



**Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística**

Vandeir Robson da Silva Matias

Análise do acesso à internet nas Unidades da Federação do Brasil: elementos para o planejamento territorial

Belo Horizonte – Minas Gerais – Brasil

2016

Vandeir Robson da Silva Matias

Monografia apresentada ao Curso de Especialização em Estatística aplicada como requisito parcial para obtenção do diploma de Especialista em Estatística aplicada.

Orientadora: Profa. Dra. Edna Afonso Reis

Belo Horizonte

2016

Monografia defendida e aprovada, em 06 de Dezembro de 2016,
perante banca examinadora

Prof.^a. Dr^a Edna Afonso Reis
(Orientadora)

Prof.^a. Dr^a Ela Mercedes Medrano de Toscano

Prof.^a. Dr^a Ilka Afonso Reis

Agradecimentos

A minha orientadora, Professora Edna, pela oportunidade de realizar esse trabalho com qualidade.

Aos meus pais pelo incentivo constante.

Aos meus colegas do Curso de Especialização em Estatística aplicada: Luiz Fernando, Luciana Monteiro e Matheus Xavier.

Aos colegas de trabalho do Centro Federal de Educação Tecnológica de Minas Gerais.

Aos professores do curso: Ela Mercedes, Gregório, Ilka, Sueli e Marcelo.

SUMÁRIO

LISTA DE FIGURAS.....	iii
LISTA DE TABELAS.....	iv
LISTA DE SIGLAS E ABREVIATURAS.....	v
RESUMO	vi
ABSTRACT.....	vii
1. Introdução.....	1
2. A sociedade em rede e o uso da Internet no planejamento territorial	3
3. Objetivos e justificativa.....	8
4. Metodologia	9
4.1. Análise de Clusters.....	9
4.2. Bancos de Dados.....	15
5. Resultados e Discussão.....	17
5.1. Análise Descritiva	17
5.2. Análise de Clusters.....	25
6- Considerações Finais.....	33
Referências.....	35

LISTA DE FIGURAS

Figura 1- Proporção de domicílio com acesso à internet	4
Figura 2-Proporção de órgãos públicos federais e estaduais que disponibilizaram alguma forma de contato com o cidadão pela internet nos últimos 12 meses, por tipo de contato.	5
Figura 3-Proporção de órgãos públicos federais e estaduais, por forma de participação do cidadão pela internet nos últimos 12 meses.....	7
Figura 4- Etapas de formação de cluster.....	11
Figura 5- Histograma das dez variáveis selecionadas.....	18
Figura 6- Boxplots das dez variáveis selecionadas.....	19
Figura 7-Distribuição espacial do percentual de domicílios com acesso à internet nas UFs.	20
Figura 8-Distribuição das pessoas de 10 anos ou mais de idade que utilizaram a internet, por grupo de idade.	21
Figura 9- Distribuição percentual das pessoas de 10 anos ou mais de idade, ocupadas na semana de referência, que utilizaram a Internet, segundo agrupamentos ocupacionais no trabalho principal- 2014	21
Figura 10- Percentual de pessoas que utilizaram a internet, no período de referência dos últimos três meses, na população de 10 anos ou mais de idade, segundo as classes de rendimento mensal domiciliar per capita Brasil- 2013-2014.....	23
Figura 11- Percentual de domicílios com utilização da internet, por tipo de equipamento utilizado para acessar a internet, no total de domicílios particulares permanentes com utilização da internet, segundo as Grandes Regiões – 2014	23
Figura 12- Percentual de domicílios com utilização da Internet somente por meio de microcomputador e somente por meio de telefone móvel celular ou tablete, no total de domicílios particulares permanente com utilização da Internet, segundo as Unidades da Federal em 2014	24
Figura 13- Percentual de pessoas que utilizaram a internet, no período de referência dos últimos três meses, na população de 10 anos ou mais de idade, por sexo, segundo as Grandes Regiões- 2014.	25
Figura 14- Dendograma: agrupamento das UF's.	26
Figura 15- Dendograma da partição final com nove clusters.....	28
Figura 16- Clusters: Acessa a internet no Brasil	29

LISTA DE TABELAS

Tabela 1-Banco de dados das variáveis selecionadas	16
Tabela 2- Estatísticas descritivas das variáveis selecionadas.....	17
Tabela 3-Percentual de pessoas que utilizaram a Internet, no período de referência dos últimos três meses, na população de 10 anos ou mais de idade, por Grandes Regiões, segundo os grupos de idade - 2014.....	22
Tabela 4-Distribuição das pessoas de 10 anos ou mais de idade que utilizaram a Internet, no período de referência dos últimos três meses, por Grandes Regiões e os grupos de anos de estudo - 2014.....	22
Tabela 5- Resultados da análise de clusters.....	26
Tabela 6- Medidas para a tomada de decisão	27
Tabela 7- Valores dos Critérios de informação Bayesiana.....	28
Tabela 8- Descritivas dos Clusters	30

LISTA DE SIGLAS E ABREVIATURAS

BIC- Critério de informação Bayesiana

CETIC- O Centro de Estudos sobre as Tecnologias da Informação e da Comunicação

IBGE- Instituto Brasileiro de Geografia e Estatística

PNAD- Pesquisa Nacional por Amostra de Domicílios

TDIC- Tecnologias Digitais da Informação e da Comunicação

UIT -International Telecommunication Union

RESUMO

O planejamento territorial que é um instrumento para organizar e ordenar o território brasileiro, harmonizando as relações econômicas, sociais e ambientais que nele acontecem. Para que ocorra sua eficácia, é necessário um efetivo esforço de compartilhamento institucional, voltado para a integração das ações e políticas públicas territoriais, bem como articulação com a sociedade civil, congregando seus interesses em torno de um pacto pela gestão do território. O uso da internet pode promover o acesso à informação e uma maior integração na sociedade. Parte significativa da prestação de serviços municipais, estaduais e federais já ocorre com aporte da rede de computadores de alguma forma. O objetivo geral desta pesquisa é realizar uma análise espacial do uso e acesso à internet pelas Unidades da Federação no Brasil a partir do aporte da técnica de análise de cluster. O objetivo específico é realizar a partição ideal, ou seja, identificar os grupos formados para subsidiar políticas públicas específicas para cada grupo formado. Foram selecionadas doze variáveis e os dados foram tratados a partir do *software* estatístico MiniTAB versão 17. Ao final, chegou-se a formação de nove clusters. A análise de cluster foi essencial para a criação de grupos e definição dos perfis. A separação das Unidades da Federação em conglomerados foi viável para conseguir agrupar estados similares entre si e divergentes em relação aos outros grupos que precisam de políticas públicas diferentes para fomentar a ampliação do acesso à internet no Brasil

Palavras-chave: Estatística multivariada, Análise de dados, Planejamento territorial e Internet.

ABSTRACT

The Territorial Planning is a tool instrument to organize and sort the Brazilian territory, harmonizing and economic relations, social and environmental. Why occurs your efficiency and required has effective institutional sharing effort, aimed paragraph Integration of territorial actions and public policies, as well as articulation with civil society, congregating interests around hum Pact for Land Management. Use the internet can promote access to information and greater integration in society. A significant part of the Provision of Municipal Services, State and Federal, occurs contribution of Computer Network Some form. The General Purpose this research and perform a spatial analysis to use and access the internet by Federative Units in Brazil From the multivariate statistical contribution. The Specific Objective and make an ideal partition, identify the groups formed paragraph subsidize specific public policy paragraph each elaborate group. Were selected twelve variables and he data were Treaties from Minitab Statistical Software to version 17. Training the end of the clusters come to nine. Cluster analysis was essential to the creation of groups and definition of profiles. Separation of the Federative Units in clusters was feasible paragraph getting group states similar and divergent in relation to others that groups need public policies different paragraph fostering Internet access expansion in Brazil.

Keywords: Multivariate Statistics, Data Analysis, Territorial Planning and Internet

1. Introdução

As *tecnologias digitais da informação e comunicação* (TDIC) já fazem parte do cotidiano na população em várias atividades como lazer, trabalho, educação, mobilização social, participação política, entre outros. O Instituto Brasileiro de Geografia e Estatística (IBGE) define como TDICs: internet, rádio, televisão, microcomputador, telefone fixo, telefone móvel, antena parabólica, televisão a cabo, fax, câmeras digitais, tablets, entre outras.

A internet, em especial, se dissemina no país desde a década de noventa do século passado e verifica-se que grande parte da sociedade utiliza essa tecnologia alguma parte do dia. O uso da internet pode promover o acesso à informação e uma maior integração do indivíduo na sociedade. Parte significativa da prestação de serviços municipais, por exemplo, já ocorre com aporte da rede de computadores de alguma forma. Assim, quanto maior o acesso da população e conhecimento para utilizar essa ferramenta, maior a promoção de cidadania e participação popular na gestão do espaço urbano (MATIAS, 2011).

Percebe-se que os avanços nos setores de telecomunicações geram um impulso de um enorme fluxo de informações, que passam a ser processadas e difundidas numa velocidade cada vez maior. Por conseguinte, observa-se o incremento das TDIC e um espaço cibernético em constante ampliação. Como a internet não está disseminada em todo o território nacional, apenas parte da sociedade utiliza essa tecnologia para interagir com o poder público. Essa interação configura um espaço público diferente, que ainda pode se tornar um ciberespaço voltado para os exercícios do lazer, experimentação, cidadania e democracia.

O acesso à Internet influencia no desenvolvimento sociocultural da população, pois se trata de um fluxo enorme de informações que divulgam ideias, comportamentos e culturas bem diferentes uma das outras. Esse fator pode colaborar para o desenvolvimento da sociedade, que depende da forma que a internet será utilizada, além da forma com que produzirá e disseminará suas informações.

A partir do seu uso intenso em vários setores, a internet pode resgatar a centralidade da ação humana e o protagonismo das práticas participativas e comunicativas. Contudo, muitas vezes esse processo de informatização social não vem acompanhando da racionalidade e bom senso. Os usos da internet experimentados até então estão mais voltados para atividades individuais de lazer do que atividades coletivas e colaborativas para fins de planejamento.

O mundo atual passa por uma forte influência da revolução informacional. As ciências começaram a dialogar umas com as outras, características essas da

Revolução Técnico-científica contemporânea. É neste cenário que se observa, no Brasil, o uso crescente da internet por parte significativa da população.

As TDIC viabilizam a comunicação e proporcionam interação e descobertas para os cidadãos. O uso da internet, por exemplo, permite monitorar os processos eleitorais, gerando um maior *accountability*¹ contra fraudes eleitorais. Na esfera do executivo, a internet facilita o acesso a informações fiscais e burocráticas, além de possibilitar a fiscalização de obras e consertos agendados pelo poder público.

A Internet possui uma história recente, mas que evoluiu muito rápido, beneficiando, assim, de maneira significativa a globalização e proporcionando um fluxo intenso de informação em alguns territórios em detrimento de outros. A questão central é como este processo se desenvolve e quem participa dele. Utilizar a internet envolve custo com um determinado equipamento (smartphone, tablet, computador de mesa, computador portátil, entre outros) assim com a forma de conexão (banda larga fixa ou móvel, linha telefônica, satélite, entre outras formas). Assim, constata-se que nem toda a sociedade a utilizará em função do seu custo. É necessário traçar o perfil dos usuários da internet para melhorar o seu acesso a partir de políticas públicas adequadas e garantir assim melhores condições de partição e acesso à informação dos habitantes no espaço público.

