

Pesquisa científica e nível de significância

Geraldina Porto Witter

Pontifícia Universidade Católica de Campinas

Analisa o conceito e as variáveis relevantes na escolha de nível de significância estatística para uma pesquisa; enfoca o papel do pesquisador e aspectos concernentes à análise, à leitura e ao uso dos resultados de pesquisa.
Palavras-chave: nível de significância, metodologia da ciência, análise de dados.

Abstract

Scientific research and significancy level

Analyse the concept and the relevant variables in the choice of the level of statistical significance to be used in one research, focus the researcher's role and aspects concerning the analysis, the reading and the use of the results of the research.

Key words: significancy level, science methodology, data analyses.

A produção de conhecimento científico está intimamente associada à pesquisa, a qual deve seguir princípios metodológicos mínimos para que a validade de seus resultados seja assegurada, para que haja alguma segurança no uso do saber gerado, entre outros aspectos, que o rigor e a ética da construção do saber requerem.

Na terminologia básica da pesquisa há uma expressão freqüentemente usada e que foi aqui destacada para consideração específica, trata-se do **nível de significância** (Winer, 1970).

Nível de significância está relacionado com estatística inferencial, prova de hipótese, poder do teste estatístico, com erro de pesquisa, generalização dos dados, confiabilidade dos dados, influência do acaso, controle de qualidade da pesquisa, interpretação dos dados, limites de confiabilidade, área de rejeição, entre outras variáveis. Isto é suficiente para que a questão mereça maior atenção por parte de pesquisador que o seu uso automático ou emprego apenas por ser habitual um dado nível em sua área de atuação. Também deve merecer atenção especial da parte dos que assumem a responsabilidade de formar novos pesquisadores e dos que ensinam estatística, quer na pós-graduação quer na graduação.

Quando uma expressão é muito usada e não se toma o devido cuidado na manutenção clara de seu

significado específico, corre-se o risco de um achatamento semântico e de ela passar a ser empregada sem significado específico, sem a precisão que a ciência requer e com isto perde o saber, perdem o produtor e o consumidor de ciências (Drew, 1980).

Isto parece estar ocorrendo, muitas vezes com o nível de significância, usado até mesmo sem contar com qualquer referencial estatístico ou fora do contexto semântico a que se refere. Uns exemplos talvez possam tornar mais clara a preocupação com inadequação de uso da expressão. Os exemplos mais grosseiros aqui apresentados foram colhidos de falas de alunos e mesmo de professores, em exposições de seminário, comunicações de trabalhos, defesas de dissertações e teses, congressos etc. Caberiam mesmo pesquisa específicas quanto a estes usos inadequados e suas conseqüências. Entre estes exemplos tem-se: "O texto tem um alto nível de significância", ora, são os dados da pesquisa que poderão ou não ter o alto nível referido, e não o texto. Acontece que o falante estava se referindo a um texto que nem sequer era de pesquisa, tratava-se de algumas ponderações sobre o papel da ciência e do cientista na pós-modernidade. O que pretendia dizer era que o trabalho era relevante ou significativo, não cabia "alto nível de significância".

Outros exemplos podem ser retomados: "O nível de significância de nosso instrumento ..."; "A qualidade da pesquisa garante-lhe um alto nível de

significância, n.sig. é o número que está sendo estudado”. “As duas pesquisas têm o mesmo nível de significância para o leitor” (uma era conduzida com n.sig de 0,05 e outra de 0,01) indicando confusões diversas ou usos inadequados da expressão aqui escolhida para consideração.

Outro erro comum é considerar que basta haver uma diferença grande entre dois valores, ou duas estatísticas, duas curvas, para se considerar a diferença significativa. Na realidade apenas observando a diferença entre duas medidas não é possível afirmar que as amostras que elas representam sejam significativamente diferentes. A significância depende de outros aspectos que esta diferença não alcança. Depende, entre outros aspectos, do número de unidades amostrais (sujeitos, objetos), da variabilidade dentro da amostra, ou seja, como o dado de cada sujeito se relaciona com a média (ou outra medida usada) em termos de afastamento. Vale dizer que é preciso levar em consideração a variabilidade dos dados. Por exemplo, quando se está trabalhando com médias é preciso no cálculo para teste de significância levar também em consideração o desvio-padrão, que mede esta variabilidade.

