

# ANÁLISE DAS POSSÍVEIS DIFERENÇAS ENTRE CONTRATANTES E CONTRATADOS EM TERCEIRIZAÇÃO DE SERVIÇOS DE SOFTWARE SEGUNDO A MÉTRICA DE ANÁLISE DE PONTO DE FUNÇÃO

*ANALYSIS OF THE POSSIBLE DIFFERENCES BETWEEN CONTRACTED AND CONTRACTORS IN THE OUTSOURCING OF SOFTWARE SERVICES ACCORDING TO THE METRIC OF ANALYSIS OF FUNCTION POINTS*

**DIRCEU DA SILVA**

dirceuds@fecap.br

**MAURO NEVES GARCIA**

mneves@gmail.com

**HILDA MARIA DA ROCHA RINALDI**

hildamar@dialdata.com.br

**CECÍLIA CARMEM CUNHA PONTES**

cpontes@fecap.br

## RESUMO

O objetivo deste estudo é a avaliação das possíveis diferenças nos contratos de terceirização para o desenvolvimento de sistemas de informação – SI, baseada em métricas de Análise de Pontos de Função – APF. Assim sendo, foi desenvolvido um instrumento de avaliação que analisasse seus aspectos mais relevantes. Procedeu-se, então, à análise de contratos e à realização de entrevistas com especialistas da área. Por fim, as considerações destes, aliadas à literatura específica de contratos, resultaram em um instrumento do tipo escala de Likert com 16 assertivas, aplicado em uma amostra de 82 respondentes que possuem certificação de APF ou que atuam na gestão de contratos. A escala foi validada através do método de análise fatorial exploratória baseada nos componentes principais. Também, procedeu-se à análise discriminante de três características dos sujeitos em relação aos fatores gerados na análise fatorial, que não apresentou diferenças significativas. Da análise das diferenças entre as variáveis categóricas dos respondentes e as assertivas, somente duas assertivas apresentaram diferenças significativas com relação à parte no contrato e natureza do trabalho (assertiva 15, nos dois aspectos e assertiva 11, somente com relação à parte no contrato). Com relação ao tempo de experiência na função, não se identificou diferença significativa em nenhuma das assertivas. Pela conclusão do estudo constata-se a necessidade de clareza e objetividade nos contratos e dos serviços de *software*, assim como a necessidade de maior definição dos envolvidos quanto aos seus papéis nas negociações.

*Palavras-chave:* métrica de *software*, análise de ponto de função, contratos de terceirização, mensuração de atitudes, análise multivariada de dados.

## ABSTRACT

This study on which this article is based analyzed possible differences in the outsourcing contracts for the development of information systems – IS on the basis of the metric of the Function Points Analysis – FPA. Thus, an evaluation instrument was developed to analyze its most relevant aspects. The next step consisted of the analysis of contracts and of interviews with IS specialists. Finally, the considerations of the latter, allied to the specific literature on contracts, resulted in an instrument similar to a Likert scale with 16 assertions, applied in a sample of 82 respondents who had FPA certification or worked in contract management. The scale was validated through the method of exploratory factorial analysis of the main components. Furthermore, a discriminating analysis of three different characteristics of the sample respondents in relation to the factors generated by the factorial analysis was made, which did not show significant differences. In the analysis of the differences between the

categorical variables of the respondents and the assertions, only two assertions presented significant differences with regard to the party in the contract and the nature of the work (assertion 15 in the two aspects and assertion 11 only with regard to the party in the contract). As far as the time of experience in the function is concerned, no significant difference was identified in any of the assertions. In terms of conclusion, the study identified the need for clarity and objectivity in the contracts and software services, as well as the need for a better definition of the stakeholders in relation to their roles in the negotiations.

*Keywords:* metric of software, function point analysis, software outsourcing contracts, measurement of attitudes, multivariate data analysis.

## INTRODUÇÃO

O aumento crescente das demandas pela utilização de sistemas de informação – SI personalizados tem levado organizações a terceirizarem os seus desenvolvimentos (Greaver, 1999). A adoção desta estratégia de contratação de produção de *softwares* implica aspectos e dificuldades para negociação, diferentes dos bens e serviços mais ortodoxos (Adamson e Dilts, 1995), pois sistemas com tamanhos iguais podem ter complexidades e custos muito diferentes (Wrigley e Dexter, 1991).

Os motivos para a atenção especial aos *softwares* e SI nas organizações residem no fato de que esses produtos, na maioria dos casos, têm valores relativamente elevados e respondem por parte significativa dos planos orçamentários das empresas. De fato, estimativas mostravam na década de 1990 que mais de 15% dos projetos terceirizados nos Estados Unidos da América – EUA foram abandonados em estágios intermediários de desenvolvimento, por razões de custos não previstos e provenientes de problemas em contratos (Mukhopadhyay *et al.*, 1992).

Dessa forma, além dos custos outras possíveis causas de dificuldades residem na negociação e na gestão de contratos de desenvolvimento de sistemas. E isso ocorre porque ainda é difícil comparar o que foi encomendado do *software* e o que foi entregue, isto é, os valores monetários apresentados na entrega da tarefa geralmente diferem dos valores acordados no momento da encomenda. O produto final de um sistema de processamento de dados (entendido como um aplicativo, desenvolvimento de um produto novo, manutenção de algum módulo de *software* já existente, ou até mesmo a correção de um erro) pode ter tamanho e complexidade diferentes daqueles que foram inicialmente negociados, comprados ou encomendados (Garmus e Herron, 2000).

Garmus e Herron (2000) apontam ainda a dificuldade de se contabilizar os investimentos em um *software* ou de um SI e se avaliar o respectivo retorno (*Return on Investment* – ROI), o que gera confusões sobre a real importância percebida pelos contratantes do desenvolvimento dos *softwares* e sua necessidade entre as prioridades das organizações. Além disso, há também a dificuldade de se avaliar se o valor a ser pago por ele é justo.

Assim, para dirimir os impasses, confrontos e "conversas surdas" entre contratantes e contratados de SI, foi criado, como

possibilidade, um padrão internacional de medida ou de métrica de *softwares* denominada de Ponto de Função ou Análise de Pontos de Função – APF. Ela representa uma medida ou métrica funcional de *software*, isto é, a avaliação do que o *software* faz com os comandos e sintaxes e não apenas o tamanho ou extensão deles. Este padrão foi conceitualmente definido pelo ISO/IEC 14143-1:1998 (ISBG, 2003) e se baseia na funcionalidade solicitada e recebida pelos respectivos usuários.

Mesmo tendo este padrão, pode haver avaliações muito diferentes. Fleck Jr. (1998), em um estudo onde avaliou o tempo previsto e o realizado em 20 projetos de manutenção de SI, usando a APF, verificou que a maioria deles apresentou erros de estimacão e problemas diversos, pois os resultados divergiram do que estava nas propostas dos contratos.

Por outro lado, não basta haver uma normalização, pois os contratos de terceirização (*outsourcing*) não parecem ser entendidos pelos profissionais envolvidos nas transações, visto que além de aspectos tecnológicos (características de aplicação de um SI e particularidades de cada projeto) há também fatores organizacionais e humanos, que influenciam sobremaneira as avaliações dos *softwares* (Bok e Raman, 2000). De fato, estudo realizado por Jones (2001) com 235 casos de contratação de empresas de SI nos EUA mostrou que existe uma diferença muito significativa entre o que foi planejado e esperado dos *softwares*, indicando que os elementos presentes nos contratos e sua interpretação geraram muitas insatisfações. Não é sem razão que todos os participantes desse estudo apresentaram a indicação da necessidade de renegociação dos contratos após a sua realização, além é claro de insatisfações.

Assim, frente à problemática exposta, este estudo pretende investigar as atitudes de contratantes e contratados de serviços de terceirização de *software* e de SI segundo os aspectos mais importantes dos contratos, que são as bases para a negociação da APF, e que podem ser sintetizados no problema de pesquisa como segue:

Quais são as opiniões de contratantes e contratados nos serviços de terceirização de sistemas de informação e *softwares*?

Para melhor contextualizar a pesquisa, objetivos e hipóteses são definidos como segue: (1) Objetivo geral: comparar as posições dos contratantes e contratados, com respeito aos

possíveis constructos gerados por uma escala de atitude com respeito a APF; (2) Objetivo específico: construir e validar uma escala de atitude para a geração, avaliação e mensuração de constructos da APF.

Com respeito às hipóteses, levou-se em consideração a opinião de dez especialistas entrevistados para a construção da escala usada neste estudo. Os aspectos referentes às entrevistas e a escala estão descritos nos procedimentos metodológicos. As hipóteses consideradas neste estudo são as seguintes:

H1: Há diferença entre as posturas das partes dos contratos, contratantes e contratados, com respeito aos aspectos da APF.

H2: Há diferenças com respeito à APF nas opiniões dos respondentes com respeito ao tempo de experiência na área.