Sabe-se que a disseminação da tecnologia não garante o seu acesso, pois, em várias partes do mundo, um quantitativo considerável não possui nem as condições básicas de saneamento, alimentação, moradia, saúde, entre outros bens caros ao bem estar de qualquer indivíduo. Embora as tecnologias digitais da informação e comunicação façam parte do nosso universo em várias situações cotidianas, elas possuem um custo oneroso, principalmente aquelas que são mais sofisticadas.

Portanto, o objetivo do trabalho é realizar uma análise espacial do uso e acesso a internet por Unidades da Federação no Brasil a partir do aporte da técnica de análise de cluster. Os principais objetivos específicos são: buscar a partição próxima do ideal, identificar os grupos formados e Identificar as características de acesso a internet dentro dos grupos.

Este texto se organiza conforme descrito a seguir. A Seção 2 contempla a revisão bibliográfica sobre o uso da internet no Brasil, relacionando o uso da internet, participação popular e gestão do território. A Seção 3 apresenta os objetivos e

¹ Entendido como princípio de gestão essencial ao desenvolvimento da democracia eletrônica, o *accountability* configura-se como uma das formas de responsabilidade democrática da administração pública. Com o objetivo de promover a transparência dos atos públicos e colaborar para a cultura da prestação de contas na política contemporânea, o *accountability*, é uma das formas mais valiosas de estreitar o laço entre cidadão e as organizações públicas. (MATIAS, 2011, p.151)

justificativa desse trabalho. A Seção 4 contempla a técnica estatística utilizada que é a análise de clusters e apresenta o banco de dados. A seção 5 faz a discussão dos resultados alcançados e, por fim, a seção 6 apresenta as considerações finais.

2. A sociedade em rede e o uso da Internet no planejamento territorial

O processo de globalização no seu modelo informacional transformou a organização espacial das relações sociais, territoriais e políticas. As mudanças contemporâneas nos territórios são pautadas a partir de sistemas técnicos, ou melhor, sistemas de engenharia materiais e imateriais que proporcionam um contato indireto com as pessoas. O advento da tecnologia possibilitou novas formas de ação política associada a diversos atores sociais e políticos no espaço urbano.

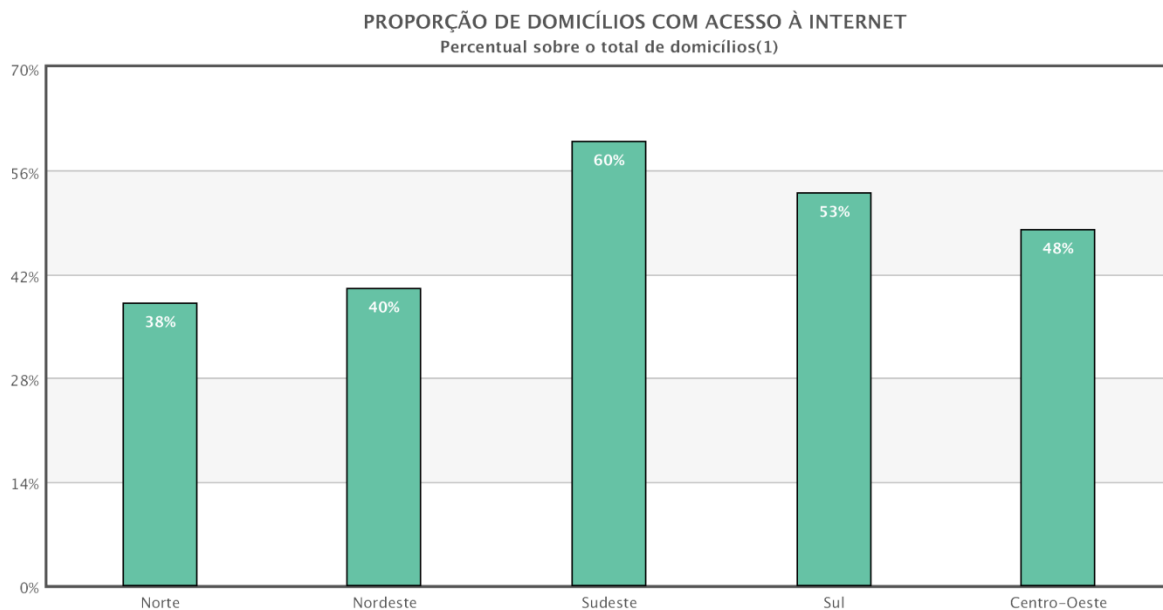
Em grande parte da sociedade mundial, verificamos a incorporação das políticas de comunicação baseadas em sistemas técnicos e em redes. Existe, então, uma virtualização da sociedade que se expande para os diversos modos de vida. A internet faz parte do dia-a-dia dos habitantes das cidades, sobretudo nos grandes centros. Percebe-se então um uso intenso da internet pela sociedade civil e também o uso pelos governos nas suas diferentes esferas (municipal, estadual e federal) e poderes (executivo, legislativo e judiciário), (MATIAS, 2011).

Dentre as TDIC atuais, destaca-se a internet, que influencia o cotidiano das relações socioterritoriais, transformando as estruturas físicas que auxiliam na formação de uma ciberdemocracia. O início do século XXI foi marcado por um imenso número de governos que passaram a operar *online*, em busca de novas oportunidades de gestão, maior eficiência e possibilitando a criação de novos canais de participação dos cidadãos. Observa-se na Figura 1 a proporção de domicílios com acesso a internet no Brasil, entre Novembro de 2015 e Junho de 2016.

A Região Sudeste é a que possui a maior proporção de domicílios com acesso à internet no território nacional, seguida da Região Sul. Com o menor acesso, destacam-se os domicílios da Região Norte. As sociedades contemporâneas mostram formas de desigualdades socioculturais fortemente associadas às desigualdades de natureza econômica e nas condições objetivas de acesso aos bens tecnológicos. (MATIAS, 2011).

A utilização do ambiente virtual pelo Estado configura-se como uma tentativa de aproximação dos cidadãos com a “democracia”, no caso uma “ciberdemocracia”, ou seja, o Estado usufrui do ciberespaço de modo a aproximar os cidadãos da gestão da

metrópole. É pertinente refletir sobre os limites e possibilidades da relação Sujeito/Cidade a partir da utilização da Internet.



(1) Base: 67.038.766 domicílios. Dados coletados entre Novembro de 2015 e Junho de 2016.
Fonte: Núcleo de Informação e Coordenação de Ponto BR (NIC.br)

Figura 1- Proporção de domicílio com acesso à internet

Assim, é importante o acesso da população à internet, pois cada vez mais os processos da administração pública realiza, consulta pública, disseminação de informação, processos deliberativos, etc., através da internet. É pertinente que a população esteja preparada para esse momento. E uma das formas de preparação é o acesso à internet. Nesse sentido, a rede assume funções de reorganização e gerenciamento e apresenta inovações para a sociedade contemporânea (GUIDI, 2002).

O Estado, através de suas ações, compreende que o período atual é da revolução informacional e utiliza a internet em diversas atividades como descrito no gráfico. Divulgação de legislação e prestação de contas são os itens mais recorrentes. Portanto é pertinente, a partir dessa constatação buscar implantar políticas que efetivem a disponibilização de informações e participação para a população.

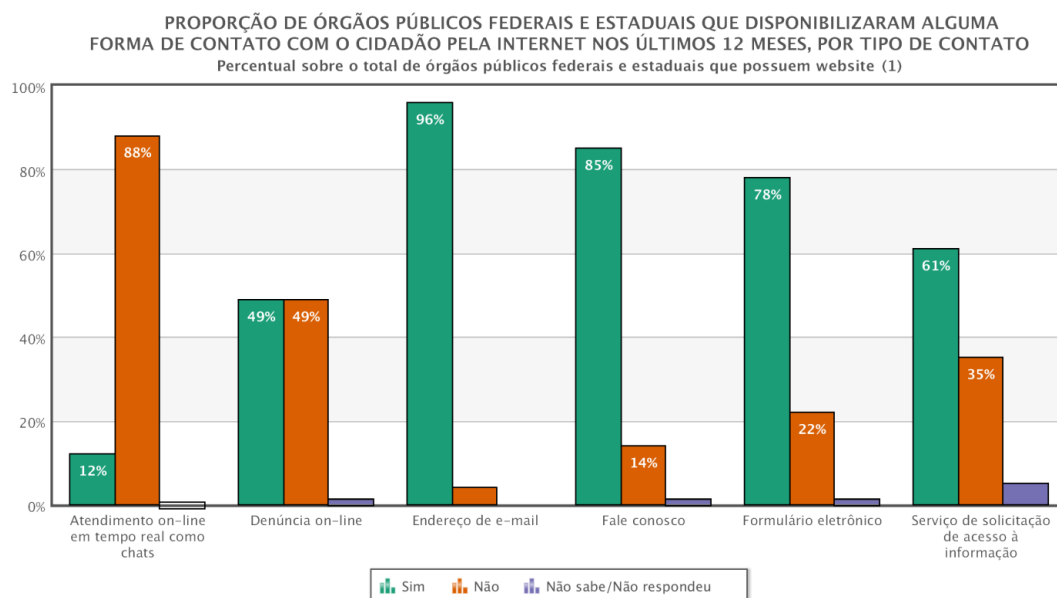
Existem algumas diferenças do que se entende por democracia e governo eletrônico. O governo eletrônico corresponde ao uso de tecnologias pelos governos como parte do esforço de modernização e racionalização da prestação de serviços públicos aos usuários, melhorando a qualidade dos serviços ofertados, reduzindo custos através do modelo virtual.

O governo avança na política pública do governo eletrônico e cria inclusive processos deliberativos digitais em algumas cidades. Os orçamentos participativos

digitais são experiências desse tipo. Para mais informações dessa prática, ver Matias, 2011. Essa é uma forma do governo manter uma interação mais ativa com o usuário do serviço público, ou seja, é um sistema de apoio a decisão.

Os gestores se preocupam substancialmente com a disseminação da informação governamental para dar visibilidade à administração. O Estado cria novas regras do jogo do urbano no contexto das transformações vigentes. É uma nova forma de fazer política, o que configura uma nova forma de analisar o espaço urbano. Não se sabe o que ainda poder está por vir, contudo é necessário ampliar a análise. Nesse contexto, alguma forma de contato via rede é importante entre gestores e usuários. A Figura 2 mostra a porcentagem de formas de contato disponibilizadas por órgãos públicos.

Dentre as formas de contato *e-mail* (96%), fale conosco (85%) e formulário eletrônico (78%) se destacam. O objeto é pedir geral do contato é tirar uma dúvida, fazer uma reclamação ou pedir uma informação. Dentro do contexto de informação governamental para promoção da cidadania e exercício da democracia, o conceito de governo eletrônico² ganha novos contornos.



(1) Base: 1.508 órgãos públicos federais e estaduais que declararam possuir website. Respostas múltiplas e estimuladas. Dados coletados entre julho e outubro de 2015. Fonte: Núcleo de Informação e Coordenação de Ponto BR (NIC.br)

Figura 2-Proporção de órgãos públicos federais e estaduais que disponibilizaram alguma forma de contato com o cidadão pela internet nos últimos 12 meses, por tipo de contato.

² O governo eletrônico é a contínua otimização da prestação de serviços do governo, da participação dos cidadãos e da administração pública pela transformação das relações internas e externas através da tecnologia, da Internet e dos novos meios de comunicação. O governo eletrônico tem assumido cada vez mais destaque pelo crescimento das expectativas dos cidadãos, globalização e progresso tecnológico e reforma do governo.