Como demonstra Jones (1995), uma diferença menor pode ser significativa enquanto outra maior pode não ser. Comparando-se duas amostras com médias respectivas de $x_1 = 12$ e $x_2 = 24$, com desvio-padrão de 2 em ambos os casos, verifica-se que a diferença de 12 é suficiente para se afirmar que as amostras são estatisticamente diferentes e que estas diferenças podem ser generalizadas para a população. Entretanto, se as medidas forem $x_1 = 40$ e $x_2 = 26$, com desvios-padrões respectivamente de 13,0 e 18,4 a diferença (14) não será significativa diante da grande variabilidade interna dos dados, não se podendo generalizar para a população.

De qualquer forma, quando se fala de diferença significativa entre amostras, entre médias, teste e reteste entre outras medidas, se está falando de diferenças que provavelmente existem não por efeito de acaso (Harris e Hodges, 1995). Mas retomando Jones (1995, p. 71) é sempre necessário lembrar, ao testar hipóteses, que a significância estatística depende “da tendência da medida central e da variabilidade dos dados que estão sendo analisados. Dados altamente

variáveis provavelmente não alcançarão diferenças significantes mesmo que as medidas de tendência central sejam muito diferentes uma da outra”.

Mas a questão da significância estatística é mais complexa e a estatística não dá respostas definitivas (sim ou não), trabalha com a probabilidade, a possibilidade (não certeza) de que uma hipótese estatística corresponde à realidade, ou seja, um certo resultado, obtido em uma pesquisa tem a probabilidade x de ser similar ao que se encontra na população.

Isto pede uma retomada do próprio conceito de nível de significância, o que pode explicar melhor os erros arrolados aqui como ilustração.

Conceito

O nível de significância (n.sig.) é uma probabilidade (α , alfa) estabelecida pelo pesquisador, a qual indica o limite aceito para flutuações ou variações dos dados que possam ser atribuídas às flutuações naturais do fenômeno estudado (Marinho, 1980). Vale dizer que é a probabilidade de ocorrência de um erro plausível, aceita como risco calculado pelo pesquisador, sendo por esta razão também denominado de erro de pesquisa, ou margem de erro da pesquisa, ou ainda nível de confiança que um resultado de pesquisa merece (Barbosa Filho, 1980; Kaplan, 1964). Conseqüentemente, tem implicações para: a análise dos dados, a interpretação, o uso dos resultados na prática profissional e em outras investigações.

Em outras palavras, trata-se da “confiança com que um pesquisador rejeita - ou aceita - uma hipótese nula” (Garret, 1958, p. 7) quando está usando estatística inferencial para teste de hipótese (Hermosa, 1982; Hays, 1973).

Também pode ser indicado pela letra p (probabilidade) seguida do número que representa o ponto de corte ou aceitação de risco, por exemplo, Odlín (1994, p. 326) assim o caracteriza: “ $p < .05$ convencional nível de significância indicando que a probabilidade (p) de se obter o resultado observado por acaso é menor do que uma em 20”. Vale lembrar que trabalha na área de aprendizagem verbal e aprendizagem de segunda língua. Portanto, uma margem de erro como a definida é aceitável, posto que não se pode ter um pleno controle das variáveis

e também porque não há grande risco para os sujeitos.

O n. sig. é uma característica que denota as limitações do estudo realizado pelo pesquisador (Baker, Pistrang e Elliott, 1994). Neste sentido também mantém relação com os chamados Erro do Tipo I e Erro do Tipo II, que podem ocorrer em ciência. Como lembra Fife-Schaw (1995), Erro do Tipo I ocorre quando se rejeita uma hipótese nula que é verdadeira e Erro do Tipo II quando não se rejeita uma hipótese que é nula e falsa. No primeiro caso pode-se levar o leitor ou outro pesquisador por vias erradas na pesquisa, ou a tomar decisões inadequadas ao usar os dados da pesquisa em sua prática profissional, por se basear em conclusões inadequadas. No segundo caso, por não se encontrar relações significantes pode-se induzir outros a não pesquisarem um fato que potencialmente pode ser importante.