H3: Há diferenças nas opiniões com respeito à APF no tocante às funções que os respondentes exercem nas empresas, técnica ou administrativa.

Uma vez apresentado o cenário geral do problema, objetivos e hipóteses da pesquisa, têm-se na seqüência os aspectos teóricos da APF visando lastrear as bases de análise deste estudo.

## ASPECTOS DA EVOLUÇÃO DAS MÉTRICAS PARA OS SOFTWARES

A demanda crescente por *softwares* específicos e SI diferenciados para a realização de diferentes funções e processos nas organizações tem levado à criação de um mercado complexo com diferentes produtos e personalizações. Essa situação faz com que empresas necessitem de produtos diferenciados, diferentes daqueles chamados de *prêt-à-porter* ou de *balcão*, tais como os aplicativos para edição de textos, planilhas, apresentações etc. Assim, a contratação de serviços terceirizados, também em oferta crescente, passa por negociações onde os padrões sobre os quais se estabelecem os contratos devem ser padronizados e explicitados com clareza para que as partes dos contratos possam estabelecer um diálogo compreensível. De fato, Santos *et al.* (2001) comentam que a motivação para a implantação de uma métrica são os vários problemas advindos da estimativa de prazos, quase sempre incorreta, e a falta de dados concretos que sirvam para controlar os projetos, tais como os riscos, preços, prazos, entre outros.

Esclarecendo mais, um *software* é medido para: (1) indicar a qualidade do produto; (2) avaliar a produtividade das pessoas que produzem o produto; (3) avaliar os benefícios, em termos de produtividade e qualidade, derivados de novos métodos e ferramentas de *software*; (4) formar uma linha básica para estimativas; (5) ajudar a justificar os pedidos de novas ferramentas e treinamento adicional.

Nesse contexto, desde a década de 1970 vêm surgindo tentativas de estabelecer padrões para as negociações. Dessa forma, o que segue são considerações para que se possa entender essa evolução.

Já faz anos que se sabe que o planejamento de *software* tem produzido prazos e preços mais justos e coerentes, mas o problema ainda presente é a dificuldade de saber qual a melhor maneira de se medir um *software* (Armour, 2004). Ainda, como apontava Azevedo (1997), existem algumas formas de medir sistemas que são aceitáveis e consistentes – contagens de linhas de código fonte, sistema métrico de Halstead (1972) e APF.

Uma possível classificação geral, conforme Pressman (1995), divide a métrica de *software* em medidas diretas e indiretas, exemplificando a primeira através do LOC (*lines of code*, ou linhas de código produzidas). Azevedo (1997) também apresenta que a vantagem é a simplicidade deste grupo de medidas diretas, porém, como as linguagens de programação podem ser diferentes em *softwares*, pode haver pouco significado para o usuário final. Por outro lado, o grupo de medidas indiretas é mais complexo, pois suas medidas residem na eficiência dos programas e, portanto, são difíceis de serem avaliadas. Mesmo assim tais medidas podem ter um padrão único para que o usuário contratante possa entender e avaliar o SI.

Existem algumas formas de medir sistemas que são aceitáveis e consistentes, tais como LOC ou SLOC, sistema métrico de Halstead, sistema métrico de McCabe, PROBE e APF (Humphrey, 1997). Cada métrica citada está baseada em diferentes aspectos dos *softwares* e dos sistemas informatizados. Por exemplo, a LOC usa apenas a contagem das linhas de programação para obter a avaliação da complexidade e do custo dos SI (Costagliola *et al.*, 2000).

Na década de 1970, com o advento do paradigma estruturado e das noções da complexidade cognitiva, foram publicados trabalhos que propunham medidas que abordassem estes aspectos, como o método de Halstead, desenvolvido em 1972 (Azevedo, 1997). Ele consiste em registrar para cada programa desenvolvido o número de operadores e operandos utilizados, permitindo o cálculo do tamanho do programa e o esforço de programação, independentemente da linguagem utilizada, mas baseada na sintaxe dos programas.

Ainda nos meados de 1970, foi desenvolvida a métrica da "complexidade ciclométrica" de McCabe, baseada no número de condições de fluxo em um programa (em um conceito estrito de complexidade). Ambas as métricas (Halstead e McCabe) traziam a proposta de independência da tecnologia empregada, ou seja, da linguagem de programação. Comparando as duas métricas, a de McCabe permite estimativas depois de uma certa etapa da programação, pois as informações necessárias podem ser extraídas a partir de diagramas de projeto e não somente de código fonte (Prantoni, 2001).

Azevedo (1997) destaca que, em 1996, Humphrey criou, como subproduto do desenvolvimento do *Personal Software Process* – PSP, um método de estimativa de tamanho de *software* chamado PROxy Based Estimation (PROBE) – Estimativa Baseada em Substituto. A sua proposta foi produzir uma estimativa de tamanho de um *software* em LOC no início de seu ciclo de desenvolvimento, exigido cada vez mais por clientes e usuários.

A idéia central era estimar o tamanho do *software* a partir do tamanho dos objetos identificados como seus possíveis componentes (Azevedo, 1997).

O método baseia-se fortemente em modelos conceituais preliminares. É fato que, no momento citado, as únicas informações disponíveis aos desenvolvedores são as extraídas de uma descrição de alto nível (muitas vezes vaga) de requisitos. Tentar produzir uma estimativa a partir destas descrições, mesmo considerando a existência de informações históricas gerenciadas para comparações, é uma atividade com poucas garantias de bons resultados, porque estes requisitos descrevem (se, nesta etapa, detalhados o suficiente) o que é desejado em relação ao produto final e não como este será construído. A pouca garantia também decorre de dois outros aspectos: (1) da grande dificuldade em encontrar similaridades suficientes em todo o universo destes requisitos; (2) identificar os requisitos utilizados em projetos anteriores para comparações mais precisas.

Mais especificamente, seguem as descrições de algumas métricas que precederam a APF e que servem de base para se ter uma idéia mais precisa do desenvolvimento dessas métricas e do processo para a busca de padrões para os SI.

### MÉTRICA POR ITENS DE SISTEMA

A introdução do ambiente gráfico, da orientação a eventos e da orientação a objetos, isto é, dos *softwares* gráficos com janelas e ícones tais como os utilizados pela maioria das pessoas que têm computadores, trouxe falhas nos modelos tradicionais. Para se ter uma visão do processo, duas situações hipotéticas podem acontecer:

Um programador gastou oito horas de trabalho reformatando as disposições visuais de uma tela, usando, por exemplo, um *software* de montagem de elementos visuais, além da criação gráfica de elementos visuais. Nenhuma linha de código foi alterada ou criada, entretanto oito horas de trabalho foram despendidas para que uma tela pudesse ser alterada e se criasse um ambiente gráfico de interação; por outro lado, pode-se ter um programador que gastou 15 minutos para a criação completa de uma outra tela simples e alterou ou reescreveu três ou quatro linhas de códigos de um *software*.

No primeiro caso, não se alteraram as linhas de código, apesar das oito horas de trabalho, mas no segundo houve alteração, apesar do tempo de trabalho de apenas 15 minutos. Dessa maneira, as introduções dos elementos gráficos nos *softwares* criaram uma nova situação de que a atenção apenas para as linhas de código ou a sintaxe não servia mais como um indicador seguro do trabalho realizado. Por outro lado, é impraticável ao gerente de desenvolvimento (da empresa contratada) ficar verificando as telas dos *softwares* e atribuindo graus de facilidade ou de dificuldade a cada uma delas, pois, além de haver o problema de indisponibilidade de tempo, tal critério ficaria no âmbito puramente subjetivo, variando de gerente a gerente.

Assim, o modelo de Métrica por Itens de Sistema foi criado para solucionar situações como as descritas, pois recomenda que se contabilize a quantidade total de itens do sistema e não mais com base na quantidade de linhas em código (LOC).

Voltando aos dois exemplos hipotéticos, no primeiro a quantidade de itens existentes é muito maior do que a do segundo. Se uma tela é mais complexa, então ela deverá ter mais controles (em termos técnicos: mais sessões, *jobs* SQL, códigos anexados etc). Portanto, a quantidade de itens indicará, por si só, que a primeira tela é mais complexa que a segunda. Aplicando o modelo de métrica sobre itens, fica patente que o primeiro programador trabalhou mais do que o segundo, apesar de não ter escrito uma única linha de código.

Também a manutenção e atualização de SI passaram a ser agora quantificadas. Tome-se um outro exemplo: um programador modificou uma tela já existente, alterando a posição de um botão de controle e mudando a cor de uma linha de *status* para, por exemplo, as cores vermelhas, destacando que a mensagem é muito importante.

A posição em tela é uma propriedade atribuída a qualquer controle. Portanto, tal propriedade indicará uma alteração em seu conteúdo quando se comparam as duas versões de sistemas. Da mesma forma, a alteração da cor de um controle também é representada pela alteração de uma propriedade correspondente. Assim, ao contabilizar-se a quantidade de propriedades alteradas, avaliar-se-á a carga de trabalho realizada.