Uma das características do governo eletrônico é a perspectiva de entrada em cena de atores nos debates sobre o urbano, com impactos importantes nas políticas públicas que tendem a colocar o cidadão, informação e a cidadania como foco central. Esse modelo de política pública é representado por um conjunto de ações sociais desenvolvidas pela administração pública direta, onde se busca o investimento em transparência e estímulo ao relacionamento com o cidadão.

A democracia eletrônica pode ser entendida como um conjunto de processos democráticos que propiciam a participação cidadã por meio das TDIC. Estas estão relacionadas com questões fundamentais acerca da natureza da governança e processos decisórios relativos ao Estado e a relação deste com os cidadãos. O uso das novas tecnologias representa uma nova fase para a comunicação democrática, uma nova possibilidade para a participação descentralizada. (MAIA, 2002).

O direito à informação é colocado em pauta pelos novos cenários da política. O acesso à informação e o direito à comunicação são direitos inalienáveis do ser humano e, por isso, o acesso e a produção devem ser compreendidos como um novo direito humano fundamental, devendo ser respeitado, garantido e promovido pelo Estado.

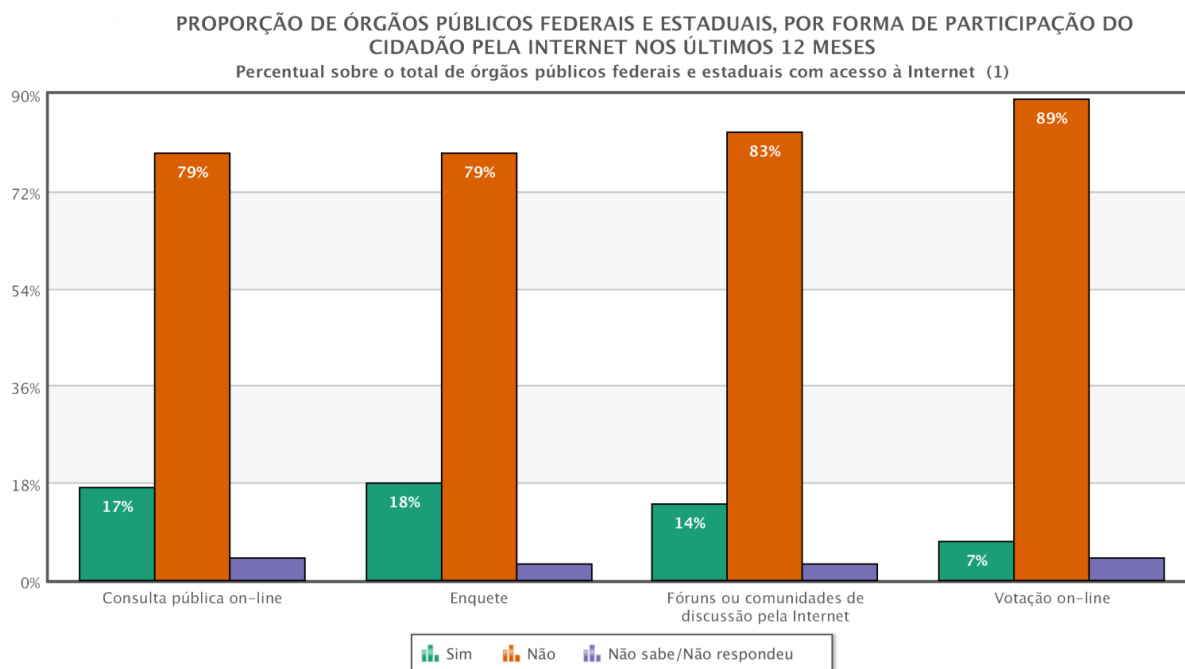
A gestão política de alguns municípios brasileiros como São Paulo, Salvador Belo Horizonte, Porto Alegre, Curitiba, entre outros, está embasada nas novas tendências e paradigmas da gestão pública contemporânea. Nessa agenda, temos a ideia de governança, eficiência, governo eletrônico, disseminação da informação, transparência, participação e descentralização. Vislumbram-se transformações dos processos espaciais e de redes sociais no contexto global a partir dos sistemas de engenharia virtuais. A economia política da informação produz uma nova espacialidade e possibilita mudanças na concepção do urbano e cidadania. A informação fortalece o papel do cidadão na sociedade contemporânea, empodera e gera conhecimento necessário às práticas democráticas que interferem no planejamento territorial (UHLIR, 2006).

Atualmente a internet é essencial para o planejamento territorial, que é um instrumento para organizar e ordenar o território brasileiro, harmonizando as relações econômicas, sociais e ambientais que nele acontecem, sobretudo nas grandes cidades do país. Para que ocorra sua eficácia, é necessário um efetivo esforço de compartilhamento institucional, voltado para a integração das ações e políticas públicas territoriais, bem como articulação com a sociedade civil, congregando seus interesses em torno de um pacto pela gestão do território.

Verifica-se que, em uma perspectiva histórica, o planejamento territorial não envolve somente o setor econômico ou mesmo a elaboração de programação

orçamentária, a partir de um planejamento governamental. O Planejamento também atua no sentido de trazer melhorias realizadas pelo governo e que garantam em certa medida uma qualidade de vida aos cidadãos. *Strictu sensu* o ato de planejamento pode ser entendido com a elaboração de um plano ou roteiro de ações que passou por dois aspectos principais. O primeiro é de que o planejador é um guia contrapondo ao imprevisto o outro aspecto é de que o planejamento pode possibilitar uma contabilização das diversas ações a se realizar.

Para romper com essa lógica perversa, as administrações públicas apostam em um modelo estruturado de gerenciamento da tecnologia e da informação, aplicados de maneira sistematizada nos processo de gestão participativa, havendo, portanto, a confiança de que esses sistemas podem ser utilizados para tomada de decisão pelos cidadãos. A Figura 3 mostra as formas de participação ofertada pelos órgãos públicos. Dentre elas destacam as enquetes (18%) e consultas públicas online (17%). São ações tímidas, mas importantes para fomentar uma maior participação dos cidadãos.



(1) Base: 1.641 órgãos públicos federais e estaduais que declararam ter acesso à Internet nos últimos 12 meses. Respostas múltiplas e estimuladas. Dados coletados entre julho e outubro de 2015.
 Fonte: Núcleo de Informação e Coordenação de Ponto BR (NIC.br)

Figura 3-Proporção de órgãos públicos federais e estaduais, por forma de participação do cidadão pela internet nos últimos 12 meses.

O desenvolvimento tecnológico representa hoje um momento importante para administração pública, pois serve para mitigar algumas deficiências do setor público, que sempre foi caracterizado por ineficiência, ineficácia, falta de transparência e

custos excessivos. O uso das tecnologias de comunicação, como a internet, nas ações governamentais vem se tornando algo essencial para corrigir essas distorções e reformular as administrações públicas, tornando-as mais eficientes e legítimas perante os atores sociais. Tal constatação traz a questão da necessidade de expansão das redes de distribuição de informações, sobretudo diante da relativa ausência de uma política de informação definida para os diferentes órgãos da administração pública.

A utilização de sistemas de engenharia imateriais pelo Estado se configura como uma tentativa de aproximação dos cidadãos com a “democracia”, no caso uma “democracia virtual”, ou seja, o Estado usufruindo do ciberespaço de modo a aproximar os cidadãos da gestão do espaço. O virtual está ligado a uma promessa de acontecimento, a uma inovação, uma criação e recriação no espaço que se metamorfoseia incessantemente.

3. Objetivos e justificativa

O objetivo geral dessa pesquisa é realizar uma análise espacial do uso e acesso à internet por Unidades da Federação no Brasil a partir do aporte da análise de clusters. O objetivo específico é realizar a partição ideal e identificar os grupos formados para subsidiar políticas públicas específicas para cada grupo apurado. Pretende-se averiguar quais Unidades da Federação estão mais avançadas quanto ao acesso à internet e em quais variáveis, assim como contrário também, quais unidades da Federação encontram-se em pior estágio de acesso e em quais variáveis precisam melhorar.

Esse trabalho de pesquisa é pertinente, pois acredita-se que quanto maior o acesso à população a essa tecnologia, maior será sua participação nos processos decisórios do planejamento territorial nas escalas local, regional e nacional. Entender esse processo da ciberdemocracia no planejamento territorial é primordial para garantir a real participação cidadã em processos democráticos.

O planejamento territorial tem por finalidade propiciar um diagnóstico preciso sobre o meio físico-biótico, socioeconômico e sobre sua organização institucional e oferece ainda diretrizes para determinadas ações, as quais deverão refletir os diferentes interesses dos cidadãos. Assim, acredita-se que o uso da Estatística Multivariada possa contribuir para um sistema de planejamento mais eficaz, onde os investimentos e esforços, tanto do governo quanto da iniciativa privada, serão aplicados de acordo com as peculiaridades das unidades de planejamento/conglomerados formados.

Sendo assim, o conhecimento sobre a distribuição espacial do dessa tecnologia é importante para fomentar políticas públicas de acesso à internet para a população do país, pois a partir de um cluster podemos ter uma melhor compreensão dos fatores da dinâmica da sua distribuição.

4. Metodologia

Os dados são provenientes da Pesquisa Nacional por Amostra de Domicílios (PNAD) de 2014 sobre Acesso à Internet e à Televisão e Posse de Telefone Móvel Celular para Uso Pessoal³. Este trabalho utiliza apenas dez variáveis referentes ao acesso à Internet associados às UFs do Brasil e o Distrito Federal.

De modo complementar ao longo do trabalho serão utilizados os dados do Centro de Estudos sobre Tecnologia da Informação e Comunicação (CETIC) relacionados aos domicílios e a dinâmica do governo eletrônico no Brasil⁴. O banco de dados está descrito na Seção 4.2.

A presente pesquisa emprega o método de investigação análise-síntese. A proposta é utilizar a corrente empírico-analítica para analisar dados secundários tratados com o aporte de uma técnica de Estatística Multivariada, chamada Análise de Cluster, descrita na Seção 4.2. As análises estatísticas foram realizadas no *software* estatístico MiniTAB versão 17. Para a validação dos Clusters utilizou se o *software* livre R versão 3.2.4 e para a elaboração do mapa utilizou se o *software* ArcGIS versão 10.2.

4.1. Análise de Clusters

Essa parte visa discutir a técnica estatística utilizada nessa pesquisa. A Estatística Multivariada apresenta muitas técnicas como: análise de componentes principais, análise fatorial, análise de correlações canônicas, análise discriminante e análise de agrupamentos. Essa pesquisa utiliza a técnica de Análise de Clusters, também conhecida como Análise de Agrupamentos ou Conglomerados (MATOS, 2007).

O objetivo da Análise de Clusters é o de organizar os elementos em grupos (*clusters*), com base nas informações quanto à proximidade de seus elementos em

³ Instituto Brasileiro de Geografia e Estatística-IBGE. (PNAD) Acesso à internet e à televisão e posso de telefone móvel celular para uso pessoal. Rio de Janeiro: IBGE, 2016.

⁴ www.cetic.br

determinadas variáveis. No processo de formação dos agrupamentos, é importante minimizar a distância intra grupos e maximizar a distância inter grupos. O resultado obtido a partir da aplicação dessa técnica é um conjunto de grupos com coesão interna e isolamento externo, ou seja, elementos dentro de um mesmo grupo são tão similares quanto possível e são, ao mesmo tempo, tão dissimilares quanto possível dos elementos presentes nos demais grupos. (MATOS, 2007).

