

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE MATEMÁTICA
BACHARELADO EM ESTATÍSTICA

**AVALIAÇÃO DAS REGIONAIS DE UMA EMPRESA DE
TELECOMUNICAÇÕES, ATRAVÉS DE ANÁLISE DE *CLUSTER*.**

JÉSSICA VIVIANA MARTINS MENDES

Uberlândia-MG

2017

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE MATEMÁTICA
BACHARELADO EM ESTATÍSTICA

**AVALIAÇÃO DAS REGIONAIS DE UMA EMPRESA DE
TELECOMUNICAÇÕES, ATRAVÉS DE ANÁLISE DE *CLUSTER*.**

JÉSSICA VIVIANA MARTINS MENDES

Orientador: Prof. Dr. Lúcio Borges Araújo

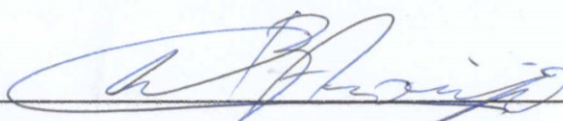
Trabalho de conclusão de curso apresentado ao curso de Estatística da Universidade Federal de Uberlândia (UFU) como parte dos requisitos para a conclusão do Curso de Bacharelado em Estatística.

Uberlândia-MG

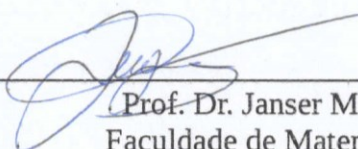
2017

JÉSSICA VIVIANA MARTINS MENDES

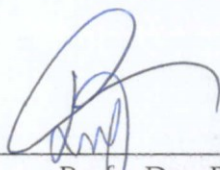
Monografia defendida em 21 de Julho de 2017 e aprovada pela seguinte Banca Examinadora:



Prof. Dr. Lúcio Borges de Araújo (ORIENTADOR)
Faculdade de Matemática – UFU



Prof. Dr. Janser Moura Pereira
Faculdade de Matemática – UFU



Profa. Dra. Priscila Neves Faria
Faculdade de Matemática – UFU

Dedico esse trabalho primeiramente
a Deus. Agradeço e ofereço esta
conquista aos meus pais e a minha
família, sem eles não conseguiria
chegar até aqui.

AGRADECIMENTOS

Primeiramente quero agradecer a Deus por ter me dado força, foco para nunca desistir do meu sonho e saúde para concluir mais uma etapa da minha longa caminhada ao topo do sucesso. Só Ele sabe o esforço e a luta que tive para chegar até aqui.

Ao meu amado pai, Márcio Mendes, pelo apoio e incentivo financeiro desde o colegial ao final da minha graduação. Agradeço imensamente a minha amada, amiga e querida mãe, Viviana Martins, por sempre estar presente em todos os momentos da minha graduação, por secar minhas lágrimas nos momentos difíceis e por sorrir comigo por pequenas conquistas ao longo dos meus seis anos e meio de graduação. Obrigada meus pais por sempre acreditarem que eu seria capaz. Essa conquista é nossa. Ao meu irmão Agner Márcio pelas suas sábias palavras de apoio. A toda minha família, por sempre me apoiarem. Em especial minha avó Maria Joana, até mesmo depois de ficar doente não deixou de me apoiar, e de me dar seu colo nos dias tristes, sorrir comigo nos dias felizes e de dizer palavras de apoio mesmo não conseguindo falar perfeitamente.

Ao meu namorado Vinicius Freitas, que a todo o momento sempre esteve ao meu lado e entendeu minha ausência.

Ao meu orientador, Lúcio Borges de Araújo, é uma alegria concluir o curso sob sua orientação. Seu conhecimento, carisma e humildade me inspiram. Obrigada por suas sábias palavras.

Aos servidores em geral, aos técnicos e a todos os professores a quem tive a honra de ser aluna durante a graduação, vocês foram muito importantes para o meu desenvolvimento e me ensinaram muito sobre Estatística e sobre a vida acadêmica. Em especial, ao Janser Moura Pereira, José Waldemar da Silva, Lúcio Borges de Araújo, Maria Imaculada, Rogério De Melo Costa Pinto e Tiago Vargas. Agradeço em especial minha querida e inesquecível professora Priscila Neves Faria, que com o seu jeito especial de nos ensinar me encantou tanto que me fez acreditar que a Estatística é a profissão dos meus sonhos.

Aos meus amigos Mariana Cruvinel, Mateus Carvalho, Wellington Paiva e Yuri Cainã pelo incentivo e momentos de descontração ao decorrer do curso. Aos meus amigos e colegas que fiz no curso, pela convivência, aprendizado e momentos inesquecíveis compartilhados. Em especial a Bruna Queiroz, Carolina Juvelina, Gabriela Bolaina, Heitor B. de Melo, Heverton Rodrigues, Laíla Campos, Luana Bahía, Luiz Carlos, Paloma Larissa, Patrícia Lorena e Vitor Hugo.

RESUMO

A telecomunicação é uma técnica que consiste na transmissão de mensagens de um ponto para o outro. No Brasil, temos diversas empresas que prestam serviços de Telecomunicações sendo capazes de oferecer diferentes planos de adesão customizando as características de uso de cada cliente. Para que uma empresa mantenha o seu cliente ativo, ela precisa oferecer um nível mínimo de qualidade nos serviços e produtos oferecidos preservando ou até mesmo aumentando a satisfação do cliente. Este estudo teve por objetivo utilizar a Análise *Cluster* para avaliar o desempenho das regionais de uma empresa de telecomunicações, nos anos 2013, 2014 e 2016, em relação a sete índices de satisfação do consumidor. O presente estudo permitiu concluir que a Análise de *Cluster* mostrou eficaz na identificação das Regionais que melhoraram seu desempenho em relação aos percentuais dos índices de satisfação do consumidor, as que mantiveram e as que pioraram seu desempenho.

Palavras Chaves: Análise multivariada, *Cluster*, índice de satisfação.

SUMÁRIO

1. INTRODUÇÃO	8
2. REFERENCIAL TEÓRICO.....	10
2.1 ANÁLISE DE <i>CLUSTERS</i>	10
2.2.2 Métodos Hierárquicos Aglomerativos	12
2.2 ANOVA.....	16
2.3 TESTE TUKEY.....	17
3. METODOLOGIA	18
4. RESULTADOS.....	18
5. CONCLUSÃO	28
6. REFERÊNCIAS	29

1. INTRODUÇÃO

Atualmente, existem várias formas das pessoas se comunicarem, sejam por telefone, rádio, jornais, revistas, televisão, internet entre outros. Os mais utilizados são as formas de comunicação à distância (telefone, internet, mídia).

A telecomunicação é uma técnica que consiste na transmissão de mensagens de um ponto para o outro. Ela constitui um ramo da engenharia elétrica que contempla o projeto, a implantação e a manutenção dos sistemas de comunicações. A principal finalidade das telecomunicações é suprir a necessidade humana de se comunicar à distância. É comum o prefixo “tele” ser omitido e, com isto, usar-se a palavra comunicações (CONCEITO DE TELECOMUNICAÇÃO, 2016).