Além disso, ainda é possível mencionar dois outros modelos: o de Putnam e o COConstructive COst MOdel – COCOMO (Pressman, 1995), também baseados em estimativas de linhas de código.

O de Putnam é um modelo dinâmico de múltiplas variáveis e foi construído a partir da distribuição de mão-de-obra encontrada em grandes projetos (esforço total de 30 pessoas-ano ou mais) e, por isso, não é muito utilizado. Já o COCOMO é um modelo estatístico de valor simples que computa o esforço e custo de desenvolvimento de sistema como uma função do tamanho de programa, expresso em linhas de código estimadas. Este modelo evoluiu para o COCOMO II, que também se utiliza da avaliação das APF.

Assim, Pressman (1995) dividiu as métricas de *software* em algumas categorias de métricas: (a) orientadas ao tamanho: medidas diretas de *software* e do processo pelo qual ele é desenvolvido; (b) orientadas à função: oferecem mediadas indiretas; (c) orientadas às pessoas: compila informações sobre a maneira segundo a qual as pessoas desenvolvem *software*; (d) da produtividade: concentra-se na saída do processo de engenharia de *software*; (e) da qualidade: quanto à adesão do uso do *software*; (f) técnicas: concentra-se nas características de *software* como complexidade lógica e não ao processo por meio do qual o *software* foi desenvolvido.

Pressman (1995) afirma ainda que as métricas orientadas ao tamanho provocam controvérsias, pois não são universalmente aceitas como a melhor maneira de se medir o

processo de desenvolvimento de *software*. Um dos maiores problemas apontados pelos opositores dessa métrica é a dependência da linguagem de programas (Azevedo, 1997).

Garmus e Herron (2001) comentam que as métricas orientadas à função foram propostas pela primeira vez por Allan Albrecht, engenheiro da IBM dos EUA, que em 1979 apresentou numa conferência, o documento chamado *Measuring Application Development Productivity – MADP*, sugerindo o uso das funcionalidades ou utilidades dos programas. Apesar de ser uma das métricas mais recentes, ela se tornou muito mais utilizada do que as outras, talvez pela respeitabilidade e pioneirismo que a IBM tem no mercado corporativo de SI, além do volume de negócios que a empresa realiza.

O crescente uso da APF e a adoção dessa métrica por grandes empresas contratantes de serviços fizeram com que houvesse o estabelecimento de regras e normalizações:

a) Foi o primeiro método a utilizar o conceito de Medição Funcional de Tamanho (*Functional Size Measurement – FSM*), estabelecido como norma ISO, sob o padrão ISO/IEC 14143-1:1998. (ISBSG, 2003). A parte 1 desse padrão foi publicada em 1998 como um padrão internacional completo e encontra-se disponível através dos organismos nacionais de padrões em cada país. A parte 1 – Definição de Conceitos – identifica as características comuns fundamentais dos métodos de medição funcional de tamanho e define um conjunto de requisitos genéricos obrigatórios para que um método possa ser chamado de Método de Medição Funcional de Tamanho (FSM).

b) O método do IFPUG foi oficializado através do padrão internacional ISO/IEC 20926 de 2002 e foi publicado pela ISO e IFPUG: ISO/IEC 20926:IFPUG (IFPUG, 1999).

c) O órgão internacional IFPUG valida a cada cinco anos as regras que serão disseminadas pela ISO.

d) Os profissionais podem pleitear a certificação CFPS – *Certified Function Point Specialist*, conferida pelo *International Function Point Users Group* – IFPUG aos aprovados no exame de certificação realizado duas vezes por ano nos Estados Unidos. Em outros países, como no Brasil, a certificação fica sob os auspícios do *Brazilian Function Point Users Group* – BFUG e é totalmente aderente ao IFPUG. A certificação é a garantia de que o profissional domina, aplica e utiliza corretamente as regras do IFPUG para a contagem de Pontos de Função. As certificações têm validade por três anos.

Recaindo sobre os aspectos da APF, Simões (1999) aponta como uma das vantagens a estimativa que é feita em função da visão do usuário, da facilidade de aprendizagem e aplicação da técnica, da independência da tecnologia e provisão de um fator de comparação de *softwares*. Como desvantagem é destacado que a aplicabilidade nas diversas fases do projeto de desenvolvimento de sistemas requer um esforço de contagens de pontos de função para cada fase. Para Dekkers (2001), a APF

se baseia em requisitos funcionais dos usuários (o que o *software* faz), independentemente da implementação física (como o *software* realiza as operações).

Também, Santos *et al.* (2001) comentam que a métrica da APF é utilizada pelo fato de conseguir um valor para a medida funcional de tamanho do sistema que será desenvolvido e, a partir dessa métrica, poderão ser derivadas as medidas de esforço, produtividade e custo. Os autores acrescentam que a APF é uma métrica que permite medir o tamanho do *software* a partir das funções que serão colocadas ao usuário, independentemente da tecnologia utilizada.

Complementando, o *The Netherlands Software Metrics Users Association* – NESMA indica que a APF já foi aplicada extensivamente no desenvolvimento de novos *softwares* e enfatiza que o método é tão objetivo e direto que os resultados obtidos são independentes do profissional que o aplica – os resultados baseados na utilização das mesmas instruções e maneiras de contagem são sempre os mesmos (NESMA, 2001)<sup>1</sup>.

## PROCEDIMENTOS METODOLÓGICOS

Para a realização do estudo, optou-se por utilizar uma escala somatória ou escala de Likert, para medir as atitudes e posicionamentos dos respondentes frente a um conjunto de assertivas que cobrem as dimensões da problemática da pesquisa. A escala é equilibrada e com ponto neutro, já que se admite a possibilidade das atitudes assumirem a posição de neutralidade dos respondentes (Malhotra, 2006). Tal escala pode ser enquadrada nos estudos de metodologia exploratória de *survey*, que é um procedimento sistemático para coletar informações que são usadas para descrever, comparar ou explicar fatos, atitudes, crenças e comportamentos (Babbie, 2001). A construção da referida escala foi realizada em três etapas:

### 1ª ETAPA

Entrevista em profundidade com dez (10) profissionais que possuem CFPS, reconhecidos pela área como atuantes em desenvolvimento de sistemas e em terceirização, e também com conhecimento de métricas de *software*. Estes profissionais tinham mais de dez anos de atuação na área de SI e tecnologia da informação, além de experiência em contratos de terceirização de serviços de *software*. Cinco deles eram contratantes e cinco de empresas contratadas. As entrevistas foram realizadas durante os meses de abril, maio e junho de 2003.

Construiu-se um conjunto de perguntas prévias, baseadas na literatura sobre APF (IFPUG, 1999; Fernandes, 1995; Micollis, 2001; Garmus e Herron, 2000; Pressmann, 1995 e Dekkers, 2001) como um conjunto de questões básicas para uma pesquisa semi-estruturada. Tomou-se o cuidado de seguir os passos descritos por Babbie (2001) para garantir-se maior confiabilidade no instrumento.

<sup>1</sup> Para maiores detalhamentos sobre a técnica APF sugerem-se as seguintes referências: IFPUG, 1999; Fernandes, 1995; Micollis, 2001; Garmus e Herron, 2000; Pressman, 1995; e Dekkers, 2001.

Também, buscou-se o levantamento e análise de 20 contratos diferentes de APF e a análise da literatura específica sobre os contratos de terceirização (Greaver, 1999), que resultou em um padrão com os seguintes itens: (1) Escopo dos serviços; (2) Padrões de performance; (3) Preço; (4) Fatores de produção – pessoas, facilidades, equipamentos, *software*, contratos com terceiros; (5) Gerenciamento e controle (processos); (6) Responsabilidades de cada parte na transição das tarefas; (7) Termos de pagamento; (8) Como o contrato será finalizado: por expiração, por causa, por conveniência, por alguma alteração; (9) Cláusulas de confidencialidade; (10) Garantias; (11) Termos contratuais; (12) Limites de obrigações; (13) Uso de subcontratados e licenças de terceiros. As entrevistas versaram sobre estes pontos e foram realizadas de forma a inserir cada ponto com a solicitação de maiores esclarecimentos e comentários, quando necessário.

As entrevistas foram gravadas, e a análise foi realizada por meio da técnica de análise de conteúdo (Bardin, 1979), onde se buscou interpretar e categorizar as respostas pelo critério de similaridades e coerências dos conteúdos das mesmas.

## 2ª ETAPA

Validação da escala de Likert mediante sua aplicação a uma amostra de profissionais que sejam CFPS e filiados ao BFPUG ou de área administrativa, mas que trabalhem diretamente com a emissão e gestão de contratos de APF.

Para a avaliação da amostra recorreu-se ao BFPUG, que apresentou a informação de que havia, em fevereiro de 2003, 156 profissionais certificados no Brasil. Em princípio tentou-se contato com todos os profissionais, mas, devido às dificuldades inerentes a esse procedimento, acabou-se optando por uma amostra de conveniência ou por julgamento, no sentido atribuído por Aaker *et al.* (2001).

