

Recebido em: 28/01/11

Aprovado em: 11/04/11

# Abordagem metodológica na análise de dados de estudos não-paramétricos, com base em respostas em escalas ordinais

Henrique Martins Rocha (UNESP – SP/Brasil) - prof.henrique\_rocha@yahoo.com.br

• Av. Dr. Ariberto Pereira da Cunha, 333, 12516-410, Guaratinguetá-SP, fone: (55)12-3123-2855

Maurício César Delamaro (UNESP – SP/Brasil) - delamaro@feg.unesp.br

## Resumo

A análise de dados não-paramétricos constitui-se um desafio para a obtenção de conclusões em pesquisas acadêmicas e científicas, pelo fato de trabalhar com dados de difícil comparação e por serem desconhecidos os parâmetros da população. O presente artigo apresenta um método de análise de dados, que foi utilizado em pesquisas sobre a percepção de gerentes de produto, sobre o nível de maturidade do processo de desenvolvimento de produtos e dos fatores críticos de sucesso em tais processos. Os dados, coletados em questionários, foram analisados, utilizando o teste da mediana e, em seguida, o teste de sinais, possível mediante a concatenação de diversos conceitos estatísticos. A combinação de tais técnicas permitiu tornar mais robustas as análises dos dados coletados, podendo ser reaplicada em outras pesquisas que precisem lidar com escalas ordinais.

Palavras-chave: escala ordinal; análise de dados; nível de confiança; testes não-paramétricos.

## Abstract

*Analysis of non-parametric data on scientific and academic research constitutes a challenge due to the fact that it deals with difficulties in comparing data and because the population parameters are unknown. This paper presents a method of data analysis which has been used for the perceptions of product manager's critical success factors in product development and the maturity level in such processes. The data collected from questionnaires has been analyzed by median test and sign test, which has become possible through the concatenation of several statistical concepts. The combination of these techniques enabled a more robust data analysis which can be replicated in other studies that have to deal with ordinal scale data.*

*Keywords: ordinal scale; data analysis; confidence level; non-parametric tests.*

## 1. INTRODUÇÃO

A análise de dados não-paramétricos, ou seja, “dados derivados de mensurações nominais e ordinais” (COOPER e SCHINDLER, 2003, p.617), constitui-se um desafio para a obtenção de conclusões robustas, em pesquisas acadêmicas e científicas, não sensíveis a julgamentos arbitrários e tendências individuais (AN e CUOGHI, 2004), pelo fato de trabalhar com dados de difícil comparação e por serem desconhecidos os parâmetros da população (MATTAR, 1998), mas, principalmente, por serem menos poderosos que os testes paramétricos, visto que estes últimos lidam com “mensurações de intervalo e de razão” (COOPER e SCHINDLER, 2003, p.397). Como citado por Montgomery e Runger (2009, p.381), “um procedimento não-paramétrico será menos eficiente que o procedimento paramétrico correspondente”.

Tal preocupação se justifica nos pensamentos de Vergara (2005, p.10), de que o pesquisador escolherá “aquele método que é mais adequado à(s) teoria(s) que suporta(m) seu estudo, ao problema que suscitou sua investigação e o fará dentro de pressupostos epistemológicos”; e também, nos pensamentos de Gil (2009), que cita a preocupação dos leitores quanto à credibilidade dos resultados, principalmente quando estes têm interesse na aplicação prática deles.

A motivação para este trabalho surgiu da necessidade de padronizar um teste estatístico de fácil aplicação, que permitisse avaliar o nível de confiança/significância dos achados em pesquisas que lidam com escalas ordinais. Como citado por Cooper e Schindler (2003), como quase toda amostra difere da população que ela representa, há a necessidade de avaliar se tais diferenças são estatisticamente significativas.

O presente artigo apresenta um método de análise de dados, que foi utilizado em duas pesquisas sobre a percepção de gerentes de produto, sobre o nível de maturidade do processo de desenvolvimento de produtos e sobre os fatores críticos de sucesso para tal desenvolvimento. Os dados, coletados em questionários, foram analisados, utilizando-se o teste da mediana e, obtidas conclusões, buscou-se validar o grau de confiança (significância) dos resultados, por meio da combinação de uma série de métodos e conceitos estatísticos. A combinação de tais técnicas permitiu tornar mais robustas as análises dos dados coletados, podendo ser replicada em outras pesquisas, que precisem lidar com escalas ordinais.

## 2. REVISÃO BIBLIOGRÁFICA

De acordo com Mattar (1998, p.65): “uma técnica estatística é chamada paramétrica, quando o modelo do teste especifica certas condições sobre parâmetros da população, da qual a amostra foi obtida, para que possa ser utilizada”. As técnicas paramétricas são escolhidas para teste, se as observações forem independentes, retiradas de populações normalmente distribuídas, com variâncias iguais (ou com relação entre as variâncias conhecidas) e com as escalas de mensuração que permitam serem efetuadas operações aritméticas com elas (COOPER e SCHINDLER, 2003; MATTAR, 1998). Quando este conjunto de exigências não é atendido, principalmente por não se saber o tipo de distribuição da população nem sua variância, sequer o fato de ser contínua, são utilizadas as técnicas não paramétricas (MATTAR, 1998; MONTGOMERY e RUNGER, 2009), as quais, para De Souza (2003, p.vii.) “não pro-

põe uma função com os parâmetros que são estimados”. Para Wilks (2006 *apud* BLAIN, 2010, p.46), “os testes não paramétricos são aqueles elaborados de forma tal que a distribuição de que os dados sob análise são oriundos, torna-se irrelevante”.