No Brasil, existem diversas empresas que prestam serviços de Telecomunicações sendo capazes de oferecer diferentes planos de adesão customizando as características de uso de cada cliente. Para clientes pessoa física oferecem serviços como: Internet Banda Larga (ADSL, VDSL e 3G), celular, TV por assinatura (cabo e via satélite) e telefonia física; para clientes micro e pequenas empresas oferecem: soluções de voz, dados, internet, TI, *outsourcing*, videoconferência e mídia de consulta; para clientes de médias e grandes empresas Soluções de voz e dados personalizados para o segmento; para clientes de outras operadoras: *Clear channel*, SMS, IP Trânsito, compartilhamento de infraestrutura.

Para que uma empresa mantenha o seu cliente ativo, ela precisa oferecer um nível mínimo de qualidade nos serviços e produtos oferecidos preservando ou até mesmo aumentando a satisfação do cliente.

Para Kotler (2007) a satisfação pode resultar num cliente encantado e altamente satisfeito. Sendo assim, a retenção dos clientes torna-se fundamental para o sucesso da empresa. O cliente pode experimentar vários graus de satisfação, se o desempenho do produto ficar abaixo de suas expectativas, ele fica insatisfeito. Se o desempenho ficar à altura de suas expectativas, fica satisfeito. Se o desempenho exceder as expectativas, fica extremamente satisfeito ou encantado (KOTLER e ARMSTRONG, 2006).

A Satisfação do Cliente depende muito da qualidade dos serviços que estão consumindo e a qualidade que eles esperam dos serviços. Dessa forma, a percepção do cliente com relação à qualidade dos serviços recebidos é proporcional às suas expectativas sobre os mesmos (PAULINS, 2005).

O trabalho na área da satisfação do cliente começou no ano de 1970, devido ao aumento do consumo de produtos e serviços. A redução da qualidade dos serviços prestados e o aumento da inflação, também colaboraram e obrigaram várias empresas a diminuir os preços dos seus produtos. Com isso, ocorreu o surgimento da insatisfação dos clientes (HOFFMAN e BATESON, 2003).

Hoje, a grande maioria das empresas tem o compromisso de identificar quais são as necessidades de cada grupo de consumidores, sendo adolescentes, jovens, adultos e idosos mediante o que a empresa oferece, pois, houve um declínio na qualidade devido à diminuição no preço dos produtos oferecidos nas empresas; então as insatisfações dos consumidores surgiram diretamente ligadas à péssima qualidade dos produtos, fazendo com que aumentasse a procura por produtos de melhor qualidade.

Examinando cuidadosamente os comportamentos dos clientes fiéis, a empresa terá as informações que necessita para permitir um melhor gerenciamento dos dados estatísticos presentes e uma projeção para o futuro. Sendo assim, a empresa irá monitorar e equilibrar as transações com seus clientes no momento da negociação, evitando a insatisfação e promovendo a satisfação, de modo que consiga novos clientes e a sua retenção.

As empresas seguem as normas estabelecidas pela ANATEL (Agência Nacional de Telecomunicações). A ANATEL tem como missão regular o setor de telecomunicações para contribuir com o desenvolvimento do Brasil e em 2015 a ANATEL definiu o Regulamento das Condições de Aferição do grau de satisfação e da qualidade percebida junto aos usuários de serviços de telecomunicações.

Diante do exposto, decidiu-se analisar o desempenho de satisfação dos clientes em relação aos produtos e serviços de uma empresa de telecomunicações em suas principais Regionais, com intuito de evidenciar onde a empresa pode melhorar para preservar ou até mesmo aumentar o número de clientes satisfeitos com os serviços e produtos oferecidos. Os produtos analisados serão os de telefonia fixa, telefonia móvel, banda larga e TV por assinatura para o segmento Residencial. Portanto, utilizou-se a Análise de *Clusters* para avaliar o desempenho das regionais de uma empresa de telecomunicações, nos anos 2013, 2014 e 2016, em relação a sete índices de satisfação do consumidor.

2. REFERENCIAL TEÓRICO

2.1 ANÁLISE DE *CLUSTERS*

A Análise de *Cluster* também conhecida por Análise de Agrupamento e de acordo com Barroso e Arte (2003) é o conjunto de técnicas utilizadas na identificação de padrões de comportamento em banco de dados através da formação de grupos homogêneos de casos. Para Khattree e Naik (2000), a Análise de *Cluster* é uma técnica multivariada de grande aplicabilidade, do qual o objetivo da classificação é repartir os indivíduos em grupos homogêneos, de modo que cada um seja bem diferenciado. Após a obtenção dos resultados, esses dados servirão para a definição do número de grupos distintos.

Segundo Mardia, Kent e Bibby (1995), a Análise de *Clustes* apresenta a vantagem de minimizar o espaço multidimensional a uma medida de distância entre os objetos, representando esta em um espaço bidimensional, muito mais simplificado do que o espaço multidimensional.

Para Barroso e Artes (2003), temos cinco etapas para a aplicação de uma Análise de Clusters:

- I. Escolha do critério de parecença (proximidades): consiste em definir se as variáveis devem ou não ser padronizadas se o critério que será utilizado na determinação dos grupos; no caso, proximidade dos pontos.
- II. Definição do número de grupos: o número de grupos pode ser definido a priori, através de algum conhecimento que se tenha sobre os dados, conveniência de análise ou ainda pode ser definido a posteriori com base nos resultados da análise.
- III. Formação dos grupos: Nesta etapa é definido o algoritmo que será utilizado na identificação dos grupos.
- IV. Validação do agrupamento: Deve-se garantir que de fato as variáveis têm comportamento diferenciado nos diversos grupos. Nesta etapa, é comum supor que cada grupo seja da amostra aleatória de alguma subpopulação e aplicar técnicas inferenciais para compará-las.

- V. Interpretação dos grupos: Ao final do processo de formação dos grupos é importante caracterizar os grupos formados. O uso de estatísticas descritivas é recomendado para esta fase da análise.

As medidas de proximidades têm um papel central nos algoritmos de agrupamentos. Através delas são definidos critérios para avaliar se dois pontos estão próximos, e, portanto podem fazer parte de um mesmo grupo ou não. (BARROSO e ARTES, 2003).

As medidas de proximidade são divididas em dois grupos: medida de similaridade (quanto maior o valor, maior a semelhança entre os objetos) e medida de dissimilaridade (quanto maior o valor, mais diferentes são os objetos).

Para as variáveis quantitativas, as medidas de proximidade entre os objetos são normalmente medidas de distância ou dissemelhança. Atualmente, diversas medidas de dissimilaridade são propostas na literatura, principalmente devido ao grande desenvolvimento e utilização das técnicas multivariadas (KHATTREE e NAIK, 2000).

As medidas de dissimilaridade mais conhecidas são a distância Euclidiana e a distância de Mahalanobis devido a sua maior utilização, segundo Khattree e Naik (2000) e Cruz e Carneiro (2006).