Os grupos devem manter o máximo de similaridade interna e dissimilaridade externa nestas variáveis. É interessante usar a Análise de Clusters quando se pretende observar a hierarquia dos dados, quando os dados forem categorizados e quando se busca estabelecer partição através do grau de similaridade e da distância entre grupos. O objetivo é criar grupos e posteriormente analisar o conjunto de dados dentro de cada grupo investigando as similaridades. A análise de agrupamento é uma técnica útil para trabalhar com dados geográficos, com índices e indicadores do IBGE por exemplo.

Nesses casos, existe uma infinidade de dados sobre municípios, estados e até mesmo países que precisam ser agrupados segundo a sua similaridade para facilitar o estudo, ou seja, é necessária uma homogeneidade dentro do grupo o que favorecer o planejamento territorial. É possível utilizar essa técnica na astronomia, separando, por exemplo, as estrelas com características similares para estudar sua evolução. No Marketing, para separar consumidores de acordo com o perfil, criando assim estratégias de publicidade; em uma escola, podemos ver através da análise de agrupamentos, aqueles alunos que são propensos a evadir.

Os dados analisados são agrupados de acordo com as características comuns observadas entre eles. O processo de construção de cluster envolve algumas etapas, conforme descreve a Figura 4 apresenta.

As etapas de formação de clusters nos fornecem um desenho dos passos a seguir até chegar à interpretação dos resultados. Após coleta dos dados primários ou secundários e a seleção das variáveis de interesse, deve se proceder à seleção do algoritmo de agrupamento e optar em usar um método hierárquico ou não hierárquico. Os métodos são complementares e não competitivos.

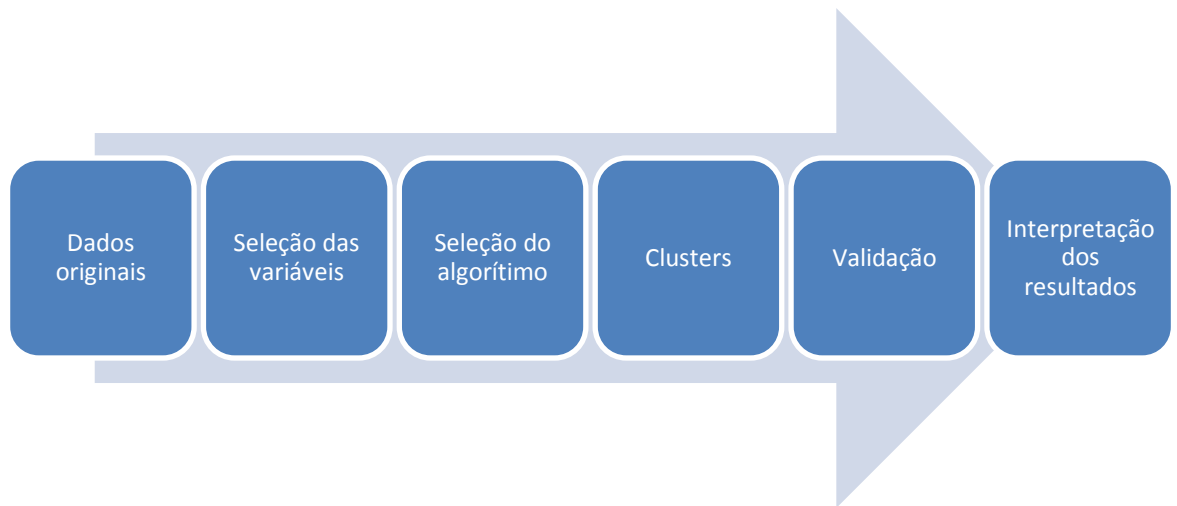


Figura 4- Etapas de formação de cluster

O método hierárquico pode ser aglomerativo ou divisivo. Esse método constrói uma hierarquia da partição. Dentro do protocolo do método aglomerativos, cada elemento começa no seu próprio grupo e, posteriormente, novos agrupamentos são formados a partir da similaridade. Assim, os grupos vão se formando por proximidade.

No método divisivo primeiramente as observações começam em um único conglomerado a partir das observações mais distantes e a partir das etapas posteriores formam finalmente um grupo isolado. (FÁVERO, 2009). Ao optar pela técnica de análise de agrupamento, também é necessário usar uma medida de distância. A mais usual e a que essa pesquisa utilizará é a distância euclidiana, que é obtida de uma distância mais geral (Distância de Minkowski).

$$d_{ab} = \left[\sum_{j=1}^p (X_{aj} - X_{bj})^2 \right]^{1/2}$$

$$p = 1, 2, \dots, j;$$

X_{aj} = valor da variável j para o indivíduo a ;

X_{bj} = valor da variável j para o indivíduo b .

$$d_{ab} = [(X_a - X_b)' \cdot (X_a - X_b)]^{1/2}$$

$X_a = [X_{a1} \ X_{a2} \ \dots \ X_{ap}]'$ = vetor de características do indivíduo a ;

$X_b = [X_{b1} \ X_{b2} \ \dots \ X_{bp}]'$ = vetor de características do indivíduo b .

Um dos momentos importantes do processo é escolher adequadamente o tipo de ligação entre os elementos que irá se estabelecer no conjunto de dados. As principais ligações são:

- **Simple:** Baseia-se na distância mínima ou vizinho mais próximo. Assim, a ligação é dada pela menor distância. É muito sensível a dados *outliers*.
- **Completa:** Baseia-se na distância máxima ou vizinho mais distante. Assim, a ligação é dada pela maior distância entre dois objetos. É mesmo sensível a *outliers*. É possível a formação de grupos mais semelhantes entre si.
- **Média:** A distância é medida pela média das distâncias de cada objeto
- **Centróide:** A distância é medida pelos vetores de médias dos clusters. É um método eficiente quando se tem um número de observações atípicas elevadas
- **Ward:** É um método derivado do centróide e forma grupos com números de observações mais próximos.

Para as variáveis quantitativas ou qualitativas, os métodos de ligação simples, completa e da média são indicados. Alguns métodos são específicos para as variáveis quantitativas, pois possuem como base a comparação de vetores de média. Entre esses métodos pode se citar o centróide e de Ward. (MINGOTI, 2013).

Essa pesquisa utilizará o método de Ward, pois, de acordo com a literatura ele tende a gerar grupos com aproximadamente o mesmo número de componentes. O método tem a tendência a produzir agrupamentos de tamanhos iguais, convexos e compactos, além de realizar um trabalho melhor de redução da heterogeneidade, isto é, da variância da partição, (LATTIN et al, 2011).

Para Rogerson (2012) o método de Ward escolhe aquela fusão que resulta no menor aumento na soma de quadrados internos, assim a variabilidade dentro do grupo permanece a menor possível. Reforçando a ideia de acordo com (Hair et al, 200), o método Ward realizar uma combinação baseada na minimização da soma interna de quadrados no conjunto completo de agrupamentos separados ou disjuntos.

Os métodos não hierárquicos necessitam de uma referência sequencial com a seleção de pontos sementes. O algoritmo mais comum nesse caso é K-means ou K-médias. É um método diferenciado do anterior, pois é necessário realizar a seleção

das sementes de agrupamentos. Portanto dependendo da posição escolhida ocorrerão resultados distintos em função de uma alteração na ordem das observações no banco de dados utilizado. (FÁVERO, 2009).

A escolha da partição final é um dos grandes gargalos da análise de agrupamentos. Existem alguns métodos para determinar a partição de grupos pertinente. O coeficiente de determinação apresenta uma medida de decisão. Ele representa a proporção da variabilidade total explicada pela partição em grupos dado pela fórmula:

$$R^2 = \frac{SSB}{SSTc}$$

Onde,
SSB: Variabilidade entre os conglomerados
SSTc: Variabilidade total

O coeficiente de determinação R^2 representa a proporção da variabilidade total explicada pela partição em grupos. Outro critério de decisão passível de ser utilizado é o pseudo-F F. Trata-se de uma medida que capta a análise de custo-benefício entre a simplicidade e adequação. Essa medida baseia-se em uma razão da soma média de quadrados entre os grupos pela soma média de quadrados dentro do grupo. (LATTIN et al, 2011).

O pseudo-F é definido como:

$$PF = \left(\frac{n - g^*}{g^* - 1} \right) \left(\frac{R^2}{1 - R^2} \right)$$

Onde.
n= Número de grupos
g*= Número de clusters da partição
R²= Variabilidade total explicada pela divisão

Após o cálculo do Pseudo-F- F deve-se buscar o maior valor, pois ele estaria relacionado com a menor probabilidade de significância do teste o que rejeita também a igualdade de vetores de médias populacionais (dos grupos) com maior significância. Assim, o resultado obtido seria a partição com maior heterogeneidade entre grupos. (MINGOTI, 2013).

O BIC - Critério de Informação Bayesiana é o valor do log-verossimilhança maximizada com uma penalidade sobre os números parâmetros do modelo. Ele realiza cálculo de diferentes modelos, o que permite a comparação de modelos de

probabilidade com diferentes parametrizações ou diferentes números de agrupamentos. Assim, o pacote *mclust* permite a especificação dos modelos e números de aglomerados. (FRALEY, 2012). Assim, o BIC pode ser representado como,

$$\text{BIC} = 2\text{loglik}_M(x, \theta * _M) - (\#\text{params})_M * \log(n), \text{ onde:}$$

$\text{loglik}_M(x, \theta * _M)$, é o valor do log-verossimilhança maximizado para o modelo e os dados

$(\#\text{params})_M$, número de parâmetros independentes para ser avaliado no modelo

n , é o número de observações

É realizado um agrupamento hierárquico, baseado em estrutura de covariâncias. Primeiro é gerado o máximo número de clusters até que se chegue em um único conglomerado. É possível utilizar o BIC no software R com o seguinte script:

```
# Model Based Clustering
>library(mclust)
>fit <- Mclust(dados)
>plot(fit) # plot results
>summary(fit) # display the best model
> summary(dados, parameters = TRUE)
>plot(dados)
```

No quesito validação do agrupamento, o resultado final é específico para um conjunto particular de objetos ou indivíduos. Assim, a validação nem sempre é pertinente, não é conveniente usar a solução de agrupamentos para dizer qualquer coisa sobre um elemento que pertencente a um agrupamento de fora da amostra. (LATTIN, et al, 2011).

Uma das principais desvantagens dessa técnica análise de clusters é a tentativa de tratar observações de diferentes medidas e a possibilidade de perda de informação. Destaca-se que ainda o aumento indesejado do número de variáveis, a ineficiência quando grande volume de informações precisa ser agrupado, a impossibilidade de se consolidar resultados discordantes e a dificuldade na determinação dos pesos da medida de proximidade conjunta. (MATOS, 2007; MINGOTI, 2013).

Fávero (2009) destaca que os outliers podem interferir nos resultados. Trata-se de uma técnica sensível à inclusão de variáveis com comportamento atípico. Portanto

o pesquisador deve decidir a manutenção ou não desses *outliers* no seu banco de dados.

Ressalta-se que vários métodos de análise de cluster focam, principalmente, em dados caracterizados por variáveis contínuas. Entretanto, na presença de dados categóricos, algumas aproximações são utilizadas, como a transformação destas em contínuas, atribuindo valores numéricos às suas categorias, ou em binárias, fazendo com que cada uma das suas categorias se torne uma variável que represente presença ou ausência desse determinado atributo (MATOS, 2007).