Mattar (1998, p.82) defende que “os métodos paramétricos são poderosos apenas quando aplicados a amostras grandes, o que não ocorre com os métodos não paramétricos”. No entanto, An e Coughi (2004, p.100) apontaram a preocupação no mau uso dos métodos estatísticos: “muitos artigos supõem que os dados sejam normais sem, entretanto, verificá-los, o que pode comprometer a veracidade dos resultados, principalmente quando a amostra for pequena”.

De acordo com Cooper e Schindler (2003, p.182), os testes não paramétricos “são abundantes, simples de calcular, têm boa eficiência e não forçam o pesquisador a aceitar suposições de testes paramétricos”. Complementando tal raciocínio, Montgomery e Runger (2009, p.381) citam que:

Em geral, procedimentos não-paramétricos não utilizam toda a informação fornecida pela amostra [...]. Essa perda de eficiência é refletida por uma necessidade de um tamanho maior de amostra para o procedimento não-paramétrico do que o requerido pelo procedimento paramétrico, de modo a encontrar a mesma potência [...]. Quando as distribuições em estudo não forem aproximadamente normais, os métodos não-paramétricos terão muito a oferecer. Eles constantemente fornecem melhoria considerável sobre os métodos paramétricos, baseados na teoria da normalidade.

Outro aspecto relevante, no que diz respeito às técnicas paramétricas ou não paramétricas, está relacionado às escalas de coletas de dados, que podem ser classificadas como: nominal (somente para identificação, como, por exemplo, numeração em camisas de jogadores de futebol), ordinal (ordenamento de preferência, como, por exemplo, preferência entre marcas), intervalo (os intervalos entre os números permite comparar diferenças, como, por exemplo, escala de temperatura) e razão (semelhante à escala de intervalo, mas contendo o “zero absoluto”, como, por exemplo, peso de produtos) (MATTAR, 1999). Se dados são coletados em escalas discretas ou ordinais, não é correto que sejam executadas operações aritméticas com elas, impedindo o uso de técnicas paramétricas.

Fazendo uso das diferenças e limitações entre testes paramétricos e não paramétricos, Silva, Santos e Ferreira (2007), em seu estudo sobre a percepção de docentes quanto às práticas de contabilidade estratégica, utilizaram o teste de Shapiro-Willk, para verificar a aderência dos dados à distribuição normal e suportar a decisão pela utilização de técnicas paramétricas ou não-paramétricas para análises das diferenças de médias ou medianas. Já na pesquisa de Ferreira *et al.* (2008), como uma das amostras não atendia aos pressupostos paramétricos, recorreram ao teste de Mann-Whitney, para analisar aspectos vitamínicos em uma população.

Rocha (2005) e Quintella e Rocha (2006), em sua pesquisa sobre o grau de maturidade no processo de desenvolvimento de produtos (PDP), em duas empresas da área automotiva, coletaram, por meio de questionários direcionados, dados, junto a 47 profissionais que atuavam diretamente no lançamento de novos produtos: gerentes, supervisores e corpo técnico de *Marketing*, Produto, Logística, Suprimentos e Manufatura das montadoras estudadas.

Para as 27 perguntas do instrumento da citada pesquisa, foi utilizada a escala somatória ou escala *Likert*, que é uma escala indireta, ao combinar um conjunto de respostas dos entrevistados, em relação ao objeto em questão, para determinar qual a sua posição na escala de atitude desenvolvida: proposta por Rensis Likert, em 1932, compreende uma série de afirmações relacionadas com o objeto pesquisado. Os respondentes são solicitados, não só a concordarem ou discordarem das afirmações, mas também, a informarem qual seu grau de concordância/discordância quanto a afirmações que auxiliavam a identificar o nível de maturidade do processo. A cada célula de resposta foi atribuído um número que refletia a direção da atitude do respondente, em relação a cada afirmação. Na citada pesquisa, pelo uso de escala ordinal e pelo fato de não haver conhecimento sobre os parâmetros da população, a escolha recaiu sobre os métodos não paramétricos.

Também, nas mesmas empresas, Rocha (2009) e Rocha *et al.* (2010) desenvolveram estudos, buscando identificar e validar os fatores críticos de sucesso (FCS) no processo de desenvolvimento de produtos. O universo da pesquisa, limitado aos gerentes diretamente relacionados ao desenvolvimento de produtos nas montadoras citadas (19, sendo 11 na empresa de veículos de passeio e 8 na empresa de caminhões e ônibus), garantiu a qualificação e significância na amostra, pois como só estes poucos gerentes definiam a operacionalização da estratégia de desenvolvimento dos novos produtos, cobriu-se assim, apenas o reduzido, mas importante, conjunto-chave de envolvidos nos PDPs.