Segundo Barroso e Arte (2003), a distância euclidiana entre os indivíduos i e k é dada por:

$$d_{ik} = \sqrt{\sum_{j=1}^p (X_{ij} - X_{kj})^2}$$

ou seja, a distância Euclidiana entre dois casos (i e k) é a raiz quadrada do somatório dos quadrados das diferenças entre os valores i e j para todas as variáveis ($j = 1, 2, \dots, p$), em que X_{ij} representa a característica do indivíduo i para a variável j ; X_{kj} representa a característica do indivíduo k para a variável j e p representa o número de variáveis na amostra.

A distância de Mahalanobis é outra medida de distância bem conhecida. Quintal (2006) diz que além de reduzir a dependência das unidades de medição, reduz também a correlação entre variáveis. A distância de Mahalanobis entre os grupos i e j é usualmente estimada segundo Rao (1952) por:

$$D_{ij}^2 = (\bar{X}_i - \bar{X}_j)' \Sigma^{-1} (\bar{X}_i - \bar{X}_j)$$

em que $\bar{X}_i$ é o vetor de médias, do i – ésimo grupo; $\bar{X}_j$ é o vetor de médias do j – ésimo Grupo; Σ é a estimativa da matriz de variância–covariância entre as variáveis.

A distância de Mahalanobis considera a variabilidade dentro de cada unidade amostral, e não somente a medida de tendência central, assim sendo, uma medida mais aceitável quando as unidades amostrais constituem um conjunto de indivíduos e, principalmente, quando as variáveis são correlacionadas (RIBOLDI, 1986).

2.2.2 Métodos Hierárquicos Aglomerativos

De acordo com Mingoti (2005), uma análise mais elaborada de cada conglomerado é mais informativa do que uma análise do conjunto de n observações como um todo, mas nada impede que outras técnicas da estatística multivariada sejam utilizadas na sumarização da informação de cada conglomerado obtido na partição. É possível utilizar todas as n observações amostrais, mesmo se incompletas, ao contrário das outras técnicas de estatística multivariada em análise de *Clusters*. Os métodos de agrupamentos são sensíveis à presença de valores discrepantes (outliers), sendo fundamental que uma análise exploratória dos dados seja feita antes para verificar a existência de observações desse tipo.

As técnicas de agrupamentos são classificadas em hierárquicas (aglomerativos e divisivas) e não hierárquicas. Ferreira (2011), afirma que inúmeros métodos de agrupamentos já foram propostos na literatura. Entre eles os hierárquicos, no qual os objetos são classificados em grupos em diferentes etapas, de modo hierárquico, produzindo uma árvore de classificação, os dendrogramas (BUSSAB et al.,1990) , são os mais importantes.

Na maioria das vezes os Métodos Hierárquicos são utilizados em análise exploratória dos dados com o intuito de identificar os possíveis agrupamentos e o valor provável do número deles. No uso dos Métodos Não Hierárquicos, é necessário que o valor do número de grupos seja pré-especificado (MINGOTI, 2005).

De acordo com Mingoti (2005), as técnicas hierárquicas aglomerativas partem do princípio de que no início do processo de agrupamento cada elemento do conjunto de dados observado é considerado como um conglomerado isolado, assim, tem-se n conglomerados. Em cada estágio do algoritmo, os elementos mais similares vão sendo

combinados e passam a constituir um único agrupamento. Cada novo cluster formado é um agrupamento de conglomerados formados nos estágios anteriores, até que todos os elementos considerados pertençam a um único grupo, diminuindo o número de conglomerados a cada estágio do processo. Já nos métodos divisivos, todos os objetos pertencem inicialmente ao mesmo grupo, que vai sendo dividido, até que cada observação forme um grupo individualmente (JOHNSON e WICHERN, 1992).

De acordo com Mingoti (2005), é possível construir um gráfico chamado Dendrograma que tem como objetivo apresentar o arranjo entre os objetos em uma escala de distância. O dendrograma é um gráfico em forma de árvore no qual a escala vertical indica o nível de similaridade (ou dissimilaridade). No eixo horizontal, são marcados os elementos amostrais numa ordem conveniente relacionada à história de agrupamento. As linhas verticais, partindo dos elementos amostrais agrupados, têm altura correspondente ao nível em que os elementos foram considerados semelhantes, isto é, a distância do agrupamento ou o nível de similaridade.

Na Figura 1 é ilustrado um exemplo de dendrograma, em que uma hierarquia significativa está indicada na figura.

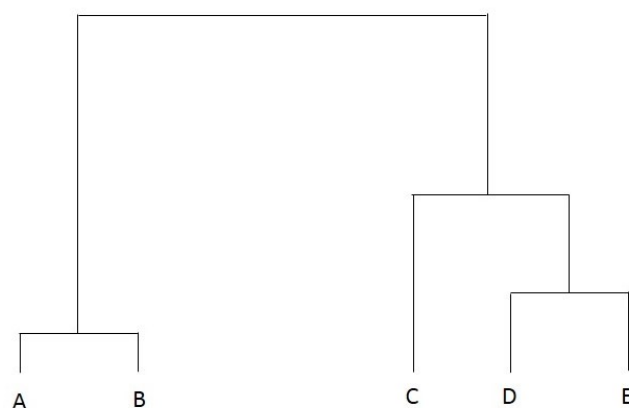


Figura 1: Exemplo de dendrograma de agrupamento de 5 indivíduos.

Os métodos de agrupamentos mais utilizados são os hierárquicos aglomerativos. De acordo com Barroso e Artes (2003) dentre os métodos hierárquicos aglomerativos mais utilizados estão: Ligação simples (conhecido também como Critério do Vizinho mais Próximo); Ligação Completa (conhecido também como Critério do Vizinho mais Distante); Ligação Média (ou UPGMA); Método do Centróide (ou UPGMC) e Critério de Ward.

No método de Ligação Simples a similaridade entre dois conglomerados é definida pelos dois elementos parecidos entre si, ou seja, a distância é a menor distância entre uma observação k e um grupo formado pelas observações i e j dada por:

$$D_{(ij)k}^2 = \min(D_{ik}^2; D_{jk}^2)$$

em que $D_{(ij)k}^2$ é a distância entre o grupo ij e a observação k ; e $\min(D_{ik}^2; D_{jk}^2)$ é a menor distância entre os grupos de observações ik e jk . (AMARAL JÚNIOR E THIÉBAUT, 1999).

Algumas características desse método são: grupos muito próximos podem não ser identificados; permite detectar grupos de formas não elípticas; apresenta pouca tolerância a ruído uma vez que tem tendência a inserir os ruídos em um grupo já existente; apresenta bons resultados tanto para distâncias Euclidianas quanto para outras distâncias; e tem tendência a formar longas cadeias. (Anderberg, 1973).

Já o Método de Ligação Completa é exatamente o oposto do Método de Ligação Simples em que no primeiro passo considera-se a distância entre dois grupos como sendo entre os objetos de maior distância, estes definindo grupos centralizados. De acordo com Quintal (2006), o método de ligação completa tende a formar grupos pequenos que depois serão aglutinados para formar grupos maiores, ou seja, quando um objeto é acrescentado a um grupo, a distância do novo grupo aos restantes aumenta ou então fica inalterada, tendo tendência para encontrar clusters compactos compostos de objetos semelhantes entre si.