Obviamente, o pesquisador precisa precaver-se e reduzir ao mínimo a probabilidade de ocorrência destes dois tipos de erro. Uma das estratégias implica uma escolha cuidadosa e criteriosa do nível de significância que estabelece para realizar sua pesquisa. O nível de significância está associado ao controle do Erro do Tipo I (Cowles e Dan's, 1995).

Escolha do Nível de Significância

A escolha do nível de significância implica a avaliação de vários aspectos da pesquisa que escapam à competência do estatístico, envolvem a qualidade dos instrumentos de medida, o delineamento da pesquisa e o conhecimento já disponível na área em que se está procedendo à investigação. Neste sentido, Gatti e Ferer (1975, p. 100) lembram que "A escolha do nível de significância não é um problema estatístico". Para definir o mesmo, o **pesquisador** deve levar em consideração os aspectos aqui referidos e o risco que corre de chegar a uma conclusão errada.

Quando se trata de alguém que está na fase de iniciação científica, ou cursando o mestrado, ou mesmo o doutorado é de se esperar que conte com o apoio de seu orientador para ponderar todas as variáveis relevantes na sua pesquisa para que possa tomar decisão quanto a este aspecto fundamental de seu trabalho - a definição do nível de significância. Com a experiência, com o enriquecimento de sua vivência de pesquisa, domínio e variação dos pro-

cedimentos de coleta e de análise irá alicerçando uma base para tomada de decisão mais autônoma. Entretanto, é importante que a opção seja consciente, ponderadas todas as variáveis; que não seja **apenas** porque tradicionalmente este ou aquele nível é aceito na área.

Mesmo o pesquisador mais experiente poderá, em dadas circunstâncias, ter dúvidas quanto a em que nível o problema ou a hipótese que está testando pode ou deve ser trabalhado. Discutir a questão com seus pares é de grande ajuda. Se a problemática envolver instrumentos de avaliação, pode ser conveniente levantar a bibliografia sobre o instrumento, discutir o seu nível de mensuração e suas características com especialistas. A dúvida pode decorrer do que já é conhecido e do que falta pesquisar em uma dada área. Que trabalhos de metaciência então disponíveis? Qual o nível de confiabilidade deles? A consulta ao "colégio invisível" é sempre útil. Também se poderá apoiar em discussões tidas com estatísticos que dominem as características da área típica de conhecimento em que se processa a pesquisa. O importante é que o nível de significância seja adequado e criteriosamente estabelecido, e não uma resposta automatizada.

Definido **quem** estabelece o nível de significância vale fazer algumas considerações sobre **quando** ele deve ser estabelecido.

Pelas dimensões conceituais já referidas, o nível de significância deve ser estabelecido antes da análise dos dados, até mesmo antes da coleta dos mesmos, para evitar que os próprios resultados influenciem a definição do n.sig. pelo pesquisador, para que esteja mais seguro de que as variáveis que poderiam influir na margem de erro foram adequadamente levadas em consideração (Ferrari, 1982).

A probabilidade de ocorrência desta influência inadequada dos resultados na definição da margem de erro é maior quando se recorre ao que Castro (1977) denominou "enlatados" ou programas de computador "sob medida" para a análise de dados. Hoje há uma grande variedade de programas, cobrindo desde a estatística descritiva à inferencial (paramétrica e não paramétrica). Como a definição do n.sig. não é um problema estatístico e sim do pesquisador, como não é possível, nem econômico,

programar para um dado tipo de frequência, para um dado problema, instrumento, área específica de conhecimento, buscou-se uma solução que viabilizasse n leituras dos resultados, diante das possíveis margens de erro que o pesquisador considerasse viável estabelecer para sua ação.

Por via de regra, os programas apresentam concomitantemente ao resultado dos cálculos qual o valor de alfa a que estão associados, isto é, se o resultado é significativo nível de 0,05; 0,01; 0,001, 0,0001 etc. Dessa forma, o programador estatístico não prejudica a ética definindo ele o nível (aliás, mesmo sem uso do computador, os que se preocupavam com este aspecto sempre tiveram este cuidado ao apresentar os dados já calculados ao pesquisador). O pesquisador é que vai ler os resultados tendo por base a margem de erro que definiu como aceitável.