Em ambas as pesquisas, as quais servirão de base para explanação sobre o método de análise proposto no presente artigo, as conclusões foram obtidas por meio da análise dos dados, que representavam as percepções dos envolvidos nas atividades estudadas, coletadas em questionários autopreenchidos, com os quais se testava o grau de concordância do respondente (p/ex: “O gerenciamento de custos e orçamentos é um FCS para os PDPs”), devendo o mesmo escolher “discordo totalmente”, “discordo parcialmente”, “não concordo nem discordo”, “concordo parcialmente” ou “concordo totalmente”, numa escala de 1 a 5. As respostas compiladas eram então, tratadas estatisticamente, calculando-se a mediana das respostas: como a escala *Likert* utilizada previa dois níveis de rejeição (1 e 2), um nível neutro (3) e dois níveis de concordância (4 e 5), estabeleceu-se que medianas inferiores a 3 caracterizariam a rejeição das afirmativas propostas nos questionários da pesquisa, referente à identificação e validação dos FCS. No que tange à avaliação do nível de maturidade dos processos, por lidar com critérios de certificação pelo CMMI (*Capability Maturity Model Integration*), os quais exigiam o atendimento pleno das metas das áreas de processo, foi entendido que o *escore* “3” (“não concordo nem discordo”) não representaria fielmente tal estado, optando o pesquisador pelo *escore* quatro, como o mínimo aceitável (ROCHA, 2005).

Ao contrário dos métodos paramétricos, no entanto, que podem facilmente indicar um grau de confiança (ou nível de significância), a utilização de métodos não paramétricos apresenta uma lacuna neste aspecto. Por exemplo, se uma das afirmativas fosse rejeitada por ter uma mediana inferior a “3”, ou aceita pelo fato de ter uma mediana maior ou igual a este valor, qual o grau de confiança de que a rejeição ou aceitação fosse acertada?

O presente artigo busca esclarecer esta dúvida, apresentando um método para estabelecer o grau de confiança nos achados em estudos não-paramétricos, com base nas respostas em escalas ordinais.

### 3. MATERIAL E MÉTODO DA PESQUISA

De acordo com Norrie (2006), a pesquisa deve buscar responder questões, em quatro áreas, sobre a solução proposta para o problema: relevância (a solução é apropriada para a situação-problema? o problema é de natureza universal ou seletiva? há necessidade de novas políticas, procedimentos ou ferramentas para suportar a solução? o problema é uniforme ao longo de toda a organização?); legitimidade (a solução é legal e ética? a solução tem consequências organizacionais ou sociais não intencionais? quais são as consequências da implementação da solução na organização?); efetividade (a solução é correta? a solução pode ser implementada sem maiores rupturas?); eficiência (a solução usa eficientemente os recursos disponíveis? ela requer recursos, competências ou habilidades novas ou únicas para implementação? existem soluções mais eficientes?).

O objetivo do artigo é, desta forma, apresentar um método para estabelecer o grau de confiança nos achados em estudos não-paramétricos, que utilizam escalas ordinais para coleta de dados, de tal forma que o método possa ser replicado em pesquisas em áreas diversas, que precisem lidar com respostas com este tipo de escala.

A coleta de dados nesta pesquisa foi realizada por meio da aplicação de questionários autoperenchidos, ou seja, questionários lidos e preenchidos diretamente pelos pesquisados, em duas montadoras instaladas na região Sul Fluminense, as quais serão brevemente descritas a seguir. A opção pelo questionário autoperenchido, em vez de entrevistas, se deve principalmente, ao tempo de aplicação e custo envolvido, que no caso de entrevistas seriam muito altos.

A fábrica do segmento de automóveis, localizada no município de Porto Real/RJ, pertencente a grupo automobilístico de origem francesa, com operações em 150 países e 207.800 colaboradores, foi inaugurada em fevereiro de 2001. Produzindo em 3 turnos, conta com quatro unidades de produção: chaparia, pintura, montagem e motores (esta última inaugurada em março 2002), numa área construída de 175.120m<sup>2</sup>, em terreno de 2 milhões de m<sup>2</sup> e população aproximada de 3.900 pessoas.

A fábrica de caminhões e ônibus iniciou suas atividades em 1995, em unidade provisória, passando à unidade definitiva em 1996, numa área de 1 milhão de m<sup>2</sup>, sendo 135.000m<sup>2</sup> construídos. Com uma população de 4.728 pessoas e capacidade de produção de 300 unidades por dia, produz atualmente diferentes modelos para o mercado local e exportação. Por conta do alto nível de customização exigido neste mercado, a empresa mantém parceria com Empresa de menor porte, dedicada ao desenvolvimento de protótipos e modelos especiais de veículos.

Na compilação de cada uma das respostas, analisando-se o grau de concordância dos respondentes, a pontuação total da atitude foi dada pelo somatório das pontuações obtidas de cada respondente. Conforme Cooper e Schindler (2003, p.181), “como o número de tais escalas tem apenas um significado de classificação, a mensuração apropriada da tendência central é a mediana”, enquanto “uma medida percentual ou quartil revela a dispersão”. Este aspecto é corroborado por diversos outros autores (HASSEGAWA, 2002; HOEL, 1984; LEVIN, 1987; MATTAR, 1998; SIEGEL, 1981; SIEGEL e CASTELLAN JR., 2006).

Como citado anteriormente, medianas menores que “3” significariam a rejeição da afirmativa, por parte dos respondentes, na pesquisa sobre FCS, enquanto para valores superiores a este, não se caracterizaria a rejeição, o que não significa, necessariamente, que estivesse confirmada: de acordo com Marconi e Lakatos (2004), uma hipótese está corroborada enquanto não excede o nível de provisoriedade, ou seja, não surgir fato que a invalide. Para tanto, foi utilizado o método hipotético-dedutivo, no qual soluções são propostas na forma de proposições, as quais são submetidas a testes de falseamento. “Se a hipótese não supera os testes, estará falseada, refutada, e exige nova reformulação do problema e da hipótese, que, se superar os testes rigorosos, estará corroborada, confirmada provisoriamente” (MARCONI e LAKATOS, 2004, p.74).