4.2. Bancos de Dados

A base de dados real utilizada neste trabalho é intitulada Acesso à Internet e à Televisão e Posse de Telefone Móvel Celular para Uso Pessoal, é oriunda da PNAD-2014. As unidades de análise são Unidades da federação, vinte seis ao todo mais o Distrito Federal. As variáveis desta base de dados referem-se ao uso, no domicílio, de internet, rádio, televisão, telefone fixo convencional e telefone fixo móvel, e à presença de micro computador, tablet, banda larga fixa e móvel, dentre outras variáveis. O IBGE seleciona essas variáveis a partir da sugestão da agência da ONU especializada em tecnologias de informação e comunicação, União Internacional de Telecomunicações (UIT, em inglês: *International Telecommunication Union - ITU*).

Essa pesquisa optou por trabalhar exclusivamente com os dados referentes à internet para cada vinte e seis Unidades da Federação (UF) além do Distrito Federal, ver Tabela 1. Foram selecionadas as seguintes dez variáveis sobre utilização da internet:

- Var. 1:** Número de pessoas de 10 anos ou mais de idade que utilizaram a Internet, no período de referência dos últimos três meses;
- Var. 2:** Rendimento médio mensal domiciliar per capita das pessoas de 10 anos ou mais de idade, por utilização da Internet, no período de referência dos últimos três meses;
- Var. 3:** Número de domicílios particulares permanentes com utilização da Internet por banda larga;
- Var. 4:** Número de moradores em domicílios particulares permanentes com utilização da Internet por banda larga;
- Var. 5:** Número de domicílios particulares permanentes com utilização da Internet;
- Var. 6:** Número de moradores em domicílios particulares permanentes com utilização da Internet;

Var. 7: Número de domicílios particulares permanentes com utilização da Internet por telefone móvel celular;

Var. 8: Número de moradores em domicílios particulares permanentes com utilização da Internet por telefone móvel celular;

Var. 9: Número de moradores em domicílios particulares permanentes com utilização da Internet por telefone móvel celular ou tablet.

Var. 10: Número de pessoas de 10 anos ou mais de idade, com ensino médio completo, que utilizaram a Internet no período de referência dos últimos três meses.

Tabela 1-Banco de dados das variáveis selecionadas

Estado	Var1	Var2	Var3	Var4	Var5	Var6	Var7	Var8	Var9	Var10
Rondônia	748	1168	260	846	190	614	82	267	167	218,9
Acre	288	1091	88	310	55	193	37	129	71	85,8
Amazonas	1631	1097	517	1985	277	1059	241	925	624	562,4
Roraima	235	1114	87	293	42	134	46	162	77	76
Pará	2593	885	914	3490	380	1394	537	2109	1151	870,8
Amapá	353	987	96	412	50	209	46	206	138	121,3
Tocantins	565	1171	166	557	117	397	60	207	115	182,1
Maranhão	1789	947	583	2177	301	1062	294	1164	698	651,6
Piauí	987	947	346	1280	176	625	164	629	305	275,8
Ceará	3131	893	1082	3824	706	2438	371	1373	748	1046,7
Rio Grande do Norte	1439	936	526	1812	346	1175	165	577	378	458,8
Paraíba	1581	950	590	2091	409	1415	166	617	372	474,6
Pernambuco	3443	955	1274	4392	969	3328	276	957	682	1134,8
Alagoas	1120	762	370	1345	254	882	115	459	266	333,3
Sergipe	866	852	351	1260	181	632	171	634	340	240
Bahia	5790	1054	2127	6995	1529	4941	581	2001	1258	2039,7
Minas Gerais	9756	1321	3753	12297	2939	9494	810	2777	1724	3064,8
Espírito Santo	1915	1340	758	2413	597	1887	159	526	351	618,9
Rio de Janeiro	9127	1609	3483	10389	3040	9006	493	1514	1387	3085,2
São Paulo	25310	1636	9787	31434	8509	27165	1341	4479	3562	8727,6
Paraná	5705	1494	2098	6626	1857	5839	283	930	750	1759,1
Santa Catarina	3427	1655	1364	4217	1201	3710	175	547	440	1037,7
Rio Grande do Sul	5708	1598	2353	7162	1917	5752	444	1437	930	1698,7
Mato Grosso do Sul	1349	1483	487	1548	351	1126	140	440	317	350,2
Mato Grosso	1515	1409	524	1691	398	1274	154	503	328	469,9
Goiás	3111	1248	1087	3656	828	2782	291	989	692	907,7
Distrito Federal	1875	2518	698	2218	600	1907	92	294	209	533,3

Fonte: PNAD: Acesso à Internet e à Televisão e Posse de Telefone Móvel Celular para Uso Pessoal, 2014.

5. Resultados e Discussão

5.1. Análise Descritiva

Inicia-se a análise apresentando a Tabela 2 com as estatísticas descritivas das variáveis selecionadas.

Tabela 2- Estatísticas descritivas das variáveis selecionadas

Variável	Média	D.P.	C.V.	Mínimo	Q1	Mediana	Q3	Máximo	Curtose
Var. 1	3532	5025	142,29	235	987	1789	3443	25310	13,96
Var. 2	1226,7	372,8	30,39	762	947	1114	1483	2518	4,2
Var. 3	1325	1952	147,33	87	351	590	1364	9787	13,99
Var. 4	4323	6210	143,66	293	1280	2177	4392	31434	14,45
Var. 5	1045	1705	163,16	42	190	398	1201	8509	14,62
Var. 6	3350	5393	161,01	134	632	1394	3710	27165	15,24
Var. 7	286,4	282,6	98,67	36,6	115,5	171	371,2	1340,9	6,85
Var. 8	995	956	96,13	129	440	629	1373	4479	6,07
Var. 9	670	718	107,17	71	266	378	750	3562	9,83
Var. 10	1149	1723	149,91	76	276	562	1135	8728	14,98

DP:desvio-padrão; C.V.: coeficiente de variação; Q1: primeiro quartil; Q3: terceiro quartil.

É necessário chamar a atenção que as unidades de medida das variáveis são diferentes, o que reflete nos valores observados. A variável 1 (Pessoas de 10 anos ou mais de idade, que utilizaram a Internet, no período de referência dos últimos três meses) apresenta média de 3.532.

Os desvios-padrão são elevados e a outra medida de dispersão, coeficiente de variação é bastante elevado também. O menor coeficiente encontra-se na variável 2 (Rendimento médio mensal domiciliar per capita das pessoas de 10 anos ou mais de idade, por utilização da Internet, no período de referência dos últimos três meses) 30,39. No geral, na maioria das variáveis selecionadas o coeficiente é alto, representando uma heterogeneidade dos dados. Na Figura 5 são apresentados os histogramas das variáveis.

A Figura 6 mostra os boxplots das variáveis. A variável 2 Rendimento médio mensal domiciliar per capita das pessoas de 10 anos ou mais de idade, por utilização da Internet, no período de referência dos últimos três meses) apresenta um outliers que é o Distrito Federal. Grande parte das variáveis apresentam valores discrepantes. Esse aspecto é bastante comum quando a unidade de análise trata-se das Unidades da Federação do país. A variável 10 (no. de pessoas que utilizaram a Internet) apresenta como discrepantes São Paulo e Minas Gerais.

Histograma das variáveis selecionadas

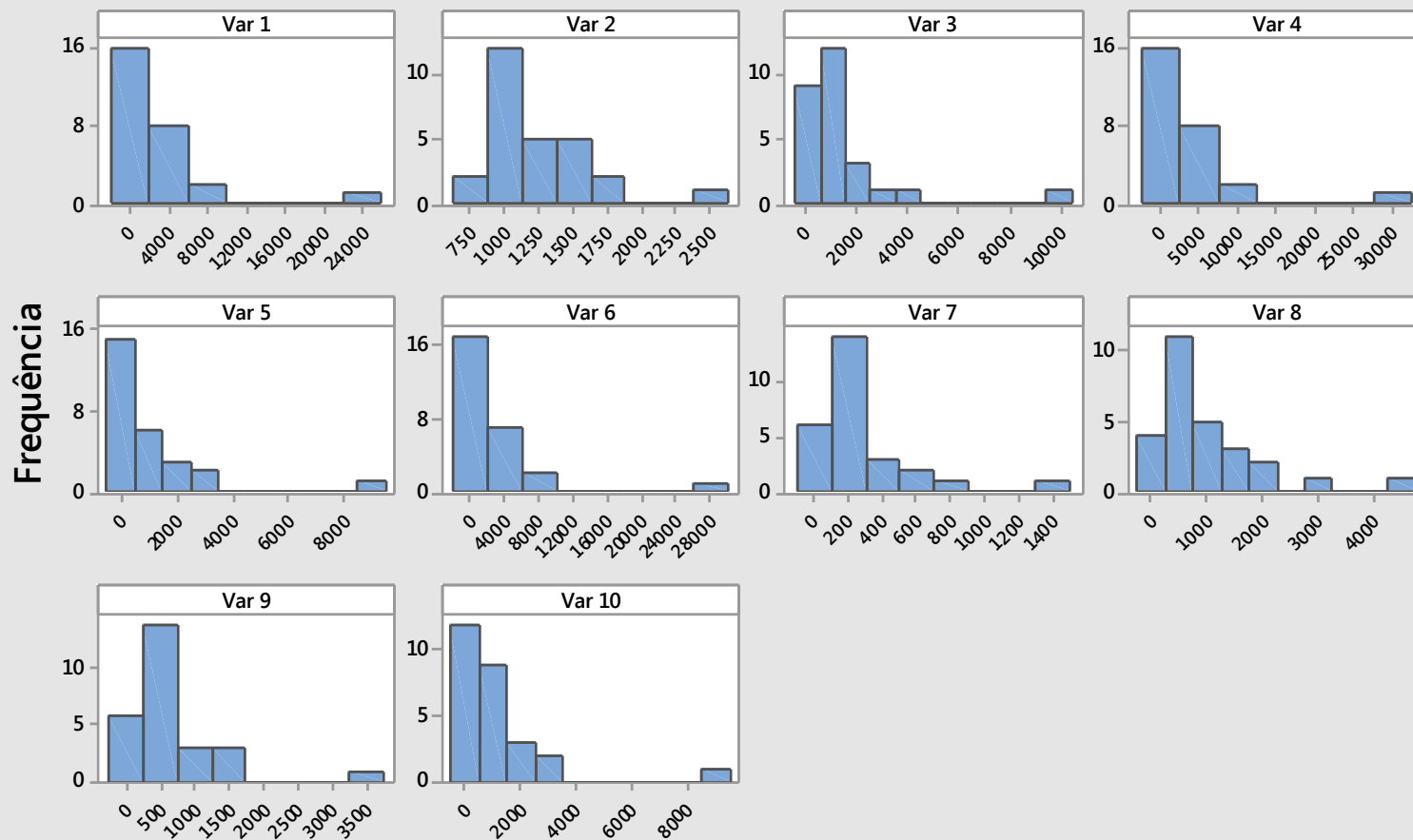
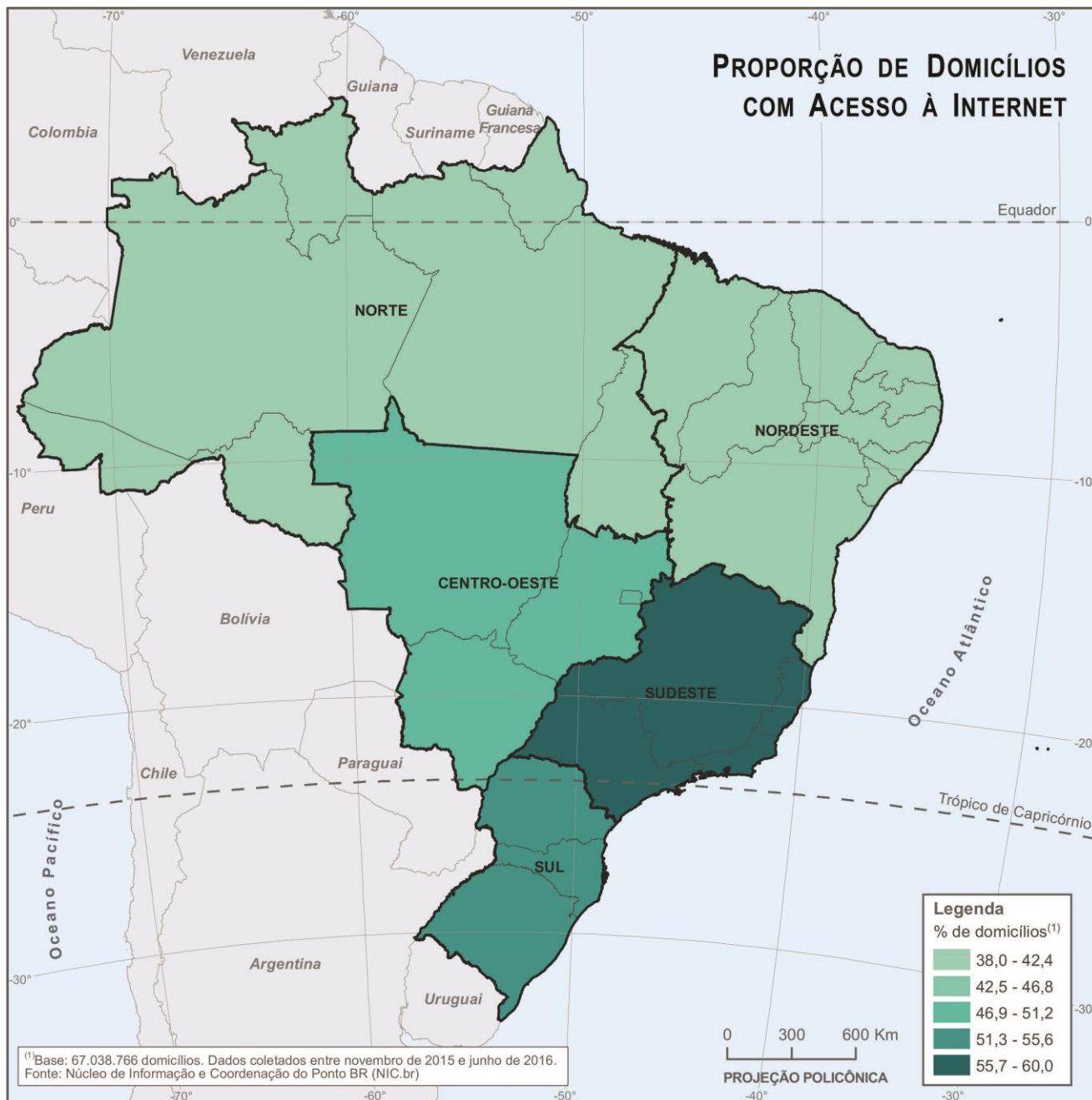


