elementos, separadamente e em conjunto, permitem ao leitor três coisas: em primeiro lugar, possibilitam experimentar de forma vicária o cenário descrito e confrontar instâncias de asserções-chave e construtos analíticos; em segundo, permitem examinar todo o espectro de evidências no qual está baseada a interpretação do pesquisador; e em terceiro, deixam que o leitor considere os fundamentos teóricos e pessoais da perspectiva do autor tal como mudou ao longo do estudo.

Esta seção deste texto foi dedicada ao tema da *fidedignidade, generalização e validade* dos estudos qualitativos. Mesmo que se pudesse argumentar que são conceitos típicos de estudos quantitativos, toda a seção foi desenvolvida com o objetivo de mostrar que tais conceitos têm sentido no contexto de uma pesquisa interpretativa, desde que se dê a eles o significado apropriado. Tal significado, como vimos, parece ser o de credibilidade, o qual nos remete a outro conceito importante na metodologia da pesquisa interpretativa: a *triangulação*, uma estratégia central para alcançar credibilidade (Sturman, 1988, p. 65).

Triangulação

A triangulação pode envolver o uso de diferentes fontes de dados, diferentes perspectivas ou teorias, diferentes pesquisadores ou diferentes métodos; é uma resposta holística à questão da fidedignidade e da validade dos estudos interpretativos (ibid.). Para Denzin (1988, p. 318), a triangulação é o emprego e combinação de várias metodologias de pesquisa no estudo de um mesmo fenômeno.

Não é uma estratégia típica da pesquisa qualitativa, também não é uma estratégia nova: o uso de múltiplas medições e métodos, de modo que se superem as debilidades inerentes ao uso de um único método ou um único instrumento tem uma longa história nas ciências naturais e sociais; na pesquisa quantitativa, a triangulação é usada, tradicionalmente, como uma estratégia de validação de observações (ibid.).

Não obstante, segundo Denzin (op. cit.), a apropriação desse conceito pelos pesquisadores interpretativos e sua aplicação a problemas típicos da pesquisa qualitativa é mais recente e representa um compromisso com um sofisticado rigor metodológico por parte dos pesquisadores, no sentido de que estão comprometidos a tornar seus esquemas empíricos e interpretativos o mais públicos possível. De acordo com esse autor (p. 319), há cinco tipos básicos de triangulação:

1. *triangulação de dados*, envolvendo tempo, espaço e pessoas;
2. *triangulação de pesquisadores*, que consiste no uso de múltiplos observadores, em lugar de um só;
3. *triangulação de teorias*, que consiste em utilizar mais de um esquema teórico na interpretação do fenômeno pesquisado;
4. *triangulação metodológica*, que envolve o uso de mais de um método e pode consistir em estratégias intra métodos ou entre métodos;
5. *triangulação de verificação por sujeitos*, na qual os pesquisados examinam e confirmam ou desconfirmam o que se escreveu sobre eles.

Existe também a *triangulação múltipla*, na qual o pesquisador combina em uma pesquisa múltiplos observadores, perspectivas teóricas, fontes de dados e metodologias.

Por outro lado, mesmo que Denzin associe a triangulação com um compromisso com o rigor metodológico, o uso dessa estratégia na pesquisa qualitativa não está livre de críticas: os argumentos são, por exemplo, que a triangulação de dados tem um viés positivista, que dois pesquisadores nunca observam o mesmo fenômeno da mesma maneira, que diferentes métodos geram diferentes imagens e recortes da realidade e que a triangulação de teorias não tem sentido epistemologicamente.

Porém, o próprio Denzin (op. cit., p. 321) contra-argumenta dizendo que a triangulação nos estudos qualitativos não deve ser comparada com a análise de correlação nas pesquisas quantitativas e que nunca deve ser uma estratégia eclética. Na triangulação de pesquisadores não se espera que observem exatamente da mesma maneira e que um corrobore o que o outro observa, mas que suas diferentes observações expandam a base interpretativa do estudo e que revelem aspectos do fenômeno pesquisado que não seriam necessariamente observados por um único pesquisador.

Com relação à *triangulação metodológica*, o importante é justamente que possam emergir diferentes imagens. Com relação à *triangulação de teorias*, Denzin (ibid.) diz que ela, em vez de requerer que as interpretações sejam consistentes com dois ou mais teóricos, simplesmente requer que o pesquisador seja consciente das diferentes maneiras através das quais pode ser interpretado o fenômeno.

O tema da triangulação, em particular as triangulações teóricas e metodológicas, está muito vinculado à questão dos paradigmas, focada desde uma perspectiva dicotômica no começo deste texto. Nessa oportunidade dissemos que o olhar dicotômico estava sendo usado por razões didáticas e que no final retomariamos o assunto dos paradigmas sob uma visão integradora. Então, chegou o momento de voltar aos paradigmas.

Os paradigmas de pesquisa em educação: rumo à acomodação

A dicotomia estabelecida anteriormente entre os “dois paradigmas clássicos” é simplificadora. A metodologia da pesquisa nas ciências sociais não pode ser pensada simplesmente como “não-positivista” em contraposição a uma suposta tradição positivista da pesquisa nas ciências naturais. Se assim fosse, não teria sentido falar de acomodação de paradigmas. Porém, considerando paradigma como “um conjunto básico de crenças que orienta a ação” (Guba, 1990, apud Alves-Mazzotti, 1996, p. 17), quer dizer, uma concepção de mundo que guia o pesquisador, não só na escolha de método, mas também em suas posições ontológicas e epistemológicas, podem-se distinguir pelo menos três paradigmas como sucessores do positivismo: o *pós-positivismo*, a *teoria crítica* e o *naturalismo/construtivismo* (ibid.).

O primeiro seria uma versão modificada do positivismo revisando pontos insustentáveis (e.g., agora se supõe que a realidade existe, mas que nunca será totalmente apreendida pela pesquisa); a teoria crítica é ideologicamente orientada uma vez que rejeita a neutralidade: o processo de pesquisa é mediado pelo pesquisador, e o termo crítica se refere tanto à crítica interna que resulta do questionamento analítico da argumentação e do método como à análise das condições de regulação social, desigualdade e poder; ao naturalismo/construtivismo subjaz à idéia de que os resultados de qualquer pesquisa são sempre influenciados pela interação pesquisador/pesquisado, de maneira que o conhecimento é sempre produto da atividade humana e, portanto, nunca pode ser visto como algo definitivo, mas como algo que está sempre se modificando (Alves-Mazzotti, 1996, pp. 17-20).

O paradigma naturalista/construtivista é o que enfatizamos neste texto: a realidade é socialmente construída (o que implica que sempre há múltiplas realidades; as idéias, os valores e a interação pesquisador/pesquisado influenciam na configuração dos “fatos”, que implica que a teoria é sub-determinada). Estas características, por sua vez, implicam um relativismo que é problemático para os outros dois paradigmas: *se alguém se propõe compreender os significados atribuídos pelos atores às situações e eventos dos quais participam, se tenta entender a “cultura” de um grupo ou organização no qual coexistem diferentes visões correspondentes aos subgrupos que os compõem, então o relativismo não constitui um problema; mas se nos propomos à construção de teorias (pós-positivismo) ou à transformação social (teoria crítica), o que exige acordo ao redor de decisões ou princípios que possibilitem a ação conjunta, o relativismo passa a ser um problema* (op. cit., p. 21).

Com a identificação destes três paradigmas "pós-clássicos", e possivelmente outros, já se percebe que a questão dos paradigmas em ciências sociais não é dicotômica, que as ciências sociais são multi-paradigmáticas e que a acomodação ou integração de paradigmas não deve ser considerada impossível ou, pelo menos, é uma questão em aberto, como disse Alves-Mazzotti (op.cit., p. 22).

Um bom exemplo de argumento em favor da integração paradigmática é dado por Bericat (1998). Este autor considera que existem três razões fundamentais que podem motivar o desenho multimétodo em uma pesquisa social: *complementação, combinação e triangulação* (p. 37).

A *complementação* existe quando, no marco de um mesmo estudo, se obtém duas imagens, uma procedente de métodos de orientação qualitativa e outra de métodos de orientação quantitativa (ibid.), resultando assim um duplo e diferenciado conjunto de asserções de conhecimento sobre o fenômeno de interesse. O que se obtém são duas perspectivas diferentes sem pretensão alguma de solapamento, ou convergência. As asserções de conhecimento são apresentadas com duas partes bem diferenciadas, cada uma das quais expõe resultados alcançados pela aplicação do respectivo método. Segundo Bericat (ibid.), *na complementação, o grau de integração metodológica é mínimo, e sua legitimidade se suporta sobre a crença de que cada orientação é capaz de revelar diferentes zonas da realidade social, assim como que é necessário contar com essa dupla visão para um melhor entendimento do fenômeno* (ibid.).

Na *combinação*, a estratégia é integrar subsidiariamente uma metodologia, seja a qualitativa ou a quantitativa, na outra, com o objetivo de fortalecer a validade desta compensando suas debilidades mediante a incorporação de informações que procedem da aplicação da outra metodologia. O que se busca não é a convergência de resultados, que finalmente procederão de uma única metodologia, mas bem mais uma adequada combinação metodológica (op. cit., p. 39).

Finalmente, na *triangulação*, o que se pretende é um solapamento ou convergência de resultados. Não se trata de complementar a visão de realidade com dois olhares, mas de utilizar duas metodologias para o estudo de um mesmo e idêntico aspecto de uma realidade social. As metodologias, tal como na complementação, são implementadas de forma independente, mas se focam em um mesmo objeto de estudo buscando resultados convergentes. *A legitimidade desta estratégia depende de se acreditamos que ambas as metodologias realmente podem captar idêntico aspecto da realidade, isto é, se o solapamento é possível. Na medida em que pensemos que conduzem a visões incomensuráveis da realidade, então estaremos no caso da complementação* (op. cit., p. 38).

Na figura 5 estas três estratégias de integração estão esquematizadas em um mapa conceitual. Trata-se sempre de captar aspectos de uma realidade social. Porém, isso se pode tentar com uma única metodologia ou integrando-as de modo complementar, triangular ou combinatório. A triangulação e a combinação só são possíveis na medida em que se aceita pelo menos certo grau de comensurabilidade paradigmática. Caso contrário, a única alternativa é a complementação.

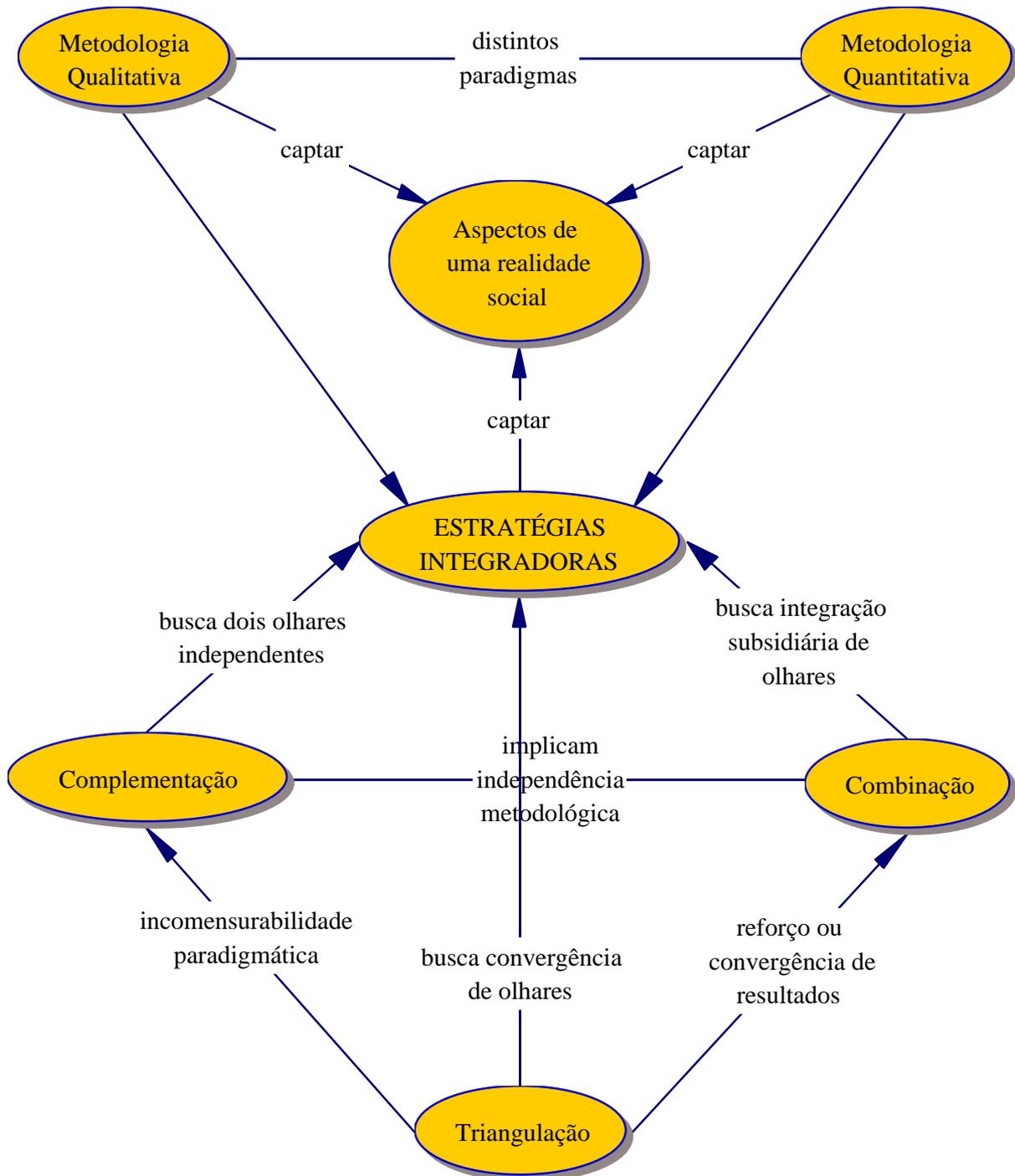


Figura 5: Um mapa conceitual para estratégias integradoras entre a metodologia qualitativa e quantitativa (Moreira, 2002).

Conclusão

Na figura 6 se apresenta, a modo de conclusão deste texto, um diagrama V. Este tipo de diagrama, também conhecido como Ve epistemológico, foi desenhado por D.B. Gowin (1981) para esquematizar a estrutura do processo de produção de conhecimento. Aqui é usado para refletir tal produção no contexto do paradigma qualitativo, particularmente na educação.

Este diagrama pretende ser uma espécie de resumo de todo o texto. Como tal, é preciso levar em conta que nenhum dos itens que aparecem no diagrama V está completo. São dados só exemplos do que poderia integrar cada item. É esta a razão dos três pontos que aparecem no final deles.

Referências

- Alves-Mazzotti, A.J. (1996). O debate atual sobre os paradigmas de pesquisa em educação. *Cadernos de Pesquisa*, São Paulo, n. 96: 15-23.
- André, M.E.D.A. (1998). *Etnografia da prática escolar*. 2ª ed. São Paulo, Papirus Editora.
- Ball, S.J. (1988). Participant observation. In Keeves, J.P. (Ed). *Educational research, methodology, and measurement. An international handbook*. Oxford, Pergamon Press. pp. 310-314.
- Bericat, E. (1988). *La integración de los métodos cuantitativo y cualitativo en la investigación social*. Barcelona, Editorial Ariel.
- Denzin, N.K. (1988). Triangulation in educational research. In Keeves, J.P. (Ed). *Educational research, methodology, and measurement. An international handbook*. Oxford, Pergamon Press. pp. 318-322.
- Eisner, E.W. (1981). On the differences between scientific and artistic approaches to qualitative research. *Educational Researcher*, 10(4): 5-9.
- Elliott, J. (1993). *El cambio educativo desde la investigación-acción*. Madrid, Ediciones Morata.
- Erickson, F. (1986). Qualitative methods in research on teaching. In Wittrock, M.C. (Ed.). *Handbook of research on teaching*. New York: Macmillan Publishing Co. p. 119-161. Traducción al español: Erickson, F. (1989) Métodos cualitativos de investigación sobre la enseñanza. In Wittrock, M.C. (Comp.). *La investigación en la enseñanza, II*. Barcelona, Paidós. pp. 195-301.
- Firestone, W.A. (1987). Meaning in method: the rethoric of quantitative and qualitative research. *Educational Researcher*, 16(7): 16-21.
- Goetz, J.P. y Lecompte, M.D. (1988). *Etnografía y diseño cualitativo en investigación educativa*. Madrid, Ediciones Morata.
- Guba, E.G. (1990). The alternative paradigm dialog. In: Guba, E.G. (Ed.) *The paradigm dialog*. London, Sage. Apud: Alves-Mazzotti (1996). O debate atual sobre os paradigmas de pesquisa em educação. *Cadernos de Pesquisa*, São Paulo, n. 96: 15-23.
- Hall, B.L. (1988). Participatory research. In Keeves, J.P. (Ed). *Educational research, methodology, and measurement. An international handbook*. Oxford, Pergamon Press. pp. 198-204.

Domínio Conceitual

Domínio Metodológico

Tipos de perguntas, temas de interesse



Figura 6. Um diagrama V para a pesquisa qualitativa, particularmente em educação (M.A. Moreira, 2002)

- Husén, T. (1988). Research paradigms in education. In Keeves, J.P. (Ed). *Educational research, methodology, and measurement. An international handbook*. Oxford, Pergamon Press. p. 16-21.
- Hyams, B.K. (1988). *Oral history*. In Keeves, J.P. (Ed). *Educational research, methodology, and measurement. An international handbook*. Oxford, Pergamon Press.
- Keeves, J.P. (1988). Towards a unified view of educational research. In Keeves, J.P. (Ed). *Educational research, methodology, and measurement. An international handbook*. Oxford, Pergamon Press. pp. 1-7.
- Kemmis, S. (1988). Action research. In Keeves, J.P. (Ed). *Educational research, methodology, and measurement. An international handbook*. Oxford, Pergamon Press. pp. 173-179.
- Kemmis, S. and McTaggart, R. (Eds.). (1988). *The action research reader*. 3rd ed. Geelong, Deakin University Press.
- Landsheere, G. de (1988). History of educational research. In Keeves, J.P. (Ed). *Educational research, methodology, and measurement. An international handbook*. Oxford, Pergamon Press. pp. 8-16.
- Marshall, C. (1984). The wrong time for mechanistics in qualitative research. *Educational Researcher*, 13(9): 26-28.
- Marton, F. (1988). Phenomenography. In Keeves, J.P. (Ed). *Educational research, methodology, and measurement. An international handbook*. Oxford, Pergamon Press. pp. 95-101.
- Miles, M.B. and Huberman, A.M. (1984) Drawing valid meaning from qualitative data: toward a shared craft. *Educational Researcher*, 13(5): 20-30.
- Moreira, M.A. (2000). Investigación en enseñanza: aspectos metodológicos. In *Actas de la I Escuela de Verano sobre Investigación en Enseñanza de las Ciencias*. Burgos, Servicio de Publicaciones de la Universidad de Burgos. pp. 13-51.
- Ödman, P.J. & Kederman, D. (1988). Hermeneutics. In Keeves, J.P. (Ed). *Educational research, methodology, and measurement. An international handbook*. Oxford, Pergamon Press. pp. 185-192.
- Ogbu, J.U., Sato, N.E. and Kim, E.Y. (1988). Anthropological inquiry. In Keeves, J.P. (Ed). *Educational research, methodology, and measurement. An international handbook*. Oxford, Pergamon Press. pp. 48-54.
- Serrano, G.P. (1998). Investigación cualitativa. Retos e interrogantes. I. Métodos. Madrid, La Muralla S.A.
- Shulman, L.S. (1981). Disciplines of inquiry in education: an overview. *Educational Researcher*, 10(6): 5-12.
- Smith, J.K. (1983). Quantitative versus qualitative research: an attempt to clarify the issue. *Educational Researcher*, 12(3): 6-13.
- Smith, J.K. and Heshusius, L. (1986). Closing down the conversation: the end of the quantitative-qualitative debate among educational inquirers. *Educational Researcher*, 15(1): 4-13.
- Stenhouse, L. (1985). Case study methods. In Husén, T. & Postlethwaite, T.N. (Eds.). *International Encyclopedia of Education*. Oxford, Pergamon Press.

- Sturman, A. (1988) Case study methods. In Keeves, J.P. (Ed). *Educational research, methodology, and measurement. An international handbook*. Oxford, Pergamon Press. pp. 61-66.
- Taft, R. (1988). Ethnographic research methods. In Keeves, J.P. (Ed). *Educational research, methodology, and measurement. An international handbook*. Oxford, Pergamon Press. pp. 71-75.
- Walker, J.C. & Evers, C.W. (1988). Research in education: epistemological issues. In Keeves, J.P. (Ed). *Educational research, methodology, and measurement. An international handbook*. Oxford, Pergamon Press. pp. 22-31.