Marconi e Lakatos (2004, p.77) apontam que “quanto mais falseável for uma conjectura, mais científica será”. Mattar (1998) prevê que isso seja feito, estabelecendo-se primeiramente, a hipótese nula ( $H_0$ ), que é a hipótese de negação (ou oposto) da ocorrência da hipótese em estudo. Em seguida, define-se a região de rejeição de  $H_0$  e, a partir daí, decide-se sobre a rejeição ou não rejeição da hipótese nula, com base na comparação do valor calculado com a região de rejeição: a rejeição de  $H_0$  significa que a hipótese (ou afirmativa) foi corroborada, enquanto a sua não rejeição, indica que a hipótese não pode ser aceita.

Para avaliar a confiança na mediana encontrada (ou seja, a confiança de que a mesma não é inferior a 3, a determinado grau de significância), utilizou-se o teste de sinais. Para que a utilização de tal teste fosse entendida como aceitável, fez-se necessária a concatenação de diferentes conceitos estatísticos, de diferentes autores, como se segue: conforme Triola (1999, p.498), “podemos aplicar o teste de sinais a uma afirmação sobre a mediana de uma única população”, enquanto Hoel (1984) cita que é possível testar valores de mediana, por meio do teste do sinal, restringindo, no entanto, a variáveis contínuas. Como Siegel (1981), admite que numa escala ordinal, com suas categorias discretas, há um *continuum* de resultados possíveis – ainda que, por conta da “falta de precisão do aparelhamento de mensuração”, não se distingua as pequenas diferenças que realmente existem entre as observações -, utilizou-se tal teste na pesquisa.

Conforme Triola (1999), o processo a ser seguido, é composto dos seguintes passos:

1. Atribuir sinais positivos e negativos, tendo como base a diferença destes para o valor de referência – na presente pesquisa, trata-se do valor três, estabelecido como pontuação mínima, para caracterizar a aceitação da afirmativa;
2. Descarte dos valores zero;
3. Compilação do número de vezes que o sinal menos frequente ocorre – identificado como “x” – e do número total combinado de sinais positivos e negativos – identificado como “n”;
4. Comparação de “x” com a expectativa de ocorrências do valor de referência – no caso da presente pesquisa, o “x” deveria ser indicativo das respostas com pontuação menor do que três;



5. Caso “x” não corresponda às expectativas, o teste se encerra, caracterizando a ocorrência da situação diversa do que era esperado ( $H_0$ ) – o que seria, no caso da presente pesquisa, encontrar uma mediana menor do que três. Conforme Triola (1999, p.495), “não há como apoiar uma afirmativa de  $p > 0,5$  com qualquer proporção amostral menor do que 0,5”;
6. Se a expectativa se confirmar, há de avaliar se o resultado é estatisticamente representativo, dentro de determinado grau de significância: para amostras de até 25 elementos, verifica-se o valor crítico tabelado para aceitação/corroboração, conforme a tabela 1;

Tabela 1 – Valores críticos para o teste de sinais.

n	$\alpha$ (bicaudal)			
	0,01	0,02	0,05	0,10
5	-	-	-	0
6	-	-	0	0
7	-	0	0	0
8	0	0	0	1
9	0	0	1	1
10	0	0	1	1
11	0	1	1	2
12	1	1	2	2
13	1	1	2	3
14	1	2	2	3
15	2	2	3	3
16	2	2	3	4
17	2	3	4	4
18	3	3	4	5
19	3	4	4	5
20	3	4	5	5
21	4	4	5	6
22	4	5	5	6
23	4	5	6	7
24	5	5	6	7
25	5	6	7	7

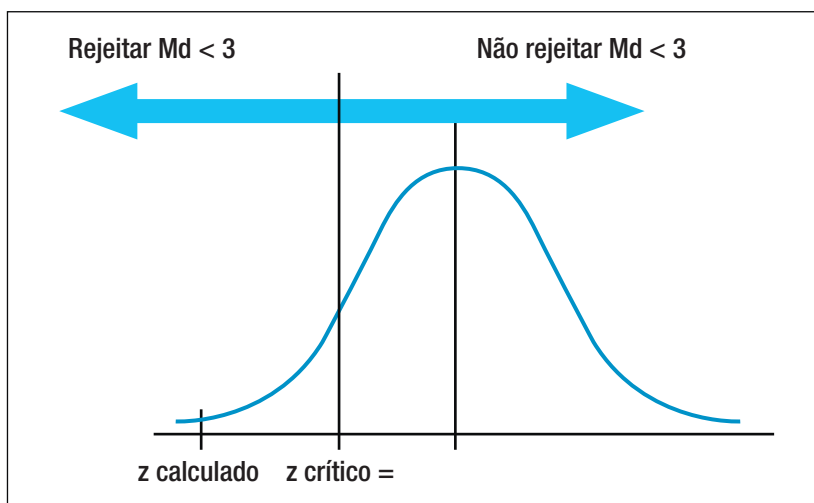
Fonte: adaptado de Triola (1999).

7. Para valores maiores do que 25, utilizar a estatística “z”: conforme Siegel (1981), a distribuição amostral – para prova de sinais – tem distribuição aproximadamente normal para  $N > 25$ , permitindo o uso da estatística “z”, com correção de continuidades, conforme a fórmula apresentada a seguir:

$$z = \frac{(x + 0.5) - (n / 2)}{n^{1/2} / 2} \quad (1)$$

Estabelecido o nível de significância (por exemplo, 5%), é calculado o valor de “z” e estabelece-se o valor do afastamento correspondente à probabilidade. A comparação entre os valores, com base na evidência amostral e ao nível de significância escolhido, leva à rejeição ou não de  $H_0$ , como pode ser observado na figura 1.