O método de Ligação Média é um método não ponderado de agrupamento aos pares, utilizando médias aritméticas das medidas de dissimilaridade, que evita caracterizar a dissimilaridade por valores extremos (máximo ou mínimo). De acordo com Mingoti (2005), a distância entre dois grupos é representada pela média da distância entre todos os pares de objetos pertencentes a cada grupo. Desta maneira, se o conglomerado C_1 tem n_1 elementos e o conglomerado C_2 tem n_2 elementos, a distância entre eles será definida por:

$$d(C_1, C_2) = \sum_{l \in C_1} \sum_{k \in C_2} \left(\frac{1}{n_1 n_2} \right) d(X_l, X_k)$$

A maior vantagem do método de Ligação Média é contornar as consequências da existência de valores extremos e considerar toda a informação dos grupos, mas em

outras situações pode ser vista como desvantagem, dependendo do tipo de *clusters* que quiser obter, pois não irá fazer a separação dos elementos.

O Método de Centróide é definido pela coordenada de cada grupo como sendo a média das coordenadas de seus objetos. Uma vez obtida essa coordenada, denominada de centróide, a distância entre os grupos é obtida através do cálculo das distâncias entre os centróides (BARROSO e ARTES, 2003). Segundo Mingoti (2005), em cada etapa do agrupamento é necessário voltar aos dados originais para o cálculo da matriz de distâncias, agrupando os conglomerados que apresentam o menor valor de distância.

O método de Ward é baseado na mudança de variação entre os grupos e dentro dos grupos que estão sendo formados em cada passo do agrupamento. É também conhecido como Mínima Variância, e baseia-se nos princípios de que inicialmente, cada elemento é considerado como um único conglomerado, e em cada passo do algoritmo de agrupamento é calculado a soma de quadrados dentro de cada conglomerado. Esta soma é o quadrado da distância Euclidiana de cada elemento amostral pertencente ao conglomerado em relação ao correspondente vetor de médias do conglomerado. De acordo com Barroso e Artes (2003) o método de Ward é um método atraente por basear-se numa medida com forte apelo estatístico e por gerar grupos que, assim possuem uma alta homogeneidade interna.

Existe uma semelhança entre o Método de Ward e o Método do Centróide no que se refere ao uso dos vetores de médias amostrais como representante da informação global dos conglomerados em cada passo do processo de agrupamento. No entanto, a distância usada no Método de Ward leva em consideração a diferença dos tamanhos dos conglomerados que estão sendo comparados, enquanto que o Método dos Centróides não possui qualquer fator de ponderação em relação a isso (MINGOTI, 2005).

O método de k-Médias é utilizado após a estipulação *a priori* do número de grupos a serem gerados (BARROSO e ARTES, 2003). É provavelmente um dos métodos mais conhecidos e mais utilizados em problemas práticos. Basicamente, cada elemento amostral é alocado àquele cluster cujo centróide (vetor de médias amostral) é o mais próximo do vetor de valores observados para o respectivo elemento. Originalmente, o método é composto por quatro passos:

- I. Primeiramente escolhe-se k centróides para dar início ao processo de partição .

- II. No segundo passo, cada elemento do conjunto de dados é, então, comparado com cada centroide inicial, através de uma medida de distância. O elemento é alocado ao grupo cuja distância é menor.
- III. No terceiro passo, para cada um dos n elementos amostrais calculam-se os valores dos centroides para cada novo grupo formado, e repete-se o segundo passo, considerando os centroides destes novos grupos.
- IV. Os passos dois e três devem ser repetidos até que todos os elementos estejam "bem alocados" em seus grupos, isto é, nenhuma realocação seja necessária.

2.2 ANOVA

A Análise de Variância (*Analysis of Variance* – ANOVA) representa um teste utilizado para comparar médias em mais de duas populações. Martins (2006) definiu a ANOVA sendo um método estatístico, que, por meio de teste de igualdade de médias, verifica se fatores (variáveis independentes) produzem mudanças sistemáticas em alguma variável de interesse (variável dependente). Os fatores propostos podem ser variáveis quantitativas ou qualitativas, enquanto a variável dependente deve ser quantitativa e observada dentro da classe dos fatores, ou seja, os tratamentos. Existem algumas considerações importantes, pressupostos, a serem feitas antes de aplicar a ANOVA:

- I. As amostras analisadas devem ser escolhidas de maneira aleatória e independentes
- II. As populações têm distribuição normal
- III. Homogeneidade das variâncias entre os grupos.

As hipóteses da ANOVA são:

H_0 : As médias dos k tratamentos são iguais;

H_1 : pelo menos duas das médias dos k tratamentos é diferente das demais.

Quando os resultados da Análise de Variância indica a rejeição de H_0 , pode concluir que o fator considerado tem influência sobre a variável estudada.

De acordo com Downing e Clark (2000), o teste ANOVA é aplicado para testar se vários grupos de observações provêm de distribuições com a mesma média; na análise de variância calcula-se uma estatística F ; se a hipótese nula é verdadeira e as médias de todos os grupos são realmente iguais, então a estatística F terá distribuição F

e; se a estatística F calculada é superior ao valor crítico, então se rejeita a hipótese nula, de que as médias dos grupos sejam as mesmas.

2.3 TESTE TUKEY

À medida que é feita a análise de variância de um experimento com apenas dois tratamentos, podemos visualizar apenas pela média qual o melhor tratamento. Porém, quando há mais de dois tratamentos, fazendo apenas o teste de F não podemos indicar qual o melhor tratamento. Sendo assim, é necessário aplicar um teste de comparação múltiplas entre pares de médias dos tratamentos, para concluir qual o melhor tratamento.

O teste de Tukey é um dos testes de comparação de médias mais utilizados, pelo motivo de ser bastante rigoroso e fácil aplicação; não permite comparar grupos de tratamentos entre si; é utilizado para testar toda e qualquer diferença entre duas médias de tratamento.

Tukey (1953) se preocupava com o erro tipo I. Ele desenvolveu um método onde o experimento é usado como unidade no estabelecimento do nível de significância, ou seja, para um conjunto de médias verdadeiramente iguais, um teste ao nível de 5% é aquele em que 5% dos experimentos darão uma ou mais diferenças significativas e 95% darão diferenças não significativas, em média.

O Teste Tukey tem como base a DMS (diferença mínima significativa) representada geralmente por Δ e calculada da seguinte forma:

$$\Delta(\alpha) = q(\alpha) \sqrt{\frac{QMR}{r}}$$

onde $q(\alpha)$ é o valor tabelado por Tukey em função do número de tratamento e dos graus de liberdade do resíduo; α o valor dado na tabela ao nível de significância estabelecido, QMR é o quadrado médio do resíduo da análise de variância e r é o número de repetições de cada um dos tratamentos. De acordo com o teste, duas médias são estatisticamente diferentes toda vez que o valor absoluto da diferença entre elas for igual ou maior que a d.m.s. (VIEIRA et. al., 1989).