Capítulo 2

Pesquisa em Ensino: Métodos Quantitativos¹

M. A. Moreira

P. R. S. Rosa

Resumo

A finalidade deste texto é a de resgatar a metodologia quantitativa aplicada à pesquisa em ensino. Não há nele nada de novo. Todos os tópicos enfocados encontram-se muito mais elaborados em livros de metodologia da pesquisa ou de Estatística. Contudo, chamar atenção para questões como tratamento de variáveis, delineamentos de pesquisa, amostragem, significância estatística, fidedignidade e validade, análise de variância, testes não-paramétricos, pode ser uma contribuição para a atual pesquisa em ensino que, há muito tempo, usa praticamente apenas métodos qualitativos e, assim fazendo, acaba enviesada.

Introdução

D. B. Gowin (1981, 2005) propôs um dispositivo heurístico – que hoje é conhecido como diagrama V (Moreira, 2006) – para desvelar a estrutura do processo de produção do conhecimento. A Figura 1 apresenta esse dispositivo aplicado à pesquisa quantitativa em educação. O centro desse diagrama é o *domínio de interesse* da pesquisa (por exemplo, o ensino e a aprendizagem) e as perguntas e hipóteses que se faz dentro desse domínio. Para respondê-las, confirmá-las ou refutá-las é preciso fazer registros de um evento que se faz acontecer ou que acontece naturalmente. O lado esquerdo do diagrama é o *domínio conceitual* (ou teórico) da pesquisa; nele estão conceitos, princípios, modelos, teorias e filosofias que fundamentam teórica e epistemologicamente a pesquisa e que interagem com os registros, transformações e asserções que constituem o *domínio metodológico* que aparece no lado direito do diagrama.

Neste texto o foco estará no domínio metodológico. Como se vê na Figura 1, o lado metodológico começa com os registros. Sem registros não se faz pesquisa empírica. A partir daí, um passo fundamental desse tipo de pesquisa é a conversão desses registros em índices numéricos.

Por exemplo, no caso de que os registros sejam mapas conceituais é preciso definir critérios como, digamos, tantos pontos para a hierarquia, tantos para conectivos, etc., para chegar a um score para cada mapa. Ou, então, definir categorias como, suponhamos, muito bom, bom, regular, falho, e atribuir pontos a cada categoria.

¹ Texto de Apoio Nº 29, Programa Internacional de Doctorado en Enseñanza de las Ciencias da Universidade de Burgos, Espanha, desenvolvido no marco de um convênio com a Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil. Publicado em Actas del PIDEAC, 2007, Vol. 9:03-56.

É certo que também se pode trabalhar quantitativamente com índices não numéricos como, por exemplo, variáveis dicotômicas do tipo sim ou não, feminino ou masculino, mas na pesquisa empírica predomina o uso de índices numéricos.

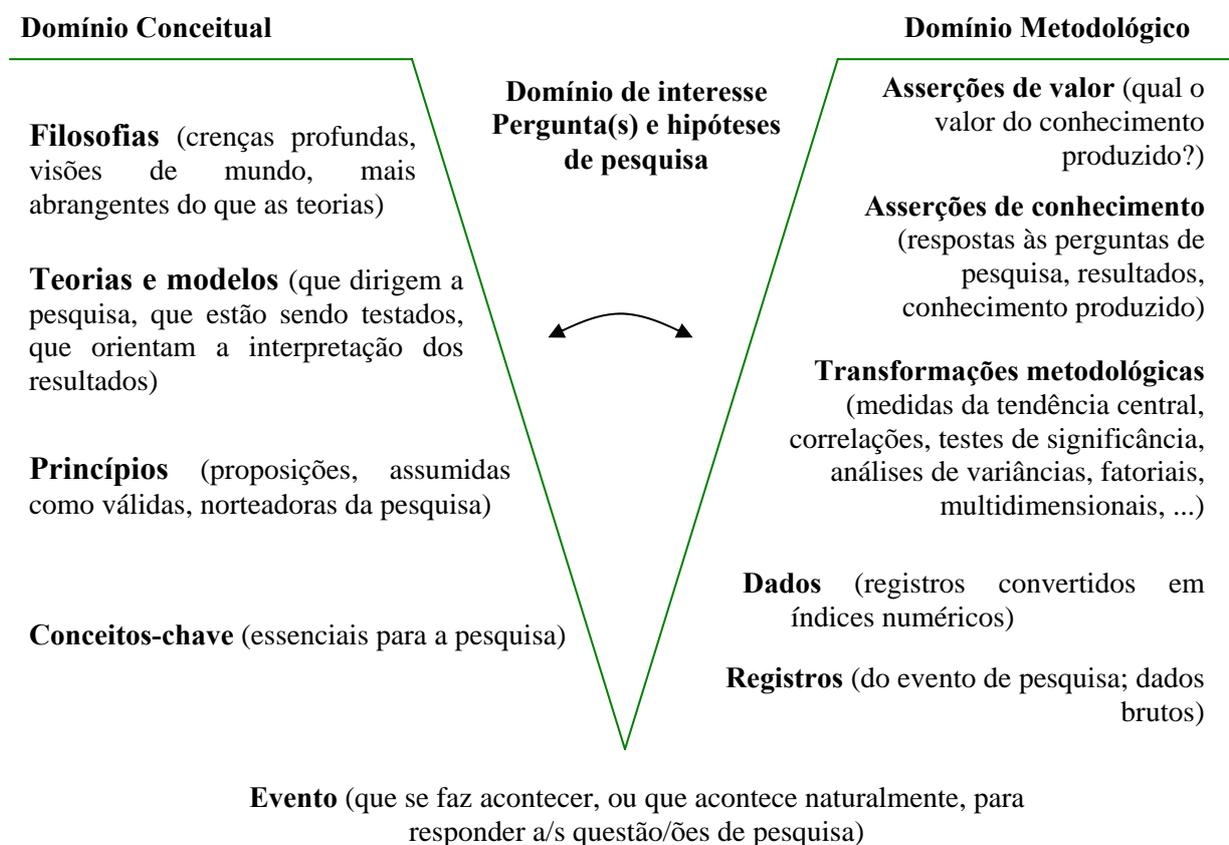


Figura 1 – O diagrama V aplicado à pesquisa quantitativa em educação.

É igualmente correto que o mais importante na pesquisa, seja ela de qualquer natureza, é a pergunta de pesquisa, aquela cuja busca de respostas gera conhecimentos. O conhecimento humano é construído, reconstruído, refutado, modificado, sempre pela busca, muitas vezes obstinada, de respostas a perguntas sobre determinados fenômenos de interesse.

A identificação de uma questão de pesquisa que valha a pena investigar, que possa gerar conhecimentos, é a parte mais difícil da pesquisa. No entanto, é também muito importante um delineamento de pesquisa que permita fazer registros relevantes que, por sua vez, originem dados (tipicamente índices numéricos, na pesquisa quantitativa) de máxima relevância à pergunta de pesquisa.

Dados e tratamento estatístico

Tipicamente, na pesquisa educacional empírica, o pesquisador se questiona sobre qual a evidência que o apoiará em relação a certas hipóteses de pesquisa. Faz, então, registros, converte-os em números e trata-os estatisticamente para ver se servem como evidência.

Costuma-se dizer que bons dados falam por si mesmos. Isso significa que se os dados forem de boa qualidade os procedimentos estatísticos são imediatos. Por outro lado, se os dados forem ruins não há tratamento estatístico que os transforme em bons.

O importante é a qualidade dos dados não as manipulações estatísticas. A relevância das conclusões estatísticas nunca será maior do que a adequação dos dados numéricos trabalhados estatisticamente. A interpretação dos resultados estatísticos depende do que está por detrás dos dados. Os números a serem analisados não são entidades sagradas representando puras abstrações. Ao contrário, sua utilidade na pesquisa empírica reside no fato que têm referentes no contexto da pesquisa, que significam algo no mundo real (Millman, 1970).

Uma boa estratégia para melhor abordar a questão da análise e qualidade dos dados é considerá-la antes, ou seja, pensar na análise dos dados antes de coletá-los evitando-se, assim, que os números obtidos não tenham muito a ver com as questões investigadas.

Outro aspecto a ser considerado nessa questão é o da fidedignidade e validade dos instrumentos. Sem instrumentos fidedignos e válidos os números que deles resultarem não serão confiáveis e de nada servirá tratá-los estatisticamente.

No que se refere a instrumentos é conveniente testá-los previamente, perguntando a alguns sujeitos como interpretam determinados itens, ou fazer uma análise do tipo “tem sentido atribuir o mesmo número a uma não-resposta e a uma

resposta neutra?”, quer dizer, “um item deixado em branco pode ser considerado como resposta neutra?”, “sem opinião é o mesmo que opinião neutra?”(op.cit.).

Significância estatística e significância prática

Na análise dos dados é importante levar em conta que o nível de significância estatística ($p < 0,05$, por exemplo) não é uma medida da importância ou da significância prática de um resultado, pois esse nível depende do número de casos e da eficiência do delineamento de pesquisa.

Quando um resultado é estatisticamente significativo o pesquisador deve analisar a magnitude dos efeitos, ou seja, qual a importância, no contexto da investigação, de uma diferença de médias, estatisticamente significativa, por exemplo, entre 7,5 e 7,2, ou qual a relevância de uma correlação, estatisticamente significativa ao nível 0,05, de um coeficiente de correlação de 0,23 entre duas variáveis?

Outros aspectos a serem considerados na questão da significância estatística e significância prática são o tamanho e a variabilidade da amostra. Mesmo que se obtenham diferenças, correlações ou fatores estatisticamente significativos para amostras de 12 ou 13 sujeitos, é preciso perguntar-se qual a significância prática desses resultados. Por outro lado, amostras pequenas e, inclusive, amostras maiores de 30 ou 40 sujeitos, por exemplo, podem apresentar grande variabilidade. Em uma amostra de 10 sujeitos se dois tiverem o escore máximo de dez pontos, dois obtiverem zero pontos e os demais alcançarem cinco ou seis pontos, a média será cinco ou mais pontos, mas o que significa essa média na prática? É claro que nesse caso é melhor trabalhar com a moda, mas mesmo assim é preciso perguntar qual a significância prática desse resultado. Analogamente, esse questionamento deve ser feito quando é grande a variabilidade dos escores em amostras maiores.

A estatística apropriada

Na escolha da técnica estatística adequada para o tratamento dos índices numéricos é conveniente considerar primeiramente se o que se quer é *descrever* características de um conjunto de números ou se o que se pretende é *estimar* valores da população. No primeiro caso, a estatística a ser usada é a descritiva, no segundo a inferencial.

É também conveniente voltar à questão de pesquisa e, tendo em conta sua natureza, considerar se medidas da tendência central, correlações, testes de significância para diferenças de médias, análises de variância, análises fatoriais ou multidimensionais são, de fato, necessárias para respondê-la.

O papel do computador

Há algoritmos de computador para realizar quaisquer análises estatísticas de dados numéricos. Basta injetar esses dados nesses algoritmos para que rapidamente saiam médias, desvios padrão, coeficientes de correlação, variâncias, fatores, etc., etc.. O computador atende comandos e seus procedimentos atuam sobre os dados que lhe são fornecidos. E aí voltamos ao começo desse assunto de análise de dados: se os dados forem ruins, não há computador que gere bons resultados. Não adianta ficar dando comandos para que ele gere mais e mais coeficientes, tabelas, fatores. Tudo lixo!

Por outro lado, se os dados forem bons é preciso saber interpretar os produtos dos procedimentos estatísticos aos quais eles são submetidos. Não é preciso usar lápis e papel, ou calculadora, para, tediosamente, calcular estatísticas (valores da amostra), parâmetros (valores da população), coeficientes, correlações, fatores. O computador faz isso muito mais rapidamente e sem erros. Mas é fundamental saber interpretar os resultados (op.cit.).

A análise dos dados é, como foi dito no início, uma etapa fundamental da pesquisa empírica. Sem dúvida. Mas mais importante do que ela é a questão de pesquisa e os registros que se faz dos eventos usados para responder essa questão. São esses registros que serão convertidos em índices numéricos que, por sua vez, serão analisados estatisticamente. Os procedimentos estatísticos estão disponíveis em profusão e o computador os executa rapidamente. Mas a análise, em si, tem que ser feita pelo pesquisador.

Passemos agora a descrever e tratar de aspectos e técnicas dos métodos quantitativos de pesquisa em educação.

Variáveis

Condições que podem ser variadas ou selecionadas pelo investigador são chamadas de **variáveis independentes**. As medidas das respostas feitas durante o experimento constituem as **variáveis dependentes**.

Os **níveis** de uma variável referem-se a categorias da variável. Sexo, por exemplo, tem dois níveis. O número de níveis de uma variável como idade pode ser arbitrariamente determinado e pode variar de dois (e.g., acima de 35 ou abaixo ou igual a 35 anos) ao infinito (quando idade é tratada como uma variável contínua).

A palavra *fator* é frequentemente usada como sinônimo da expressão variável independente. Uma **variável independente manipulada** é uma condição que está sob controle direto do experimentador. **Variáveis de tratamento** são, normalmente, variáveis independentes manipuladas pelo pesquisador, cujos efeitos ele está querendo observar. Uma **variável independente normativa** é aquela na qual o pesquisador não está livre para produzir a condição em si mesma, embora ele seja livre para decidir quais níveis da variável serão incluídos na investigação². Sexo e idade são exemplos de variáveis deste tipo.

Medidas iniciais, ou seja, obtidas antes do início da investigação, que são usadas para formar grupos homogêneos (relativamente à variável dependente) são chamadas de **variáveis de agrupamento**. Observações suplementares às observações antes do tratamento relativamente a possíveis diferenças são chamadas de **observações concomitantes** ou **covariáveis**. Uma observação concomitante pode ser usada como uma alternativa ao agrupamento em delineamentos experimentais³ ou, em alguns delineamentos quase-experimentais, ela pode ser usada como um esforço para superar as deficiências do delineamento.

Outro tipo de caracterização de variáveis diz respeito à natureza do processo de contagem da variável. Podemos definir quatro grupos de variáveis, com respeito à escala usada para medir as variáveis:

1. **Variáveis nominais** - estamos lidando com esse tipo de variável quando apenas apontamos se a grandeza medida pela variável está presente ou não. Assim, por exemplo, uma pessoa pode ser professor ou não. Normalmente, atribui-se valor 1 se o atributo medido pela variável está presente e atribui-se o valor 0 se o atributo não está presente. Ou seja, neste caso, professor (1) – não professor (0).
2. **Variáveis ordinais** - são aquelas onde os dados apresentam algum tipo de ordenação. É atribuído um grau, conforme alguma escala, a cada sujeito a medida em que ele apresenta o atributo que está sendo medido. Por exemplo, um teste classificatório de proficiência em matemática, onde os sujeitos são ordenados de acordo com as notas obtidas no teste.
3. **Variáveis intervalares** - este tipo de variável se caracteriza por valores que estão distribuídos em uma escala com uma diferença constante entre dois valores consecutivos. Idade pode ser um exemplo deste tipo de variável, se contarmos apenas os anos efetivamente completados.
4. **Variáveis racionais** - são aquelas onde a escala de medida é composta por números racionais e, além disso, existe um zero que define a ausência da propriedade medida pela variável. Por exemplo, notas em um teste são variáveis desse tipo se atribuirmos valores fracionários às respostas⁴.

Controle de variáveis

Variáveis que não são de interesse direto do pesquisador podem ser removidas ou terem a sua influência minimizada por vários métodos:

² Este tipo de variável também é chamado de parâmetro.

³ Aqueles nos quais a amostragem é aleatória; ver p. 38.

⁴ Se, por outro lado, apenas considerarmos cada questão como certa (1 ponto) ou errada (0 pontos) então teremos uma escala intervalar.

1. **Remoção de variáveis** - algumas variáveis podem ser eliminadas selecionando-se casos com características uniformes (usando-se, por exemplo, apenas mulheres para eliminar-se a influência da variável sexo).
2. **Pareamento de casos** - selecionando-se pares ou conjuntos de indivíduos com características idênticas (ou quase idênticas) e distribuindo-os aos grupos experimental e de controle. Entretanto, o pareamento não é considerado satisfatório a menos que os sujeitos dos pares ou conjuntos sejam distribuídos aleatoriamente aos grupos experimental ou de controle. Uma limitação desse método ocorre pela dificuldade em parear-se sujeitos usando-se duas ou mais variáveis.
3. **Balanceamento de casos** - distribui-se os sujeitos aos grupos experimental e de controle de tal modo que médias e variâncias dos grupos sejam semelhantes, tanto quanto possível. Este método também apresenta uma dificuldade similar àquela observada no pareamento de casos: a dificuldade de equacionar grupos com base em mais do que uma variável.
4. **Análise de covariância** - este método permite ao experimentador eliminar diferenças iniciais em várias variáveis entre os grupos experimental e de controle por métodos estatísticos. Usando escores de pré-testes como co-variáveis, este método é considerado preferível ao convencional pareamento de grupos.
5. **Aleatoriedade** - a aleatoriedade pode ser obtida através da seleção ao acaso dos sujeitos, dentre aqueles da população que se quer estudar, que vão participar dos grupos de controle e experimental. A aleatoriedade nos dá um método efetivo de eliminar erros sistemáticos e de minimizar o efeito de variáveis externas. O princípio da aleatoriedade é baseado na hipótese de que através da seleção aleatória quaisquer diferenças entre os grupos seja simplesmente devido à amostragem ou ao acaso. Essas diferenças são conhecidas como **erros de amostragem** ou **erros de variância** e suas intensidades podem ser estimadas pelo pesquisador. Em um experimento, diferenças na variável dependente que podem ser atribuídas ao efeito da variável independente são conhecidas como **variância experimental**. A significância de um experimento pode ser testada comparando-se a variância experimental com o erro de variância. Se ao término do experimento as diferenças entre os grupos experimental e de controle forem muito grandes para serem atribuídas ao erro de variância pode-se presumir que estas diferenças são atribuíveis à variância experimental. A aleatoriedade é o método mais efetivo de formar-se grupos e controlar-se variáveis externas e deve ser usado sempre que as circunstâncias o permitirem. (Best, 1970).

Delineamentos de pesquisa

Entende-se por **delineamento** de uma pesquisa ao conjunto composto pelo plano de trabalho do pesquisador, a maneira como este seleciona as suas amostras e analisa os seus dados. Pode-se dizer que de nada valem a observação cuidadosa e a exaustiva e detalhada análise estatística se isto for feito para um plano de pesquisa inadequado à situação em estudo. Convém lembrar o que já foi ressaltado neste texto: **não** é uma boa estatística que torna boa uma pesquisa.

O assunto do delineamento experimental encontra-se muito bem desenvolvido e apresentado na obra de Campbell e Stanley (1963, 1991), livro, aliás, de leitura obrigatória

para o pesquisador interessado em métodos quantitativos. Aqui será feito apenas um resumo da classificação de Campbell e Stanley.

Seguir-se-á aqui a notação clássica de Campbell e Stanley no que se refere à designação de observações e tratamentos. Designaremos pela letra O uma observação. Um sub-índice na letra O indica uma observação particular de uma série, *não necessariamente em ordem cronológica*. O índice funciona apenas como um rótulo para uma dada observação. Designaremos pela letra X a um tratamento. Então, como exemplo, a sequência abaixo:

$$O_1 X O_2$$

indica que foi feita uma observação (denotada por O_1) em seguida foi aplicado um tratamento X e então feita uma segunda observação (denotada por O_2). Quando na frente de uma sequência de observações e tratamentos estiver a letra A isto significa que as amostras foram selecionadas aleatoriamente. Cada sequência que se encontra em uma linha diz respeito a um mesmo grupo de sujeitos (a amostra). Assim a sequência abaixo:

$$\begin{array}{cccc} A & O_1 & X & O_2 \\ A & O_3 & & O_4 \end{array}$$

é lida como: observa-se um grupo de sujeitos uma vez (O_1) submete-se então o grupo de sujeitos a um determinado tratamento X e então observa-se o mesmo grupo de sujeitos uma segunda vez (O_2). Este grupo de sujeitos é chamado de *grupo experimental*. A segunda linha do delineamento significa que observa-se um segundo grupo uma vez (O_3), não se aplica o tratamento⁵ e então observa-se o grupo de sujeitos uma segunda vez (O_4). A esse segundo grupo é dado o nome de *grupo de controle*. A letra A indica que os dois grupos, de controle e experimental, foram escolhidos aleatoriamente. Outra convenção adotada é a de que quando duas letras se encontram na mesma coluna significa que os eventos aconteceram simultaneamente no tempo. Assim, no exemplo, as observações O_1 e O_3 aconteceram no *mesmo instante de tempo*⁶.