Usou-se uma amostra avaliada pela regra empírica de se considerar, pelo menos, quatro a cinco sujeitos para cada variável independente analisada (Hair *et al.*, 2005). Esta regra deve ser seguida para evitar-se o fenômeno estatístico de sobre-ajuste de dados (*over fit*), conforme Hair *et al.* (2005). O instrumento foi aplicado em dois momentos: (1) maio e junho de 2003; (2) janeiro de 2004.

Como o instrumento final possuía 16 assertivas (ou variáveis independentes) inferiu-se, segundo (Hair *et al.*, 2005), que seria necessária uma faixa de valores para a amostra de 64 e 80 respondentes. Apesar de se ter distribuído 100 questionários, teve-se o retorno de 82, considerado adequado, pois estavam ligeiramente acima do nível superior da faixa de valores preconizados. As respostas das assertivas foram codificadas em uma escala de cinco pontos crescentes (de um a cinco) para aquela que representavam aspectos positivos e em uma escala decrescente (de cinco a um) para aquelas que encerravam aspectos negativos.

## 3ª ETAPA

Corresponde à análise das diferenças e testes de hipóteses segundo os métodos de Análise Discriminante e testes não paramétricos nas respostas da escala aplicada.

A seguir se explicita os testes e métodos realizados, nas etapas dois e três, com a massa de dados e com o uso do *software* SPSS base 10.0:

- a) Análise Fatorial Exploratória – AFE: visando à redução das variáveis em agrupamentos de fatores com forte correlação interna, para validação estatística da escala, através dos métodos de matriz rodada Varimax, Quartimax ou Equamax; para buscar-se um melhor ajuste de correlação, com a Normalização de Kaiser, que consiste em se considerar o "valor próprio" (*eigenvalor*) superior a 1,0 e limitação da carga fatorial no mínimo em 0,300 (Hair *et al.*, 2005);
- b) Coeficientes Alfa de Cronbach: para verificar a consistência e confiabilidade interna dos dados obtidos. Foram calculados os referidos coeficientes para cada fator obtido pelo método anterior (AFE), como recomenda Malhotra (2006);
- c) Teste de Kolmogorov-Smirnov: para a verificação da aderência à distribuição normal (Siegel, 1979). Tal teste tem grande importância, na medida em que, para ser realizada a análise discriminante, há a necessidade das variáveis terem distribuições que podem ser representadas pela curva de Gauss (Maroco, 2005);
- d) Análise Discriminante – AD: constructos ou fatores gerados na AFE, em relação às variáveis nominais (categóricas) que caracterizam os respondentes, ou seja: (1) Parte nos contratos (contratante ou contratada); (2) Natureza das suas funções ou cargos exercidos nas organizações (técnica, administrativa, ou ambas); (3) Tempo de experiência profissional (até dez anos e mais de dez anos), para separar as possíveis tendências de possibilidades que a amostra apresentou como um todo;
- e) Análise das possíveis diferenças entre as variáveis da escala de Likert (16 assertivas) com respeito às mesmas variáveis categóricas da AD: foram utilizados testes estatísticos uni e bivariados (t de Student ou Mann-Whitney e ANOVA ou Kruskal-Wallis) nas versões paramétricas e não paramétricas, escolhendo-se pela aderência ou não à distribuição normal.

Para efeito de esclarecimento geral, acrescenta-se que a realização dos testes "a" e "b" visam ao processo de validação da escala construída. Nesse sentido, a validação é a análise da capacidade de um teste medir aquilo a que se propõe. É também um processo laborioso de análise e de correção do instrumento de medida (Cronbach, 1996).

## APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Mostrados os aspectos teóricos e metodológicos da pesquisa, têm-se a seguir os resultados obtidos e suas análises, divididas em duas partes: (1) sobre as entrevistas com especialistas, com análise de conteúdo; (2) sobre a aplicação da escala de Likert desenvolvida, por meio de métodos estatísticos multivariados e testes bivariados.

### SÍNTESE DAS ENTREVISTAS PARA AJUSTE DO INSTRUMENTO

Como comentado, as análises das entrevistas por meio da técnica de análise de conteúdo (Bardin, 1979) mostraram muitos pontos importantes ressaltados pelos especialistas da área e permitiram uma síntese de 16 assertivas para cobrir todos esses pontos importantes. Estas análises serviram para a construção do instrumento final (escala de Likert) e estão condensadas a seguir:

- a) Houve discordância da existência de uma taxa predefinida de entrega em PF por mês. PF deve ser avaliada no início, no meio e determinada no final. Um dos exemplos indicava como maneira de se definir a taxa de entrega: 10% após a análise; 20% após o *design* e restante na entrega. Nessa questão ficou ressaltada que a discordância ocorre somente quando há mudança de escopo. No contrato deve haver cláusula para incrementos que não são medidos em Pontos de Função, como, por exemplo, acompanhamento de testes, geração de dados para teste, treinamento. Ficou destacada a necessidade de se ter medidas no decorrer e na conclusão dos módulos para ajuste do número de Pontos de Função;
- b) Deve ser prevista uma taxa/hora para atividades adicionais solicitadas e que não estão incluídas numa métrica como APF. Um dos comentários mais interessantes foi sobre o fato de que o SLA (*Service Level Agreements – Acordo de Nível de Preço*) causa tensão numa relação que precisa ser saudável e que as empresas sabem que, se não cumprirem os SLAs, perderão o contrato rapidamente. Esclarecendo, o SLA é um acordo estipulado pela contratante, a fim de trabalhar com transparência em relação a seus produtos entregues, sua produtividade, sua taxa de erros, sua pontualidade, sua qualidade, seus índices de retrabalho e replanejamento e a avaliação de satisfação do cliente;
- c) Ressalta-se uma resposta em que foi comentado que as regras são discutidas no momento da aceitação da proposta. Uma situação apontada foi que os projetos em geral são mal especificados (massa de testes, atendimento a agendas de reuniões, resposta e definições às dúvidas funcionais etc);
- d) Notou-se que os respondentes, em linhas gerais, são contrários à colocação do fator de produtividade no contrato;
- e) Os respondentes, na maioria, acham que o fundamento teórico deve ser explicitado para diminuir diferenças de interpretação, que podem ser causadas até por versões diferentes;
- f) A maioria dos respondentes acredita que isso melhora o entendimento do preço pedido, pois o mesmo pode variar conforme o ambiente. A contratada ficaria mais segura para determinar um preço, embora tenha ficado registrado que sempre existe um risco de ambos os lados;
- g) Os respondentes em geral mostraram que a inclusão desta cláusula implica conscientização dos dois lados;
- h) Reparou-se que os respondentes concordam com a multa para a contratada, mas alguns comentaram que deveria haver uma contrapartida por parte da contratante. Os respondentes, na sua maioria, acataram a existência de cálculo para multa, mas exigem maior clareza nas sanções que

poderão ser aplicadas. Houve uma contribuição interessante que seria a especificação de um limite máximo para a multa a fim de que não haja perda de atrativo numa concorrência. A multa deve ser medida em outra unidade que não a APF.

A partir dessas considerações foi elaborada a escala de Likert usada para a tomada de dados na amostra definitiva. Para tal foram relacionados os pontos mencionados pelos entrevistados, juntamente com outros aspectos presentes em contratos de serviços de *software* e SI. A primeira versão da escala foi submetida a outros cinco profissionais com experiência na área de mais de dez anos, para a validação teórica da escala (Pasquali, 2003). Estes fizeram algumas observações com respeito a termos técnicos, estrutura de algumas das questões e inclusão de duas assertivas não contempladas anteriormente.

### ANÁLISE DE DADOS DA APLICAÇÃO DO INSTRUMENTO

A massa de dados foi submetida ao critério de exclusão de casos discrepantes (*outliers*) como sugerido por Hair *et al.* (2005) e destacado pelo SPSS (1999), onde se compararam os critérios *listwise* (eliminação total do respondente com dados faltantes) e *pairwise* (eliminação apenas da célula em branco da matriz de dados). Como recomendam Hair *et al.* (2005), a escolha do critério *listwise* é sempre mais recomendada, mas é necessário saber-se se há diferenças significativas entre os dois critérios. Como os resultados das matrizes rodadas não mostraram diferenças entre os fatores e apenas pequenas diferenças nos valores das cargas fatoriais, optou-se pelo critério *pairwise*, pois este preserva todos os sujeitos da pesquisa.

Os resultados seguem apresentados em três momentos: (1) AFE para validação da escala; (2) Análise discriminante para exploração das características dos respondentes; (2) Análise das possíveis diferenças entre as escolhas dos respondentes das variáveis da escala.