3. METODOLOGIA

Para esta pesquisa tomou-se como base o banco de dados fornecido por uma empresa de Telecomunicações no segmento residencial, com o objetivo de identificar pontualmente aspectos na qual a empresa está sendo bem ou mal sucedida, mas principalmente, indicar onde a empresa deve intervir e realizar mais investimentos e/ou melhorias em relação aos anos de 2013, 2014 e 2016 das Regionais de Uberlândia, Uberaba, Franca, Ituiutaba, Itumbiara, Patos de Minas, Pará de Minas e Divinópolis.

O banco de dados tem como variáveis sete perguntas: Imagem da Empresa (P1); Esforço em Melhorias (P2); Cumprimentos de Promessas (P3); Eficiência dos Serviços (P4); Retorno as Solicitações (P5); Relatório Custo versus Benefício (P6) e Agilidade no Atendimento (P7). O banco de dados em 2013, 2014 e 2016 contém respectivamente 1.248, 1.131 e 1.456 clientes.

Primeiramente realizou-se um estudo descritivo utilizando estatísticas básicas: média, desvio padrão, mínimo, máximo e coeficiente de variação. Após a Análise Descritiva dos dados, foi aplicada a técnica de estatística multivariada, Análise de Clusters, com o objetivo de analisar a similaridade da combinação do ano e Regionais e posteriormente identificar quais Regionais melhoraram seu desempenho em relação aos percentuais dos índices de satisfação dos clientes, as que mantiveram e as que pioraram seu desempenho. Para a obtenção dos resultados, os procedimentos estatísticos foram realizados no software XLSTAT um suplemento do Excel e no software R.

Como os dados já estão padronizados a melhor matriz de dissimilaridade entre as variáveis foi a distância euclidiana e para a formação dos grupos foram aplicados os seguintes métodos com a distância euclidiana: Ligação Simples, Ligação Completa e Critério de Ward.

Para identificar quais Regionais melhoraram seu desempenho em relação aos índices de satisfação dos clientes, as que mantiveram e as que pioraram seu desempenho, é necessário o uso da ANOVA. Como uma técnica confirmatória em relação aos resultados obtidos da análise multivariada, ou seja, verificando-se se existe diferença entre as médias das variáveis entre os grupos formados (VICINI, 2005).

4. RESULTADOS

Analisando os resultados obtidos, foi observado que na Tabela 1 a variável Eficiência do Serviço (P4) apresentou os maiores valores de desvio padrão nos anos de 2013 e 2014 e em 2016 o terceiro maior. Comparando os anos de 2013 e 2016, observa-se que o ano de 2016 apresentou desvios padrões entre 4 e 7 e em 2013 apresentou entre 6 e 10. Isto indica que em 2016 ocorreram respostas mais homogêneas quando se compara ao ano de 2013. Além disso, em relação a média, observa-se que em 2013 todas as Regionais tiveram índices de satisfações altos, em 2014 o desempenho caiu em relação a todas as perguntas e em 2016 o desempenho aumentou (Figura 2).

Tabela 1: Análise Descritiva do Percentual dos Índices de Satisfação de todas as Regionais.

Ano	Variável	Mínimo	Máximo	Média	Desvio padrão
2013	P1	67,00	90,00	78,88	7,97
	P2	56,00	81,00	67,00	7,86
	P3	47,00	75,00	61,75	8,48
	P4	60,00	84,00	71,75	9,08
	P5	58,00	80,00	68,38	6,80
	P6	55,00	80,00	68,50	7,71
	P7	52,00	76,00	65,13	8,01
2014	P1	64,00	81,00	74,75	6,30
	P2	52,00	66,00	59,88	4,64
	P3	55,00	64,00	58,38	3,46
	P4	58,00	78,00	65,63	6,80
	P5	60,00	71,00	67,50	6,30
	P6	57,00	71,00	63,75	4,83
	P7	57,00	78,00	64,75	6,25
2016	P1	73,00	88,00	80,25	5,87
	P2	57,00	77,00	66,50	6,99
	P3	61,00	75,00	66,13	4,19
	P4	65,00	84,00	73,38	5,78
	P5	60,00	76,00	71,29	5,94
	P6	60,00	76,00	66,38	5,63
	P7	65,00	83,00	72,75	5,34

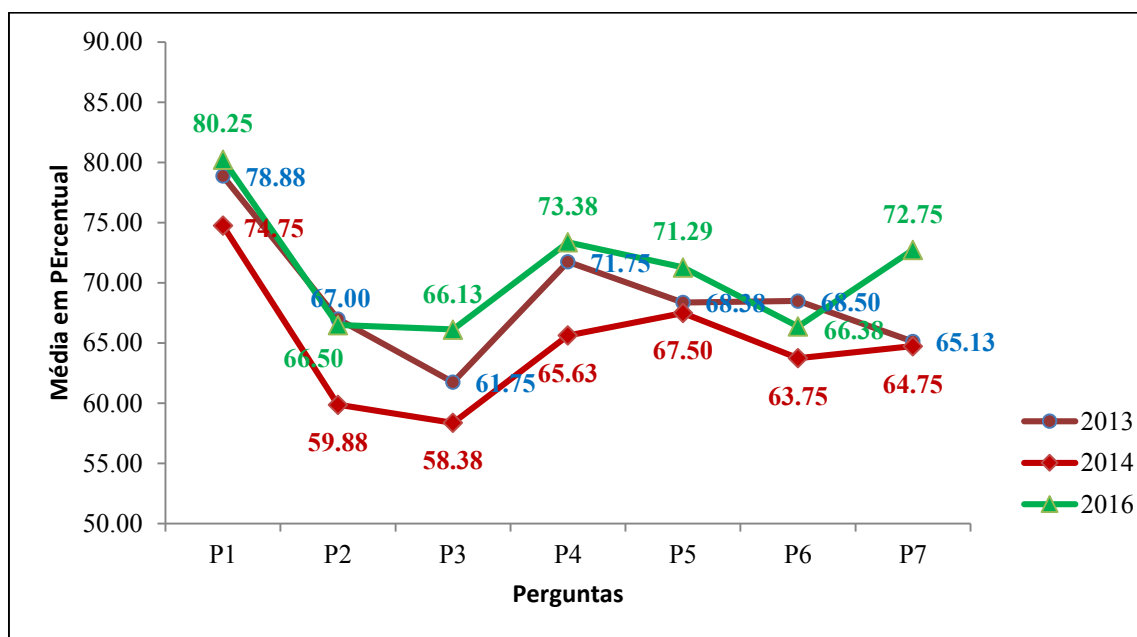
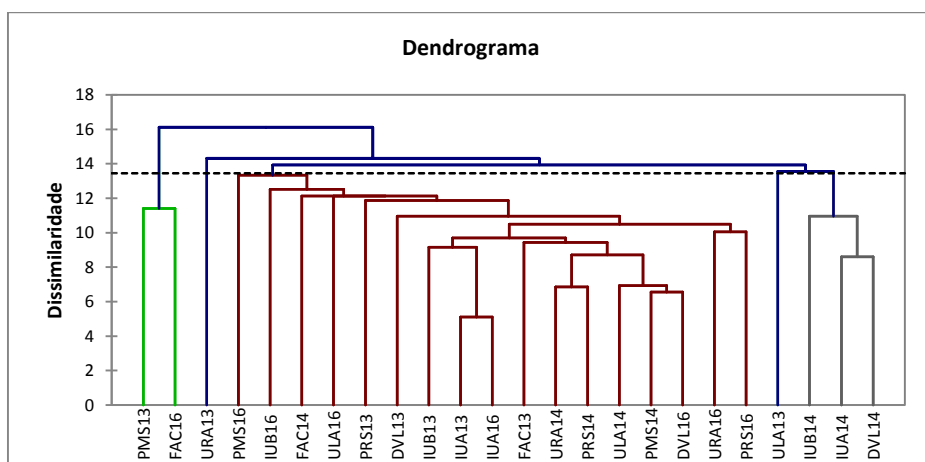