A margem de erro mais comumente usada na área de ciências humanas é 0,05, embora também possam aparecer em subáreas, instrumental e metodologicamente mais sofisticadas, ou de maior risco para os sujeitos, a definição de níveis mais exigentes como 0,02 ou 0,01. Já em áreas de conhecimento em que há maior avanço na mensuração, ou em que podem ocorrer riscos para sujeitos, ou ainda grandes prejuízos materiais, a tendência é definir níveis de significância mais exigentes (0,001; 0,0001). É o que ocorre na medicina, na engenharia, na astronáutica etc.

Em sua revisão da história do uso predominante do nível de 0,05, Cowles e Davis (1995) demonstram que foi apenas aparentemente que Fisher indicou este como um nível “conveniente e preferível” (p. 286), pois, na realidade ele está sustentado na teoria da probabilidade, com destaque para o erro provável.

Mais especificamente, por trás da escolha de um nível de significância está a teoria da probabilidade da matemática formal. A escolha deve também refletir a ligação entre realidade pesquisada e matemática - estatística, que dão também sustentação à estimativa e à previsão.

De qualquer forma, o n.sig. deve ser estabelecido antes da coleta de dados. Neste sentido, é de se esperar que no projeto de uma pesquisa, no tópico relativo ao plano de análise dos dados, esteja clara-

mente especificada a margem de erro em que a pesquisa será conduzida e terá seus dados analisados e interpretados. Deve ser um nível único para a pesquisa e não flutuar ao sabor do número obtido como resultado dos cálculos. O fixar do n.sig. precede e está acima dos dados, deve nortear a análise, dar-lhe validade, e não variar como se as suas bases fossem mutáveis a cada cálculo.

Visto **quem** e **quando** se estabelece o n.sig. é momento para se especificar as variáveis que influem e devem ser consideradas ao estabelecê-lo, ou seja, os **critérios** que devem nortear esta decisão.

Critérios para escolha do Nível de Significância

Ao definir o nível de significância com que irá trabalhar seus dados o autor deve estar atento aos pontos fracos e às limitações de sua pesquisa, levando em consideração vários aspectos, ou critérios.

Desenvolvimento de conhecimento na área, subárea ou tema específico que se está investigando para saber se a pesquisa está voltada para uma temática nova, na qual há poucas pesquisas já relatadas, muitas dúvidas, é preferível assumir um nível de exigência menor, aceitar que a probabilidade de erro é maior. Assim uma margem de 0,05 será mais adequada, mais honesta para o desenvolvimento do saber responsável e de menor risco na transferência para a prática do que um nível estabelecido em 0,001 (Berkowitz, 1983).

Instrumento de medida é outra variável a considerar ao se definir a margem de erro tolerável em uma pesquisa. Os instrumentos usados pelo cientista na coleta de dados variam quanto a aspectos qualitativos e quantitativos. Vale dizer que a qualidade da medida varia com a capacidade do instrumento, com as características de mensuração que o mesmo apresenta. Quando o pesquisador trabalha com instrumentos de alta precisão e fidedignidade pode assumir que o risco de erro em seu trabalho é menor. Portanto, pode estabelecer uma probabilidade menor de erro, ou seja, nível de significância mais exigente. Por exemplo, ao usar uma para medida de peso uma balança de alta precisão, o pesquisador pode estabelecer até $p = 0,0001$ como a probabilidade de erro com que trabalha. Caso seu instrumento tenha um nível de mensuração mais

precário, por exemplo, um questionário de levantamento de opiniões, ou um teste de criatividade em leitura, é de se esperar que seja mais cauteloso, aqui $\alpha = 0,05$ será mais adequado.

Controle de Variáveis ao realizar uma pesquisa, ao definir o tipo de delineamento que será executado o pesquisador define e estabelece estratégias para o controle de variáveis com que irá trabalhar, variáveis que pretende estudar quanto às possíveis relações entre si, variáveis que controlará experimentalmente para verificar seu efeito sobre outra ou outras. A natureza, estabilidade, frequência e outras características destas variáveis também devem ser consideradas pelo pesquisador.