Seguindo a classificação de Campbell e Stanley os delineamentos de pesquisa podem ser divididos segundo três classes:

- **Delineamentos não-experimentais.**
- **Delineamentos experimentais.**
- **Delineamentos quase-experimentais.**

Por *experimental*, que é o adjetivo comum a todas as classes colocadas acima, entendemos condições *controladas* de pesquisa. Ou seja, o pesquisador deve ser capaz de controlar ou de levar em conta de forma apropriada todas as variáveis pertinentes a um determinado estudo. Os fatores de validade de cada experimento podem ser classificados como fatores de validade interna, os quais dizem respeito às variáveis que se não controladas tornam sem significância quaisquer asserções de conhecimento sobre os resultados do experimento, ou fatores de validade externa, os quais se não controlados invalidam a

⁵ De fato essa denominação é arbitrária uma vez que um não-tratamento também é um tratamento.

⁶ Claro que esta afirmação deve ser entendida como aproximadamente no mesmo instante de tempo, podendo haver um certo intervalo de tempo entre as observações.

generalização dos resultados do experimento para uma determinada população. Um determinado experimento somente é verdadeiramente experimental se controlar a totalidade das variáveis que influenciam os fatores de validade, interna e/ou externa. Obviamente, quando se trata de pesquisa em Ciências Sociais, nem sempre é possível controlar-se apropriadamente todas as variáveis envolvidas em determinada situação. Portanto, o estudo será tão mais próximo de um experimento verdadeiro quanto mais controle o experimentador tiver sobre estas variáveis. A linha básica de raciocínio para classificar-se determinado delineamento numa ou noutra categoria é o quanto mais controle o delineamento oferece dos fatores de validade interna e externa.

Delineamentos que oferecem pouco ou nenhum controle das variáveis pertinentes são chamados de *delineamentos não-experimentais* ou *pré-experimentais*. Por outro lado, delineamentos que oferecem alto grau de controle são chamados de *delineamentos experimentais*. Por fim, delineamentos que oferecem grau de controle em nível médio, porém sem oferecer controle nos níveis da categoria anterior, são chamados de *delineamentos quase-experimentais*.

Na óptica quantitativa, o pesquisador deve **sempre** procurar um delineamento experimental para o seu trabalho. Na impossibilidade de um tratamento deste tipo é aceitável um delineamento quase-experimental. Um tratamento não-experimental **nunca é aceitável**.

Delineamentos não-experimentais ou pré-experimentais

Dentro da classificação de Campbell e Stanley os delineamentos não experimentais ou pré-experimentais são de três tipos:

Delineamento de tipo 1 - Neste delineamento é observado apenas um grupo sob a ação do tratamento X. Esse delineamento é esquematizado como:

$$X \quad O_1$$

ou seja, nesse delineamento o grupo que experimentou o tratamento X, e apenas ele, é observado uma única vez.

As desvantagens desse delineamento são evidentes. Talvez a maior delas seja o fato de que não se ter controle algum sobre as variáveis externas que atuam concomitantemente com X. Variáveis tais como *história*, *maturação*, *interação do experimentador com o tratamento*, etc., não são de forma alguma controladas. Não há razão plausível para a utilização desse delineamento, devendo-se evitá-lo a todo o custo. Um exemplo desse tipo de delineamento é aquela situação onde o professor aplica um novo método de ensino e então uma prova. Nenhuma consequência que se tire do resultado da prova é válida devido às deficiências desse delineamento.

Delineamento de tipo 2 - Um delineamento muito usado na pesquisa em ensino, mas que na verdade é um delineamento pré-experimental, é o seguinte:

$$O_1 \quad X \quad O_2$$

Nesse delineamento, aplica-se um pré-teste O_1 a um grupo, submete-se o grupo a um tratamento X e aplica-se, então, um pós-teste O_2 . O_1 e O_2 significam que o mesmo grupo é

observado antes e depois do tratamento que pode ser, por exemplo, um novo método de ensino ou um recurso didático alternativo. Diferenças entre O_1 e O_2 (que podem ser simples testes de conhecimento) evidenciarão a eficácia ou ineficácia do tratamento X . O problema com esse delineamento é que não controla outras variáveis, além de X , que poderiam explicar as diferenças entre O_1 e O_2 . Por exemplo, os alunos poderiam ter melhores resultados no pós-teste porque algum evento ocorreu entre a aplicação do pré-teste e do pós-teste (variável história) e não porque o tratamento X tenha sido eficiente.

Delineamento de tipo 3 – Deve-se ter cuidado em não confundir este delineamento com um dos delineamentos experimentais que vão ser descritos mais adiante. Nesse delineamento, temos dois grupos, experimental e de controle, mas a seleção dos sujeitos que pertencem aos dois grupos **não é aleatória**. Dessa forma esse delineamento não controla a variável *seleção*. Esse delineamento tem a forma:

$$\begin{array}{c} X \quad O_1 \\ \quad O_2 \end{array}$$

onde a ausência da letra A à esquerda de X significa que não houve aleatoriedade no processo de seleção. Nesse tipo de delineamento não há evidência alguma da equivalência entre os dois grupos antes do início do experimento. Um exemplo desse tipo de delineamento ocorre quando seleciona-se dois grupos de sujeitos de determinada escola, para pertencerem aos grupos experimental e de controle, pelo simples fato de pertencerem à mesma turma. Se as turmas forem formadas por alunos que no ano anterior foram bons alunos em Matemática ou maus alunos nesta disciplina então a variável seleção com certeza influenciará qualquer tratamento alternativo sobre o ensino de Matemática que for aplicado a um ou outro grupo.

Delineamentos experimentais

Os delineamentos experimentais são aqueles onde consegue-se controlar a maior parte, senão todas, as fontes de invalidade interna e externa. Seguindo a tradição de Campbell e Stanley, estes delineamentos podem ser classificados em três categorias.

Delineamento de tipo 4 - Um delineamento experimental muito usado é o seguinte:

$$\begin{array}{cccc} A & O_1 & X & O_2 \\ A & O_3 & & O_4 \end{array}$$

Neste delineamento trabalha-se com dois grupos e os sujeitos da pesquisa são designados *aleatoriamente* a um deles (este é o significado de A). Observa-se ambos os grupos antes da aplicação do tratamento X , por exemplo, aplicando-se um pré-teste a ambos os grupos ($O_1 = O_3$). Um dos grupos (grupo experimental) é então submetido ao tratamento X , enquanto o outro (grupo de controle) não recebe o tratamento. Após, observa-se os grupos, aplicando-se, por exemplo, um pós-teste ($O_2 = O_4$) a ambos os grupos. Na prática, os pré e pós-teste podem ser iguais.

Um erro comum no uso de delineamentos desse tipo é analisar-se o resultado para determinação da eficácia do tratamento do seguinte modo: tomam-se as diferenças entre os resultados do pré e pós-teste em ambos os grupos ($O_2 - O_1$ e $O_4 - O_3$), aplicando-se a seguir um teste estatístico. Se a diferença entre as médias do grupo experimental antes e depois da aplicação do tratamento X , for estatisticamente significativa e a diferença entre as médias do

grupo de controle não for significativa, toma-se então o tratamento como eficaz. Esta é uma forma errada de analisar-se a eficácia do tratamento e não fornece evidência alguma sobre o efeito do tratamento X . A forma correta de proceder-se é comparar-se o resultado *final* (as médias finais em um teste de conhecimento por exemplo) entre os grupos experimental e de controle entre si.

Este delineamento controla variáveis na medida em que elas influenciarão igualmente ambos os grupos, exceto X , obviamente, e, portanto, os efeitos dessas variáveis não pesarão na comparação das diferenças $O_2 - O_1$ e $O_4 - O_3$.

Além disso, a aleatoriedade da designação dos sujeitos a um dos grupos, embora não garanta equivalência entre os grupos em 100 %, reduz ao mínimo a probabilidade de que sejam diferentes. Segundo Kerlinger (1980, p. 102):

“Casualização é a designação de objetos (sujeitos, tratamentos, grupos) de um universo a subconjuntos do universo de tal maneira que, para qualquer designação dada a um subconjunto, todo membro do universo tem igual probabilidade de ser escolhido para a designação. Não há total garantia de que a casualização ‘igualará’ os grupos, mas a probabilidade de igualar é relativamente alta. Há outra forma de expressar essa idéia: [...] já que em procedimentos aleatórios todo membro de uma população tem igual probabilidade de ser escolhido, membros com certas características distintas – homem ou mulher, alto ou baixo grau de inteligência, dogmático ou não dogmático, e assim por diante – se selecionados, provavelmente serão contrabalançados a longo prazo pela seleção de outros membros da população com a quantidade ou qualidade ‘opostas’ da característica.”

Delineamento tipo 5⁷ - A aleatoriedade da designação de sujeitos aos grupos de controle e experimental é, portanto, a mais adequada segurança de que não existam diferenças ou vieses iniciais entre os grupos. Nesse caso, o pré-teste não é condição essencial para que um delineamento seja verdadeiramente experimental. Assim, o delineamento anteriormente apresentado poderia ser simplesmente:

A	X	O_1
A		O_2

De fato, esse delineamento não só pode ser usado ao invés do anterior como também é mais adequado pois elimina qualquer influência do pré-teste no experimento. Entretanto, talvez por razões psicológicas, muitos pesquisadores não abrem mão de saber “com certeza” se os grupos experimental e de controle eram iguais no início do experimento, de modo que o quarto exemplo de delineamento aqui apresentado é provavelmente mais usado que o quinto, embora menos apropriado logicamente.

Delineamento tipo 6 (Delineamento de quatro grupos de Solomon) - Este delineamento é a soma das vantagens dos delineamentos quatro e cinco. Seu esquema é:

⁷ Aqui se está fazendo uma inversão entre a denominação dada por Campbell e Stanley aos delineamentos 5 e 6. Para Campbell e Stanley, o que está sendo chamado de delineamento 5 é o delineamento 6 e vice-versa.

A	O_1	X	O_2
A	O_3		O_4
A		X	O_5
A			O_6

Este tipo de delineamento controla variáveis como *interação do pré-teste com o tratamento, maturação e história*. A desvantagem desse tipo de delineamento é a dificuldade em obter-se tantos grupos para participar da pesquisa.

Delineamentos quase-experimentais

Um terceiro grupo de delineamentos identificado por Campbell e Stanley é o dos delineamentos quase-experimentais, ou seja, aqueles em que falta ao pesquisador “o pleno controle da aplicação dos estímulos experimentais - quando e quem expor e a capacidade de casualizar exposições” (op. cit., p. 61). Todos os delineamentos pertencentes a esse grupo carecem do rigor e controle existentes nos delineamentos pertencentes ao grupo dos delineamentos experimentais, mas podem ser usados quando a situação não permitir o uso de delineamentos verdadeiramente experimentais.

Delineamento tipo 7 (Série temporal) - O delineamento “série temporal” exemplifica essa situação:

$$O_1 O_2 O_3 O_4 X O_5 O_6 O_7 O_8$$

Neste delineamento, os sujeitos são observados várias vezes antes de se aplicar o tratamento X e várias vezes após a aplicação. Supondo-se que antes do tratamento as observações fossem quase que homogêneas, sem variações, apresentando um padrão bem definido e que houvesse um salto nos escores das observações feitas após o tratamento e que, a partir daí, houvesse nova estabilização nos escores das observações, com a apresentação de um outro padrão, esse salto quantitativo na série temporal seria tomado como evidência do efeito X.

Observe-se que esse delineamento é semelhante ao primeiro apresentado como exemplo, porém implica em muito mais observações, o que minimiza, embora não exclua, as deficiências do primeiro. Note-se também que implica na existência de um só grupo, o que, na prática, é uma vantagem pois muitas vezes é difícil obter-se dois grupos de sujeitos.

Um exemplo simples de aplicação desse delineamento seria aquele em que o professor observa cuidadosamente seus alunos durante algumas semanas do curso, fazendo várias medições (que podem ser testes de aproveitamento ou de atitude) antes de fazer uso de uma nova estratégia de ensino. Da mesma forma, após o uso da estratégia, volta a observar seus alunos, durante algum tempo, fazendo novos registros. Diferenças, qualitativas ou quantitativas, no desempenho dos alunos após o uso da estratégia, e que se mantêm ao longo do tempo, podem ser tomadas como evidência do efeito da estratégia sobre a aprendizagem cognitiva ou afetiva dos alunos.

Delineamento de tipo 8 (Amostras temporais equivalentes) - Este delineamento é, de fato, uma variação do delineamento anterior. Neste delineamento introduz-se a variável experimental (o tratamento X) alternadamente e observa-se o grupo. O seu esquema é o seguinte:

$$O_1 X O_2 X_0 O_3 X O_4 X_0 O_5 X O_6 X_0 O_7 X O_8$$

Como se pode ver nesse delineamento o mesmo grupo de sujeitos é observado alternadamente na presença do tratamento e sem a presença do tratamento (aqui simbolizada pelo símbolo X_0). A análise é feita a partir da comparação dos valores médios do grupo com e sem tratamento experimental. Nesse ponto esse delineamento se assemelha a um delineamento com dois grupos.

Delineamento de tipo 9 (Grupo de controle não-equivalente) - Este delineamento tem a seguinte estrutura:

$$\begin{array}{ccc} O_1 & X & O_2 \\ O_3 & & O_4 \end{array}$$

Neste caso o grupo de controle e o grupo experimental não possuem equivalência amostral pois não foi usada a aleatoriedade na escolha das amostras. Neste tipo de delineamento, os grupos constituem coletivos reunidos naturalmente, tais como classes escolares já compostas previamente à ação do pesquisador. O controle do pesquisador reside unicamente na decisão sobre qual dos grupos vai receber o tratamento e quando.

Seguramente, neste tipo de delineamento haverá problemas sérios derivados do fator *seleção* e de sua interação com outros fatores importantes tais como *história*, *maturação*, etc..

Campbell e Stanley propõem ainda cinco outros delineamentos quase- experimentais que não serão aqui apresentados.

Nesta seção foi dada ênfase bastante grande ao delineamento porque esta é uma questão crucial na realização de uma pesquisa quantitativa em ensino. Assim como o pesquisador deve formular uma questão de pesquisa clara, orientadora e relevante, ele deve também investigá-la usando um delineamento adequado.

Um mau delineamento pode invalidar as asserções de conhecimento (resultados) e de valor de uma pesquisa, jogando por terra todo o trabalho realizado, seja por não controlar as fontes de invalidade interna seja por não controlar as de invalidade externa.

Sempre que possível um dos delineamentos experimentais deve ser utilizado. Quando isso não é possível, a alternativa é o uso de um dos delineamentos quase-experimentais aqui expostos (e discutidos em muito maior extensão em Campbell e Stanley, 1963, 1991) tendo em mente as deficiências que estes delineamentos oferecem. A pesquisa não deve deixar de ser feita se a situação não permitir o uso de um delineamento puramente experimental, mas o pesquisador deve deixar claro para si e para os demais pesquisadores a limitação de escopo, no que tange à validade, de seus resultados bem como quais são os pontos onde novos trabalhos devem ser realizados de modo a estudar a interferência de fatores os quais não puderam ser controlados naquele experimento específico. Cabe lembrar que não é uma única pesquisa que constrói o corpo de conhecimentos de uma área mas, sim, um conjunto delas.

Um pouco de estatística

O objetivo desta seção e das próximas é o de discutir os principais tópicos relacionados com a análise de experimentos quantitativos em ensino. Como foi dito anteriormente, a principal ferramenta para esse tipo de análise é a Estatística. Justifica-se, portanto, dedicar algum espaço a ela.

Amostragem aleatória

O termo técnico **aleatório** indica que a amostra é selecionada de tal modo que cada elemento da população tem uma chance igual de entrar na amostra. O pesquisador deve ter uma lista completa de todos os elementos da população e então selecionar sua amostra de modo tal que nenhum elemento da população seja privilegiado pelo procedimento de escolha.

O propósito da aleatoriedade não é garantir que os dois grupos se comportarão igualmente bem na ausência do tratamento. A aleatoriedade não garante igualdade. A aleatoriedade permite evitar aquele tipo de resultado que poderia ser atribuído à variabilidade da amostra. Aleatoriedade é um procedimento para selecionar amostras e não uma característica da amostra. Ela também não assegura representatividade e nem dá indicativo de como a amostra se comportará.

Define-se por **erro de amostragem** àquele tipo de erro que se comete ao selecionar amostras aleatórias para representar a população. Em virtude deste tipo de erro, é virtualmente impossível para um grupo pequeno ser exatamente representativo de outro muito maior. Esse erro de amostragem está presente toda vez que se selecionar amostras, não importa o quão cuidadosa seja feita a seleção aleatória. A seguir são definidos alguns tipos de procedimentos de seleção de amostras:

- **Amostragem aleatória simples** - é o processo de selecionar observações de um grupo maior de tal modo que cada sujeito na população de onde se está selecionando a amostra tenha uma probabilidade igual e independente de ser selecionada.
- **Amostragem estratificada** - é, algumas vezes, um modo recomendado de proceder ao escolher amostras. Divide-se a população em grupos menores homogêneos de modo a obter-se uma melhor representação. Com cada subgrupo algum processo de seleção aleatória pode ser usado. Este processo dá ao pesquisador uma amostragem mais significativa da que seria obtida diretamente da comunidade inteira. De modo a ter-se uma representação mais fiel da população como um todo pode-se, ainda, fornecer pesos ao número de sujeitos pertencentes aos diferentes grupos de modo a ter-se uma representação proporcional à distribuição na população.
- **Amostragem estratificada proporcional** - este tipo de processo de amostragem ocorre quando se toma uma percentagem de cada grupo no processo de composição da amostra.
- **Amostragem sistemática** - tem-se esse tipo de amostragem caso uma população tenha sido listada e então algum tipo de seleção seja feita segundo algum critério tal como, por exemplo, tomar-se o *enésimo* elemento da lista.
- **Amostragem por agrupamento** - é uma variação da amostragem aleatória simples, particularmente apropriada quando a população é grande ou quando a distribuição geográfica da população é espalhada. A cada agrupamento é dado um número e selecionados grupos escolhidos aleatoriamente. O uso da amostragem por agrupamento é geralmente escolhido por razões econômicas e aspectos administrativos.

Em qualquer tipo de amostragem, a característica da amostra inevitavelmente diferirá em algum grau, pequeno no entanto, da característica da população. Mas, quando a amostragem

aleatória é usada, as chances de que a ocorrência do erro de amostragem influencie a variável dependente em uma direção particular são as mesmas que existem de influenciar a mesma variável em outra direção qualquer. *Amostragem aleatória é a única forma de amostragem pela qual alguma quantidade de erro específica pode ser estimada. Pode-se dizer que é a amostragem aleatória o que diferencia as pesquisas experimentais das pesquisas não-experimentais.*

Medidas de tendência central

Entende-se por **distribuição de frequências** a uma tabulação (listagem) dos escores obtidos em uma certa amostra com o número de vezes em que esses escores aparecem na amostra. Assim, uma listagem com as notas obtidas pelos alunos em determinada avaliação com o número de vezes com que cada nota aparece é uma distribuição de frequências.

Quando o número de sujeitos pertencentes à(s) amostra(s) é pequeno pode-se ter uma idéia de como se comportam os escores. No entanto, se tivermos amostras muito grandes fica difícil ter uma idéia do comportamento da amostra e temos, então, que recorrer a transformações sobre os escores. Uma das formas de fazer isto é usando as chamadas **medidas de tendência central**.

Define-se uma medida de tendência central como sendo um número em torno do qual os valores da distribuição de frequências se distribuem. Em Estatística, usa-se basicamente três tipos de medidas de tendência central: a média, a moda e a mediana.

- **A Moda** de uma distribuição de frequências é definida como o escore que tem a mais alta frequência.
- **A Mediana** é definida como aquele escore que marca o ponto médio do conjunto de dados, ou seja, aquele escore para o qual tem-se 50 % dos escores com valores superiores 50 % dos escores com valores inferiores.
- **A Média aritmética simples** de uma distribuição de frequências é definida como aquele número obtido a partir da divisão da soma de todos os escores obtidos pelo número de elementos na distribuição de frequências. Matematicamente:

$$\bar{X} = \frac{\sum X_i n_i}{N}$$

onde os X_i são os escores obtidos, n_i é o número de vezes em que cada escore aparece e N é o número de elementos na distribuição de frequências. O símbolo \sum significa que se está somando.

Exemplo: seja a distribuição de frequências constante Tabela 1.