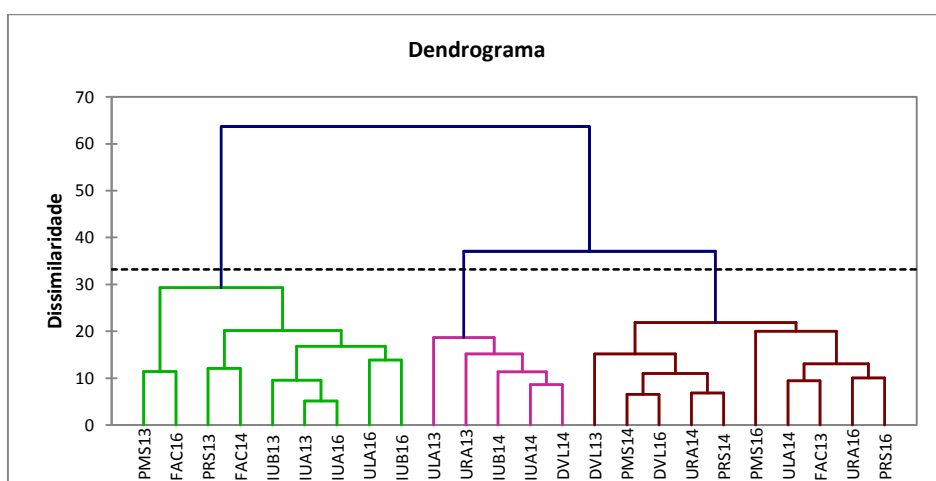
Figura 2: Gráfico das médias percentuais em relação às perguntas dos anos de 2013, 2014 e 2016.

Os resultados dos agrupamentos por cada método estão resumidos em dendrogramas na Figura 3. O critério de agrupamento escolhido foi o método de Ward obtida pela matriz de Distância Euclidiana (Figura 3c), pois esta mostra uma melhor distinção dos grupos. O critério resultou em três grupos, sendo que os grupos retratados relacionaram as Regionais em cada ano.

a)



b)



c)

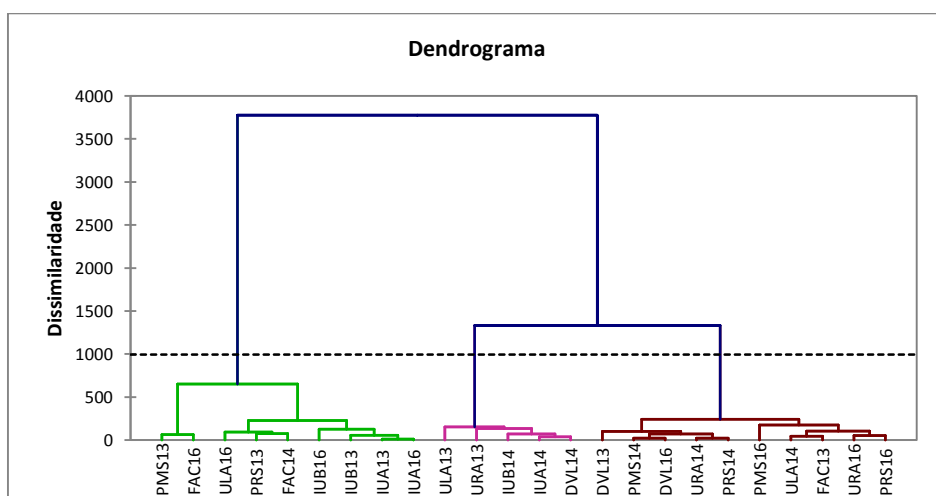


Figura 3: Dendrogramas obtidos pelos métodos de agrupamento: a) Ligação Simples; b) Ligação Completa; c) Critério de Ward.

Os três grupos formados pelo método de K-medias, são retratados na Tabela 2, em que é possível observar que somente uma das oito regionais conseguiu manter os índices dos três anos em um único grupo, a Regional de Divinópolis (DVL). As Regionais de Uberlândia (ULA), Itumbiara (IUB), Pará de Minas (PRS) e Patos de Minas (PMS) foram as que mantiveram seus índices nos diferentes grupos. Entre estas quatro Regionais, Itumbiara, Pará de Minas e Patos de Minas tiveram os anos de 2014, 2013 e 2016 alocados nos grupos 1, 2 e 3, respectivamente, e Uberlândia manteve os anos de 2013, 2014 e 2016 nos grupos 1, 2 e 3, respectivamente. A Regional de Uberaba (URA) manteve os anos 2013 e 2014 no grupo 1 e o ano de 2016 no grupo 3. Já a Regional de Ituiutaba (IUA) manteve anos 2013 e 2016 no grupo 2 e o ano 2014 no grupo 1. A Regional de Franca (FAC) manteve dois dos três anos (2014 e 2016) no grupo 3 e o índice de 2013 no grupo 2.

Tabela 2: Grupos obtidos por meio da Análise de Clusters, pelo K-Medias.

Grupo 1	Grupo 2	Grupo 3
ULA13	ULA14	ULA16
URA13		URA16
URA14		
IUA14	IUA13	
	IUA16	
IUB14	IUB13	IUB16
PMS14	PMS13	PMS16
PRS14	PRS13	PRS16
	FAC13	FAC14
		FAC16
DVL13		
DVL14		
DVL16		

Tendo em vista que a satisfação é um estado mental e emocional transitório, resultante de uma experiência única ou acumulada com um serviço, consumo ou utilização de um produto, onde um cliente satisfeito está convencido de que a aquisição de determinado produto/serviço foi algo positivo, o que facilita a hipótese de uma relação de continuidade entre o cliente e a empresa. Ou seja, a satisfação é posterior à compra: é impossível dar o seu nível de satisfação em relação a um produto ou um

serviço se o cliente não o utilizar consumir ou experimentar. A avaliação pode assentar sobre o conjunto das experiências de consumo ou sobre partes deste processo, como o feito de compra/aquisição propriamente dito, o consumo, a utilização do produto ou o serviço, também conhecida por abordagem analítica (SOUSA, 2011).