Mesmo tendo um bom controle sobre as variáveis, como ocorre em pesquisas experimentais realizadas em laboratórios, podem ocorrer interferências e ocorrências não previstas. Evidentemente, espera-se que o pesquisador as relate pois poderão ter afetado seus resultados. É preciso que seu relato do método descreva estas ocorrências, da melhor e mais completa forma que o possa fazer (Barkes, Pistrang e Elliot, 1994) Estas limitações não esperadas podem pedir uma revisão no nível de significância.

Mas, dependendo do tipo de pesquisa, do local onde ela será conduzida, o pesquisador, ainda que esteja realizando uma pesquisa experimental, antecipadamente pode prever que não terá controle sobre as possíveis variáveis que estarão presentes ou poderão surgir no processo de realização da pesquisa. Este é o caso, por exemplo de quando vai testar a eficiência de dois procedimentos de alfabetização em duas classes. Por mais que controle as variáveis do professor (pode até ser o mesmo para as duas classes), do material didático, das relações interpessoais em sala, dos procedimentos de ensino, da administração da sala de aula, do arranjo do ambiente etc., sempre poderão ocorrer variáveis não previstas como um surto de catapora, ou outra moléstia, um movimento de terraplanagem ao lado de uma das classes, ou outra atividade não prevista. Sabendo dessa possibilidade e que não terá controle sobre todas as variáveis que influem no processo ensino-aprendizagem, como por exemplo as variáveis da família, é uma temeridade o pesquisador pretender

que seus resultados sejam enfocados pressupondo uma margem de segurança superior a 0,05.

Uso que se pretende fazer dos dados da pesquisa é outra variável que o pesquisador deve considerar ao escolher o nível em que irá trabalhar. Se este uso posterior for ocorrer em áreas de grande risco, que possam prejudicar pessoas, evidentemente precisará ser mais rigoroso.

Este é o caso em que pode haver risco para outras pessoas quando se usa o resultado de uma pesquisa, por exemplo, quando o médico receita um remédio, quando um psicoterapeuta usa uma dada técnica ou procedimento. Certamente a amplitude que se pretende dar ao uso dos resultados e o frequentemente pouco preparo do consumidor não podem ser ignorados pelo pesquisador (Bryant e cols. 1995).

Nível de significância e interpretação da pesquisa

Como já foi dito anteriormente, o nível de significância está relacionado com a análise e interpretação dos dados da pesquisa.

Goodwin e Klausmeier (1975) incluem a questão do nível de significância como um tópico relevante da interpretação que inclui o conhecimento: das relações população => amostra => estatística => parâmetro; da Formulação e das alternativas de testes de hipóteses e a significância das diferenças no nível pré-fixado.

Interpretar os dados da pesquisa é essencial não só para o pesquisador - produtor de conhecimento, mas também para o leigo - consumidor de ciências. Evidentemente, todo pesquisador é também consumidor de ciências (Witter, 1992). Isto implica que o pesquisador o use não só para interpretar os seus dados, mas também os que obtém através da leitura de textos que relatam pesquisas. O pesquisador deve ser um consumidor sofisticado, exigente, capaz de interpretar os dados independentemente da leitura que o autor do texto tenha feito.

Goodwin e Klausmeier (1975) lembram que para formar um bom consumidor de ciência, em nível de pesquisador, é preciso todo um curso específico, além de uma ampla experiência em leitura, análise e interpretação de pesquisas.

A isto pode ser acrescida a vivência no uso de instrumentos de avaliação de pesquisas. A formação

deste consumidor de alto nível espera-se que seja estabelecida em termos básicos nos cursos de pós-graduação. Entretanto é necessário que o pesquisador-consumidor se mantenha atualizado quanto aos vários aspectos metodológicos envolvidos na interpretação.

O leigo - consumidor de ciência também precisa aprender a interpretar os resultados para poder compreender melhor o impacto da ciência no mundo moderno, aquilatar a conveniência ou não da incorporação de algum produto ou procedimento; defender-se do uso inadequado da ciência; tomar decisões quanto a como fazer e usar os dados da ciência, bem como, para seu desenvolvimento pessoal.

Esta aprendizagem deve ser preocupação do cientista. É preciso divulgar melhor as bases do saber - poder - fazer ciência; elaborar textos para publicação de divulgação das bases de segurança e das limitações do conhecimento científico.