*Tabela 1 - Distribuição de frequências
para uma prova de uma disciplina
hipotética.*

Turma 1	9,8
Turma 2	7,4
Turma 3	6,2

Turma 4	6,0
Turma 5	5,9
Turma 6	4,5
Turma 7	3,4
Turma 8	3,4
Turma 9	3,4
Turma 10	1,0
Turma 11	0,5

A **moda** desta distribuição é dada pelo valor 3,4 pois é o escore que aparece o maior número de vezes na distribuição (3 vezes). A **mediana** da distribuição é o valor 4,5 pois, para esse valor, 50 % (5 escores) são maiores que ele e 50 % são menores. Por fim, a **média aritmética simples** da distribuição é dada por:

$$\bar{X} = \frac{9,8 + 7,4 + 6,2 + 6,0 + 5,9 + 4,5 + 3 \times 3,4 + 1,0 + 0,5}{11} = \frac{51,5}{11}$$

$$\bar{X} = 4,7$$

Outros tipos de média podem ser definidos como, por exemplo, a **média aritmética ponderada** e a **média geométrica**.

A média aritmética ponderada é usada quando deseja-se ter uma medida de tendência central de certa distribuição de frequências onde os escores contribuem com pesos diferentes. Matematicamente ela é definida por:

$$\bar{X}_p = \frac{\sum p_i n_i X_i}{\sum p_i}$$

onde \bar{X}_p é o valor da média ponderada, p_i é o peso atribuído a cada escore X_i e n_i é o número de vezes que o escore aparece na distribuição de frequências.

A **média geométrica** é definida por:

$$\bar{X}_g = \sqrt[N]{\prod n_i X_i}$$

onde \bar{X}_g é a média geométrica, o símbolo Π significa que devemos multiplicar os elementos que vêm a seguir e os X_i e n_i foram definidos anteriormente. Esse tipo de média é usado quando os escores da nossa distribuição são obtidos uns dos outros a partir de um fator multiplicativo.

Medidas de variabilidade

A média, seja ela de que tipo for, não diz tudo a respeito de uma distribuição de frequências. Aqui cabe um comentário a respeito dos processos estatísticos e a perda de informação que necessariamente ocorre quando usa-se números, tais como a média, para representarem distribuições de frequências. No processo de mediação ocorre uma perda de informação uma vez que substituí-se a informação total, ou seja, a distribuição, por algo que

pretende ser representativo dessa distribuição. Ao fazê-lo, perde-se a estrutura fina de informação providenciada pelo conjunto completo de escores. Assim, por exemplo, consideremos as hipotéticas Tabelas 2 e 3:

<i>Tabela 2 - Notas na Turma 1.</i>		<i>Tabela 3 - Notas na Turma 2.</i>	
<i>aluno 1</i>	<i>5,1</i>	<i>aluno 1</i>	<i>9,0</i>
<i>aluno 2</i>	<i>5,9</i>	<i>aluno 2</i>	<i>5,0</i>
<i>aluno 3</i>	<i>7,2</i>	<i>aluno 3</i>	<i>6,0</i>
<i>aluno 4</i>	<i>5,9</i>	<i>aluno 4</i>	<i>8,0</i>
<i>aluno 5</i>	<i>5,9</i>	<i>aluno 5</i>	<i>2,0</i>

Se for calculada a média aritmética simples para essas duas distribuições de frequência, obter-se-á para ambas a média 6,0. Entretanto, se essas distribuições forem representativas de notas de alunos em duas turmas diferentes, em uma disciplina cuja média de aprovação seja 6,0, observa-se, então, que na turma 1 há um sujeito aprovado enquanto que na turma 2 há três. Olhando somente para a média das duas turmas essa informação seria perdida.

Uma forma de minimizar essa perda de informação, decorrente do uso de uma medida de tendência central, são as **medidas de variabilidade**. Uma medida de variabilidade indica o quão espalhados estão os escores na distribuição. Ou seja, uma medida de variabilidade é uma forma de ter uma idéia do quanto os escores se afastam da medida de tendência central que se está utilizando.

A mais simples das medidas de variabilidade é o **intervalo (I)**. O intervalo diz entre quais valores se distribuem os escores da distribuição que está sendo analisada. Assim, no exemplo, o intervalo para a distribuição 1 é dado por $[7,2;5,1]$ enquanto que na distribuição 2 o intervalo é dado por $[9,0;2,0]$.

Outra medida de variabilidade é o **desvio da média (dm)**. Essa quantidade diz o quanto os escores se desviam da média. No exemplo dado, o escore 9,0 da Tabela 3 possui um desvio da média de 3,0 ($9-6=3$).

Uma medida de variabilidade das mais utilizadas em análises de distribuições de frequência é o **desvio padrão (dp)**. O desvio padrão indica o quão *espalhada* é uma distribuição. O desvio padrão tem uma interpretação muito simples, originada da equação que define a *distribuição normal* (que será analisada na próxima seção). Tomando um intervalo definido por $[\bar{X} - dp; \bar{X} + dp]$ tem-se dentro desse intervalo em torno de 68 % dos valores da distribuição. Fazendo uma nova medida, então, há uma *probabilidade* de 68 % de que essa nova medida pertença a esse intervalo. Considerando intervalos definidos por múltiplos do desvio padrão se estará englobando um número cada vez maior de escores dentro do intervalo em questão.

Matematicamente, o desvio padrão é dado por:

$$dp = \sqrt{\frac{\sum_1^N (X_j - \bar{X})^2}{N}}$$

onde N é o número de escores, X_j significa o j -ésimo elemento da distribuição de escores e \bar{X} é o valor médio da distribuição.

Curva normal

Define-se como **curva da distribuição** a representação gráfica de uma frequência de distribuição de escores onde os valores dos mesmos são indicados no eixo horizontal e os valores das frequências dos escores particulares no eixo vertical. Curvas de distribuição podem vir em diferentes formas e tamanhos. No entanto, muitas frequências de distribuição tendem a seguir um certo padrão chamado de **distribuição normal**, especialmente quando existem muitos escores envolvidos. A forma da curva que pode ser traçada destas distribuições é chamada **curva normal** (Figura 1). Uma das características da curva normal é a *simetria*; outra característica importante é que a média, a mediana e a moda são *idênticas*.

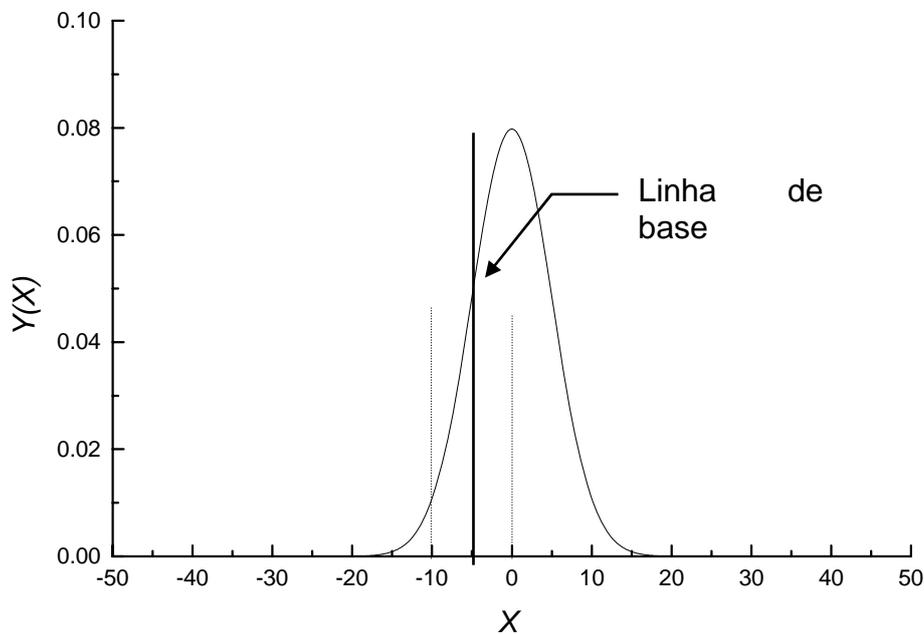


Figura 1 - A curva normal

Na curva normal, a distância da linha de base até a primeira linha pontilhada é o **desvio padrão** da distribuição, por definição⁸.

A equação que define a curva normal é dada por:

$$Y(X) = \frac{1}{dp\sqrt{2\pi}} \exp\left[-\frac{1}{2} \frac{(X - \bar{X})^2}{dp^2}\right]$$

onde aqui dp é o desvio padrão, \bar{X} é o valor médio e X é o valor para o qual se quer calcular Y .

⁸ No presente exemplo, a curva foi gerada com um desvio padrão assumido de 5.

Intervalos de confiança

Quando se diz que um escore aleatoriamente selecionado cairá dentro de um intervalo específico dos valores dos escores obtidos, é preciso fazê-lo com algum grau de confiança, ou seja, sabendo quão provável é que se esteja correto. O intervalo de 95 % é chamado de **intervalo de confiança** porque, se a distribuição for normal, pode-se estar certo de que 95 % das vezes em que um valor da variável em questão for escolhido aleatoriamente ele estará dentro desse intervalo. Em termos de probabilidades, o intervalo de confiança de 95 % designa os dois valores entre os quais existe a probabilidade $p=0,95$ de que um escore selecionado aleatoriamente pertença a esse intervalo. O intervalo de confiança de 95 % é dado por:

$$I_{95\%} = \left[\bar{X} - 1,96dp; \bar{X} + 1,96dp \right]$$

onde dp são o desvio padrão e \bar{X} a média calculados para a distribuição.

O intervalo de confiança de 99 % é outro intervalo normalmente utilizado. Os limites desse intervalo estão entre:

$$I_{99\%} = \left[\bar{X} - 2,58dp; \bar{X} + 2,58dp \right]$$

Distribuição de médias amostrais

Suponha-se que são selecionadas de uma população hipotética um grande número de amostras, cada uma das quais, digamos, com 50 sujeitos, e seja calculado o escore médio para cada grupo em alguma variável de interesse. Pode-se então traçar uma curva representando a distribuição desses escores. Nessa distribuição de médias amostrais, a **média de todas as médias** é a média da população e as médias amostrais são distribuídas em torno da média da população seguindo a distribuição normal.

Quando se lida com distribuições de médias, o desvio padrão é chamado de **desvio padrão da média (dpm)**. A interpretação dessa quantidade é semelhante àquela do desvio padrão de medida: tomando-se outra amostra da população a probabilidade de que a **média** obtida por essa nova amostra esteja dentro do intervalo de confiança de 95% é dada por:

$$I_{95\%} = \left[\bar{X} - 1,96dpm; \bar{X} + 1,96dpm \right]$$

Inferências a respeito do valor da média da população

Quando se sabe a média de uma amostra e toma-se uma estimativa do desvio padrão da média não é possível inferir o valor da média da população a partir da média que se tem para aquela amostra, mas pode-se fazer hipóteses sobre o valor para a média da população e, usando a estimativa do desvio padrão da média, determinar a probabilidade de obter-se uma média amostral que difira da média hipotetizada da população tanto quanto se queira.

Suponhamos que uma média amostral seja 97 e que o desvio padrão da média estimado seja de $dpm=2$. Suponhamos também que a hipótese feita foi de que a média da

população é 100. Pode-se então determinar a probabilidade de obter nossa média da amostra de 97.

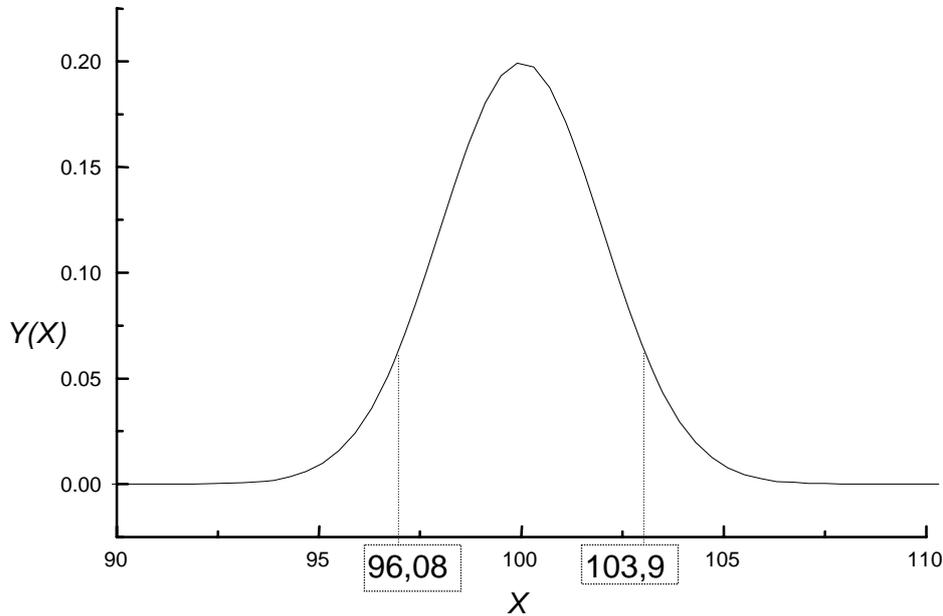


Figura 2 - Curva normal indicando intervalo de confiança de 95 %.

A Figura 2 indica que o intervalo de confiança de 95 % varia de 96,08 a 103,92. Isto diz que se a média da população for 100, então, a probabilidade de selecionar-se uma amostra cuja média esteja dentro do intervalo de confiança considerado é $p=0,95$. Outra forma de dizer isso é que a probabilidade de obter-se uma média amostral menor que 96,08 ou maior que 103,92 é $p=0,05$. Portanto, pode-se aceitar a hipótese de que a amostra, cuja média é 97, vem de uma distribuição de médias amostrais, retiradas de uma população cuja média é 100.

Comparações entre múltiplas amostras

Em situações de pesquisa, usualmente deseja-se comparar duas ou mais amostras. Por exemplo, pode-se querer determinar se existe uma diferença em aquisição de conhecimentos entre estudantes que são ensinados por um método A e aqueles que são ensinados por um método B. A questão é: *qual é a probabilidade de que diferenças entre as duas médias das amostras seja devido simplesmente a erro de amostragem?* Em outras palavras, *pode a diferença de médias entre as duas amostras ser atribuída ao erro aleatório nas amostras, ou os alunos ensinados por um método realmente aprendem mais que aqueles ensinados pelo outro método? Temos de fato duas amostras pertencentes a populações diferentes, representadas por duas distribuições normais de escores?*

Ao testar a **hipótese nula** se está assumindo que a diferença entre as duas amostras é devida simplesmente ao erro de amostragem.

Tipos de hipóteses

Quando se quer decidir se determinado procedimento é melhor que outro, formula-se a hipótese de **não existe diferença** entre os procedimentos (i.e., qualquer diferença observada é meramente devida a flutuações ao tomar duas amostras **da mesma** população). Este tipo de hipótese é a chamada **Hipótese Nula** denotada por H_0 . A hipótese alternativa à H_0 , ou seja, de que a diferença observada **não é** meramente devido a amostragem, é denotada por H_1 .

Ao decidir se rejeita-se ou não a hipótese nula dois tipos de erros podem acontecer. O primeiro tipo de erro acontece quando rejeita-se a hipótese nula com base em dados de amostras que de fato vêm da mesma população, diz-se que se comete um erro **Tipo I**. Se por outro lado, aceita-se a hipótese nula quando, de fato, as amostras vêm de populações diferentes, diz-se que foi cometido um erro do **Tipo II**⁹.

Na prática, conforme o caso, um tipo de erro pode ser mais sério do que o outro e, assim, uma solução de compromisso deve ser obtida em favor de uma limitação do erro que, naquela situação, seja considerado mais sério. A única forma de diminuir a influência desses dois tipos de erro é aumentar o tamanho da amostra, o que nem sempre é possível.

Ao testar a hipótese nula, a probabilidade máxima com a qual o pesquisador aceita correr o risco de cometer um erro de *tipo I* é chamada de **nível de significância estatística** da pesquisa. Quando o pesquisador decide o nível de probabilidade que usará ao rejeitar a hipótese nula ele estará dando a probabilidade com a qual arriscará estar errado em sua decisão. Se ele seleciona o nível de significância como $0,05$, está dizendo que há uma probabilidade de $0,05$ de que ele esteja errado. Se não quiser correr um risco de erro tão significativo, ele pode então escolher para nível de significância $p=0,01$. Neste nível é menos provável que esteja cometendo um erro do *tipo I*, entretanto, estará incrementando a probabilidade de cometer um erro do *tipo II*.

Na prática da pesquisa em educação os níveis de significância de $0,05$ e $0,01$ são usuais, embora outros níveis de significância sejam também utilizados.

Uma hipótese que não indica a direção da diferença esperada, mas meramente estabelece que existe uma diferença, é chamada de uma **hipótese bilateral** (*two-tailed*). Esse tipo de hipótese é assim designada porque ela está preocupada com ambas as caudas da distribuição normal das diferenças entre médias amostrais.

Uma hipótese que afirma qual tratamento é melhor que o outro é chamada de **hipótese unilateral** (*one-tailed*) porque ela está somente preocupada com um dos lados da distribuição de diferenças entre médias amostrais.

Distribuição de médias de pequenas amostras

Até aqui foram consideradas amostras que contêm um grande número de sujeitos (trinta ou mais). As propriedades da distribuição normal são válidas para grandes amostras, mas não quando há um número pequeno de sujeitos em cada amostra. A distribuição tende a ficar achatada quando, em cada amostra, o número de sujeitos é pequeno.

Para fins estatísticos isso significa que, para dados onde as amostras são pequenas, não é possível usar-se as propriedades da curva normal para decidir a favor ou contra a aceitação

⁹ Tecnicamente, o pesquisador não deveria aceitar a hipótese nula mas, sim, falhar em rejeitar a hipótese nula.

da hipótese nula. Ao invés disso, deve-se usar valores que reflitam esse achatamento da curva normal. Esses valores são chamados de valores **t** para os quais também foram calculados valores para os níveis de significância $p=0,05$ e $p=0,01$ para amostras de qualquer tamanho. Existem tabelas estatísticas preparadas (uma das quais é apresentada na Tabela 4) para estes valores **t** para todos os tamanhos de amostras sendo comparadas, de modo que sabendo-se quantos sujeitos estão em cada amostra que se está comparando poderemos facilmente determinar o valor de **t** necessário ao nível de significância escolhido (normalmente $0,05$ ou $0,01$).

Se um valor **t** indica diferenças dentro do intervalo de confiança de 95 %, seu valor normalmente **não é comunicado**. Ao invés, o pesquisador afirma que o valor **t** não é significativo. Neste caso, ele aceita a hipótese nula e atribui a diferença observada entre suas amostras ao simples erro de amostragem.

Este tipo de teste estatístico é chamado de **teste t** e é utilizado para comparação entre médias de amostras pequenas **quando, por hipótese, as amostras foram escolhidas aleatoriamente e os escores vieram de populações distribuídas segundo a distribuição normal**. Outros testes estatísticos são disponíveis se a hipótese de normalidade não puder ser feita.