Figura 1 – Teste da mediana.



Fonte: adaptado de Triola (1999).

## 4. ANÁLISE DE DADOS E RESULTADOS

Por limitações de espaço e por não ser o escopo do presente artigo, o aprofundamento na comparação entre as empresas e na abrangência de quais afirmativas referentes aos FCS ou níveis de maturidade foram validadas, serão apresentadas tão somente as análises e as etapas referentes à análise dos dados para compreensão e acompanhamento da aplicação, passo a passo, do método desenvolvido.

O resultado da tabulação dos dados da questão, que tinha como objetivo verificar o grau de concordância, por parte dos respondentes na pesquisa sobre FCS, com determinada afirmativa, é apresentado na tabela 2. Os números mostrados representam as respostas, com base na escala *Likert* utilizada (1 a 5), enquanto os FCS foram apresentados de forma simplificada (FCS1, FCS2, etc.).



Tabela 2 – Respostas da questão (empresa de veículos de passeio).

	RESPONDENTES											Mediana
	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	
FCS1	4	3	3	5	4	3	3	3	2	3	3	3
FCS2	3	5	3	5	4	4	5	3	5	4	5	4
FCS3	3	4	4	3	3	4	5	4	3	5	5	4
FCS4	3	5	5	4	4	5	5	5	3	4	5	5
FCS5	5	3	4	3	4	5	5	4	5	5	4	4
FCS6	2	4	5	5	4	2	4	3	4	3	4	4
FCS7	3	5	4	3	3	3	3	4	4	4	3	3
FCS8	4	1	5	2	4	2	4	4	5	3	3	4

Fonte: adaptado de Rocha (2009).

Os resultados obtidos na questão, que apura o grau de concordância dos respondentes com os FCS apresentados, após sua tabulação, tiveram sua mediana calculada. Como pode ser observado na tabela 2, nenhum dos FCS teve uma mediana inferior a “3”. Tais resultados foram então submetidos ao teste de sinais, de forma a se avaliar o grau de confiança na mediana encontrada, conforme processo definido anteriormente.

Os valores maiores e menores que “3”, bem como os valores  $x$  e  $n$  e os valores críticos correspondentes (com base na tabela 1), são vistos na tabela 3 (os diferentes valores de  $\alpha$  mostrados nas tabelas indicam os níveis de significância mais severos, em que era possível rejeitar  $H_0$ , ou hipótese nula).

Tabela 3 – Teste de sinais (empresa de veículos de passeio).

	Respostas		Comparação com expectativa	x	n	Valor crítico
	>3	<3				
FCS1	3	1	Ok	1	4	-
FCS2	8	0	Ok	0	8	0 ( $\alpha = 1\%$ )
FCS3	7	0	Ok	0	7	0 ( $\alpha = 1\%$ )
FCS4	9	0	Ok	0	9	0 ( $\alpha = 1\%$ )
FCS5	9	0	Ok	0	9	0 ( $\alpha = 1\%$ )
FCS6	7	2	Ok	2	9	1 ( $\alpha = 10\%$ )
FCS7	5	0	Ok	0	5	0 ( $\alpha = 10\%$ )
FCS8	6	3	Ok	3	9	1 ( $\alpha = 10\%$ )

Fonte: adaptado de Rocha (2009).

Como se testa a situação da mediana ser maior ou igual a três, tem-se como  $H_0$  a mediana ser inferior a esse valor. Pelo teste de sinais, se o “x” for menor ou igual ao valor crítico, rejeita-se  $H_0$  (ou seja, é aceita – corrobora-se – a afirmação/conjectura). Observa-se na tabela 3, que não foi possível fazer qualquer inferência quanto ao FCS1, visto não ser possível obter uma região crítica (TRIOLA, 1999), com base nos valores da tabela 1, a qual é repetida a seguir, reordenada e apresentada de forma resumida (somente com os campos de interesse para análise da tabela 3). Da mesma forma, no FCS7, foi possível rejeitar  $H_0$ , mas somente a um grau de significância de 10% (não foi possível rejeitar com  $\alpha = 1\%$ ).

Tabela 4 (tabela 1 resumida e adaptada) – Valores críticos para o teste de sinais.

n	$\alpha$ (bicaudal)				FCS
	0,01	0,02	0,05	0,10	
5	-	-	-	0	7
7	-	0	0	0	3
8	0	0	0	1	2
9	0	0	1	1	4, 5, 6, 8

Fonte: adaptado de Triola (1999).

Observa-se também, que os FCS6 e FCS8, apesar de terem mediana de “4”, não foram validados ( $x = 2$  e  $3$ , respectivamente), pois não foi possível rejeitar a  $H_0$  deles, mesmo a um grau de significância de 10%.

Faz-se ainda necessária uma verificação adicional sobre os demais FCS (ou seja, 2, 3, 4 e 5), antes que possam ser efetivamente aceitos: se realmente as medianas deles não são inferiores a três, elas não podem ser menor ou igual a 2,5. A análise desta nova situação pode ser vista na tabela 5 (somente para os FCS não rejeitados anteriormente).

Tabela 5 – Teste de sinais (empresa de veículos de passeio).