É preciso que o leigo-consumidor saiba ler a margem de erro implícita na pesquisa quando pretende apropriar-se de dados, posto que requerem cuidados no seu uso e na sua generalização para outras realidades. Cabe aos pesquisadores cuidar desta educação científica da comunidade, e conscientizar-se de que precisam desempenhar um papel ativo na alfabetização científica da população de um modo geral.

Na alfabetização científica é fundamental cuidar da leitura da pesquisa, devendo haver particular atenção em relação às limitações do conhecimento, com a margem de erro implícita em cada trabalho. Nos últimos anos, já aparece nos jornais até mesmo em pesquisas sobre eleições o cuidado em indicar o nível de significância e mesmo em auxiliar o leitor a interpretar com mais segurança os resultados divulgados e suas limitações.

Nestas circunstâncias, cabe ao pesquisador a tarefa de educar cientificamente a comunidade, dar base para o desenvolvimento deste aspecto da cultura. Certamente ensinar os limites da interpretação, a margem de segurança do conhecimento científico é um tema com que se deve ocupar em suas relações de comunicação com o grande público.

Como bem se refere MacRae (1995 p. 47) “as limitações das inferências a serem extraídas dos testes de significância são sérias e pouca atenção tem sido dada na psicologia às principais formas de lidar com o problema”. Pode-se acrescentar que este descuido não é apenas da Psicologia, sendo comum a muitas ciências.

Além de todos os aspectos já referidos, ao se tratar da significância de um teste estatístico, o pesquisador deve levar em consideração o intervalo de confiança ou de confiabilidade e a inferência bayesiana. Esta última diz respeito à questão que realmente se está pretendendo responder, sendo raros os autores que trabalham efetivamente esta questão. MacRae (1995) localizou na literatura científica apenas um texto para psicólogos (Phillips, 1973) que trata do assunto.

O método bayesiano ou clássico corresponde, na inferência estatística convencional, à p de obtenção de resultados que assegurem que H_0 está correta (não rejeitada, isto é, provavelmente é uma das explicações corretas para o evento pesquisado). Porém implica um ajuste, introduzido no século XVIII, por Bayes, o qual introduziu a **Regra de Bayes** expressa na fórmula:

$$p(H_A / D) = p(H_A) \frac{p(D/H_A)}{p(D)}$$

A fórmula quer dizer que “a probabilidade de a H_A ser verdadeira, diante de dados observados, é encontrada multiplicando-se a probabilidade de a Hipótese A ser obtida de qualquer forma, $p(H_A)$, pela probabilidade de se obter o dado se a Hipótese A for verdadeira, $p(D/H_A)$, e dividindo-se a resposta pela probabilidade de se obter dados sem o estabelecimento de que a hipótese é verdadeira, $p(D)$ ” (MacRae, 1975, p. 48). Como foi apresentada a fórmula é de pouca utilidade, porque seu resultado somente poderia ser calculado em casos de adoção de uma hipótese em particular, quando seria útil dispor da informação para hipóteses análogas. Ao

longo do tempo isso foi se aprimorando e se chegou à fórmula:

$$\frac{p(HA/D)}{p(H_B/D)} = \frac{p(HA)}{p(H_B)} \cdot \frac{p(D/HA)}{p(D/H_B)}$$

na qual se contrasta a H_A com a H_B , mas sempre enfocando uma hipótese nula precisa, específica, para verificar se os resultados encontrados têm uma probabilidade suficientemente pequena para ser significativa.

A perspectiva oferecida pela adoção de p como marca de região de rejeição não é pontual, específica da hipótese sendo testada, mas segue aproximadamente a lógica subjacente à proposta clássica e pode-se trabalhar com uma variedade de resultados que podem cair ou não na região de rejeição. Isso amplia as possibilidades de análise, pois outras H_s são consideradas concomitantemente e não há que se esperar um único resultado específico, variações decorrentes de outras variáveis casuais são também consideradas. Entretanto, ao se interpretar os resultados isso não pode ser ignorado. Não se está trabalhando nestes casos com a exata ou precisa probabilidade de ocorrência de que o resultado seja significativo, mas sim com uma amplitude de variação em que a probabilidade de o resultado cair nela é definida pelo nível de significância em termos de acaso, mas outras hipóteses podem estar também atuando. Nestas circunstâncias a conclusão pode ser de **rejeição** ou **não rejeição**, e nunca de **aceitar** ou de considerar **verdadeira** a hipótese, como já se fez menção antes. Usar as duas últimas expressões pede que se tenha trabalhado pontualmente, usando a fórmula bayseana e não uma das muitas tabelas que aparecem nos livros e computadores e que fixam áreas de rejeição, ou melhor, o limite da área de rejeição. É preciso considerar que se está, neste último caso, vale destacar, trabalhando com “a variação dos possíveis resultados que constituem a região de rejeição e a probabilidade de cair nela somente por acaso -em qualquer ponto dela- é o nível de significância do resultado” (MacRae, 1995, p. 49), portanto não deve flutuar com o valor numérico, mas