Amostras com número de sujeitos menor que 30 são chamadas pequenas amostras. Um estudo estatístico de distribuições amostrais no qual as amostras são pequenas é chamado *Teoria de Pequenas Amostras*. Entretanto, um nome mais apropriado seria *Teoria Exata da Amostragem*, uma vez que os resultados obtidos mantêm-se tanto para pequenas como para grandes amostras. Uma distribuição importante é a distribuição **t de Student**¹⁰. Esta distribuição é dada, matematicamente, por:

$$Y = \frac{Y_0}{\left(1 + \frac{t^2}{N-1}\right)^{N/2}}$$

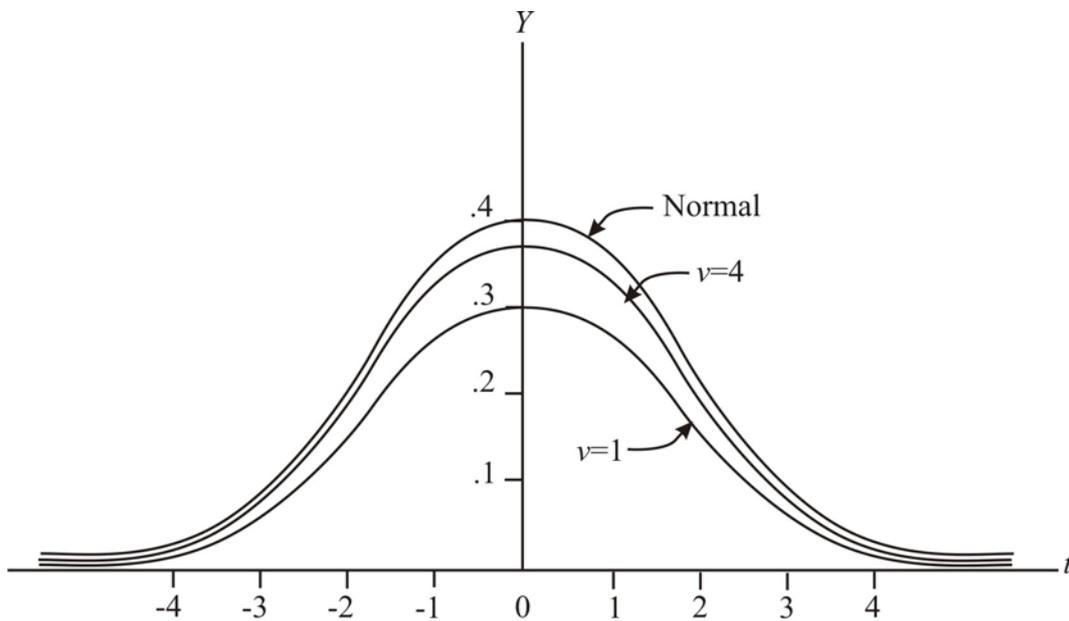
$$Y = \frac{Y_0}{\left(1 + \frac{t^2}{gl}\right)^{(gl+1)/2}}$$

onde Y_0 é uma constante que depende em N de tal modo que a área total sobre a curva é 1¹¹. A quantidade $N-1$ é chamada de *grau de liberdade* (gl). Para grandes valores de gl ou N ($N > 30$) as curvas da figura se aproximam da curva normal padrão. A Figura 3 mostra várias curvas desta distribuição para vários valores de graus de liberdade gl .

$$gl=N-1$$

¹⁰ Mantém-se aqui o nome em inglês por ser consagrado na literatura e pelo uso na área.

¹¹ A isto chama-se de condição de normalização.



Student's t distributions for various values of v .

Figura 3 - A distribuição t de Student para distintos graus de liberdade.

Para fins de cálculo entre duas amostras, com N_1 e N_2 sujeitos, de médias e desvios padrão dados por \bar{X}_1 , dp_1 , \bar{X}_2 e dp_2 respectivamente, o valor de t é dado por:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sigma \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

onde a quantidade σ é definida por:

$$\sigma = \sqrt{\frac{N_1 dp_1^2 + N_2 dp_2^2}{N_1 + N_2 - 2}}$$

Valores de t foram calculados, para níveis de significância de $0,05$ e $0,01$, para qualquer tamanho de amostra. Os estatísticos prepararam tabelas estatísticas destes valores t para todos os tamanhos das amostras sendo comparadas, de modo que, se soubermos quantos sujeitos existem em cada amostra podemos compará-las facilmente e determinar o valor t necessário para o nível de significância desejado ($0,05$ ou $0,01$). Tabela desse tipo está exemplificada na Tabela 4.

Tabela 4 - Tabela t para hipóteses unilaterais e bilaterais.

gl	Nível de significância para hipótese unilateral									
	.40	.25	.10	.05	.025	.01	.005	.0025	.001	.0005
	Nível de significância para hipótese bilateral									
	.80	.50	.20	.10	.05	.02	.01	.005	.002	.001
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.326	31.598
3	.277	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	.271	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	.265	.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	.263	.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	.262	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	.261	.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	.260	.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	.259	.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	.259	.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	.258	.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.133	2.602	2.947	3.286	3.733	4.073
16	.258	.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	.257	.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	.257	.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	.257	.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	.257	.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	.256	.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	.256	.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	.256	.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	.256	.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	.256	.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	.256	.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	.256	.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	.255	.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	.254	.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	.254	.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	.253	.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

Fidedignidade e validade

Antes que qualquer análise dos dados possa ser feita é necessário que o pesquisador se pergunte: *a forma de obtenção dos dados me dá segurança sobre a sua validade?*

Em seções anteriores foi discutida a estrutura de um experimento analisando os vários tipos de delineamentos possíveis (destacando o que diferencia um delineamento do tipo experimental dos que não possuem esta propriedade) bem como aspectos relacionados com as características de um bom teste. No entanto, de nada adianta ser cuidadoso na escolha do

delineamento da pesquisa se os instrumentos de coleta de dados forem inadequados. Inadequação significa dizer que o instrumento não mede corretamente o que se propõe a medir (questão relacionada com a **fidedignidade** do instrumento) ou mede outra coisa diferente daquela que se propõe a medir (questão relacionada com a **validade** do instrumento). Para usar uma analogia comum quando se fala em fidedignidade e validade, pode-se imaginar a seguinte situação: um atirador de dardos acerta repetidamente o mesmo lugar do alvo. Neste caso, diz-se que há fidedignidade, pois em várias repetições o atirador acerta sempre no mesmo lugar ou, em outras palavras, consegue reproduzir o mesmo resultado (posição) com o mesmo instrumento. Porém se a posição acertada não for o centro do alvo, dizemos que não há validade, pois o objetivo do jogo é acertar o centro. Se a posição acertada for o centro, sempre, então dizemos que há fidedignidade e validade.

A ferramenta básica para a análise de fidedignidade é a correlação estatística entre variáveis. Vejamos como se calcula esta quantidade e qual é o seu significado.

O que é a correlação entre duas variáveis

Conceitualmente a *correlação* ou o *coeficiente de correlação* (r) indica como se comportam duas ou mais variáveis umas em relação às outras. Quando temos uma correlação alta isto indica que o crescimento de uma variável é acompanhado pelo crescimento (no caso de um coeficiente de correlação perto de $+1$) ou pelo decréscimo (no caso de um coeficiente de correlação perto de -1) da outra variável. Convém lembrar aqui que o fato de duas variáveis serem correlacionadas (tanto positiva quanto negativamente) *não* implica uma relação causal entre as duas variáveis. Para o estabelecimento de uma relação causal entre elas deve-se recorrer a outras ferramentas de análise, ou seja, devemos procurar na teoria as razões dessa dependência e os fatores de comprovação dessa dependência.

O grau de correlação é indicado pelo valor do coeficiente de correlação. Para uma correlação perfeitamente positiva tal como mostrado na *Figura 4.a* o coeficiente de correlação r tem valor $+1$. O coeficiente para uma correlação perfeitamente negativa é mostrado na *Figura 4.b*, o coeficiente de correlação r tem valor de -1 . Estes dois valores são os valores máximos para r . O coeficiente $r=0$ indica a inexistência de correlação. Neste caso o comportamento de uma variável não é relacionado de qualquer modo ao comportamento da outra variável. Isso é mostrado na *Figura 4.c*

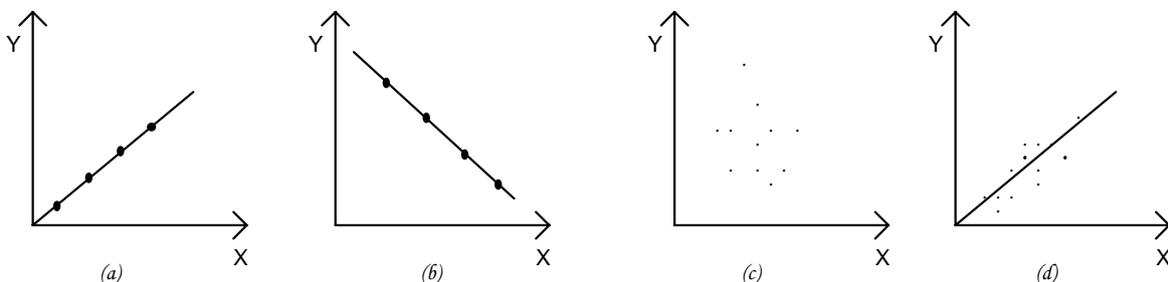


Figura 4 - Representação gráfica dos vários tipos de correlação possíveis entre variáveis.

A *Figura 4.d* nos mostra um caso onde a correlação existe mas é imperfeita. Pela simples visão dessa figura, podemos ver que os pontos tendem a se alinhar ao longo de uma direção específica, embora não estejam todos sobre uma mesma reta. Isto nos indica que a correlação é positiva, embora não seja perfeita.

Tal como ocorre com os escores médios de amostras, os coeficientes de correlação, calculados a partir de dados amostrais, são afetados por erros de amostragem. Assim, a exemplo do que foi feito com as médias amostrais, uma pergunta se impõe: *quão provável é que o coeficiente de correlação obtido a partir dos dados amostrais não seja fruto do erro de amostragem e reflita um verdadeiro relacionamento existente na população?* É razoável se tomar por hipótese de que, como em toda inferência feita a partir de um processo de amostragem, exista um erro devido ao próprio processo de amostragem. Entretanto, a exemplo do que acontece para outros tipos de testes estatísticos, existem tabelas para vários valores de tamanho de amostra, a qualquer nível de significância estatística desejado¹².

Cálculo do coeficiente de correlação

A forma de cálculo do coeficiente de correlação é função do tipo de variável com a qual estamos lidando. Como já foi discutido, estas podem ser divididas em quatro grupos: nominais, ordinais, intervalares ou racionais. Para cada pareamento de variáveis a forma de cálculo do coeficiente de correlação é diferente, devendo-se levar em conta os tipos de variáveis envolvidos. A fórmula de cálculo do coeficiente de correlação, definida a seguir, é válida somente quando as variáveis envolvidas são ambas dos tipos *intervalares* ou *racionais*. Para outros tipos de variáveis sugere-se a consulta a Glass e Stanley, (1970).

O termo *correlação* tal como está sendo usado aqui significa *correlação linear*. Nesse caso, o coeficiente de correlação entre duas variáveis X e Y é dado por:

$$r_{xy} = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}}$$

onde $x = X - \bar{X}$ e $y = Y - \bar{Y}$ ¹³.

Esta equação é chamada de *fórmula produto-momento*¹⁴. Deve-se observar a simetria entre as variáveis que compõe essa equação: se trocarmos x com y o resultado é o mesmo. Com isso $r_{xy} = r_{yx}$ ou seja, tanto faz calcularmos o coeficiente de correlação chamando uma das variáveis de X e a outra de Y como calcularmos denominando inversamente as variáveis.

Validade e fidedignidade de testes

Testes de conhecimento tentam medir o que um indivíduo apreendeu em uma certa área, seu nível atual de domínio de um certo conteúdo ou seu desempenho. Muitos testes usados em escolas são testes de conhecimento. Frequentemente, escores em testes de conhecimento são usados na avaliação de cursos, professores, métodos de ensino e outros fatores considerados significativos na prática educacional. São usados na classificação, promoção ou retenção de estudantes em níveis particulares de ensino. Eles são usados como ferramentas para diagnosticar pontos fracos e fortes dos estudantes e como uma base para prêmios, recompensas, etc., dentro do ambiente escolar.

¹² Usualmente os níveis 0,01 e 0,05.

¹³ Os valores com barra, seguindo a convenção anterior, são as médias em X e Y .

¹⁴ Observe-se que essa equação fornece automaticamente o sinal correto para o coeficiente de correlação.

Testes de aptidão tentam prever o grau de conhecimento que pode ser esperado de indivíduos em uma atividade particular. Estes testes tentam prever a capacidade de um indivíduo em particular para adquirir um melhor desempenho a partir de treino adicional. De fato, capacidade (ou aptidão) não pode ser medida diretamente. Aptidão somente pode ser inferida com base no desempenho apresentado. Testes de aptidão podem ser usados para dividir os estudantes em grupos relativamente homogêneos com fins instrucionais de modo a identificar estudantes para bolsas escolares ou para elaborar guias individuais dentro de áreas onde elas mais provavelmente terão sucesso.

Na pesquisa, testes são instrumentos de medida usados pelos pesquisadores para, de um modo geral, coletar dados. Deste modo, se a pesquisa for de qualidade é de se supor que os procedimentos usados para a coleta de dados (e sua análise) devem ser bons. A qualidade da pesquisa não pode ser melhor do que a qualidade dos procedimentos usados para coletar e analisar os dados.

Como já foi dito, esses instrumentos devem ter, duas características são indispensáveis:

1. **Fidedignidade:** um teste é fidedigno se suas medidas são acuradas e consistentes quando aplicado em tempos diferentes.
2. **Validade:** em geral, um teste possui validade se mede o que se propõe a medir.

Um teste pode ser fidedigno, mesmo se não for válido. Um teste válido sempre é fidedigno. Não existe uma única forma de validade ou fidedignidade de um instrumento. Existem muitos tipos de validade e fidedignidade. Em geral, fidedignidade está relacionada com consistência e validade está relacionada com a interpretação do teste. Um teste pode ser válido para um objetivo e não ser válido para outro.

Fidedignidade¹⁵

Reiterando, por fidedignidade entendemos o quão acurados os dados são no sentido de sua estabilidade, reproduzibilidade ou precisão. Um instrumento de coleta de dados fidedignamente perfeito é um que se fosse aplicado duas vezes sob as mesmas circunstâncias nos forneceria os mesmos resultados. Como foi visto, a *correlação* é a ferramenta estatística básica na análise de fidedignidade de testes. Uma correlação de *1,00* indicaria perfeita fidedignidade enquanto que correlação *0,00* indicaria nenhuma fidedignidade. Correlações intermediárias indicariam diferentes níveis de fidedignidade.

São várias as formas de medir-se a fidedignidade de um teste:

1. **Teste-reteste** - é exatamente o que o nome implica; produzimos o primeiro conjunto de dados através da aplicação do teste em um dado instante de tempo e, após decorrido algum intervalo de tempo longo o suficiente para esquecer o teste mas curto o suficiente para que não sejam produzidas alterações significativas nos respondentes, um segundo conjunto de dados é obtido pela aplicação do mesmo teste ao mesmo conjunto de respondentes (o *reteste*). Após a segunda aplicação os dois conjuntos de dados são correlacionados e a correlação obtida estima a fidedignidade do teste.

¹⁵ Parte dessa seção é baseada em *The Research Process in Education*, de D.J. Fox, 1969.

O maior problema nesta estratégia envolve o intervalo de tempo. Como definir o que se entende por intervalo de tempo grande o suficiente para esquecer do teste mas curto o suficiente de modo a que não se produzam alterações nos respondentes? A resposta a esta questão depende do tipo de pesquisa que está sendo feita.

O procedimento de *teste-reteste* é mais aplicável àquelas situações onde a aptidão, habilidade ou conhecimento que está sendo avaliado possui estabilidade, mudando em escalas de tempo longas frente ao período entre o *teste* e o *reteste*. Esta, obviamente é uma limitação de ordem prática sendo, sob o ponto de vista técnico, muito difícil garantir a integridade dos respondentes entre o *teste* e o *reteste*. Outro aspecto perigoso da estratégia do *teste-reteste* é que estamos calculando a correlação entre dados provenientes de duas exposições dos sujeitos ao mesmo conteúdo. Desse modo, se existir um direcionamento do *teste* em alguma direção será, naturalmente, um direcionamento consistente em cada uma das exposições ao instrumento.

2. Forma alternada: nesta forma de medição de fidedignidade o pesquisador deve desenvolver duas formas paralelas ou equivalentes de seu instrumento, digamos formas *A* e *B*, administrar ambas às mesmas pessoas e correlacionar os dois conjuntos de dados obtidos. A administração das duas formas pode ser com um intervalo de tempo entre as duas aplicações ou de uma única vez, com as duas formas sendo aplicadas consecutivamente. Em ambos os casos é aconselhável alternar-se a ordem de aplicação entre as duas formas do teste, isto é, metade dos respondentes respondem à forma *A* do teste e depois à forma *B* e a outra metade responde primeiro à forma *B* e depois à forma *A* do teste. Esta maneira de aplicar o teste elimina efeitos que poderiam mascarar os resultados obtidos tais como cansaço ou aborrecimento durante a aplicação da segunda forma do teste.

Este tipo de procedimento possui as seguintes dificuldades:

- Em muitos casos é difícil desenvolver-se uma forma do instrumento, o que dirá duas.
- Há situações onde é difícil, senão impossível, perguntar-se a mesma coisa duas vezes, de forma equivalente ou similar.
- Se o procedimento envolver certo intervalo de tempo entre a aplicação do teste e de sua forma alternada, da mesma forma que ocorreu para a forma teste-reteste, pode surgir o problema de reagrupar o mesmo grupo de sujeitos para responder o teste pela segunda vez.
- Administrar consecutivamente duas formas do mesmo teste exige duas vezes mais tempo e exige a manutenção de um interesse suficiente por parte dos respondentes uma vez que eles trabalharão, basicamente, sobre os mesmos conteúdos.

3. Método da metade ou par-ímpar: um terceiro procedimento chamado *da metade* ou *par-ímpar* soluciona cada um dos problemas apontados anteriormente, quando da discussão das outras formas de cálculo de fidedignidade. Ela envolve somente a administração de um instrumento em apenas uma única forma. O instrumento é aplicado uma única vez a um grupo de respondentes, entretanto, ele é analisado de forma a separar os escores de cada respondente em duas metades. Estes dois conjuntos de escores são então correlacionados. Desse modo, se um instrumento possui 100 itens, para obter-se a estimativa de fidedignidade na forma *par-ímpar*, obter-se-ia os escores de cada respondente nas questões ímpares e então um escore separado para as outras 50 questões pares. É possível, mas não o mais aconselhável, usar o procedimento de separação em *metades*, isto é, obter um escore para cada respondente baseado na primeira metade do teste e outro baseado na segunda metade do teste.

A forma *par-ímpar* é a preferível devido às seguintes vantagens:

- Normalmente, um instrumento de medida cobre diferentes áreas do conhecimento em diferentes seções as quais geralmente são estanques e bem diferenciadas.
- Fatores tais como fadiga ou perda de interesse poderiam causar omissão por parte do respondente nas questões finais do teste.

No entanto, não importa qual seja o procedimento usado, estimativas de fidedignidade obtidas a partir do uso de metades de um instrumento também apresentam seus problemas, principalmente relacionados com o fato de que a fidedignidade é relacionada com o número de itens de um instrumento.

O seguinte procedimento foi desenvolvido de modo a calcular a fidedignidade de um instrumento como um todo a partir do cálculo da fidedignidade usando metades do teste. É a chamada **fórmula de Spearman-Brown**, que recebe esse nome em homenagem aos pesquisadores que, de forma independente, a desenvolveram:

$$\alpha_{SB} = \frac{2 \times |\alpha|}{1 + |\alpha|}$$

onde α_{SB} é a chamada *estimativa de fidedignidade de Spearman-Brown*, α é a correlação entre as duas metades do teste. O que esta fórmula nos dá é somente uma predição ou estimativa da fidedignidade que o pesquisador poderia esperar para o instrumento como um todo a partir dos valores de fidedignidade obtidos para cada metade do teste.

A maior vantagem de usar-se correlação entre metades como uma estimativa de fidedignidade do teste total é de natureza prática: necessita-se de apenas um teste e uma sessão de testagem por parte dos respondentes. Suas desvantagens são as mesmas.

Expectativas para a fidedignidade de um instrumento diferirão dependendo da natureza da informação que está sendo procurada. Se o que se estiver procurando é uma informação tipo demográfica, como, por exemplo, lugar de nascimento, escolaridade e experiência profissional poderíamos esperar uma alta fidedignidade de um instrumento. Em termos de correlações isto implica em coeficientes de correlação da ordem ou superiores a 0,90. Por outro lado, se a procura for sobre conhecimento e habilidades, as quais são informações de caráter não tão fixo quanto o exemplo anterior, a expectativa de fidedignidade será menor e um coeficiente de fidedignidade de 0,85 é aceitável. Por fim, se o tipo de informação procurada for mais movediça, como por exemplo atitudes e interesses, um coeficiente da ordem de 0,70 será aceitável.

Uma pergunta que surge naturalmente nesse ponto é a seguinte: *como melhorar a fidedignidade de um teste?* Como comentado anteriormente, a fidedignidade de um teste está diretamente relacionada com a extensão do teste. Portanto, um procedimento óbvio de modo a melhorar a fidedignidade de um teste é alterar a sua extensão. *Mas de quanto a fidedignidade poderia ser melhorada com o acréscimo de uma ou mais questões ao teste?* A resposta a essa pergunta encontra-se embutida na fórmula de cálculo do coeficiente de Spearman-Brown, a qual fornece a fidedignidade máxima que poderia ser obtida pelo aumento no tamanho do teste. Essa estimativa é dada pela equação abaixo:

$$r_k = \frac{kr}{1 + (k - 1)r}$$

onde k é o número de vezes que o novo teste é maior que o anterior, r é a fidedignidade apresentada pela forma atual do teste e r_k é a estimativa de fidedignidade da nova forma do teste.