	Respostas		Comparação com expectativa	x	n	Valor crítico
	>2,5	<2,5				
FCS2	11	0	Ok	0	11	0 ( $\alpha = 1\%$ )
FCS3	11	0	Ok	0	11	0 ( $\alpha = 1\%$ )
FCS4	11	0	Ok	0	11	0 ( $\alpha = 1\%$ )
FCS5	11	0	Ok	0	11	0 ( $\alpha = 1\%$ )

Fonte: elaboração dos autores

Percebe-se, nesta análise, que houve corroboração dos achados, no que tange aos FCS 2, 3, 4 e 5. Na pesquisa referente ao nível de maturidade do processo de desenvolvimento, com o nível de maturidade indicado, como “2” (ML2) nas duas organizações estudadas, novamente foi verificado o nível de confiança sobre tal resultado. Avaliando a tendência central e a dispersão com a função “QUARTIL” do MS Excel, com base nas pontuações compiladas dos respondentes, foram obtidos os valores apresentados na tabela 6.

Tabela 6 – Cálculo de tendência central e dispersão no nível de maturidade.

Organização (tipo de produto)	Veículos de passeio	Caminhões e ônibus
Mediana	4	4
Desvio médio	0,665	0,643
Intervalo interquartil	1	1

Fonte: adaptado de Rocha (2005).

Define-se intervalo interquartil, como a diferença entre o primeiro e o terceiro quartil de distribuição, sendo uma medida utilizada em conjunto com a mediana (COOPER e SCHINDLER, 2003). O desvio médio, calculado para ambas as organizações, é definido, segundo Levin (1987), pela seguinte fórmula.

$$DM = \frac{\sum |X|}{N} \quad (2)$$

sendo  $\sum |X|$  a soma dos valores absolutos das discrepâncias, enquanto N é o número de dados.

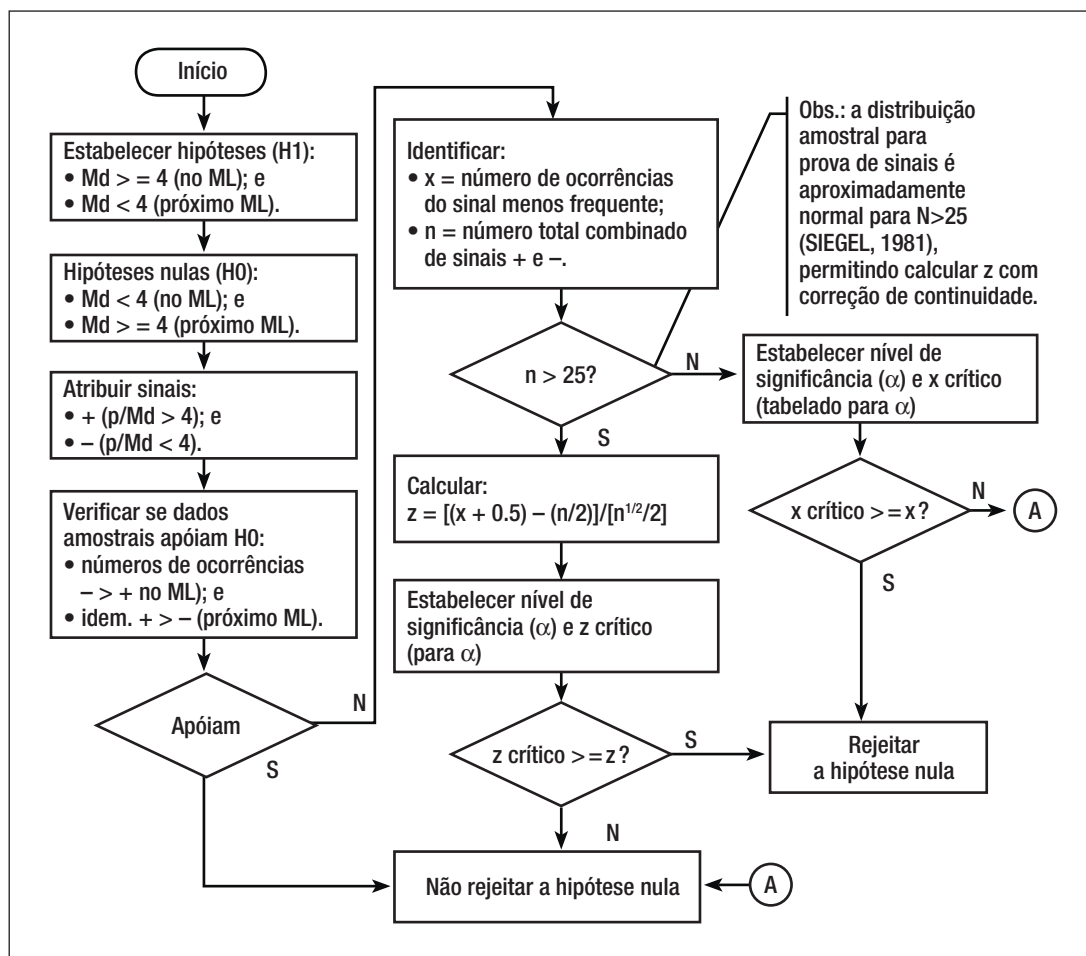
Os valores calculados podem ser utilizados para avaliar a variabilidade – ou a sensibilidade – do valor de tendência, central, com relação à dispersão, utilizando-se o coeficiente de variação relativa, o qual, conforme Toledo e Ovalle (1985), pode ser calculado com base no desvio médio ou desvio padrão e a média ou mediana. Ao lidar-se com a mediana na pesquisa citada, optou-se por utilizar a fórmula a seguir:

$$CV = \frac{DM}{\text{mediana}} \times 100 \quad (3)$$

Encontra-se neste caso, o valor de 16,62% para a empresa de veículos de passeio, enquanto, para a de caminhões e ônibus, é encontrado o valor de 16,08%, como indicador do desvio médio por unidade da mediana, que é um indicador de risco análogo ao coeficiente da variação, usualmente calculado para variáveis contínuas – neste caso, utilizando-se o desvio padrão e a média.