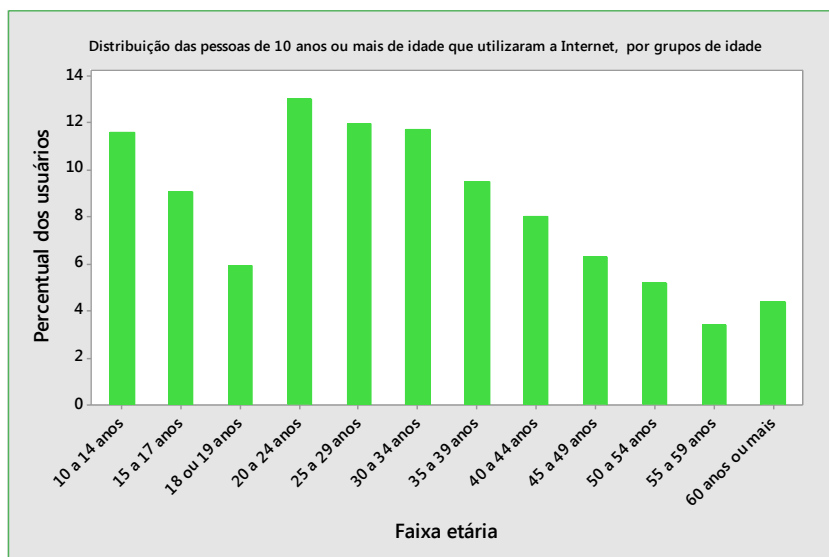
Figura 5- Histograma das dez variáveis selecionadas.



Percentual sobre o total de domicílios

Base: 67.038.766 domicílios. Dados coletados entre Novembro de 2015 a Junho de 2016.

Figura 7-Distribuição espacial do percentual de domicílios com acesso à internet nas UFs.



Fonte: IBGE, Diretoria de Pesquisas, Coordenação de Trabalho e Rendimento, Pesquisa Nacional por Amostra de Domicílios 2014.

Figura 8- Distribuição das pessoas de 10 anos ou mais de idade que utilizaram a internet, por grupo de idade.

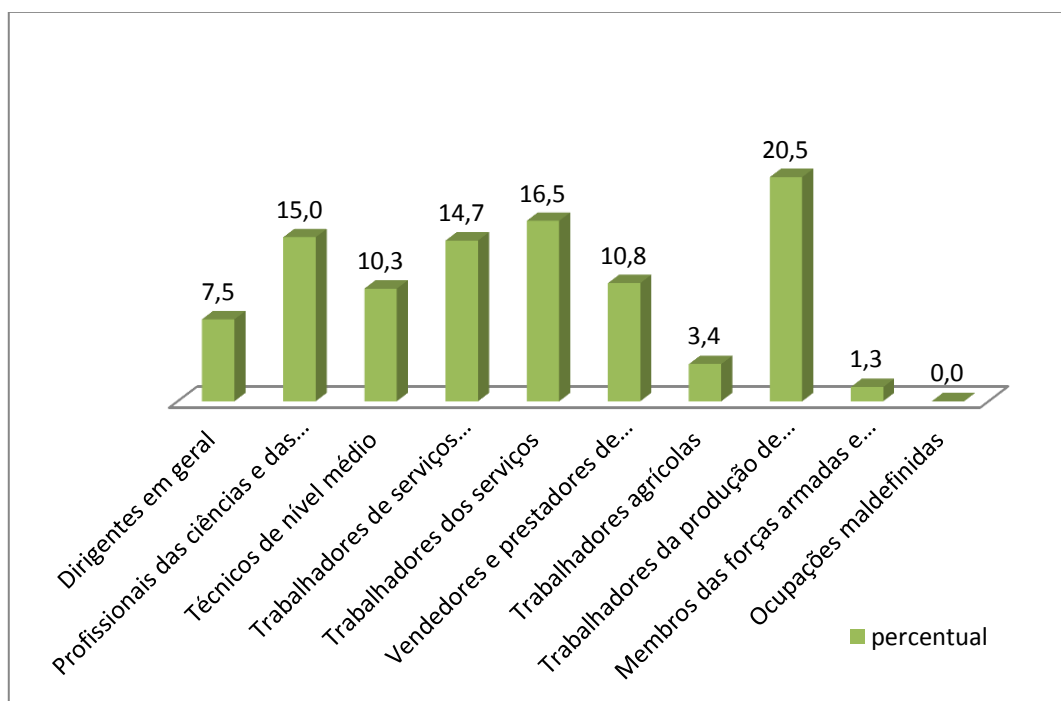


Figura 9- Distribuição percentual das pessoas de 10 anos ou mais de idade, ocupadas na semana de referência, que utilizaram a Internet, segundo agrupamentos ocupacionais no trabalho principal- 2014

Nesse contexto, é possível dizer que o uso das TDICs, faz-se presente no cotidiano da população, mas também estão associadas a algumas características como à renda, e outras diferenças adquiridas pelos indivíduos (escolaridade e

experiência), diferenças de remuneração no emprego, a faixa etária, ao tipo de ocupação dentre outros fatores. Ver Tabelas 3 e 4 e Figura 10.

Tabela 3-Percentual de pessoas que utilizaram a Internet, no período de referência dos últimos três meses, na população de 10 anos ou mais de idade, por Grandes Regiões, segundo os grupos de idade - 2014.

grupos de idade	Percentual de pessoas que utilizaram a Internet, no período de referência dos últimos três meses, na população de 10 anos ou mais de idade (%)					
	Brasil	Grandes Regiões				
		Norte	Nordeste	Sudeste	Sul	Centro-Oeste
Total	54,4	45,2	42,1	61,8	58,2	60,0
10 a 14 anos	69,1	46,6	55,0	81,2	80,5	81,0
15 a 17 anos	81,8	67,5	71,8	90,6	87,4	87,2
18 ou 19 anos	81,1	70,2	71,2	88,4	87,0	86,6
20 a 24 anos	78,5	67,0	66,4	86,3	84,0	85,7
25 a 29 anos	72,7	59,1	57,7	82,4	80,1	78,8
30 a 34 anos	67,4	52,3	50,6	77,2	76,4	72,5
35 a 39 anos	60,0	47,4	41,5	70,3	68,4	65,9
40 a 44 anos	53,4	40,1	35,7	63,6	60,3	55,7
45 a 49 anos	45,1	32,6	28,3	54,9	49,3	48,6
50 a 54 anos	39,1	28,0	23,9	48,5	40,9	39,4
55 a 59 anos	31,3	19,3	19,3	38,6	33,0	29,9
60 anos ou mais	14,9	8,0	7,3	19,5	16,5	14,9

Fonte: IBGE, Diretoria de Pesquisas, Coordenação de Trabalho e Rendimento, Pesquisa Nacional por Amostra de Domicílios 2014.

Tabela 4-Distribuição das pessoas de 10 anos ou mais de idade que utilizaram a Internet, no período de referência dos últimos três meses, por Grandes Regiões e os grupos de anos de estudo - 2014

Grupos de anos de estudo	Distribuição das pessoas de 10 anos ou mais de idade que utilizaram a Internet, no período de referência dos últimos três meses (%)					
	Brasil	Grandes Regiões				
		Norte	Nordeste	Sudeste	Sul	Centro-Oeste
Total	100,0	100,0	100,0	100,0	100,0	100,0
Sem instrução e menos de 1 ano	0,8	1,1	0,8	0,7	1,0	1,1
1 a 3 anos	3,8	3,9	4,4	3,4	3,8	4,1
4 a 7 anos	18,4	19,8	21,4	16,8	17,8	19,6
8 a 10 anos	19,5	21,1	20,1	18,7	20,6	19,3
11 a 14 anos	41,4	41,6	41,2	42,5	39,8	38,4
15 anos ou mais	16,1	12,5	12,0	17,9	16,8	17,4
Não determinados	0,1	0,1	0,2	0,0	0,1	0,1

Fonte: IBGE, Diretoria de Pesquisas, Coordenação de Trabalho e Rendimento, Pesquisa Nacional por Amostra de Domicílios 2014.