Exemplo: suponhamos que a fidedignidade de um teste é estimada em $0,50$. Este teste tem a sua extensão duplicada pelo acréscimo de itens. Qual será a estimativa da nova fidedignidade?

$$r_k = \frac{kr}{1 + (k - 1)r}$$

$$r_k = \frac{2 \times 0,5}{1 + (2 - 1) \times 0,5}$$

$$r_k = 0,67$$

Um algoritmo para o cálculo do coeficiente de fidedignidade de testes - Análise de Consistência Interna

Uma das aplicações do coeficiente de correlação é na **Análise da Consistência Interna** de testes. É muito comum na prática docente que o professor some escores de itens isolados de testes compondo, assim, um escore bruto o qual é usado para análises e inferências. No entanto, esse procedimento somente é aceitável quando todos os itens do teste se referem a um mesmo conjunto de conceitos e/ou habilidades. A análise de consistência interna de um teste tem por objetivo verificar o quanto de verdade existe nessa hipótese a respeito de determinado teste. A idéia geral é comparar-se o desempenho em cada item dos respondentes ao teste com o desempenho dos respondentes ao teste como um todo. Se o desempenho dos respondentes a um item não se correlaciona com o desempenho dos respondentes ao teste como um todo isto significa que aquele item em particular não está avaliando as mesmas características das demais questões que compõem o teste e, portanto, deve ser descartado (ou modificado). A Análise de Consistência Interna é parte indispensável do processo de pesquisa. O pesquisador não pode usar um teste sem verificar a sua consistência interna. Sem essa etapa, a soma de escores atribuídos a itens particulares não pode ser feita e toda inferência obtida a partir desse escore total será sem significado.

Apresenta-se a seguir um roteiro¹⁶ para a execução de análise de consistência interna. Os dados utilizados no exemplo são retirados da Tabela 5:

Tabela 5. - Dados para o exemplo de análise de consistência interna.

ITEM → INDIVÍDUO ↓	1	2	3	4	5	6	7	8	Total
1.	5	5	5	5	5	5	5	2	37
2.	5	4	4	5	3	5	5	3	34
3.	5	5	4	5	3	4	5	3	34
4.	4	4	5	4	4	5	5	2	33
5.	5	4	5	5	2	4	4	3	32
6.	4	5	4	4	3	5	5	2	32

¹⁶ Este roteiro foi elaborado pelo professor Fernando Lang da Silveira do Instituto de Física da UFRGS.

7.	4	4	5	5	5	5	3	1	32
8.	4	5	5	4	4	4	5	1	32
9.	3	4	5	4	3	5	5	2	31
10.	4	3	4	5	2	5	5	3	31
11.	5	5	5	4	1	4	3	4	31
12.	4	4	3	4	5	4	4	3	31
13.	4	4	4	3	3	3	4	4	29
14.	4	4	4	4	2	4	3	3	28
15.	3	4	4	4	3	4	4	2	28
16.	3	4	5	3	2	4	5	2	28
17.	4	3	3	5	4	3	2	4	28
18.	4	4	4	3	3	5	4	1	28
19.	3	3	4	4	4	4	3	2	27
20.	4	4	3	3	2	3	4	4	27
21.	4	3	3	3	4	3	4	3	27
22.	3	4	3	3	4	4	4	2	27
23.	3	4	3	3	1	3	4	5	26
24.	3	3	3	4	3	4	3	3	26
25.	2	3	3	3	5	4	3	1	24
26.	3	2	3	3	4	3	4	2	24
27.	1	2	3	2	5	3	3	5	24
28.	4	3	3	3	4	1	1	4	23
29.	2	2	2	1	4	3	3	4	21
30.	3	3	1	2	3	2	2	2	18
31.	1	1	2	1	3	1	2	4	15
<i>F(1)</i>	2	1	1	2	2	2	1	4	
<i>F(2)</i>	2	3	2	2	5	1	3	10	
<i>F(3)</i>	9	8	11	10	10	8	8	8	
<i>F(5)</i>	13	14	9	10	9	12	10	7	
<i>F(6)</i>	5	5	8	7	5	8	9	2	

1. Transformar a resposta de cada indivíduo a cada item em um escore.
2. Calcular o escore total de cada indivíduo somando os escores obtidos em cada item por aquele indivíduo.
3. Ordenar os indivíduos, em *ordem decrescente*, pelo escore total.
4. Construir a matriz dos resultados (ver a *tabela 5*).
5. Determinar a frequência (número de vezes em que o escore aparece) de cada escore em cada item (ver as últimas cinco linhas da *tabela 5*).
6. Calcular a média do escore total. Isto é feito somando-se os escores totais de cada um dos sujeitos e dividindo-se pelo número de sujeitos (N). Assim, no exemplo, torna-se a soma da última coluna da *tabela 5* e divide-se por 31 (número de sujeitos):

$$\bar{T} = \frac{37 + 34 + \dots + 15}{31} = \frac{868}{31} = 28,00$$

7. Calcular a variância do escore total. No exemplo:

$$V_i = \frac{\sum T^2}{N} - (\bar{T})^2$$

$$V_i = \frac{24986}{31} - (28,00)^2$$

$$V_i = 22,00$$

8. Calcular a média e a variância de cada item. Por exemplo, para o *item 1* tem-se:

$$I = (1 \times 2) + (2 \times 2) + (3 \times 9) + (4 \times 13) + (5 \times 5) = 110 \text{ (soma dos escores)}$$

$$I^2 = (1^2 \times 2) + (2^2 \times 2) + (3^2 \times 9) + (4^2 \times 13) + (5^2 \times 5) = 424 \text{ (soma dos quadrados dos escores)}$$

$$\bar{I} = \sum I/N = 110/31 = 3,548 \text{ (média dos escores atribuídos ao item)}$$

$$V_i = 424/31 - (3,548)^2 = 1,089 \text{ (variância no item).}$$

A *Tabela 6* traria os valores da média e da variância para cada item do nosso exemplo:

ITEM	1	2	3	4	5	6	7	8
MÉDIA	3,548							
VARIÂNCIA	1,089	0,946						1,273

9. Calcular o somatório da variância dos itens. Para os dados do exemplo (ver a *tabela 5*):

$$V_i = 1,089 + 0,946 + \dots + 1,273 = 9,144$$

10. Calcular o **coeficiente de fidedignidade (coeficiente α de Cronbach)**:

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum V_i}{V_T} \right)$$

$$\alpha = \frac{8}{7} \left(1 - \frac{9,144}{22,00} \right)$$

$$\alpha = 0,668$$

onde k é o número de itens do teste.

Como foi dito, o coeficiente de fidedignidade é um número entre -1 e 1 . Quanto mais próximo de 1 (em módulo) melhor é o teste. No exemplo, o coeficiente de fidedignidade obtido está bastante longe de um valor aceitável. A fim de melhorá-lo é preciso descobrir quais os itens do teste estão levando a esse valor baixo. Para isso, deve-se calcular o *índice de discriminação* de cada item, descartando-se aqueles itens que possuem baixo índice de discriminação. A lógica por trás desse procedimento é a seguinte: se uma questão do teste tem, aproximadamente, a mesma frequência de escores em dois grupos com escores totais diferentes (por exemplo os 25 % com escores superiores e os 25% com escores inferiores) então esse não é um bom item do teste. Essa é a próxima fase.

11. Dividir os sujeitos em três grupos (superior, médio e inferior) através do escore total. Os grupos superior e inferior devem perfazer, aproximadamente, 25 % dos indivíduos.

12. Calcular a média de cada item no grupo superior e no grupo inferior.

No exemplo, média do item 1 no grupo superior:

$$\bar{I}_s = (4 \times 5 + 4 \times 4) / 8 = 4,5$$

Média do item 1 no grupo inferior:

$$\bar{I}_i = (2 \times 3 + 4 + 2 \times 2 + 2 \times 1) / 7 = 2,286$$

A Tabela 7 traria os valores calculados para cada um dos itens do exemplo.

Tabela 7 - Valores médios para cada item relativos aos grupos superior G_s e inferior G_i para os dados do exemplo.

1. ITEM	1	2	3	4	5	6	7	8
2. MÉDIA G_s	4,5							
3. MÉDIA G_i	2,286							

13. Calcular o **índice de discriminação** de cada item. O índice de discriminação é definido como a diferença entre os valores médios dos escores do item nos grupos superior e inferior:

$$D_i = \bar{I}_s - \bar{I}_i$$

$$D_i = 4,5 - 2,286 = 2,214$$

no exemplo, para o item 1.

A Tabela 8 apresentaria os valores do índice de discriminação para todos os itens de nosso exemplo.

Tabela 8 - Valores do índice de discriminação para cada item do exemplo.

ITEM	1	2	3	4	5	6	7	8
D_i	2,214							

14. Calcular a seguir o **coeficiente de correlação item-total**. Esse coeficiente é o que indicará quais os itens que estão prejudicando a fidedignidade do teste e que devem ser eliminados. O **coeficiente de correlação item-total** é definido por:

$$r_{IT} = \frac{D_i}{K \times \sqrt{V_i}}$$

A constante K que aparece nesta equação depende da extensão dos grupos extremos; para os dados do exemplo seu valor é de 2,57. Valores dessa constante para várias combinações de grupos encontram-se tabelados em livros de Estatística.

Para o *item 1* do exemplo, o coeficiente de correlação item-total é dado por:

$$r_{IT} = 2,214 / [2,57 \times \sqrt{(1,089)}] = 0,826$$

Para os demais itens organizar-se-ia uma *tabela como a Tabela 9*:

<i>ITEM</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>
<i>r_{IT}</i>	<i>0,826</i>							

15. Verificar se existem itens com coeficiente de correlação item-total próximo de zero ou negativo. Se existirem esses devem ser eliminados. No exemplo os itens 5 e 8 devem ser eliminados.

16. Recalcular o escore total de cada indivíduo eliminando os itens deficientes. Note-se que agora os itens a serem levados em conta são seis e não mais oito. Assim, por exemplo, o escore total do primeiro sujeito passará ser 30, do segundo 28 e assim por diante.

17. Repetir as etapas de 6 a 9 com o objetivo de encontrar o novo coeficiente de fidedignidade. Para os dados do exemplo, após a eliminação dos itens 5 e 8 e da realização do novo cálculo, o coeficiente de fidedignidade passará a ser 0,904.

Na prática esse procedimento é feito, em segundos, por um programa de computador apropriado. O exemplo detalhado, passo a passo, foi dado com a intenção de facilitar a compreensão do que significa “calcular o coeficiente de fidedignidade”.

Validade¹⁷

Enquanto a fidedignidade é pré-requisito básico para qualquer instrumento de pesquisa, a validade é a característica mais importante que ele deve possuir. Isto porque a validade lida com o relacionamento entre os dados obtidos e o propósito para o qual eles foram coletados. Deste modo, validade é definida como o grau com que o procedimento realmente mede o que se propõe a medir.

A fidedignidade é uma etapa que precede a da validade, sendo essencial à mesma, e coloca um limite superior à validade do instrumento. Assim, um instrumento com fidedignidade 0,00 não pode ter qualquer validade; no outro extremo, um instrumento com validade 1,00 pode *possivelmente* ser perfeitamente válido. Para os graus intermediários de fidedignidade a validade máxima pode ser estimada a partir da raiz quadrada do coeficiente de fidedignidade. Entretanto, enquanto a fidedignidade coloca limites na validade, ela não é garantia desta última. Por exemplo, pode acontecer que um instrumento com uma fidedignidade de 0,60 tenha uma validade muito menor que $0,77(\sqrt{0,60} = 0,77)$ e, de fato, pode ter nenhuma validade.

Portanto, fidedignidade é pré-requisito para que um instrumento seja válido, garantindo que ele mede de forma correta, mas isto não é garantia de que ele meça o que está se propondo a medir. A validade deve ser estimada separadamente uma vez que a fidedignidade do instrumento já tenha sido estabelecida e que os valores encontrados sejam satisfatórios.

São vários os tipos de análise que podem ser feitas de modo a estabelecer a validade do instrumento. Contudo, não há, como para a fidedignidade, uma forma matemática de estabelecer a validade de um instrumento. A seguir são apresentados os vários tipos de análises possíveis que levam ao estabelecimento da validade de um instrumento:

¹⁷ Adaptado de Fox, 1969.

1. **Validade de face:** este tipo de validade é estabelecido a partir da análise superficial da natureza do instrumento, isto é, pela apresentação do instrumento. Obviamente, este é o tipo de análise de validade mais fraco.

2. **Validade de conteúdo:** mais apropriada que a anterior, esta técnica de análise de validade é, muitas vezes, a forma mais adequada disponível ao pesquisador para analisar certos tipos de instrumentos, tais como questionários e entrevistas. Entretanto, é uma técnica que depende do julgamento de quem faz a análise de validade o que, sem dúvida, é seu ponto fraco. Esta técnica verifica se o instrumento está medindo o que se propõe a medir através da análise, feita por especialistas no conteúdo do instrumento, da existência de razões racionais para a escolha dos itens do instrumento ou de uma base, lógica ou empírica, para esta escolha.

3. **Validade de construção:** esta técnica é definida como a habilidade do instrumento de distinguir grupos para os quais se sabe previamente que se comportam de forma diferentes na variável ou construto sob estudo. Em nível de procedimento, determinar a validade de construção tem dois estágios. O primeiro consiste na definição de um *critério* para identificar os grupos que diferem no construto que o novo instrumento se propõe a medir. O segundo estágio consiste em administrar o instrumento a estes grupos e determinar se eles diferem significativamente no novo instrumento também. Se assim for, então tem-se argumento para defender a validade do novo instrumento.

4. **Validade congruente ou concorrente:** o pesquisador que oferece validade concorrente ou congruente apresenta dados de natureza correlacional mostrando que o desempenho no seu novo instrumento correlaciona-se com o desempenho de algum instrumento de medir a variável já existente e aceito como válido. Se o critério que está sendo empregado pelo novo instrumento é da mesma família que o instrumento antigo se fala em validade congruente. Se, por outro lado, o critério empregado é de natureza diferente entre os dois instrumentos fala-se em validade concorrente.

O cerne das validades concorrente e congruente está na relevância e validade do critério.

5. **Validade preditiva:** existe validade preditiva quando o pesquisador pode prever comportamentos dos respondentes, na área de interesse da pesquisa, a partir dos dados obtidos pelo instrumento. Este tipo de processo de validação implica que o pesquisador deva esperar algum tempo para saber se as predições feitas se concretizaram ou não e em que extensão. Estes dados podem ser apresentados de várias formas, dentre elas a correlação entre os escores preditos e o escore real, porcentagens de predições corretas, etc..

Análise da Variância

Como foi visto nas seções sobre Estatística, a tendência central (a média por exemplo) não é suficiente para distinguir distribuições de frequência. A média aritmética de duas amostras, por exemplo, pode ser praticamente a mesma, mas a variabilidade dos escores pode ser muito diferente.

O teste F - análise da variância (anova)

Quando se quer determinar se os escores em uma amostra são mais variáveis que os escores em outra amostra, pode-se usar a técnica chamada de *teste F*. Usando o *teste F* podemos determinar se a variabilidade em um conjunto de dados é significativamente maior

que a variabilidade em outro conjunto de dados. Ao conduzir um *teste F*, usa-se uma medida da variabilidade *chamada de variância*, ao invés do desvio padrão. **A variância, em termos simples, é o quadrado do desvio padrão.**

Para executar um *teste F* entre duas variâncias, simplesmente divide-se a maior variância pela menor. Isto dá o que é chamado de *razão F* entre as duas variâncias. A questão a responder é se a variância obtida de uma amostra difere significativamente da variância obtida em outra amostra. Neste caso, a hipótese nula é que não existe diferença entre a variabilidade dos escores em uma amostra em relação a variabilidade dos escores da outra amostra. Usa-se então uma tabela de valores *F*, existente em livros de Estatística, para determinar se rejeita-se ou não a hipótese nula, no nível de significância escolhido.

O *teste F* pode também ser usado para analisar a variabilidade entre médias de escores de três ou mais amostras quando se pode assumir que as amostras foram obtidas através de seleção aleatória e a partir de uma população distribuída normalmente¹⁸. O *teste F* usado para comparação entre várias médias de escores é chamado de **Análise da Variância (ANOVA)** e envolve a comparação de duas variâncias estimadas.

O que se pretende é comparar uma estimativa da variância da população obtida a partir dos escores dentro de cada amostra com uma estimativa da variância obtida dos escores médios das várias amostras. Uma das variâncias estimadas é obtida pela comparação da variância estimada para cada uma das amostras separadamente e então combinando-as de modo a obter uma estimativa única chamada de **estimativa de variância dentro dos grupos**.

A outra variância estimada é calculada a partir dos escores médios para cada uma das amostras e calculando-se a variância estimada usando-se estes escores médios e o tamanho da amostra no cálculo. Esta variância é chamada de **estimativa de variância entre os grupos**.

O objetivo é determinar se a estimativa de variância entre grupos é significativamente maior que a estimativa de variância dentro dos grupos. Se a estimativa de variância entre os grupos for significativamente maior que a estimativa de variância dentro dos grupos poderemos rejeitar a hipótese nula e dizer que as amostras não provêm da mesma população. De modo a aplicar a análise da variância, uma *razão F* é calculada entre as duas variâncias estimadas: usando a variância entre grupos estimada como numerador e a variância dentro dos grupos como denominador. A partir daí, usando-se uma tabela de valores *F*, podemos determinar, para qualquer tamanho das amostras, a *razão F* necessária para rejeitar a hipótese nula, no nível de significância especificado.

Exemplo: suponhamos que queiramos determinar se os níveis de iluminação afetam a produtividade no trabalho em uma empresa de produtos eletrônicos. Para estudar isto seleciona-se, aleatoriamente, quatro amostras de quarenta empregados cada e os distribuí-se em diferentes níveis de iluminação. Mede-se então, a produtividade de cada grupo e obtem-se os dados da *Tabela 10*.

É evidente que a produtividade média das quatro amostras é diferente. Mas, é preciso saber se a variabilidade entre as médias amostrais (i.e., a diferença entre as médias) ocorreu como resultado de erro de amostragem ou se essa variabilidade pode ser atribuída à quantidade de iluminação.

¹⁸ De fato, poderíamos usar o *teste F* para comparação entre duas amostras, mas nesse caso os resultados seriam idênticos àqueles do *teste t*.

Tabela 10 - Dados para o exemplo 1.

<i>Nível</i>	<i>Valor médio</i>
<i>I</i>	<i>40</i>
<i>II</i>	<i>38</i>
<i>III</i>	<i>27</i>
<i>IV</i>	<i>26</i>

Neste exemplo, a hipótese nula a ser testada é a de que *não há diferença na produtividade dos empregados como resultado dos diferentes níveis de iluminação*. Para determinar isto, ou seja, se pode-se ou não desprezar a hipótese nula, examina-se este conjunto de dados usando a análise da variância. A interpretação dos resultados da análise da variância é feita do mesmo modo que aquela do teste *t*. Suponhamos que a *razão F* no exemplo dado seja significativa ao nível *0,01*. Poder-se-ia então rejeitar a hipótese nula e concluir que o nível de iluminação é relacionado à produtividade.

A inspeção das médias das amostras diz que a produtividade entre os níveis *I* e *II* difere somente por dois pontos. Da mesma forma, a diferença entre os níveis *III* e *IV* é somente de um ponto. A diferença maior aparece entre os níveis *II* e *III*. A análise da variância somente nos diz que existe uma diferença geral entre as quatro médias, porém, não informa qual dos grupos é o responsável pela diferença significativa. Como será visto mais adiante, a análise de qual é a causa da diferença observada é feita através da **Análise Fatorial da Variância**. A técnica de análise de variâncias pode ser usada para a análise de diferenças entre qualquer número de amostras e é também aplicável para a análise de diferenças entre grupos dentro de amostras, tais como masculino - feminino ou agrupamentos por idade. No exemplo, as amostras poderiam ter sido divididas em grupos de empregadas mulheres e grupos de empregados homens e, além disso, seria possível ainda tê-los agrupado por idade. Naturalmente, neste caso seriam necessários muito mais empregados na amostra. Usando a análise de variâncias poderiam ter sido examinadas diferenças em produtividade entre sexos, de acordo com a idade dos empregados e entre diferentes níveis de iluminação. Isto poderia ser chamado de análise de variâncias tri-lateral uma vez que poderia ter sido analisada a produtividade como função de idade, sexo e quantidade de iluminação.

O método da análise da variância

De modo a descrever o método de cálculo da *razão F* será utilizado outro exemplo. Suponhamos que um investigador educacional esteja interessado na eficácia relativa de dois métodos de ensino, denotados por A_1 e A_2 . Após selecionar dez estudantes o pesquisador os divide em dois grupos, aleatoriamente, um experimental e outro de controle. Após algum tempo, ele mede a aprendizagem dos sujeitos de ambos os grupos, usando algum tipo de teste. Os resultados são mostrados na *Tabela 11*:

Tabela 11 - Dois conjuntos de dados experimentais para um exemplo hipotético.