Para interpretar os grupos formados é necessário analisar os grupos relacionando as sete perguntas que resultaram os índices nos três anos. Na Tabela 3 observa-se que as Regionais pertencentes ao primeiro, segundo e terceiro grupo tiveram melhor média percentual de satisfação na pergunta 1 – Imagem da Empresa. Ou seja, a imagem da empresa é essencial para fazerem uso dos produtos. Em relação às piores médias percentuais de satisfação dos grupos foram as perguntas Cumprimento de Promessas (P3) para o grupo 1 e 2 e Relatório Custo versus Benefício para o grupo 3.

Um coeficiente de variação (CV) é considerado baixo (indicando um conjunto de dados razoavelmente homogêneo) quando for menor ou igual a 25%. Entretanto, esse padrão varia de acordo com a aplicação. O coeficiente de variação é interpretado como a variabilidade dos dados em relação à média. Quanto menor o CV mais homogêneo é o conjunto de dados (Martins, 2006). Os grupos apresentaram coeficientes de variações abaixo de 25%, considerado um valor baixo. Isto é explicado pela não discrepância entre os índices máximos e mínimos encontrados para cada pergunta.

Tabela 3: Estatística Descritiva das Perguntas em cada Grupo

Perguntas	Grupos	Média	Desvio Padrão	Mínimo	Máximo	Coeficiente de Variação
1	1	72,0	4,6	64	78	6,36%
	2	83,6	3,3	80	90	3,99%
	3	80,9	5,0	74	88	6,15%
2	1	59,2	3,8	52	65	6,36%
	2	69,1	5,9	61	81	8,52%
	3	67,3	6,6	57	77	9,85%
3	1	56,8	4,5	47	63	7,87%
	2	65,3	4,7	59	75	7,17%
	3	66,4	4,1	61	75	6,12%
4	1	63,0	3,6	58	69	5,77%
	2	75,3	5,7	69	84	7,50%
	3	75,6	4,6	69	84	6,04%

5	1	63,8	3,8	58	71	5,90%
	2	72,7	2,4	70	77	3,35%
	3	72,3	6,0	60	80	8,26%
6	1	62,5	4,7	55	70	7,48%
	2	71,4	4,4	66	80	6,12%
	3	66,3	5,5	60	76	8,29%
7	1	61,8	5,3	52	69	8,62%
	2	68,4	4,2	63	76	6,09%
	3	74,9	4,4	70	83	5,82%

Na Tabela 4 estão dispostas as médias de cada um dos três grupos e a significância de cada pergunta segundo, a ANOVA. A 5% de significância todas as perguntas possuem *valores – p* significativos. Além disso, as letras “a”, “b” dispostas na tabela representaram os grupos que obtiveram diferença não significativa ($p > 0.05$), segundo o teste de Tukey.

Tabela 4: Média das Regionais em cada pergunta nos anos de 2013, 2014 e 2016.

Variável	Clusters			<i>valor – p</i> (ANOVA)
	1	2	3	
P1	72,0 b	83,6 a	80,9 a	0,0001
P2	59,2 b	69,1 a	67,3 a	0,0038
P3	56,8 b	65,3 a	66,4 a	0,0006
P4	63,0 b	75,3 a	75,6 a	0,0000
P5	63,8 b	72,7 a	72,3 a	0,0006
P6	62,5 b	71,4 a	66,3ab	0,0081
P7	61,8 b	68,4 a	74,9 a	0,0001

Observação: médias seguidas de letras distintas nas linhas, indica que as médias diferem entre si pelo Teste Tukey, ao nível de significância de 5%

Observa-se que os grupos 2 e 3 não diferem de acordo com o teste Tukey. Por isso, é necessário realizar a análise de cluster para 2 grupos e verificar o comportamento das Regionais em cada grupo.

Os 2 grupos formados pelo método de K-medias, são retratados na Tabela 5, em que é possível observar que as Regionais de Uberaba (URA), Divinópolis (DVL) e

Franca (FAC) mantiveram os índices dos três anos no mesmo grupo, sendo que as Regionais de Uberaba e Divinópolis tiveram os índices dos três anos alocados no grupo 1 e a Regional de Franca teve o índice dos três anos no grupo 2. A Regional de Uberlândia (ULA), Ituiutaba, (IUA) Itumbiara (IUB), Pará de Minas (PRS) e Patos de Minas (PMS) foram as que mantiveram seus índices nos diferentes grupos. As Regionais Ituiutaba, Itumbiara tiveram os anos de 2013 e 2016 alocados no grupo 2 e o ano de 2014 alocado no grupo 1. Uberlândia teve os anos de 2014 e 2016 alocados no grupo 2 e o ano de 2013 alocado no grupo 1. Patos de Minas e Pará de Minas tiveram o alocados no grupo 1 os anos de 2014 e 2016 e no grupo 2 o ano de 2013.

Tabela 5: Grupos obtidos por meio da Análise de Clusters, pelo K-Médias.

Grupo 1	Grupo 2
ULA13	ULA14
	ULA16
URA13	
URA14	
URA16	
IUA14	IUA13
	IUA16
	IUB13
IUB14	IUB16
PMS14	PMS13
PMS16	
PRS14	PRS13
PRS16	
	FAC13
	FAC14
	FAC16
DVL13	
DVL14	
DVL16	

A Tabela 6 apresenta a análise descritiva dos grupos relacionado às setes perguntas resultantes dos índices nos três anos sendo que o grupo 1 pode ser

considerado o pior grupo por ter médias menores comparadas ao grupo 2. Foi observado que as Regionais pertencentes aos 2 grupos tiveram melhor média em relação à pergunta 1- Imagem da Empresa, enquanto que, a pior média foi em relação à pergunta 3 – Cumprimento de Promessas. Os grupos apresentaram coeficientes de variações abaixo de 25%, considerado um valor baixo.

Tabela 6: Estatística Descritiva das Perguntas em cada Grupo.

Perguntas	Grupos	Média	Desvio Padrão	Mínimo	Máximo	Coefficiente de Variação
1	1	72,9	4,4	64	78	5,97%
	2	84,0	3,1	80	90	3,73%
2	1	59,6	3,7	52	65	6,16%
	2	70,2	5,5	61	81	7,89%
3	1	58,5	5,2	47	67	8,85%
	2	66,3	4,7	59	75	7,14%
4	1	65,5	5,6	58	76	8,55%
	2	75,9	5,6	69	84	7,38%
5	1	64,9	4,8	58	74	7,38%
	2	73,6	3,3	70	80	4,42%
6	1	62,5	4,2	55	70	6,75%
	2	70,6	4,9	61	80	6,92%
7	1	63,9	3,6	52	65	5,62%
	2	71,9	5,5	61	81	7,89%

Na Tabela 7 foram dispostas as médias de cada um dos dois grupos e a significância de cada pergunta segundo, a ANOVA. A 5% de significância todas as perguntas possuem *valores p* significativos, portanto em nenhuma das variáveis os clusters serão semelhantes.

Tabela 7: Média das Regionais em cada pergunta nos anos de 2013, 2014 e 2016.