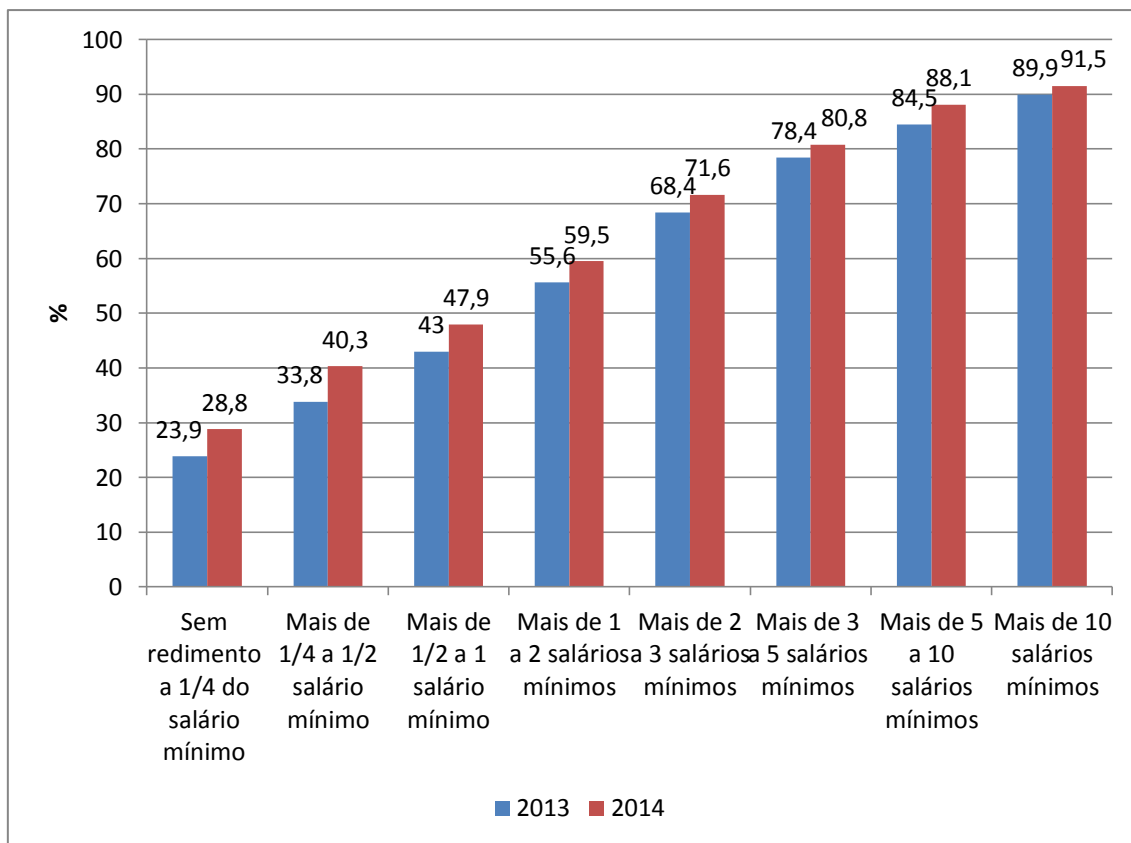


Figura 10- Percentual de pessoas que utilizaram a internet, no período de referência dos últimos três meses, na população de 10 anos ou mais de idade, segundo as classes de rendimento mensal domiciliar per capita Brasil- 2013-2014

A Figura 11 apresenta o percentual do tipo de equipamento utilizado pelos domicílios para realizar o acesso à internet, segundo as regiões do país.

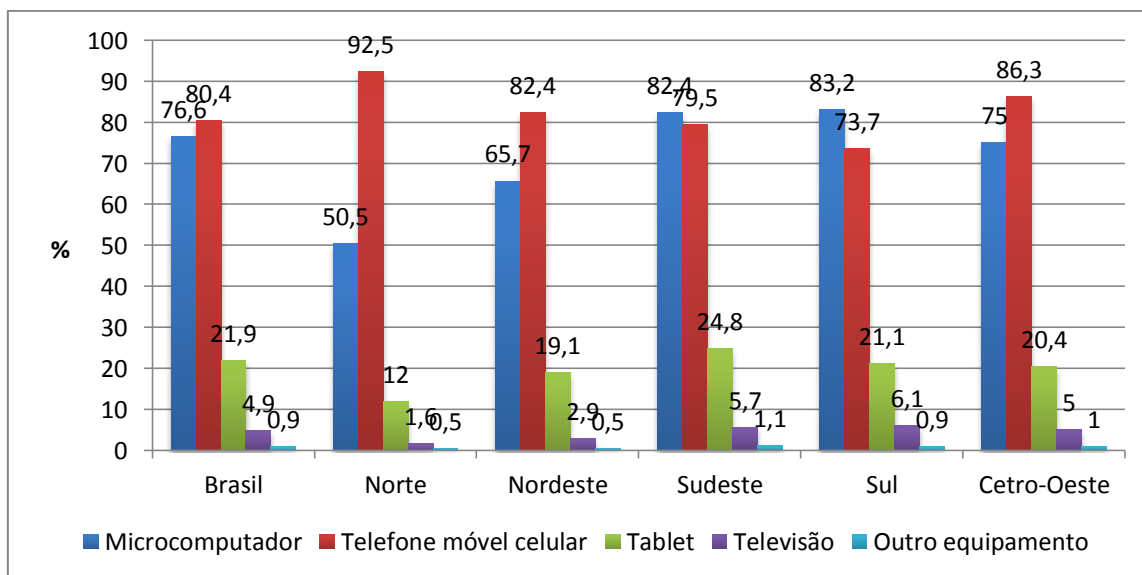


Figura 11- Percentual de domicílios com utilização da internet, por tipo de equipamento utilizado para acessar a internet, no total de domicílios particulares permanentes com utilização da internet, segundo as Grandes Regiões – 2014

A utilização do microcomputador recebe destaque nas Regiões Sudeste e Sul, 82,4% e 83,2 % respectivamente. Nas demais regiões predomina o acesso via telefone móvel celular, mostrando a popularização desse tipo de equipamento no Brasil.

A Figura 12 confirma a disseminação do uso do telefone móvel celular para ao acesso a internet.

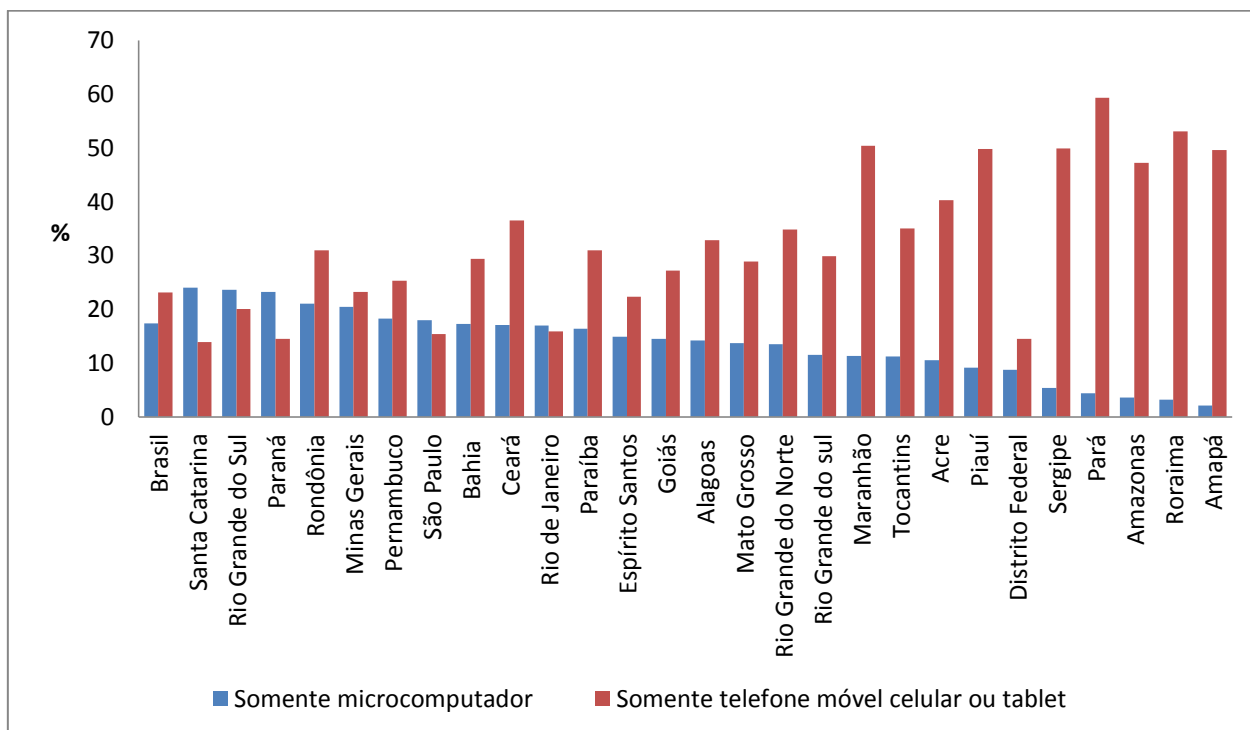


Figura 12- Percentual de domicílios com utilização da Internet somente por meio de microcomputador e somente por meio de telefone móvel celular ou tablete, no total de domicílios particulares permanente com utilização da Internet, segundo as Unidades da Federal em 2014

Verifica-se que vinte e dois estados e o Distrito Federal utilizam o telefone móvel celular ou tablet para o acesso a internet e quatro estados utilizam apenas o microcomputador para o acesso à internet são eles: Santa Catarina, Rio Grande do Sul, Paraná e Rio de Janeiro.

Quanto ao acesso de acordo com o sexo, apresenta-se a Figura 13 em seqüência. Dentro da amostra selecionada, as mulheres são as que mais acessam no geral (54,7%) e Homens (54,4%). As mulheres também representam a maioria nas Regiões Centro-Oeste (60,7%), Nordeste (43,3%) e Norte (47,1%).

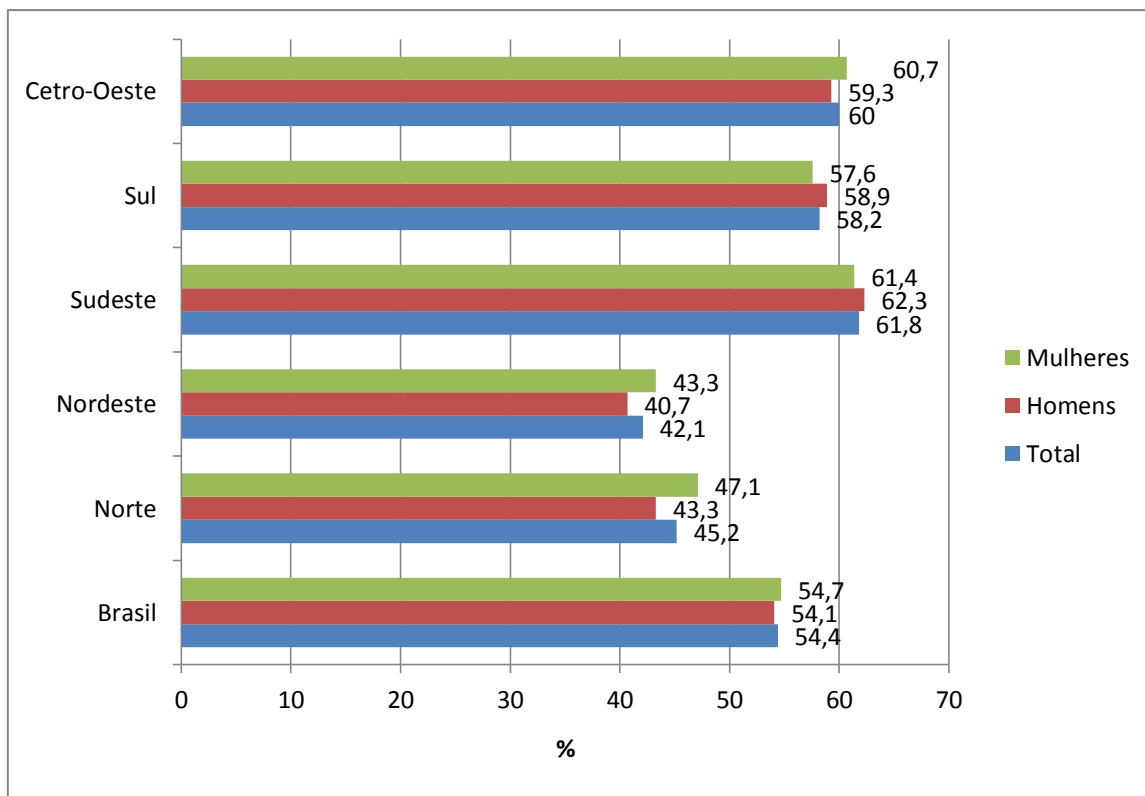


Figura 13- Percentual de pessoas que utilizaram a internet, no período de referência dos últimos três meses, na população de 10 anos ou mais de idade, por sexo, segundo as Grandes Regiões- 2014.