A_1	x	x^2	A_2	x	x^2
<i>4</i>	<i>0</i>	<i>0</i>	<i>3</i>	<i>0</i>	<i>0</i>
<i>5</i>	<i>1</i>	<i>1</i>	<i>1</i>	<i>-2</i>	<i>4</i>
<i>3</i>	<i>-1</i>	<i>1</i>	<i>5</i>	<i>2</i>	<i>4</i>

	2	-2	4		2	-1	1
	6	2	4		4	1	1
ΣX	20				15		$\Sigma X_t=35$
Σx^2			10				10
M	4				3		$M_t=3,5$

O trabalho a ser feito com os dados acima é o de localizar e computar as diferentes variâncias que compõem a variância total. Para calcular a *variância total* usa-se a fórmula:

$$V_t = \frac{\sum x^2}{N - 1}$$

onde $\sum x^2$ é a soma de quadrados (*sq*), $x = X - \bar{X}$ é o desvio dos escores médios e N é o número de casos na amostra como um todo. Deste modo, reagrupando os dados da *tabela 6.2*, sem a preocupação quanto ao grupo a que o sujeito pertence, obtém-se para a variância total o valor de $V_t=2,5$ (ver a *Tabela 12*).

Tabela 12 - Cálculos de V_t para os dados do exemplo.

	X	x	x^2
	4	0,5	0,25
	5	1,5	2,25
	3	-0,5	0,25
	2	-1,5	2,25
	6	2,5	6,25
	3	-0,5	0,25
	1	-2,5	6,25
	5	1,5	2,25
	2	-1,5	2,25
	4	0,5	0,25
ΣX	35		
M	3,5		
Σx^2			22,5

Observe-se que existe também uma variância entre os grupos, a qual é, presumivelmente, devida às manipulações experimentais. Isto é, o experimentador fez algo para um grupo e algo diferente para o outro. Estes tratamentos diferentes poderiam ocasionar diferenças entre os grupos, expressas pelas médias diferentes. Deste modo se os grupos forem diferentes haverá uma *variância entre os grupos*:

$$V_b = \frac{\sum x_b^2}{k - 1}$$

onde k é o número de grupos e x_b é a diferença entre a média do grupo e a média entre os grupos. O cálculo da variância entre os grupos, V_b é mostrado na *Tabela 13*.

Tabela 13 - Cálculos de V_b para os dados do exemplo 1.

X	x	x^2
4	0,5	0,25
3	0,5	0,25
ΣX	7	
M	3,5	
Σx^2		0,50

Existe ainda uma fonte remanescente de variância devida ao erro de amostragem. Leva-se em conta esse tipo de erro quando calcula-se a *variância dentro de cada grupo* separadamente e então toma-se a média desses valores de modo a obter V_w . Para o exemplo:

$$\frac{\sum x_{A1}}{n_{A1} - 1} = 10 / 4 = 2,5$$

$$\frac{\sum x_{A2}}{n_{A2} - 1} = 10 / 4 = 2,5$$

A média desses valores dá $V_w = 2,5$.

Foi dito antes que a variância total é composta de fontes separadas de variância: a variância entre grupos (V_b) e a variância dentro dos grupos (V_w). Logicamente, essas variâncias deveriam adicionar-se de modo a fornecer a variância total (V_t). A equação teórica seria então: $V_t = V_b + V_w$. Como pode-se ver dos dados acima não é isso que acontece. A razão é que os graus de liberdade é que foram usados nos cálculos feitos acima, como denominadores, ao invés de N , n e k . O cálculo das variâncias usando-se N , n e k é matematicamente correto porém estatisticamente incorreto. Outro aspecto importante da análise é a estimativa dos valores para a população. Pode ser mostrado que usando-se os graus de liberdade nos denominadores a fórmula da variância dará estimativas imparciais dos valores da população, um tópico de grande preocupação estatística. Deve-se, portanto, lembrar que variâncias, como usadas na análise de variâncias não são aditivas. Somas de quadrados por outro lado são sempre aditivas. As somas de quadrados também são, naturalmente, uma medida da variabilidade. Exceto na fase final da análise as somas de quadrados (sq) são calculadas, estudadas e analisadas.

Comparando a análise de variância com o teste t , pode-se dizer que a forma de encarar o problema é conceitualmente similar, com diferenças na metodologia. O método é geral: enquanto diferenças de mais que dois grupos podem ser testadas em termos de significância estatística usando a análise de variância o teste t aplica-se somente a dois grupos. A análise de variâncias para dois grupos fornece os mesmos resultados que o teste t .

A razão formada pela divisão da variância entre grupos (V_b) pela variância dentro dos grupos (V_w) é chamada de razão F :

$$F = \frac{V_b}{V_w}$$

Os valores F dos dados experimentais são calculados e comparados contra uma tabela de valores F . Se os valores obtidos são maiores ou muito maiores que os valores tabelados, naquele nível de significância estatística e graus de liberdade, as diferenças expressas por V_b refletem diferenças significativas. Neste caso, a hipótese nula, de que não há diferença alguma entre as médias é rejeitada naquele nível de significância.

Na prática podemos calcular a *razão F* a partir dos escores brutos. No exemplo precedente, variâncias padrão foram usadas de modo a mostrar os aspectos fundamentais do método. Naturalmente, todo esse cálculo pode ser feito rapidamente com um aplicativo apropriado.

Análise fatorial da variância

Na análise fatorial da variância duas ou mais variáveis variam independentemente ou interagem uma com a outra para produzir variações na variável dependente. A análise fatorial da variância é o método estatístico que analisa os efeitos independentes e interativos de duas ou mais variáveis independentes em uma variável dependente.

No passado, muitos pesquisadores (e isto ainda é válido para muitos pesquisadores nos dias de hoje) acreditavam que o método de pesquisa mais efetivo era permitir que uma variável independente evoluísse enquanto as demais variáveis independentes eram controladas, tanto quanto possível. A análise fatorial da variância veio mudar esse quadro, permitindo que se possa analisar o efeito de várias variáveis independentes ao mesmo tempo.

A análise fatorial serve para vários propósitos. Primeiro, o delineamento fatorial e a análise fatorial da variância permitem ao pesquisador manipular e controlar duas ou mais variáveis simultaneamente¹⁹. Por exemplo, podemos não somente estudar os efeitos de determinado método de ensino sobre a aprendizagem mas também analisar os efeitos, digamos, de tipos de reforço sobre as respostas. Além disso podemos controlar as variáveis tais como sexo, inteligência e classe social.

Variáveis que não são manipuladas podem ser controladas. Ao invés do procedimento disseminado de parear os sujeitos para testes sobre inteligência ou atitudes, pode-se construir estas variáveis (e muitas outras) em delineamentos com um caráter fatorial. Não somente controlar-se estas variáveis como também obtém-se informação adicional de grande valor e significância.

Interação estatística

Relativamente à interação estatística, a hipótese nula (H_0) é de que **não há interação entre as variáveis independentes**, ou seja, de que não existe influência da combinação de variáveis. Esta hipótese poderia ser chamada de *hipótese da diferença constante ou hipótese dos escores relativos constantes*. O que se quer dizer com isto é que as diferenças entre células de linhas diferentes permanecerão constantes de coluna para coluna; equivalentemente

¹⁹ Embora seja possível o controle de mais que três variáveis, estes delineamentos se tornam pouco práticos devido à dificuldade de obter-se um número suficiente de sujeitos de modo a ocupar todas as células. A forma mais simples de uma análise fatorial de variância é $2 \times 2 \times 2$.

poder-se-ia dizer que as diferenças entre células de colunas diferentes permanecerão constantes de linha para linha.

Como um exemplo, considere-se a *Tabela 14*, uma hipotética tabela de médias, de 12 células.

*Tabela 14 - Dados de um exemplo onde **inexiste** interação.*

	<i>Colunas</i>				
<i>L</i>		1	2	3	4
<i>i</i>	1	3	6	5	7
<i>n</i>	2	2	5	4	6
<i>h</i>	3	5	8	7	9
<i>a</i>					
<i>s</i>					

Este é um exemplo típico de falta de interação. Observe-se que a diferença das médias em células na segunda linha em relação à primeira linha é constante e igual a 1 u enquanto que a diferença entre os valores na primeira linha e os da terceira também é constante e igual a 2. As diferenças entre linhas sendo constantes, então as diferenças entre colunas também o serão. A *Figura 5* mostra graficamente o que significa não haver interação entre as variáveis.

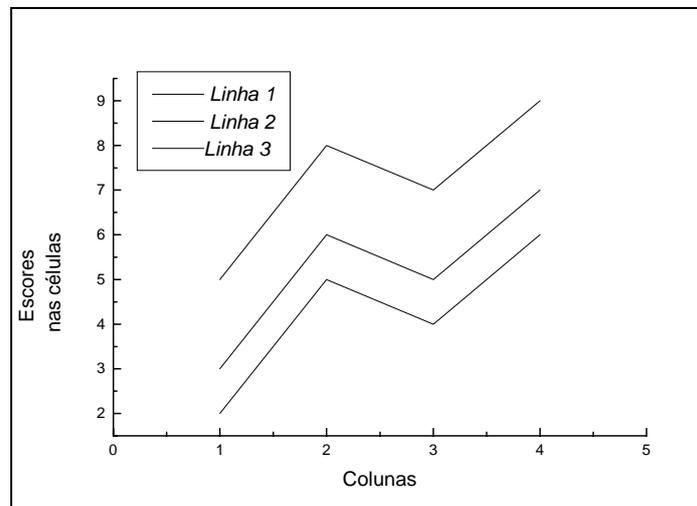


Figura 5 - Um exemplo onde não há interação.

Como já comentado anteriormente, graficamente a não-interação aparece como linhas paralelas em um gráfico onde representa-se os valores médios das células em cada linha. Marca-se no gráfico, para cada linha da tabela, os valores médios e então são ligados por uma linha. Se não houver interação (*hipótese nula*) então as linhas obtidas para cada linha da tabela não se cruzarão, ou mais precisamente, serão paralelas. No caso de haver algum tipo de interação então as linhas não mais serão paralelas, não havendo necessidade de que elas se cruzem. Um exemplo de interação (fictício) vai mostrado na *Figura 6* a qual é uma representação gráfica dos dados da *Tabela 15*:

Tabela 15 - Dados de um exemplo onde existe interação.

		Colunas			
		1	2	3	4
Linhas	1	4	5	7	5
	2	3	1	4	4

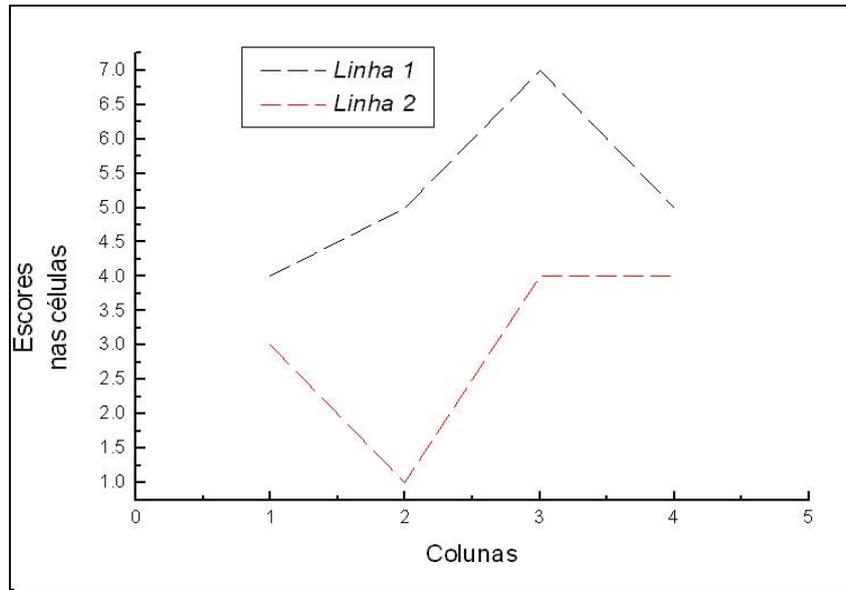


Figura 6 - Um exemplo onde existe interação.

Como um último comentário desta seção, convém discutir o que se entende por análise de variância unilateral e o que se entende por análise de variância bilateral. Não é o **número de variáveis** que se está utilizando que define o tipo de análise de variância, é o **número de classes** de variáveis utilizadas. Assim, por exemplo, em um determinado experimento analisa-se a influência de três métodos de ensino diferentes. Então, nesse caso faz-se uma análise de variâncias unilateral, uma vez que as variáveis analisadas (**métodos de ensino**) pertencem a uma única classe (**métodos**). Por outro lado, se além de métodos de ensino estiverem sendo analisados diferentes tipos de motivação o caso será de uma análise de variâncias bilateral, pois, nesse caso, haverá **duas** classes: **métodos de ensino e tipos de motivação**. Esse exemplo será discutido na próxima seção de modo a discutir o método propriamente dito.

O método da análise fatorial da variância

De modo a estudar-se o método usado na análise fatorial da variância apresenta-se a seguir um exemplo hipotético²⁰.

Um pesquisador está interessado no estudo da influência de dois fatores na aprendizagem. O primeiro desses fatores é o método de ensino, que pode ser um de dois métodos a serem estudados, e que serão denotados por A_1 e A_2 e o segundo fatores de motivação, os quais poderão ser um de dois possíveis, e serão denotados por B_1 e B_2 . Desse

²⁰ Extraído de Kerlinger, 1964.

modo os sujeitos são distribuídos em células para o estudo onde interagem um método de ensino e um fator de motivação como, por exemplo $A_1 B_2$ que nos indica que os sujeitos serão submetidos ao método de ensino A_1 e ao fator de motivação B_2 . As possibilidades estão dispostas na *Tabela 16*. A nossa amostra hipotética é composta por oito sujeitos, distribuídos na forma de dois sujeitos por célula.

Tabela 16 - Delineamento fatorial para variáveis método de ensino e motivação.

<i>Métodos</i>			
		A_1	A_2
<i>Motivação</i>	B_1	$A_1 B_1$	$A_2 B_1$
	B_2	$A_1 B_2$	$A_2 B_2$

Testes estatísticos paramétricos e não paramétricos

Um tópico central na moderna teoria estatística é a *Estatística Inferencial*. A estatística inferencial está preocupada em resolver dois tipos de problemas: a estimativa dos parâmetros da população e testes de hipóteses. Na inferência estatística a preocupação é como tirar conclusões sobre um grande número de eventos com base em observações de uma porção deles. A Estatística dá ferramentas com as quais formaliza-se e padroniza-se os procedimentos para tomarmos decisões.

Um problema comum em estatística inferencial é o de determinar, em termos de probabilidades, se as diferenças observadas entre duas amostras significam que as populações a partir das quais as amostras são retiradas são realmente diferentes. As diferenças podem ocorrer apenas devido ao acaso durante o processo de amostragem.

No desenvolvimento dos métodos estatísticos, as primeiras técnicas estatísticas de inferência que apareceram foram aquelas as quais faziam muitas hipóteses sobre a natureza da população da qual os escores eram retirados. Uma vez que os valores da população são parâmetros, estas técnicas estatísticas foram chamadas de *paramétricas*²¹. Por exemplo, a técnica de inferência pode ser baseada na hipótese de que os escores foram retirados de uma população cujos escores seguem a distribuição normal. Ou a técnica pode ser baseada na hipótese de que os conjuntos de escores foram retirados de populações tendo a mesma variância ou o mesmo espalhamento dos escores.

Mais tarde surgiu um grande número de técnicas estatísticas de inferência que não fazem hipóteses por demais numerosas ou restritivas sobre os parâmetros da população. Essas técnicas, as quais são *independentes de distribuição* ou *não-paramétricas*, levam a conclusões que apresentam poucas limitações. Algumas técnicas não-paramétricas são chamadas de *testes de ordenamento*. Este nome tem origem no fato de que enquanto as técnicas paramétricas têm o seu foco central nas diferenças de médias e variâncias, as técnicas não-paramétricas têm seu foco no *ordenamento* dos escores e não em seus valores numéricos.

²¹ A média, o desvio padrão e a variância de uma população, ou qualquer outra medida da população, são *parâmetros*.

Enquanto um parâmetro é um valor de uma determinada população, uma estatística por outro lado é uma medida calculada de uma amostra. *Um teste estatístico não-paramétrico é um teste cujo modelo não especifica condições sobre os parâmetros da população da qual as amostras foram retiradas.*

Quando faz-se alguma afirmativa a respeito da natureza da população e sobre o processo de amostragem se está estabelecendo um *modelo estatístico*. Associado com *todo* teste estatístico existe um modelo e uma prescrição de medida; o teste estatístico em questão é válido sobre certas condições, e o modelo e a prescrição de medida especificam estas condições. Algumas vezes algumas vezes é possível testar se as condições de um modelo estatístico particular se encontram presentes mas, muitas vezes, tem-se que tomar por hipótese a presença dessas condições. Deste modo, as condições de validade de um teste, o modelo estatístico (ou seja, as hipóteses feitas quando da construção do teste), são muitas vezes chamadas de *hipóteses* do teste.

É óbvio que quanto menos ou mais fracas forem as hipóteses por trás do modelo subjacente a um teste estatístico menos restrições é preciso que fazer sobre as conclusões obtidas pelo teste estatístico associado com o modelo. Ou seja, quanto menor ou menos restritivas forem as hipóteses por trás do modelo mais *gerais* as conclusões e/ou resultados obtidos.

Entretanto, os testes mais poderosos são justamente aqueles que têm as hipóteses mais fortes ou condições mais restritivas. Os testes paramétricos, por exemplo, o teste *t* ou o teste *F*, têm uma variedade de hipóteses fortes embasando o seu uso. Quando aquelas hipóteses são válidas estes testes são os que mais provavelmente rejeitarão a hipótese H_0 quando esta hipótese for falsa.

As condições as quais devem ser satisfeitas para fazer o teste *t*, o teste mais poderoso e, de fato, necessárias para que se tenha confiança em qualquer inferência feita a partir de resultados obtidos com o teste são:

1. **As observações devem ser independentes** – com isto se quer dizer que a seleção de qualquer caso da população para inclusão na amostra não deve influenciar a probabilidade de inclusão de qualquer outro caso. Da mesma forma, o escore atribuído a um caso não deve influenciar o escore atribuído a outro caso.
2. **As observações devem ser retiradas de uma população normal** - Como visto anteriormente, uma das hipóteses por trás do teste *t* é que a população segue a distribuição normal.
3. **As variâncias das populações de onde as amostras são retiradas devem ser iguais.**
4. **As escalas devem ser intervalares ou racionais** - isso é necessário para que se possa executar operações de caráter aritmético sobre os números representativos dos escores.
5. **Aditividade (condição para validade do teste F)** – as médias das distribuições normais devem ser combinações lineares dos efeitos devidos a colunas e/ou linhas. Isto é, os efeitos devem ser aditivos.

Deste modo, um teste estatístico paramétrico é um teste cujo modelo especifica certas condições sobre os parâmetros da população da qual a amostra foi retirada. Uma vez que essas condições normalmente não são testadas, elas são assumidas como verdadeiras. A

significância e validade de um teste paramétrico depende do grau de acerto ao supor-se corretas estas hipóteses.

Um teste estatístico não-paramétrico é um teste cujo modelo não especifica condições sobre os parâmetros da população da qual as amostras foram retiradas.

Vários critérios poderiam ser considerados na escolha de um teste estatístico para uso quando um processo de tomada de decisão sobre a hipótese de pesquisa está em curso. Estes critérios são:

1. **O poder do teste:**

Poder = 1 – probabilidade de um erro tipo II.

Isto é, a probabilidade de aceitar-se H_0 quando de fato ela é falsa.

2. **A aplicabilidade do modelo estatístico no qual o teste é baseado aos dados da pesquisa.**

3. **Poder-eficiência** - O conceito de poder - eficiência é relativo à quantidade de aumento no tamanho da amostra a qual é necessária para tornar um teste B tão poderoso quanto um teste A . O poder-eficiência do teste B em relação ao teste A é definido por:

$$P_B = 100 \times \frac{N_A}{N_B}$$

onde N_A e N_B são, respectivamente, os tamanhos das amostras submetidas aos testes A e B .