Variável	Clusters		<i>valor p</i>
	1	2	(ANOVA)
P1	72,9	84,0	< 0,0001
P2	59,6	70,2	<0,0001
P3	58,5	66,3	0,0015
P4	65,5	75,9	0,0002
P5	64,9	73,6	<0,0001
P6	62,5	70,6	0,0004
P7	63,9	71,9	0,0047

Através dos 2 grupos formados pelo método de K-Médias, foi possível identificar que relacionando os índices das Regionais agrupadas com as Perguntas o grupo 2 obteve maiores médias percentuais comparadas ao grupo um. Os clientes da empresa em relação ao grupo 2 estão muito satisfeitos quanto à imagem da Empresa, Eficácia dos serviços, Retorno as Solicitações, Retorno Custo versus Benefício Agilidade no Atendimento. Os clientes estão insatisfeitos em relação ao cumprimento das promessas da empresa nos 2 grupos, sendo um ponto negativo para a empresa, pois um cliente satisfeito é fiel à empresa, volta a comprar e comunica as suas experiências positivas ao seu entornos. É importante ressaltar que das três Regionais que tiveram os índices dos três anos no mesmo grupo, a Regional de Uberaba e Divinópolis não tiveram melhora significativa em relação ao ano de 2014 e ao ano de 2016 por estar alocada no grupo um. Embora Franca esteja no grupo de melhores médias, a Regional não obteve melhora. As Regionais de Pará de Minas, Uberlândia, Patos de Minas, Ituiutaba e Itumbiara tiveram melhoras significativas comparando o ano de 2014 com o ano de 2016.

A empresa precisa intervir e realizar mais investimentos no cumprimento de suas promessas para garantir uma melhora significativa em relação à satisfação do cliente e retenção dos mesmos, pois um cliente satisfeito influencia positivamente cinco novos clientes. Um cliente insatisfeito influencia negativamente vinte clientes atuais.

5. CONCLUSÃO

O presente estudo permitiu concluir que a Análise de Clusters mostrou eficaz na avaliação das regionais de uma empresa de telecomunicações. A princípio o Teste K-médias identificou 3 grupos e com o teste Tukey foi possível identificar que dois dos três grupos eram estatisticamente iguais, o que justifica a utilização de 2 grupos.

Verificou-se ainda que as regionais Uberlândia, Ituiutaba e Itumbiara, apresentaram uma melhora significativa; Patos de Minas e Pará de Minas pioraram significativamente; Franca não apresentou melhoras, mas foi alocada no grupo com melhores avaliações; Uberaba e Divinópolis não apresentaram melhoras significativas, sendo alocadas nos grupos com piores avaliações.

6. REFERÊNCIAS

- AMARAL JÚNIOR, A.T.; THIÉBAUT, J.T.L. **Análise multivariada na avaliação da diversidade em recursos genéticos vegetais**. Campos dos Goytacazes - Universidade Estadual do Norte Fluminense - UENF, CCTA, p. 55, 1999.
- ANDERBERG, M. R. **Cluster analysis for applications**. New York: Academic Press, 1973.
- BARROSO, L.P; ARTES R. **Análise Multivariada**. Lavras: UFLA, 2003
- BUSSAB, W. de O.; MIAZAKI, S. E.; ANDRADE, D. F. **Introdução à análise de agrupamento**. In: IX SIMPÓSIO BRASILEIRO DE PROBABILIDADE E ESTATÍSTICA, 1990, IME-USP São Paulo, 105 p.
- CALINSKI, T.; HARABASZ, J. **A Dendrite Method for Cluster Analysis. Communications in Statistics**, v. 3, n. 1, p. 1-27, 1974.
- CONCEITO DE TELECOMUNICAÇÃO. Disponível em: <<http://conceito.de/telecomunicacao>>. Acesso em 27 de Setembro de 2016.
- CRUZ, C. D.; CARNEIRO, P. C. S. **Modelos Biométricos aplicados ao melhoramento genético**. Viçosa: UFV, 2006. 585 p.
- DOWNING, D.; CLARK, J. **Estatística Aplicada**. 1ª Ed. São Paulo: Saraiva, 2000
- FERREIRA, D.F. **Estatística Multivariada** – 2.ed.rev e ampl. – Lavras : Ed.UFLA , 2011.
- HOFFMAN, K.D.; BATESON, J.E.G. **Princípios de marketing de serviços, conceitos, estratégias e casos**, 2 ed. São Paulo: Pioneira Thomson Learning, 2003.
- KHATTREE, R.; NAIK, D.N. **Multivariate data reduction and discrimination with SAS Software**. New York: BBU Press and John Wiley Sons Inc., 2000. 574p

KOTLER, P.; ARMSTRONG, G. **Princípios de marketing**. 9. ed. São Paulo: Pearson Prentice Hall, 2006.

KOTLER, P.; KELLER, K.L. **Administração de marketing**. 12. ed. São Paulo: Pearson Prentice Hall, 2007.

MARDIA, K. V.; KENTK, J. T.; BIBBYB, J. M. **Multivariate analysis**. London: Academic Press, 1995. 518p.

Martins, G.A. **Estatística Geral e Aplicada**. 3. Ed. 2 reimpr. – São Paulo: Editora Atlas, 2006.

MILLIGAN, G. W.; COOPER, M. C. **An examination of procedures for determining the number of *clusters* in a data set**. Psychometrika, v. 50, n. 2, p. 159-179, 1985.

MINGOTI, S.A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 2005.

PAULINS, V.A. (2005), **An analysis of customer service quality to college students as influenced by customer appearance through dress during the in-store shopping process**. Journal Retailing Consumer Service, v. 12, p. 345-355, 2005.

QUINTAL, G. **Análise de *clusters* aplicada ao Sucesso/Insucesso em Matemática**. 2006. Dissertação (Mestrado em Matemática para Ensino) – Universidade da Madeira, Funchal.

RAO, C. R. **An advanced statistical method in biometric research**. New York: Ed. John Wiley e Sons, 1952. 390 p.

R Development Core Team (2017). R: **A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <<http://www.rproject.org>>. Acesso em: 5 de abril de 2017.

RIBOLDI, J. **Análise de agrupamento "*Cluster Analysis*" e suas aplicações.** Piracicaba: ESALQ, 1986. 33 p

SOUSA, F.S.F.J. **Satisfação de Clientes - O Caso de Uma Empresa Industrial – 2011 – 125 f** (Dissertação de Mestrado em Marketing), Instituto de Economia, Universidade de Coimbra, Coimbra. 2011.

VIEIRA, S.; HOFFMANN, R. **Estatística experimental.** São Paulo: Atlas, 1989. 175p.

VICINI, L. **Análise multivariada da teoria à prática.** Santa Maria: UFSM, CCNE, 2005. 2015p.

TUKEY, J.W. **The problem of multiple comparisons.** Mimeographs Princeton University, Princeton, N.J., 1953.

WIKIPÉDIA. Disponível em: < https://pt.wikipedia.org/wiki/Ag%C3%A2ncia_Nacional_de_Telecomunica%C3%A7%C3%B5es >. Acesso em 27 de Setembro de 2016.