### ANÁLISE FATORIAL EXPLORATÓRIA – AFE

Para que possa ser verificado se o método de AFE pode ser utilizado e se ele é adequado aos nossos propósitos, foram empregados dois testes estatísticos robustos: (1) o teste KMO (Kaiser – Mayer – Oklin), que mede o ajuste de possibilidades dos dados ao método em questão. Obteve-se o valor de 0,780, que mostra um ajuste aceitável, acima do limite inferior de 0,60 (Hair *et al.*, 2005); (2) o teste de Esfericidade de Bartlett, que avalia a mesma possibilidade do teste anterior. Neste caso, foi obtido um valor de significância menor que 0,0001 ( $\chi^2 = 160,387$ ;  $p < 0,0001$ ), indicando que há grande probabilidade dos dados se ajustarem ao método em questão (Hair *et al.*, 2005).

Uma vez confirmada a possibilidade de uso do método citado, testaram-se três critérios (Varimax, Quartimax e Equamax). O critério Equamax das Componentes Principais (melhor ajustado – valores dos resíduos padronizados maiores que 5% em 23,3% dos dados) revelou melhor ajuste e

mostrou seis fatores, que seguem em tabelas separadas com as cargas fatoriais. Obteve-se com os seis fatores um valor de variância de 66, 57%, que é considerado aceitável, pois valores acima de 60% em estudos exploratórios mostram que os dados são adequados às análises multivariadas. (Hair *et al.*, 2005). Os fatores foram nomeados segundo a análise do conteúdo das assertivas carregadas em cada um deles.

### FATOR 1 - OBJETIVIDADE E CLAREZA NAS DEFINIÇÕES

As variáveis 5, 7, 11 e 14 (vide tabela 1) ficaram aglutinadas num resultado que tem como linha mestra os preceitos básicos do documento de negociação e a teoria de medição, revelando que os respondentes querem que o contrato de desenvolvimento de sistemas seja um documento mais objetivo e que tenha definições descritas de maneira precisa. A objetividade ficou evidenciada na clareza com que responderam (e vale ressaltar aqui que foram profissionais que se posicionaram tanto como contratadas quanto contratantes) e que deve haver sempre um *Service Level Agreement*. Ou seja, um Acordo de Nível de Serviço estipulado entre as partes que norteie o trabalho, e, caso não seja cumprido, que se aplique algum tipo de multa. Pelo estudo das respostas ficou também evidenciado que não há expectativa de uma recompensa monetária para uma empresa que entregue o trabalho antes do prazo combinado.

Ressalta-se aqui a necessidade de explicitação da objetividade e transparência que os respondentes procuram nesses contratos com a citação positiva sobre a existência de uma auditoria que verifique e comprove a pertinência dos pagamentos efetuados.

**Tabela 1** – Assertivas reunidas no fator 1 pela AFE e cargas.

Assertivas	Fator 1
5. Um contrato de desenvolvimento de sistemas deve expor a teoria de medição de desenvolvimento de sistemas, ou seja, a teoria sobre a unidade utilizada para quantificar o trabalho a ser desenvolvido na qual se baseou.	0,731
7. A CONTRATANTE sempre deverá cobrar algum tipo de multa, se os acordos regidos pelo SLA não forem cumpridos.	0,694
14. O trabalho da Auditoria, de verificação da pertinência dos pagamentos efetuados à CONTRATADA, é facilitado quando se baseia em um contrato de desenvolvimento de sistema que teve seus pagamentos orientados pela quantidade de Pontos de Função implantado.	0,585
11. Quando a CONTRATADA entrega o trabalho antes do prazo combinado deve ser estipulada uma maneira da CONTRATANTE recompensá-la monetariamente.	0,512

Este fator responde por 15,47% da variância dos dados e apresentou valor do alfa de Cronbach de 0,876, mostrando que os dados não possuem vieses significativos (Pestana e Gageiro, 2000), isto é, que o espalhamento ou incidência de respostas atribuídas pelos sujeitos da pesquisa se mostrou coerente com as expectativas de não haver escolhas únicas nas assertivas. Tem-se como critério que valores maiores que 0,60 são aceitáveis para estudos exploratórios (Churchill Jr., 1999).

### FATOR 2 - EXPLICITAÇÃO DAS METODOLOGIAS NOS CONTRATOS

As variáveis 6, 8 e 10 (Tabela 2) evidenciaram, neste agrupamento, que para o bom desenrolar de um trabalho de desenvolvimento de sistemas deve haver, no contrato, a explicitação da metodologia a ser empregada.

Deve-se deixar claro o quanto vai ser cobrado e pago, através de medidas universalmente conhecidas e reconhecidas que concretizem o esforço despendido no desenvolvimento de sistemas e a remuneração adequada.

Os respondentes mostraram-se convergentes no fato de que deve haver multa como penalidade para desvios do processo e essa multa deve ser explicitada na mesma unidade da cobrança para que haja uma correspondência entre os valores, não ficando arbitrária a imputação da mesma. Essa conclusão difere das respostas obtidas no pré-teste, pois este teve como amostra um número pequeno de sujeitos.

Se a metodologia estiver bem delineada, o contrato será redigido de maneira mais clara e não haverá tantas discussões futuras. Deve haver uma metodologia na contratação desses serviços para que seja possível estudarem-se padrões e desvios.

Por fim, este fator responde por 13,51% da variância dos dados e apresentou valor do alfa de Cronbach de 0,705, que, como o anterior, é muito adequado.

**Tabela 2** – Assertivas reunidas no fator 2 pela AFE e cargas fatoriais.

Assertivas	Fator 2
8. Quando estiver evidenciada a existência de multa em um contrato de desenvolvimento de sistemas, ela deverá ser explicitada na mesma unidade de medida utilizada para quantificar o desenvolvimentos de sistemas.	0,798
6. A menção e a utilização explícita de Ponto por Função num contrato de desenvolvimento de sistema deixa transparente o preço total desse contrato.	0,790
10. Todo desenvolvimento de sistemas deve estar relacionado a um contrato que explicita: cláusulas legais, sobre o objeto do contrato, confidencialidade responsabilidades, garantias, obrigações, penalidades, preço e condições de pagamento, vigências e prazos.	-0,662

### FATOR 3 - PRODUTIVIDADE E TRANSPARÊNCIA

Este fator agrupa as variáveis 1 e 2 (Tabela 3) e mostra que os respondentes apresentam a necessidade de deixar detalhado o ambiente do trabalho em que o serviço está sendo contratado e também que as condições devem ser discutidas e acertadas antes da assinatura do contrato para que não haja distorções na hora do pagamento. Isso redundará numa prévia da produtividade que deverá ser praticada para, conseqüentemente, influenciar o preço e o prazo do projeto.

**Tabela 3** – Assertivas reunidas no fator 3 pela AFE e cargas fatoriais.

Assertivas	Fator 3
2. É imprescindível que, para se praticar um determinado preço, haja uma associação entre o ambiente onde vai ser desenvolvido, testado e executado o sistema e o tipo de unidade utilizada para se quantificar o <i>software</i> (comumente designado "medição de <i>software</i> ")	0,868
1. O montante a ser pago, quando da entrega dos produtos especificados no contrato de desenvolvimento de sistema não deverá apresentar distorções em relação ao estabelecido no momento da assinatura do contrato (estabelecido que não houve alteração de escopo)	0,754

Este fator apresentou um valor da variância de 11,02% e um valor do teste alfa de Cronbach de 0,691%, que também se revela adequado.

### FATOR 4 - ASPECTOS DIFERENCIADORES DO USO APF

O valor calculado da variância para este fator foi de 9,79% e o alfa de Cronbach de 0,652. As variáveis 4 e 9 (Tabela 4) mostraram que produtividade em Pontos de Função é um diferencial entre as empresas concorrentes e a inserção dessa informação no contrato pode assegurar essa distinção.

**Tabela 4** – Assertivas reunidas no fator 4 pela AFE e cargas fatoriais.

Assertivas	Fator 4
4. O contrato de desenvolvimento de sistemas deve deixar clara a adoção e o valor do fator de produtividade (quantas horas serão utilizadas por 1 profissional de sistemas para desenvolver 1 Ponto de Função).	0,874
9. Os contratos de desenvolvimento de sistemas devem se basear em número de horas para cada módulo/fase.	0,792

### FATOR 5 - TARIFAÇÃO DOS CONTRATOS

As variáveis 13 e 15 (Tabela 5) indicaram a preocupação dos respondentes com relação à tarifação. A maneira como ela

é expressa no contrato de desenvolvimento de sistemas pode tornar o contrato inviável para uma das partes em relação ao valor a ser despendido. Para este fator a variância obtida foi de 8,75%, e o valor do teste alfa de Cronbach mostrou o valor de 0,613, que pode ser considerado como adequado.

**Tabela 5** – Assertivas reunidas no fator 5 pela AFE e cargas fatoriais.

Assertivas	Fator 5
13. Não se deve basear um contrato de desenvolvimento de sistemas numa quantidade fixa de horas para o desenvolvimento total do sistema.	0,784
15. Nunca existirá uma redução no gasto total de um desenvolvimento de sistema, por parte da CONTRATANTE, quando a proposta estiver expressa em quantidade de Pontos de Função ao invés do número de horas.	0,594

### FATOR 6 - FORMAS DE COBRANÇA DOS CONTRATOS

As variáveis 12 e 3 (Tabela 6) confirmaram a preocupação com a maneira com que o contrato de desenvolvimento de sistemas será cobrado.