Para se avaliar a confiança no nível de maturidade identificado, seguiram-se os passos indicados anteriormente, para avaliar o grau de confiança no atingimento da pontuação para o enquadramento no nível específico (quatro). No entanto, com uma amostra superior a 25, é possível, agora, demonstrar a utilização do método, com o uso da fórmula (1), no lugar da tabela de valores críticos (tabela 1). A figura 2 apresenta o fluxo do teste, mostrando as duas opções possíveis, com base no tamanho da amostra.

Figura 2 – Fluxograma para teste de sinais.



Fonte: Quintella e Rocha (2007), adaptado de Triola (1999).

O levantamento das respostas às questões referentes ao nível de maturidade 2, apresentou os resultados mostrados na tabela 7.

Tabela 7 – Compilação das pontuações referentes às perguntas do ML 2.

Segmentação das pontuações das questões	Empresa	
	Veículos de passeio	Caminhões e ônibus
= 4	63	76
< 4	36	29
> 4	62	63
n	98	92
x	36	29

Fonte: elaboração dos autores

Como o número de vezes que o sinal menos frequente ocorre (x), corresponde à expectativa que se tem – da mediana ser inferior a quatro ( $62 > 36$ ; e  $63 > 29$ ), dá-se prosseguimento ao teste: o cálculo de “z”, com base nos valores de “x” e “n” para as organizações, resultou, para a empresa de veículos de passeio, em  $-2,53$  e, para a de caminhões e ônibus,  $-3,44$ . Utilizando-se a função “INV.NORMP” do MS Excel, é possível calcular o valor do afastamento correspondente à probabilidade de 1% (0,01) em uma distribuição normal com média zero e desvio padrão 1, para o qual é encontrado o valor  $-2,326$ . Desta forma, conclui-se que, sendo o “z” de ambas as empresas mais afastado do valor central que o z crítico, com base na evidência amostral, a um nível de significância de  $\alpha=0,01$ , ambas as organizações têm uma mediana igual ou maior que quatro, nas perguntas referentes ao nível de maturidade 2.

O teste exposto pode ser aplicado individualmente a cada uma das perguntas utilizadas para identificação do nível de maturidade, conforme estudo desenvolvido por Quintella e Rocha (2007), justificado pelo fato de que a determinação do nível de maturidade passava pelo atingimento ou satisfação das metas das diversas áreas de processo: se ao menos uma das metas fosse avaliada como “não-satisfeita”, a área de processo como um todo seria avaliada como “não-satisfeita”.

## 5. CONSIDERAÇÕES FINAIS

Testes estatísticos são fundamentais para comparar condições e obter conclusões, a partir de dados coletados em pesquisas. As análises de dados em escalas ordinais, de uma forma geral, exigem tratamentos diferentes dos dados intervalares: como não há uma razão definida entre estes, os métodos matemáticos ficam limitados. Por exemplo, uma resposta “2”, não necessariamente significará o dobro de uma resposta “1”. Ou ainda, como citado por Cooper e Schindler (2003, p.181), “a diferença real entre a categoria 1 e 2 em uma escala de satisfação pode ser mais ou menos do que a diferença entre a categoria 2 e 3”. Desta forma, elementos, como média e desvio padrão não devem ser aplicados, visto que qualquer função que considere operações aritméticas com os dados, perderia o sentido.

Assim sendo e, por conta do usual menor poder nos testes não paramétricos, (quando comparados com testes paramétricos equivalentes), a significância estatística dos resultados é um fator de preocupação dos pesquisadores, que buscam identificar aspectos quanto à credibilidade das conclusões das análises de dados, por métodos não-paramétricos.

O presente artigo expôs o uso de um método que permitiu avaliar o nível de significância dos resultados, em duas pesquisas, que lidaram com percepções dos agentes envolvidos em processos de desenvolvimento de produtos, podendo, no entanto, ser utilizado em qualquer outra área de conhecimento, na qual as pesquisas lidem com o mesmo tipo de instrumento de coleta de dados.

Por meio da concatenação de fundamentos e conceitos estatísticos desenvolvidos por Triola (1999), Hoel (1984) e Siegel (1981), foi possível utilizar o teste de sinais, como instrumento para tal avaliação, utilizando-se cálculos simples e aplicativos de uso geral, como o MS Excel. Foi também, elaborado um fluxo explicativo do processo, detalhando passo a passo, as etapas e pontos de decisão.

Como sugestão para novos trabalhos, destaca-se que, com a existência de diversos testes não-paramétricos, dos quais os mais comumente observados são os testes de Wilcoxon, Mann-Whitney, Kruskal-Wallis, Kolmogorov-Smirnov e Qui-quadrado, a identificação de métodos, para avaliar o nível de significância destes testes, poderia ser um campo a ser explorado em novos estudos, reduzindo a lacuna existente nesta área.