sim, de acordo com as variáveis já enfocadas neste trabalho, deve ser pré-fixado pelo pesquisador. Se este último considera que a região deve ser de 0,05 esta sua definição deve manter-se, seja qual for o valor numérico encontrado, ainda que o mesmo possa estar tabelado em 0,01 (que estará contido no primeiro).

Outra relação freqüentemente descuidada na leitura dos dados e que tem muito em comum com o **nível de significância** é a **importância dos resultados**. Um dos primeiros autores a tratar a matéria foi Bolles (1962) quando discutiu as diferenças entre hipóteses estatísticas e hipóteses científicas. Aliás, tanto produtores como consumidores de ciência tendem a focar apenas a questão da significância, o que é uma falha de leitura dos dados que tem implicações para o uso da ciência.

Embora um resultado não seja significativo ele pode ser importante. Por exemplo quando o resultado de uma pesquisa mostra que não há diferenças significativas entre meninos e meninas em uma área do comportamento, isso pode ser importante teoricamente e para o planejamento de programas educacionais, de prevenção e todas as atividades práticas dos profissionais e pais responsáveis por crianças no que diz respeito ao aspecto focado na pesquisa. Vale dizer o que é errado porque um resultado não é significativo também não seja importante. Significância e importância são leituras distintas dos resultados.

Também um resultado pode ser significativo e não ser importante, mesmo que o assunto seja relevante. Isto porque quando a quantidade de dados é muito numerosa, quando o número de sujeitos é muito alto, pode-se demonstrar significância mesmo quando a magnitude do efeito ou da relação entre as variáveis for muito fraca. Por exemplo, se o pesquisador estiver trabalhando com $N = 100$ e correlacionando os dados de inteligência e criatividade, tendo definido como nível de significância 0,05, mesmo $r = 0,19$ será significativo (Fisher & Yates, 1971) o resultado pode ser significativo mas a relação não ser realmente importante, pois com este N o conjunto de variáveis intervenientes e as possibilidades de hipóteses explicativas do pesquisador (não estatísticas) são muitas (efeito da medida, da situação, da cultura, do sujeito).

Se em uma pesquisa são calculadas 100 correlações (itens de testes) e apenas umas poucas são significantes, isso pode ser resultado do acaso, mais do que das relações entre as variáveis em estudo. Assim, as poucas correlações significantes podem não ser importantes tanto para a teoria como para o uso posterior do conhecimento.

Neste contexto surgiu na estatística uma preocupação a que se passou chamar **magnitude do efeito**. Não está nos objetivos deste artigo tratar deste aspecto da pesquisa, ele foi aqui mencionado apenas por estar relacionado com nível de significância e para alertar o leitor quanto à confusão entre **significância** e **importância** de um resultado.

Entretanto, vale lembrar que a estatística também desenvolveu recursos para a avaliação da magnitude referida e que corresponde à proporção de pessoas, sujeitos, ou unidades de pesquisa que serão corretamente classificadas de acordo com as diretrizes tomadas como referências na pesquisa. Quando se generaliza para a população, a preocupação é com o tamanho do efeito (N de unidades atingidas) mais do que com a significância. Um dos primeiros trabalhos na área, freqüentemente referido na literatura, é o de Levy (1967), mas, hoje, há muitas formas para recalculá-la esta importância (impacto) sendo usual o uso do intervalo de confiança. São objeto de atenção na maioria dos manuais de estatística, especialmente em relação à análise de variância como a ANOVA, talvez pela facilidade de cálculo, embora possa ser calculado para qualquer prova estatística (Colman, 1995).