Se estiver expresso, de uma forma bem conhecida para ambas as partes envolvidas, como vai se medir o sistema desenvolvido, o contrato terá uma tarifação mais segura e será menos trabalhosa a sua gestão. Este fator é explicado por 8,00% da variância dos dados. O valor do teste de confiabilidade interna dos dados (alfa de Cronbach) foi de 0,589, muito próximo de 0,600, o que caracteriza um fator sem vieses significativos nas respostas dos sujeitos da amostra.

**Tabela 6** – Assertivas reunidas no fator 6 pela AFE e cargas fatoriais.

Assertivas	Fator 6
12. A existências de aditivos ao contrato inicial não torna a gestão de um contrato de desenvolvimento de sistema mais trabalhosa.	0,786
3. Os contratos de desenvolvimento de sistemas sempre devem fazer menção a meios de medição de <i>software</i> , isto quer dizer, à unidade em que se pode quantificar o trabalho, como por exemplo: os Pontos de Função ou o número de linhas de código.	0,728

A Variável 16 não foi composta em termos de correlação com os outros fatores, mostrando-se isolada. O texto desta variável era: *Deve ser levada em conta, para a composição do custo do projeto, como os profissionais são contratados (por exemplo: empresa, cooperado, funcionário contratado através da CLT, freelancer etc).* Uma possível explicação para esse não alinhamento foi o conteúdo da mesma, pois são comuns na área contratos de

vários tipos. Também a idéia subjacente à assertiva é nova e não representa um segmento consolidado em termos profissionais. Assim, os respondentes, em sua maioria, concordaram com a assertiva (média de 4,3 e desvio-padrão de 0,14), mostrando que ela não é adequada para a avaliação das opiniões sobre a APF. Em outras palavras, mostrou-se como uma "verdade auto-evidente".

Pela coerência lógica entre os fatores e pela possibilidade de sugestão de nomeações coerentes com eles, além da garantia de não violação dos testes estatísticos específicos, pode-se afirmar que a escala possibilita medir aquilo a que se destina e, portanto, pode ser considerada validada. (Cronbach, 1996).

### ANÁLISE DISCRIMINANTE

Uma vez tendo-se validado a escala de Likert utilizada, procedeu-se a três análises discriminantes, um método multivariado de larga utilização (Maroco, 2005) com respeito às variáveis categóricas, classificadoras ou nominais, a saber: (1) Posição nos contratos (contratante ou contratado); (2) Natureza das suas funções – administrativa, técnica ou ambas; (3) Tempo de atuação na área – até dez anos e mais de dez anos.

Tais análises buscaram testar as hipóteses apresentadas no início deste estudo. Para a análise discriminante, como comentado anteriormente, há a necessidade das variáveis independentes serem aderentes à distribuição normal (Maroco, 2005). Assim sendo, aplicou-se o teste de Kolmogorov-Smirnov (KS) nas retas de regressão de Bartlett dos fatores – estas retas são ajustes de regressão categóricos gerados pelo SPSS como opção de uso para outras análises (SPSS, 1999). Tais retas de regressão representam os fatores ou constructos gerados pela AFE, que mostram ou caracterizam conjuntos de assertivas da escala original.

Explicando melhor, procurou-se identificar como a análise fatorial estabelece um conjunto de equações lineares onde os parâmetros destas são as cargas fatoriais (os coeficientes de ajuste da variável ao todo), e as retas de regressão de Bartlett são as classificações em eixos arbitrários de cada sujeito ao fator. Assim, com esse procedimento, os dados originais, que eram as respostas às assertivas, passam a ser os valores de classificação do conjunto de respostas às novas variáveis que formam cada um dos fatores, de cada sujeito. Em outras palavras, o valor da regressão em questão para cada sujeito da pesquisa é a sua projeção no modelo fatorial obtido.

**Tabela 8** – Valores do teste M de Box para a avaliação da igualdade das variâncias-covariâncias das variáveis categóricas sobre os fatores.

Variável Classificadora	Discriminação	M' Box test	Significância (p)
Posição nos contratos	contratante contratada	29,34093322	0,270
Tempo de profissão	até 10 anos acima de 10 anos	32,50590219	0,161
Função nas Empresas	Administrativas Técnica Mista Adm. e Téc.	78,60697372	0,093

O teste KS apresentou valores da significância acima de 0,05 ( $p > 0,05$ ) (Tabela 7), mostrando que as variáveis representativas dos fatores são aderentes à distribuição normal (SPSS, 1999). Assim, pode-se tratá-las segundo o método de análise discriminante.

Além do teste KS, procedeu-se à aplicação do teste M de Box, para a avaliação da hipótese nula ( $H_0$ ), padrão do teste, de igualdade das matrizes de variância-covariância de cada grupo das variáveis categóricas discriminantes sobre as das retas de regressão de Bartlett dos fatores. Tal teste é importante para verificar se os dados, além de normais, podem ser submetidos à análise de discriminante, uma vez que esse método usa as técnicas de regressão lineares múltiplas e esta técnica necessita destas duas condições para ser empregada (Kazmier, 1982). Os valores da significância (*p-value*) para aceitação do teste devem ser superiores a 0,10 (SPSS, 1999). Observando os resultados do referido teste para as três variáveis classificadoras dos respondentes (Tabela 8), verifica-se que duas delas (Tempo de profissão e Posição nos contratos) apresentam valores de *p* maiores que 0,10 e que a variável "Função nas Empresas" tem valor muito próximo ao estabelecido e, assim, pode ser considerada para o tratamento de discriminação, podendo ser considerada significativa ( $p \approx 0,10$ ). Assim, os testes (KS e M de Box) mostraram que o método de análise discriminante é adequado aos dados selecionados.

**Tabela 7** – Valores obtidos no teste de Kolmogorov-Smirnov para análise de aderência à distribuição normal das retas de regressão de Bartlett dos fatores.

Variáveis dos fatores	Kolmogorov-Smirnov Z	Significância (p)
REGR factor score 1	0.970	0.304
REGR factor score 2	0.549	0.924
REGR factor score 3	1.129	0.157
REGR factor score 4	0.708	0.698
REGR factor score 5	0.520	0.950
REGR factor score 6	0.646	0.798

Legenda: REGR factor score 1, por exemplo, significa a reta de regressão de Bartlett do fator 1 e assim por diante.

Uma vez confirmada a possibilidade de uso da AD, procedeu-se às referidas análises para cada grupo indicado.

### **Análise Discriminante com respeito à Posição dos Respondentes nos Contratos (Contratante ou Contratado)**

Em princípio, realizou-se o teste Lambda de Wilks (LW) para se avaliar as possibilidades de discriminação dos seis fatores. Tal teste, baseado na estatística da distribuição do chi-quadrado, é normalizado entre 0 e 1 e, para que possa mostrar indícios de discriminação, deve possuir um valor menor que 0,200 (Dillon e Goldstein, 1984), e a significância do teste deve ser menor que 0,10 para a mesma finalidade. Neste caso, o teste LW resultou em um valor de 0,904 com significância  $p = 0,656$  e, portanto, encontra-se muito acima do indicado. Tal teste revela que não há discriminação significativa entre os níveis da variável "Posição nos Contratos".

### **Análise Discriminante com respeito à Natureza das suas funções**

Com respeito à análise discriminante da variável categórica "Naturezas das suas funções", solicitou-se aos respondentes escolher a opção da área de atuação com maior frequência de vínculo, isto é, de natureza administrativa, técnica ou ambas.

Assim, passou-se a ter uma variável com três dimensões e duas funções discriminantes, pois no modelo sempre uma variável é considerada fixa, resultando em um número  $k-1$  de graus de liberdade.

Os resultados dos testes LW mostram valores de 0,746 ( $p = 0,458$ ) para a primeira função e 0,914 ( $p = 0,603$ ) para a segunda função discriminante. Em virtude desses valores, também se pode inferir que não há discriminação entre as declarações de funções nas organizações dos respondentes.

### **Análise Discriminante com respeito ao Tempo de Atuação dos Respondentes**

Também com respeito à variável "Tempo de Atuação", dividiu-se os respondentes em dois grupos, até 10 anos de experiência e mais que 10 anos de experiência. O teste LW apresentou o valor de 0,466 com significância de 0,527, resultando na não discriminação das amostras. Uma possível explicação para a não discriminação pode estar referida à novidade da APF no Brasil, além do fato de que a APF é uma prática consagrada e muitas vezes imposta por algumas empresas de grande porte.

Os três casos de AD mostram que as diferenças são devido ao acaso e não tendências marcantes nas respostas dos sujeitos da pesquisa.