## 6. REFERÊNCIAS BIBLIOGRÁFICAS

AN, T. L.; CUOGHI, O. A. A utilização da estatística na Ortodontia. *Revista Dental Press Ortodontia e Ortopedia Facial*. 2004, Vol 9, nº 6, pp. 97-108.

BLAIN, G. C. **Precipitação pluvial e temperatura do ar no Estado de São Paulo: periodicidades, probabilidades associadas, tendências e variações climáticas**. 2010, 194f. Tese (Doutorado em Agronomia). Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2010.

COOPER, D. R.; SCHINDLER, P. S. **Métodos de pesquisa em administração**, 7ª ed. Porto Alegre: Bookman, 2003, 640p.

DE SOUZA, D. P. H. **Avaliação de métodos paramétricos e não paramétricos na análise de eficiência da produção de leite**. 2003, 136f. Tese (Doutorado em Ciências). Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2003.

FERREIRA, H. S.; CAVALCANTE, S. A.; CABRAL JR., C. R.; PAFFER, A. T. Efeitos do consumo da multimistura sobre o estado nutricional: ensaio comunitário envolvendo crianças de uma favela da periferia de Maceió, Alagoas, Brasil. *Revista Brasileira de Saúde Materno Infantil*, Vol. 8 nº 3 Recife July/Sept. 2008.

GIL, A. C. **Estudo de caso**. São Paulo: Atlas, 2009, 148p.

HASEGAWA, L. N. **O papel das interfaces no sucesso de projetos utilizando equipes virtuais**. São Paulo; 2002, 138f. Dissertação (Mestrado em Administração de Empresas). Faculdade de Economia, Administração e Contabilidade. Universidade de São Paulo, São Paulo, 2002.

HOEL, P. G. **Estatística matemática**, 4ª ed. Rio de Janeiro: Guanabara Dois, 1984.

- LEVIN, J. **Estatística aplicada a ciências humanas**, 2ª ed. São Paulo: Habras, 1987.
- MARCONI, M. A.; LAKATOS, E. M. **Metodologia científica**, 4ª ed. São Paulo: Atlas, 2004, 305p.
- MATTAR, F. N. **Pesquisa de Marketing – execução, análise**. 2ª ed. São Paulo: Atlas, 1998, 224p.
- \_\_\_\_\_. **Pesquisa de Marketing – metodologia, planejamento**. 5ª ed. São Paulo: Atlas, 1999, 337p.
- MONTGOMERY, D. C.; RUNGER, G. C. **Estatística aplicada e probabilidade para engenheiros**, 4ª ed. Rio de Janeiro: LTC, 2009, 496p.
- NORRIE, J. **Improving results of project portfolio management in the public sector using a balanced strategic scoring model**. 2006, 173f. Tese (Doctor of Project Management – DPM). Royal Melbourne Institute of Technology, School of Property, Construction and Project Management, Design and Social Context, RMIT University, 2006.
- QUINTELLA, H. L. M. M.; ROCHA, H. M. Medindo o Nível de Maturidade dos Processos de Desenvolvimento de Produtos das Montadoras de Veículos com o CMMI. **Revista MundoPM**, Curitiba, Vol. 1, nº 06, pp. 20-27, 2006.
- \_\_\_\_\_. Nível de maturidade e comparação dos PDPs de produtos automotivos. **Produção** (São Paulo), Vol. 17, pp. 199-215, 2007.
- ROCHA, H. M. **Fatores críticos de sucesso do start up de veículos e a qualidade (CMMI) do desenvolvimento de produtos no Sul fluminense**. 2005, 353f. Dissertação (Mestrado em Sistemas de Gestão). Universidade Federal Fluminense, Niterói, 2005.
- \_\_\_\_\_. **Fatores críticos de sucesso no processo de desenvolvimento de produtos**. 2009, 272f. Tese (Doutorado em Engenharia Mecânica). Universidade Estadual Paulista Júlio de Mesquita Filho, Guaratinguetá, 2009.
- ROCHA, H. M.; DELAMARO, M. C.; QUINTELLA, H. L. M. M; AFFONSO, L. M. F. Fatores críticos de sucesso no processo de desenvolvimento de produtos na indústria automotiva Sul-Fluminense. In: VII Simpósio de Excelência em Gestão e Tecnologia, 2010, Resende. **Anais do VII Simpósio de Excelência em Gestão e Tecnologia**, Resende: AEDB, 20-22 out. 2010, p.1-15.
- SIEGEL, S. **Estatística não-paramétrica para ciências do comportamento**. São Paulo: McGraw-Hill do Brasil, 1981, 288p.
- SIEGEL, S.; CASTELLAN JR, N. J. **Estatística não-paramétrica para ciências do comportamento**, 2ª ed. Porto Alegre: Artmed, 2006, 448p.
- SILVA, P. D. A.; SANTOS, O. M.; FERREIRA, A, C. S. A percepção dos docentes quanto às práticas de contabilidade estratégica: um estudo comparativo. **Revista Contabilidade & Finanças**, São Paulo, Vol. 18, nº 44, pp. 44 – 59, Maio/Agosto 2007.
- TOLEDO, G. L.; OVALLE, I. I. **Estatística básica**, 2ª ed. São Paulo: Atlas, 1985, 459 p.
- TRIOLA, M. **Introdução à estatística**, 7ª ed. Rio de Janeiro: LTC, 1999, 410 p.
- VERGARA, S. C. **Métodos de pesquisa em administração**. São Paulo: Atlas, 2005, 287 p.