4. **O nível da medida obtido na pesquisa.**

Vantagens de testes estatísticos não-paramétricos

1. Afirmações probabilísticas obtidas a partir de testes estatísticos não-paramétricos são **probabilidades exatas**.
2. Se os tamanhos das amostras são pequenos ($N \leq 6$) não existe alternativa ao uso de testes estatísticos não-paramétricos a menos que se conheça **exatamente** as características da distribuição seguida pela população.
3. Existem testes estatísticos apropriados para tratamento de amostras as quais provêm de várias populações diferentes. Nenhum dos testes paramétricos pode manipular dados deste tipo sem exigir que se acredite em hipóteses irrealísticas.
4. Há testes estatísticos não-paramétricos para tratar dados os quais são inerentemente *ordinais*, isto é, o pesquisador é apenas capaz de dizer que um sujeito apresenta mais ou menos de determinada característica, mas não sabe dizer *o quanto* a mais ou a menos em relação a outro sujeito.
5. Os testes não-paramétricos são capazes de tratar dados os quais são simplesmente classificatórios, seguindo uma escala tipo *nominal*. Nenhum dos testes paramétricos é capaz de lidar com dados com essas características.

6. Por fim, os testes não-paramétricos são mais fáceis de aprender e aplicar que os testes paramétricos.

Entretanto, se todas as hipóteses do modelo estatístico paramétrico são de fato encontradas nos dados e se as medidas são do tipo exigido pelo teste, então testes estatísticos não-paramétricos são desnecessários. Uma medida da necessidade do uso de um teste não-paramétrico em uma situação desse tipo é dada pelo *poder-eficiência* do teste não-paramétrico. Suponhamos que o poder-eficiência do teste não-paramétrico seja de 90%. Isto significa que um teste não-paramétrico na mesma situação exigiria somente 10 % da amostra para ser tão efetivo quanto o teste paramétrico.

Testes não-paramétricos - o caso de uma amostra

Esta situação acontece quando, aleatoriamente, retira-se uma amostra de determinada população e testa-se a hipótese de que essa amostra vem de uma população com uma distribuição especificada.

Uma técnica paramétrica comum no caso de uma amostra é usar-se o teste **t** para a diferença entre as médias observada (a da amostra) e esperada (a da população). Entretanto, existem muitos tipos de dados para os quais o teste **t** é inaplicável. Esses fatores de inaplicabilidade do teste **t** podem ter várias fontes, dentre as quais cabe salientar:

1. As hipóteses e exigências para a aplicabilidade do teste não são satisfeitas para os dados do experimento particular.
2. Pode ser preferível frente a determinada situação evitar-se as hipóteses sobre as quais o teste **t** é construído e ganhar-se com isto em generalidade das afirmações resultantes da pesquisa.
3. Os dados da pesquisa são, inerentemente, dos tipos nominal ou ordinal e, portanto, não são passíveis de análise pelo teste **t**.
4. O pesquisador não está de fato interessado somente em diferenças na localização, mas sim em qualquer tipo de diferença não importando a sua origem.

Neste caso o experimentador pode escolher usar um dos seguintes testes estatísticos não-paramétricos:

1. **O teste binomial:** a distribuição binomial é a distribuição de amostragem de proporções que podemos observar em amostras retiradas de uma população composta por duas classes. Isto é, ela fornece os vários valores os quais podem ocorrer sob H_0 . Portanto, quando os escores da pesquisa estão em duas classes, a distribuição binomial pode ser utilizada para testar H_0 .
2. **O teste χ^2 para uma amostra:** este teste é utilizável quando queremos analisar dados os quais caem em mais que duas categorias. Este teste é útil para decidir se a diferença observada entre o número de objetos que recaem em cada categoria e o número esperado com base na hipótese nula é significativa.
3. **O teste de Kolmogorov-Smirnov para uma amostra:** este teste está preocupado com o grau de concordância entre a distribuição de um conjunto de escores observados e

alguma previsão teórica específica. Ele determina se os escores na amostra podem razoavelmente serem pensados como vindo de uma população a qual obedece à distribuição predita teoricamente.

4. **Teste do período para uma amostra:** neste caso se está interessado em saber se os dados sob análise provêm de uma distribuição aleatória. Ele é baseado na sequência em que os escore aparecem originalmente, isto é, o número de períodos que a amostra apresenta.

Testes não-paramétricos - o caso de duas amostras relacionadas

Testes estatísticos de duas amostras são usados quando o pesquisador deseja estabelecer se dois tratamentos são diferentes ou se um tratamento é ‘*melhor*’ do que outro. Neste tipo de comparação, algumas vezes diferenças significativas são observadas as quais não resultam do tratamento. Uma maneira de transpor-se a dificuldade imposta por diferenças estranhas entre os grupos é usar duas amostras relacionadas no experimento. Isto é, podemos parear ou relacionar de alguma outra maneira as duas amostras estudadas. Este pareamento pode ser atingido usando cada sujeito como seu próprio controle, ou pelo pareamento de sujeitos e então distribuindo os membros do par aleatoriamente aos grupos de controle e experimental. Quando o sujeito serve como seu próprio controle ele é exposto a ambos os tratamentos em tempos diferentes. Quando o método do pareamento é utilizado, o esforço é para selecionar para cada par de sujeitos indivíduos que sejam tão iguais quanto possível na (s) variável (eis) estranha (s) a(s) qual (is) pode (m) influenciar o experimento.

Quando possível, o método de usar cada sujeito como seu próprio controle é preferível ao método do pareamento. A razão para isto é que é difícil parear pessoas por desconhecimento de variáveis relevantes as quais determinam o comportamento. O delineamento de pareamento é somente uma boa opção na medida em que o pesquisador seja hábil em parear os sujeitos, sendo esta habilidade, frequentemente, bastante limitada.

A técnica paramétrica mais usada para comparações entre duas amostras correlacionadas é o teste **t**. Uma diferença de escores pode ser obtida de dois escores provenientes de cada um dos membros dos sujeitos pareados ou dos escores de cada sujeito sob as duas condições. O teste **t** assume que estas diferenças nos escores são normalmente distribuídas em uma população da qual a população foi retirada.

Em inúmeros casos o teste **t** é inaplicável. Nestas situações o pesquisador pode escolher um dos seguintes testes estatísticos não-paramétricos:

1. **O teste de McNemar para a significância de variações:** este teste é particularmente aplicável para aqueles delineamentos de “*antes*” e “*depois*” nos quais cada pessoa é usada como seu próprio controle e em medidas nas quais os escores são obtidos usando-se variáveis nominais ou ordinais. Deste modo, por exemplo, o teste pode ser usado para testar a efetividade de um tratamento particular (encontro, editorial de um jornal, propaganda via correio, visita pessoal, etc.) nas preferências de votos entre vários candidatos. Ou ele pode ser usado para averiguar a influência das migrações do campo para a cidade na preferência política, e assim por diante.
2. **O teste do Sinal:** o teste do sinal recebe o seu nome do fato que ele usa como dados os sinais de mais e menos mais do que medidas quantitativas. Este teste é particularmente

útil para a pesquisa na qual medidas quantitativas são impossíveis ou impraticáveis, mas na qual é possível ordenar com respeito um ao outro os membros de pares.

3. **O teste de Walsh:** se o experimentador puder assumir que a diferença nos escores que ele observa, em duas amostras relacionadas, provém de populações que obedecem a distribuições simétricas, ele pode usar um teste bastante poderoso desenvolvido por Walsh. Note-se que a hipótese não é que as diferenças nos escores são de populações normais (para as quais pode-se usar o teste paramétrico t); note-se também que as diferenças também não provêm da mesma população. O que o teste faz é assumir que as populações são simétricas, de modo que a média é uma adequada medida de tendência central e é igual à mediana.

Testes não-paramétricos - o caso de duas amostras não relacionadas

Quando o uso de duas amostras relacionadas é impraticável ou inapropriado, pode-se fazer uso de duas amostras independentes. Neste delineamento, as duas amostras podem ser obtidas por um dos dois métodos:

1. as amostras podem ser retiradas aleatoriamente de duas populações,
2. as amostras podem originar-se devido à atribuição, aleatória, de dois tratamentos a membros, cujas origens são arbitrárias, de uma mesma amostra.

Nos dois casos não é necessário que as duas amostras tenham o mesmo tamanho.

A técnica paramétrica normal de averiguar-se a diferença entre duas amostras não relacionadas é também aplicar-se o teste t nas médias das amostras. No caso da não aplicabilidade do teste paramétrico (por exemplo, não ter certeza que as distribuições são normais) o pesquisador pode escolher para analisar os seus dados um dentre vários testes não-paramétricos, os quais são apresentados a seguir.

Teste da probabilidade exata de Fisher

Esta é uma técnica extremamente útil para analisar dados discretos quando as duas amostras independentes têm tamanhos pequenos. Este teste é usado quando os escores das duas amostras independentes, escolhidas aleatoriamente, caem em uma de duas categorias mutuamente exclusivas. Em outras palavras, todo sujeito em ambos os grupos obtém um de dois escores possíveis. Os escores são representados por frequências em uma tabela de contingência 2×2 , como a Tabela 17.

Tabela 17 - Tabela de contingência para o teste de Fischer.

	-	+	
<i>Grupo I</i>	<i>A</i>	<i>B</i>	<i>A+B</i>
<i>Grupo II</i>	<i>C</i>	<i>D</i>	<i>C+D</i>
<i>Total</i>	<i>A+C</i>	<i>B+D</i>	<i>N</i>

Para os dados nessa tabela (onde A , B , C e D são frequências) poderia ser determinado se o *Grupo I* e o *Grupo II* diferem significativamente na proporção de mais ou menos atribuídos a eles.

A probabilidade exata de observar-se um particular conjunto de frequências em uma tabela 2×2 , quando os totais marginais são mantidos fixos, é dada pela *distribuição hipergeométrica*:

$$p = \frac{\binom{A+C}{A} \binom{B+D}{B}}{\binom{N}{A+B}}$$

$$p = \frac{(A+C)! (B+D)!}{A! C! B! D!} \times \frac{(A+B)! (C+D)!}{N!}$$

Deste modo:

$$p = \frac{(A+B)! (C+D)! (A+C)! (B+D)!}{N! A! B! C! D!}$$

Ou seja, a probabilidade exata da ocorrência observada é obtida tomando-se a razão entre o produto de fatoriais de quatro totais marginais e o produto dos fatoriais dos escores encontrados em cada célula por fatorial de N , o número total de observações independentes²¹.

Exemplo: Considere-se os dados mostrados na *Tabela 18*:

	-	+	
<i>Grupo I</i>	10	0	10
<i>Grupo II</i>	4	5	9
<i>Total</i>	14	5	19

Nesta tabela, $A=10$, $B=0$, $C=4$ e $D=5$. Os totais marginais são $A+B=10$, $C+D=9$, $A+C=14$ e $B+D=5$. N , o número total de observações independentes, é 19. A probabilidade exata de que estes 19 casos recaíssem nas células do modo como o que foi observado é dada por:

$$p = \frac{10!9!14!5!}{19!10!0!4!5!} = 0,0108$$

Portanto determinamos que a probabilidade de obter-se essa distribuição dos escores, sob H_0 , é $p=0,0108$.

²¹ O fatorial de um número a , denotado por $a!$ (lê-se a fatorial), é obtido pelo produto dos inteiros, começando em 1, até o número a . Assim, por exemplo: $3! = 1 \times 2 \times 3 = 6$. O $0!$ é igual a 1, por definição.

O teste χ^2 para duas amostras independentes

Quando os dados consistem de categorias discretas esse teste pode ser usado para determinar a significância estatística de diferenças entre dois grupos independentes. As medidas envolvidas podem ser inclusive as de uma escala nominal. Por exemplo, se deseja-se saber se dois grupos de professores, de Física e de Química, diferem quanto a sua opinião em relação a uma certa estratégia de ensino, pode-se medir essa opinião com um simples “a favor” ou “contra” calcular as frequências e aplicar o teste χ^2 .

O teste da Mediana

Este é um procedimento para testar se duas amostras independentes diferem em tendência central. Ele fornecerá informação de quão provável é que as duas amostras independentes (não necessariamente de mesmo tamanho) tenham sido retiradas de populações com a mesma mediana.

O teste U de Mann-Whitney

Quando dispõe-se de pelo menos uma medição ordinal esse teste pode ser usado para verificar se duas amostras independentes foram tiradas da mesma população. É um dos mais potentes testes não-paramétricos e é uma das melhores alternativas ao teste paramétrico t quando o pesquisador quer evitar suposições subjacentes ao teste t ou quando a medição feita é mais fraca do que uma escala intervalar (Siegel, 1956, p. 116).

O teste de duas amostras de Kolmogorou-Smirnov

Este é também um teste para verificar se duas amostras independentes foram retiradas da mesma população. A forma bilateral do teste é sensível somente a qualquer tipo de diferença na distribuição da qual as duas amostras foram retiradas: diferenças em localização (tendência central), em dispersão, em simetria, etc.. O teste unilateral é utilizado para testar se os escores de um grupo experimental serão “melhores” que aqueles do grupo de controle.

O teste de Wald-Wolfowitz

Se quisermos testar a hipótese nula de que duas amostras independentes foram selecionadas da mesma população contra a hipótese alternativa de que os dois grupos diferem completamente poderemos utilizar este teste. Isto é, com amostras suficientemente grandes este teste pode rejeitar H_0 se as duas populações diferirem em qualquer aspecto: tendência central, variabilidade, simetria ou algum outro fator. Deste modo, este teste pode ser usado em uma larga classe de hipóteses alternativas. Enquanto muitos testes são endereçados a tipos específicos de diferenças entre dois grupos o teste de Wald-Wolfowitz analisa qualquer tipo de diferença.

O teste de Moses de reações extremas

Em ciências do comportamento algumas vezes espera-se que uma condição experimental cause em alguns sujeitos o aparecimento de comportamentos extremos em uma determinada direção enquanto que em outros sujeitos o comportamento será extremado na direção oposta. Deste modo, pode-se pensar que depressão econômica e instabilidade política provocarão em algumas pessoas reações extremamente conservadoras enquanto que outras reagirão de uma forma extremamente progressista, em termos de opiniões políticas.

O teste de Moses é especificamente projetado para uso com dados coletados para testar esse tipo de hipóteses. Ele poderia ser usado quando é esperado que a condição experimental afete alguns sujeitos de um modo e outros de maneira oposta.

O teste da aleatoriedade para duas amostras independentes

Esta é uma técnica não-paramétrica poderosa e útil para a testagem da significância da diferença entre as médias de duas amostras independentes quando N_1 e N_2 são pequenos. Com o teste da aleatoriedade podemos determinar a probabilidade exata, sob H_0 , associada com nossas observações e podemos fazê-lo sem assumir a distribuição normal ou homogeneidade da variância nas populações envolvidas (as quais devem ser assumidas se o teste paramétrico equivalente, o teste t , for usado).

Discussão

Todos os testes não-paramétricos para duas amostras independentes testam se é provável que as duas venham da mesma população. Entretanto, os vários testes apresentados são mais ou menos sensíveis aos diferentes tipos de diferenças entre as duas amostras. Por exemplo, quando se quer testar se duas amostras representam populações que diferem em localização (tendência central), existem testes os quais são mais sensíveis a este tipo de diferença e portanto poderiam ser escolhidos: o teste da mediana, o teste de Fisher (para N pequeno), o teste U de Mann-Whitney, o teste de Kolmogorov-Smirnov (para duas amostras, unilateral) e o teste da aleatoriedade. Por outro lado, se fosse desejo do pesquisador determinar se suas duas amostras provêm de populações as quais diferem em qualquer aspecto em geral, i.e., localização ou dispersão ou simetria, etc., ele poderia escolher um dos seguintes testes: o χ^2 , o teste de Kolmogorov-Smirnov (bilateral) ou o teste de Wald-Wolfowitz. A técnica restante, o teste de Moses, serve unicamente para testar se um grupo experimental está exibindo reações extremas, sejam extremistas ou defensivas, em comparação com as reações exibidas por um grupo de controle independente.

A escolha entre os testes que são sensíveis a diferenças em localização são determinadas pelo tipo de medida obtida na pesquisa e pelo tamanho das amostras. O teste mais poderoso em termos de localização é o teste da aleatoriedade. Entretanto, este teste pode ser somente usado quando os tamanhos das amostras forem pequenos e quando tivermos confiança na natureza numérica da medida obtida. Com grandes amostras ou medidas fracas (medidas ordinais) a alternativa sugerida é o teste U de Mann-Whitney o qual, quase sempre, é mais poderoso que o teste da aleatoriedade. Se as amostras são muito pequenas, o teste de Kolmogorov-Smirnov é levemente mais eficiente que o teste U. Se a medida for de um tipo tal que é significativo somente dicotomizar as observações acima ou abaixo da mediana, então o teste da mediana é aplicável. Este teste não é tão poderoso quanto o teste U de Mann-

Whitney em termos de proteção contra diferenças de localização mas é mais apropriado que aquele quando os dados das observações não podem ser completamente hierarquizados. Se o tamanho das amostras sob consideração for muito pequeno, quando aplicando o teste da mediana, o pesquisador poderia fazer uso do teste de Fisher.

A escolha entre os testes que são sensíveis a todas as diferenças é decidida a partir da intensidade das medidas obtidas, o tamanho das amostras e o poder relativo dos testes disponíveis. O teste χ^2 é apropriado para dados os quais são medidos em escalas nominais ou mais fortes. Quando N é pequeno e os dados estão em uma tabela de contingência 2×2 poderíamos usar o teste de Fisher ao invés do teste χ^2 . Em muitos casos o teste χ^2 pode não fazer uso eficiente de toda a informação contida nos dados. Se os escores das populações forem continuamente distribuídos, nós podemos escolher ou o teste de Kolmogorov-Smirnov (bilateral) ou o teste de Wald-Wolfwitz ao invés do teste χ^2 . De todos os testes para qualquer tipo de diferença, o teste de Kolmogorov-Smirnov é o mais poderoso. Se ele for usado com dados os quais não assumem a hipótese de continuidade ele ainda é aceitável mas opera mais conservadoramente, isto é, os valores de p obtidos serão levemente maiores. Se a hipótese nula for rejeitada a partir desses dados, pode-se seguramente ter confiança na decisão. O teste de Wald-Wolfwitz também protege contra todos os tipos de diferenças mas não é tão poderoso quanto o anterior.

Dois pontos devem ser enfatizados sobre o uso de testes do segundo grupo. Primeiro, quando se está interessados em testar hipóteses alternativas de que os grupos diferem em tendência central, isto é, de que um grupo tem média maior que o outro, então deve-se usar um teste especificamente projetado para capturar diferenças em localização, um dos testes do primeiro grupo listado acima. Segundo, quando rejeita-se a hipótese nula com base em um teste que capta qualquer tipo de diferença (um dos testes do segundo grupo) deve-se assegurar de que os dois grupos provêm de populações diferentes, porém não pode-se dizer em que elas diferem.

Bibliografia

- BEST, J. W. (1970). *Research in education*. 2. ed. Englewood Cliffs: Prentice Hall.
- CAMPBELL, D. R.; STANLEY, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In Gage, N. L. *Handbook of research on teaching*, Chap. 5. Chicago: Rand Mc Nally.
- CAMPBELL, D. R.; STANLEY, J. C. (1991). *Diseños experimentales y cuasiexperimentales en la investigación social*. Buenos Aires: Amorrortu Editores.
- DARLINGTON, R. B. (1975). *Radicals and squares*. Ithaca, N. Y.: Logan Hill Press.
- ELSEY, F. F. (1967). *A first reader in statistics*. Belmont, CA: Brooks/Cole Publishing Co.
- FOX, D. J. (1969). *The research process in education*. New York: Holt, Rinehart and Winston.
- GLASS, G. V.; STANLEY, J. C. (1970). *Statistical methods in education*. Englewood Cliffs, N. J.: Prentice Hall.
- GOWIN, D. B. (1981). *Educating*. Ithaca, N.Y.: Cornell University Press.
- GOWIN, D. B.; ALVAREZ, M. (2005). *The art of educating with V diagrams*. New York: Cambridge University Press.

- KERLINGER, F. N. (1964). *Foundations of behavioral research*. New York: Holt, Rinehart and Winston.
- KERLINGER, F. N. (1980). *Metodologia da pesquisa em ciências sociais*. São Paulo: E.P.U., EDUSP, INEP.
- MILLMAN, J. (1970). *Data analysis*. Conferência convidada proferida no Simpósio Nacional de Professores de Pesquisa Educacional, St. Louis, USA.
- MOREIRA, M. A. (2006). *Mapas conceituais e diagramas V*. Porto Alegre: Editora do Autor.
- RUNKEL, P. J.; MC GRATH, J. E. (1972). *Research on human behavior*. New York: Holt, Rinehart and Winston.
- SIEGEL, S. (1956). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill Book Co.
- SPIEGEL, M. R. (1973). *Statistics*. New York: Schaum Publishing Co.