### **ANÁLISE DAS DIFERENÇAS DAS VARIÁVEIS DA ESCALA COM RESPEITO ÀS VARIÁVEIS CATEGÓRICAS**

Aplicou-se o teste de Kolmogorov-Smirnov, como já abordado anteriormente, para a avaliação da aderência das 16 variáveis da escala de Likert à distribuição normal. Os resultados apresentam valores da significância muito abaixo de 5% ( $p <$

0,0001). Esses valores indicam que as variáveis (assertivas) da escala não têm distribuições gaussianas ou normais e que os dados devem ser tratados por testes não paramétricos (livres de distribuição), conforme Siegel (1979). Assim, o nome deste item foi caracterizado como "análise das diferenças", para não gerar confusões entre o procedimento estatístico anterior (AD).

Assim, aplicaram-se testes não paramétricos, teste de U Mann-Whitney para as variáveis categóricas de dois níveis e o teste H de Kruskal-Wallis (ANOVA por postos) para a variável categórica de três níveis.

Como resultado (Tabela 9), em ambos os testes, as significâncias se apresentaram muito acima do valor do nível de significância adotado ( $\alpha = 5\%$  ou 0,05) para a rejeição da hipótese de nulidade do teste ( $H_0 =$  as amostras não apresentam diferenças significantes), indicando assim que não há diferenças entre os respondentes. A exceção (vide tabela 9 – células em cinza) são três casos que apresentaram significâncias abaixo do valor 0,05. Estes três casos dizem respeito à variável categórica "parte nos contratos" sobre as variáveis da escala 11 e 15 e a categórica natureza do trabalho sobre a mesma variável 15 da escala.

Como há distinção nessas variáveis, deve-se buscar conhecer as tendências dos grupos ou amostras de respondentes, e isto pode

**Tabela 9 – Valores da significância dos testes não-paramétricos para a análise das diferenças das variáveis da escala pelas classificadoras.**

Variável da escala	Significância dos testes não paramétricos		
	Parte contrato*	Tempo exper_ função*	Natureza do trabalho*
1	0,74	0,37	0,31
2	0,96	0,67	0,45
3	0,13	0,87	0,35
4	0,36	0,63	0,98
5	0,96	0,42	0,15
6	0,31	0,34	0,41
7	0,08	0,64	0,57
8	0,51	0,13	0,27
9	0,72	0,96	0,39
11	0,02	0,86	0,44
12	0,12	0,85	0,15
13	0,53	0,19	0,81
14	0,45	0,14	0,11
15	0,03	0,61	0,03

Legenda: \* significância do teste U de Mann-Whitney – 2 amostras independentes. \*\* significância do teste H de Kruskal-Wallis – 3 ou mais amostras independentes.

ser feito pela observância dos postos médios, que são entendidos como sendo a soma normalizada da quantidade de postos (Maroco, 2005). Estes valores não têm relação direta com os valores da escala ou com o conceito de média, mas são números relativos que devem ser interpretados por si mesmos, como distâncias em uma escala arbitrária. Assim, a Tabela 10 mostra os valores dos postos médios.

Na Tabela 9, vê-se que a variável "tempo de experiência na função" não diferencia as amostras, pois os valores da significância do teste de Mann-Whitney são superiores a 5%, e, assim, não se pode rejeitar a hipótese de nulidade de não igualdade entre as amostras.

**Tabela 10** – Valores dos postos médios das variáveis que se mostraram distintas.

Variável da escala	Variável categórica	Postos Médios
11	Parte contrato	
	Contratados	27,27
	Contratantes	17,03
15	Parte contrato	
	Contratados	21,19
	Contratantes	30,00
15	Natureza do trabalho	
	Administrativa	29,04
	Técnica	24,61
	Ambas	12,93

A análise da Tabela 10 revela que há diferenças entre as opções dos respondentes. No caso da variável 11, há uma maior concordância dos "contratados" (27,27) e na variável 15, a situação é contrária, pois os "contratantes" concordam mais com esta variável (30,00). Essa situação mostra que podem ser estes dois pontos aqueles que têm conflito.

Ainda, na variável 15, com respeito à "natureza do trabalho", profissionais com funções mais administrativas (29,04) tendem a concordar mais com o conteúdo da assertiva, seguidos daqueles com funções técnicas (24,61) e, por último, pelos que desempenham as duas funções (12,93).

Observando novamente os textos das assertivas em questão, nota-se que a assertiva 11 diz respeito a um aumento do valor de remuneração se o *software* for entregue em menor prazo do que aquele limite estipulado. Esta posição privilegiada pelos contratados parece óbvia, mas reflete um ponto de conflito quando comparada com a variável 15, pois esta afirma que: *Nunca existirá uma redução no gasto total de um desenvolvimento de sistema, por parte da CONTRATANTE, quando a proposta estiver expressa em quantidade de Pontos de Função em vez de número de horas.* Portanto, espelha mais a visão do contratante do que a do contratado evidenciando assim o conflito entre as partes.

Por fim, profissionais com formação administrativa mostram maior concordância nesta última assertiva, já que, na maioria dos casos, mantêm-se na posição de contratantes.

## CONCLUSÕES

Nesta pesquisa buscou-se analisar quais são as opiniões de contratantes e contratados nos serviços de terceirização de serviços de sistemas de informação e *softwares*. Para tal, foi desenvolvido um instrumento de avaliação para a análise dos aspectos relevantes dos contratos na modalidade de terceirização para o desenvolvimento de SI, baseados em métricas e padrões de APF.

Para a construção do instrumento, além da revisão da literatura, desenvolveu-se um conjunto de dez entrevistas com especialistas e análise de 20 contratos de terceirização de serviços e desenvolvimento de SI. A partir dos resultados construiu-se uma escala de Likert que foi aplicada a 82 profissionais da área. Os dados foram analisados segundo os métodos de AFE e Análise Discriminante, além de testes univariados e bivariados.

A escala foi validada, e os seis fatores encontrados na AFE explicaram 66,57% da variância dos dados, isto é, a maioria dos dados foi incluída no modelo adotado para a redução dos dados em fatores (Pasquali, 2003).

Os fatores 1, 2 e 3 apresentam valores da variância superiores a 10% e são responsáveis pela explicação de 50,02%, de um total de 66,57%. Assim sendo, apresentaram as necessidades de haver maior clareza e objetividade nos contratos e na explicitação das metodologias utilizadas para a avaliação dos sistemas de informação, além da necessidade de deixar detalhado o ambiente do trabalho em que o serviço está sendo contratado e as condições que devem ser discutidas e acertadas, antes da assinatura do contrato, para que não haja distorções na hora do pagamento.

Esses aspectos reforçam a literatura utilizada, quando autores como Fleck Jr. (1998) e Jones (2001), entre outros, apontam problemas nas realizações dos contratos. Então, a necessidade de haver maior clareza possibilita negociações mais justas e com maior grau de lisura por ambas as partes.

Por outro lado, a não discriminação das variáveis categóricas ou qualificadoras dos respondentes, segundo os fatores, além de rejeitar as três hipóteses formuladas, pode estar indicando que tanto o perfil profissional do técnico, como do pessoal administrativo, ou ainda das partes dos contratos de terceirização, são muito parecidos. Assim, pode-se estar detectando que os conflitos citados na literatura não estão nas APF, mas sim na operacionalização e execução dos contratos.

Também, diferenças foram observadas apenas nas assertivas 11 e 15 que dizem respeito à premiação fora dos contratos e à possível redução de custos nas contratações. Estes aspectos reforçam e caracterizam que os pontos de possíveis impasses e conflitos estão em aspectos específicos e locais, pois estas assertivas colocam posições contrárias entre os contratantes e contratadas.

Quanto à assertiva 15, aqueles que têm prática nas APF sabem que em muitos casos os valores dos contratos, quando aplicados às métricas, têm seus valores majorados e raramente diminuídos.

Dessa forma, como apenas duas assertivas ou variáveis apresentaram diferenças, as hipóteses formuladas (H1, H2 e H3) não podem ser confirmadas, e isto reforça a idéia de que as partes nos contratos ainda não têm posições muito definidas, talvez pelo fato de que a indústria de *software* ainda é uma novidade. Aliás, como apontaram Garmus e Herron (2000), não se sabe ao certo como contabilizar os custos dos produtos de computação personalizados.

Assim, este estudo apresenta algumas possibilidades de contribuições: a) a própria escala pode servir para avaliar as percepções de profissionais quanto às negociações de serviços de terceirização de contratos de sistemas; b) mostra que outros pontos, tais como os detectados, não se diferenciam das amostras, geram polêmicas e devem merecer melhor negociação antes das assinaturas dos contratos.

Ainda, as outras variáveis categóricas mostram que ainda há uma fronteira tênue e visivelmente confusa entre contratante e contratado. Isso mostra que os dois lados devem estar mais atentos ao que estão negociando e celebrando em um contrato. Uma possível explicação poderia ser dada em função da natureza dinâmica da área, posto que muitos profissionais, agora contratantes, já estiveram muito tempo como contratados, pois muitos profissionais de empresas que terceirizam estes serviços têm formação e experiência no desenvolvimento de *softwares* e, portanto, têm uma base de experiências profissionais ligadas à área técnica.