Conclusão

O Nível de significância é um conceito relevante no contexto da pesquisa científica, que precisa ser bem conhecido pelos pesquisadores, os quais devem defini-lo de acordo com critérios específicos e interpretar e usar os dados das pesquisas tendo por base os limites estabelecidos.

Referências

- BARBOSA FILHO, M. (1980). *Introdução à pesquisa: método, Técnica e instrumentos*. Rio de Janeiro: Livros Técnicos e científicos.
- BARKER, C; PISTRANG, N. & ELLIOTT, R. (1994). *Research methods in clinical and counselling Psychology*. New York: John Wiley & Sons.
- BERKOWITZ, C. (1983). *Advances in Experimental Social Psychology*. New York: Academic Press.
- BOLLES, R.C. (1962). The difference between statistical hypotheses and scientific hypotheses. *Psychological Reports*, 11, (4): 639-645.
- BRYANT, F.B.; EDWARDS, J.; TINDALES, R.S.; POSAVAC, E.J.; HEATH, L.; HENDERSON, E. & SUAREZ-BALCAZAR, Y. *Methodological issues in applied Social Psychology*. New York: Plenum Press.
- CASTRO, C. de M. (1977). *A prática da pesquisa*. São Paulo: Mc Graw-Hill.
- COLMAN, A.M. (1995). *Psychological research methods and statistics*. London: Longman.
- COWLES, M. & DAVIS, C. (1995). On the origins of the 05 level of statistical significance. In A.E. KAZDIN. *Methodological Issues & Strategies in Clinical Research*. Washington (D.C.): American Psychological Association.
- DREW, C.J. (1980). *Introduction to designing and conducting research*. St. Louis: CV Mosby.
- FERRARI, A.T. (1982). *Metodologia de Pesquisa Científica*. São Paulo: McGraw-Hill.
- FIFE-SCHAN, C. (1995). Bivariate Statistical Analyses. In G.M, BROAKWELL; S. HAMMOND & C. FIVE-SCHAN. *Research Methods in Psychology*. London: Sage.
- FISHER R.A. & YATES, C.B.E. (1971). *Tabelas Estatísticas para pesquisa em biologia medicina e agricultura*. Tradução do original inglês s/d por S.L. Haim. São Paulo: Pligono.
- GARRET, H. (1962). *Estatística na Psicologia e na Educação*. Tradução do original inglês de 1958 por R. Rocha. Rio de Janeiro: Fundo de Cultura.
- GATTI, B.A. & FERER, N.L. (1975). *Estatística básica para Ciências Humanas*. São Paulo: Alfa-Omega.
- GOODWIN, W.L. & KLAUSMEIER, H.J. (1975). *Facilitating student learning: an introduction to educational psychology*. New York: Haper & Row.
- HAYS, W.L. (1973). *Statistics: For social sciences*. New York: Holt, Rine hart and Winston.

- HARRIS, T.L. & HODGERS, R.E. *The literacy dictionary: the vocabulary of reading and writing*. Newark (Del). IRA.
- HERMOSA, L.P. (1982). *Investigation en Ciências Humanas y Educación*. Lima: EIPAC.
- JONES, J. L. (1995). *Understanding Psychological Science*. New York: Harper.
- KAPLAN, A. (1975). *A conduta na pesquisa: metodologia para as ciências do comportamento*. Tradução de original inglês de 1964 por L. HEGENBERG e O.S. da META. São Paulo: EPU.
- LEVY, P.M. (1967). Substantive significance of significant differences between two groups. *Psychological Bulletin*. 67 (1): 37-40.
- MACRAE, A.W. (1995). Descriptive and Inferential statistics. In A.M. COLMAN. *Psychological research methods and statistic*. London: Longman.
- MARINHO, P. (1980). *A pesquisa em Ciências Humanas*. Petrópolis: Vozes.
- ODLIN, T. (1994). (Org.) *Perspectives on Pedagogica and Grammar*. Cambridge: Cambridge University Press.
- WINER, B.J. (1970). *Statistical principles in experimental design*. New York: Mc Graw-Hill.
- WITTER, G.P. (1992). Produção e leitura de texto científico. *Estudos de Psicologia*, 9 (1): 19-26.