Por outro lado, a não diferenciação da variável "tempo de experiência na função" pode estar indicando os aspectos dinâmicos e a atualização constante dos profissionais, além das exigências que a área de alta tecnologia impõe aos mesmos, tornando-a *sui generis* e diferenciada de áreas mais cristalizadas, tal como nos contratos de fornecimento de autopeças para as montadoras de veículos.

Sugere-se que outros estudos sejam desenvolvidos nessa temática. Um deles pode ser a busca das causas das diferenças e semelhanças entre os aspectos que, neste estudo, se mostraram discriminantes e não discriminantes. Sugere-se, também, a realização de uma pesquisa com amostra maior, para poder desenvolver, por exemplo, um estudo causal entre os fatores, ou seja, a modelagem de equações estruturais ou a análise fatorial confirmatória.

## REFERÊNCIAS

- AAKER, D.; KUMAR, V. e DAY, G. 2001. *Pesquisa de marketing*. São Paulo, Atlas, 745 p.
- ADAMSON, I.L. e DILTS, D.M. 1995. Development of an accounting object model from accounting transactions. *Journal of Information Systems*, 1(9):43-64.
- ARMOUR, P.G. 2004. The business of software. *Communications of the ACM*. 47(1):21-27.
- AZEVEDO, D.J.P. 1997. FPA – Function Point Analysis – Sistemática de Métrica. *Bate Byte*, 68(1):11-19. Disponível em <http://www.pr.gov.br/celepar/celepar/batebyte/>. Acesso em 05/08/2006.
- BABBIE, E. 2001. *Método de pesquisas de Survey*. Belo Horizonte, Editora da UFMG, 519 p.
- BARDIN, L. 1979. *Análise de conteúdo*. Lisboa, Edições 70, 229 p.
- BOK, H.S. e RAMAN, K.S. 2000. Software engineering productivity measurement using function points: a case study. *Journal of Information Technology*, 4(15):79-90.
- COSTAGLIOLA, G.; FERRUCCI, F.; TORTORA, G. e VITIELLO, G. 2000. A metric for the size estimation of objects-oriented graphical user interfaces. *International Journal of Software Engineering and Knowledge Engineering*, 10(5):581-603.
- CHURCHILL Jr., G. 1999. *Marketing research: methodological foundations*. 7ª ed., New York, Thomson Publishing, 1017 p.
- CRONBACH, L. 1996. *Fundamentos da testagem psicológica*. 5ª ed., Porto Alegre, Artes Médicas, 575 p.
- DEKKERS, C. 2001. *Demystifying function points: clarifying common terminology*. Disponível em <http://www.qualityplustech.com>. Acesso em 05/01/2006.
- DILLON, W. e GOLDSTEIN, M. 1984. *Multivariate analysis: methods and applications*. New York, John Wiley & Sons, 587 p.
- FLECK Jr., R.A. 1998. Managing programmer resources in a maintenance environment with function points. *Industrial Management & Data Systems*, 9(2):63-70.
- FERNANDES, A. 1995. *A gerência de software através de métricas: garantindo a qualidade do projeto, processo e produto*. São Paulo, Atlas, 421 p.
- GARMUS, D. e HERRON, D. 2000. *Function point analysis: measurement practices for successful software projects*. Boston, Addison Wesley Professional, 385 p.
- GARMUS, D. e HERRON, D. 2001. *Function point analysis: measurement practices for successful software projects*. Boston, Addison Wesley Professional, 400 p.
- GREAVER, M. 1999. *Strategic outsourcing*. New York, Amacom, 314 p.
- HAIR, J.F.; TATHAM, R.L.; ANDERSON, R.E. e BLACK, W. 2005. *Análise multivariada de dados*. 5ª ed., Porto Alegre, Bookman, 593 p.
- HALSTEAD, M.H. 1972. Natural Laws Controlling Algorithm Structure. *ACM SIGPLAN Notices*, 7(2):19-26.
- HUMPHREY, W. 1997. *A discipline for software engineering*. 5ª ed., New York, Addison-Wesley Publishing Company, 816 p.
- IFPUG – THE INTERNATIONAL FUNCTION POINT USERS GROUP. 1999. *Function point counting practices manual*. Release 4.1. Westerville, IFPUG Standards, 321 p.
- ISBSG – INTERNATIONAL SOFTWARE BENCHMARKING STANDARD GROUP. ISO/14143:1998. 2003. *Standard for functional size measurement*. Disponível em <http://www.isbsg.org.au/html/index2.html>. Acesso em 29/01/2006.
- JONES, C. 2001. Conflict and litigation between software clients and developers. *Software Productivity Research*, 6(2):98-114.
- KAZMIER, L.J. 1982. *Estatística Aplicada a economia e administração*. São Paulo, McGraw-Hill, 376 p.
- MALHOTRA, N. 2006. *Pesquisa de marketing: uma orientação aplicada*. 4ª ed., Porto Alegre, Bookman, 720 p.
- MAROCO, J. 2005 *Análise estatística: com a utilização do SPSS*. 2ª ed., Lisboa, Ed. Silabo, 508 p.
- MICOLLIS, A. 2001. *Análise de Pontos de Função*. Rio de Janeiro, Integrated Solutions, 298 p.
- MUKHOPADHYAY, T.; VICINANZA, S.S. e PRIETULA, M.J. 1992. Examining the feasibility of a case-based reasoning for software effort estimation. *MIS/Quarterly*, 16(2):155-171.

- NESMA - THE NETHERLANDS SOFTWARE METRICS USERS ASSOCIATION. 2001. *Function point analysis for software enhancements*. Netherlands, versão 1.0, Disponível em <http://www.nesma.org>. Acesso em 10/01/2006.
- PASQUALI, L. 2003. *Psicometria: teoria dos testes na psicologia e na educação*. São Paulo, Ed. Vozes, 397 p.
- PESTANA, M.H. e GAGEIRO, J.N. 2000. *Análise de dados para ciências sociais: a complementaridade do SPSS*. 2ª ed., Lisboa, Edições Silabo, 570 p.
- PRANTONI, G. 2001. *Estimativa de Tamanho de Softwares Baseada em Objetos*. Rio de Janeiro, Infochoose, 67 p.
- PRESSMAN, R. 1995. *Engenharia de software*. São Paulo, Makron Books, 1095 p.
- SANTOS, F.; COSTA, M.; MICCOLIS, A. e SANTOS, J. 2001. Implantação da métrica análise de pontos de função segundo uma abordagem de aprendizagem organizacional. In: CONFERÊNCIA INTERNACIONAL DE TECNOLOGIA DE SOFTWARE, XII, Curitiba, 2001. *Anais...*, Curitiba, CITS (Centro Internacional de Tecnologia de Software), p. 1-13.
- SIEGEL, S. 1979. *Estatística não-paramétrica para ciências do comportamento*. São Paulo, McGraw-Hill, 350 p.
- SIMÕES, C. 1999. Sistemática de métricas, qualidade e produtividade. *Developer's Magazine*, 37(1):24-26.
- SPSS – Statistical Package for the Social Sciences. 1999. *Guide base 10.0*. Chicago, SPSS, 537 p.
- WRIGLEY, C. D. e DEXTER, A. S. 1991. A model for measuring information system size. *MIS/Quaternary*, 15(2):245-257.

Submissão: 24/03/2006

Aceite: 09/03/2007

#### **DIRCEU DA SILVA**

Doutor em Educação – FEA/USP. Professor do Programa de Mestrado em Controladoria e Contabilidade Estratégica da UNIFECAP.  
E-mail: [dirceuds@fecap.br](mailto:dirceuds@fecap.br)  
Avenida Liberdade, 532 CEP 01502-001 São Paulo – SP

#### **MAURO NEVES GARCIA**

Doutor em Administração – FGV-SP. Professor do Programa de Mestrado em Administração do IMES.  
E-mail: [mneves@gmail.com](mailto:mneves@gmail.com)  
Avenida Goiás, 3400, Barcelona CEP: 09550-051 São Caetano do Sul – SP

#### **HILDA MARIA DA ROCHA RINALDI**

Mestra em Administração – UNIFECAP. Professora do Centro Universitário Ibero-Americano – UNIBERO.  
E-mail: [hildamar@dialdata.com.br](mailto:hildamar@dialdata.com.br)  
Avenida Brigadeiro Luiz Antonio, 871 CEP 01317-001 São Paulo – SP

#### **CECÍLIA CARMEM CUNHA PONTES**

Doutora em Sociologia – USP. Professora do Programa de Mestrado em Controladoria e Contabilidade Estratégica da UNIFECAP.  
E-mail: [cpontes@fecap.br](mailto:cpontes@fecap.br)  
Avenida Liberdade, 532 CEP 01502-001 São Paulo – SP