5.2. Análise de Clusters

A ampliação do acesso as TDIC, o diagnóstico e a analisar a dinâmica da internet por UFs auxilia no entendimento dos pontos fracos dessa dinâmica, percebendo quais são os desafios e onde é necessário avançar para melhorar o acesso da população. A Análise de Clusters é uma técnica que ajuda nesse processo.

A partir do referencial teórico, analisado optou-se: pelo método hierárquico, por não conhecer a partição (número de grupos) ideal; pela distância euclidiana, por ser a mais utilizada; e pelo método de ligação de Ward, por garantir maior homogeneidade entre os grupos formados. Os resultados da análise realizada são mostrados no Dendograma da Figura 14 e no Tabela 5.

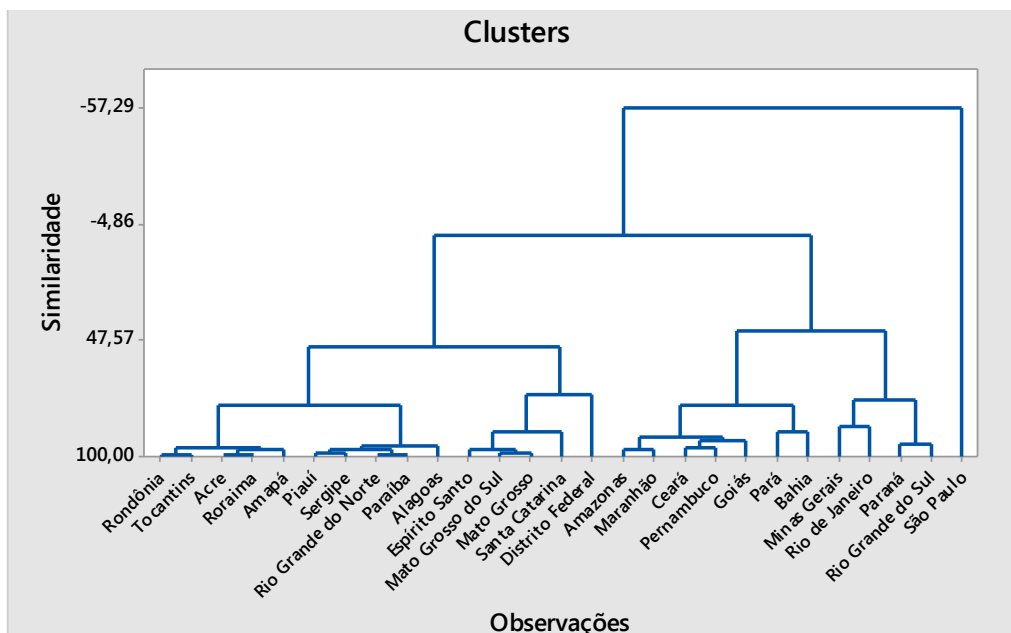


Figura 14- Dendrograma: agrupamento das UF's.

Tabela 5- Resultados da análise de clusters

Passo	Número de agrupados	Nível de similaridade	Nível de distância	Agrupados reunidos		Novo agrupado	Número de obs. no novo agrupado
1	26	99,4573	0,0800	2	4	2	2
2	25	99,3020	0,1029	11	12	11	2
3	24	98,9332	0,1572	1	7	1	2
4	23	98,4110	0,2342	24	25	24	2
5	22	98,2174	0,2627	9	15	9	2
6	21	97,1707	0,4169	2	6	2	3
7	20	96,6810	0,4891	18	24	18	3
8	19	96,5901	0,5025	9	11	9	4
9	18	96,4317	0,5258	3	8	3	2
10	17	95,8600	0,6101	1	2	1	5
11	16	95,6298	0,6440	10	13	10	2
12	15	95,3656	0,6829	9	14	9	5
13	14	94,0292	0,8799	21	23	21	2
14	13	92,8158	1,0587	10	26	10	3
15	12	91,4195	1,2644	3	10	3	5
16	11	89,0952	1,6069	18	22	18	4
17	10	88,8049	1,6497	5	16	5	2
18	9	86,4821	1,9920	17	19	17	2
19	8	77,0966	3,3751	3	5	3	7
20	7	76,5508	3,4555	1	9	1	10
21	6	74,4463	3,7656	17	21	17	4
22	5	71,9566	4,1325	18	27	18	5
23	4	50,2385	7,3329	1	18	1	15
24	3	43,6322	8,3064	3	17	3	11
25	2	0,1387	14,7157	1	3	1	26
26	1	-57,2923	23,1788	1	20	1	27

		Dentro da soma de	Distância	Distância
		quadrados	média do	máxima do
		do	centróide	centróide
Agrupado	Número de observações	agrupado		
1	27	260	2,18188	12,6216

A análise do Dendograma aponta para a formação de aproximadamente seis a nove grupos, que serão confirmados por medidas do R² e o pseudo-F. A área delimitada de amarelo corresponde aos agrupamentos que serão testados para os cálculos do R² e Pseudo-F com o auxílio da soma do quadrado total expresso em sequência. A análise do nível de similaridade nos permite concluir que o intervalo que varia entre 6 a 9 grupos é um marco na queda dos valores apresentados.

A Tabela 6 em sequência apresenta os resultados do R² e pseudo-F para as partições selecionadas para a análise: 9, 8, 7 e 6.

Tabela 6- Medidas para a tomada de decisão

Número de clusters	Soma dos Quadrados Total	Soma dos Quadrados entre clusters	R ²	Pseudo F
9	260	7,08806	0,97	72,75
8	260	11.30345	0,96	65,14
7	260	13.41435	0,95	63,33
6	260	19.51918	0,93	55,80

A análise do nível de similaridade nos permite concluir que a partição 9 apresenta o maior R² (94%), valor também próximo da partição 8 (92%). A distância aumenta consideravelmente da partição 9 para a 8. De 2,5071 passa para 3,5118. O pseudo F da partição 9 é de 32,91 e da partição 8 é 29,79. A partir da análise realizada optou-se pelo K= 9.

Com forma de validação do número de clusters gerado utilizou o BIC - Critério de Informação Bayesiana gerado a partir no software R a partir do pacote *mclust*. Esse pacote fornece, além do BIC, também a possibilidade de realizar uma análise discriminante dos dados. O principal resultado gerado foi o descrito em sequência:

```
> summary(fit) # display the best model
-----
Gaussian finite mixture model fitted by EM algorithm
-----

Mclust EEE (ellipsoidal, equal volume, shape and orientation) model with 9 comp$

log.likelihood  n  df          BIC          ICL
-1456.882 27 173 -3483.943 -3483.965

Clustering table:
 1  2  3  4  5  6  7  8  9
12  4  1  4  1  1  1  1  2
```

Ele mostra que a melhor partição para esse banco de dados deve ser em 9 componentes (clusters), pois temos nesse caso o maior valor de BIC.

Assim, neste caso, o melhor modelo de acordo com a BIC é um modelo igual covariância com 9 componentes ou clusters. Opta-se pelo maior BIC sugerido que é um modelo de mistura multivariada Mclust EEE (elipsoidal, igual volume, corpo e orientação), com nove componentes. Apenas para efeito de constatação segue o quadro com os valores de BIC, gerados a partir do código (:

```
> dadosBIC = Mclust(dados, G = número de clusters 1 a 8)
> summary (dadosBIC)
```

Tabela 7- Valores dos Critérios de informação Bayesiana

Número de clusters	Valor do BIC
1	-3541.103
2	-3557.656
3	-3576.723
4	-3508.64
5	-3516.896
6	-3540.164
7	-3541.097
8	-3492.058
9	-3483.943

A Figura 15 apresenta o Dendrograma com a partição em 9 clusters e a Figura 16 o mapa elaborado a partir do número de clusters sugeridos.

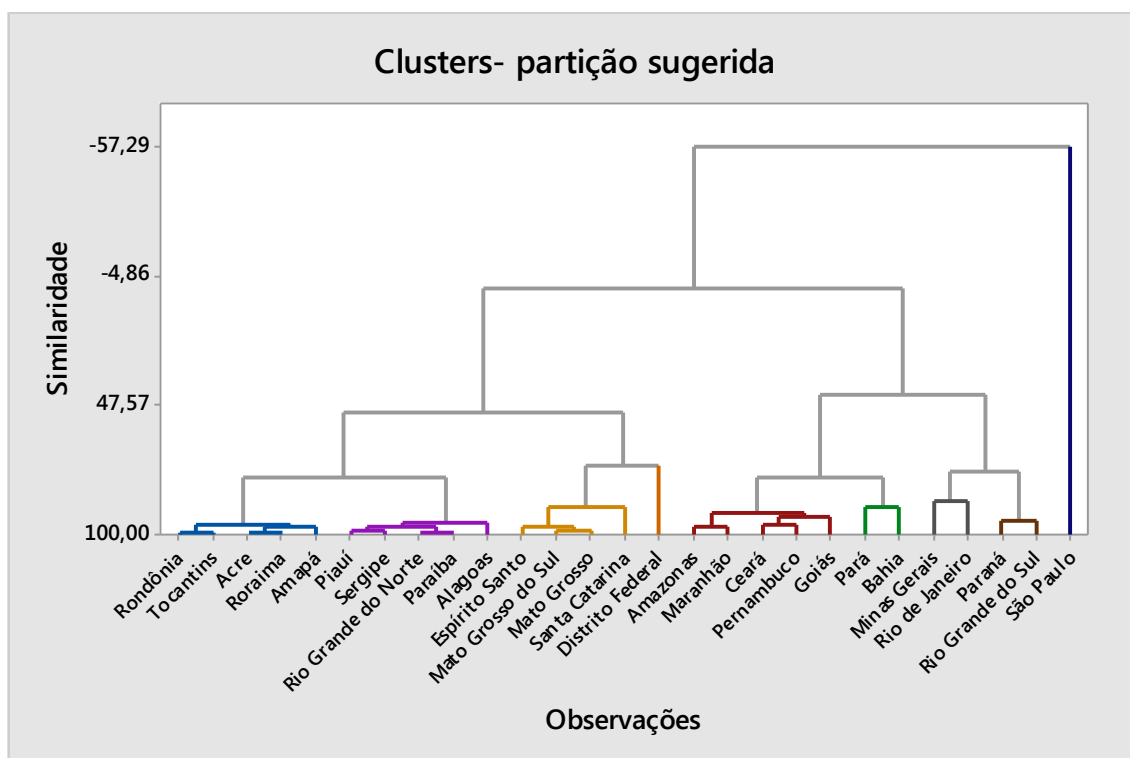


Figura 15- Dendrograma da partição final com nove clusters.

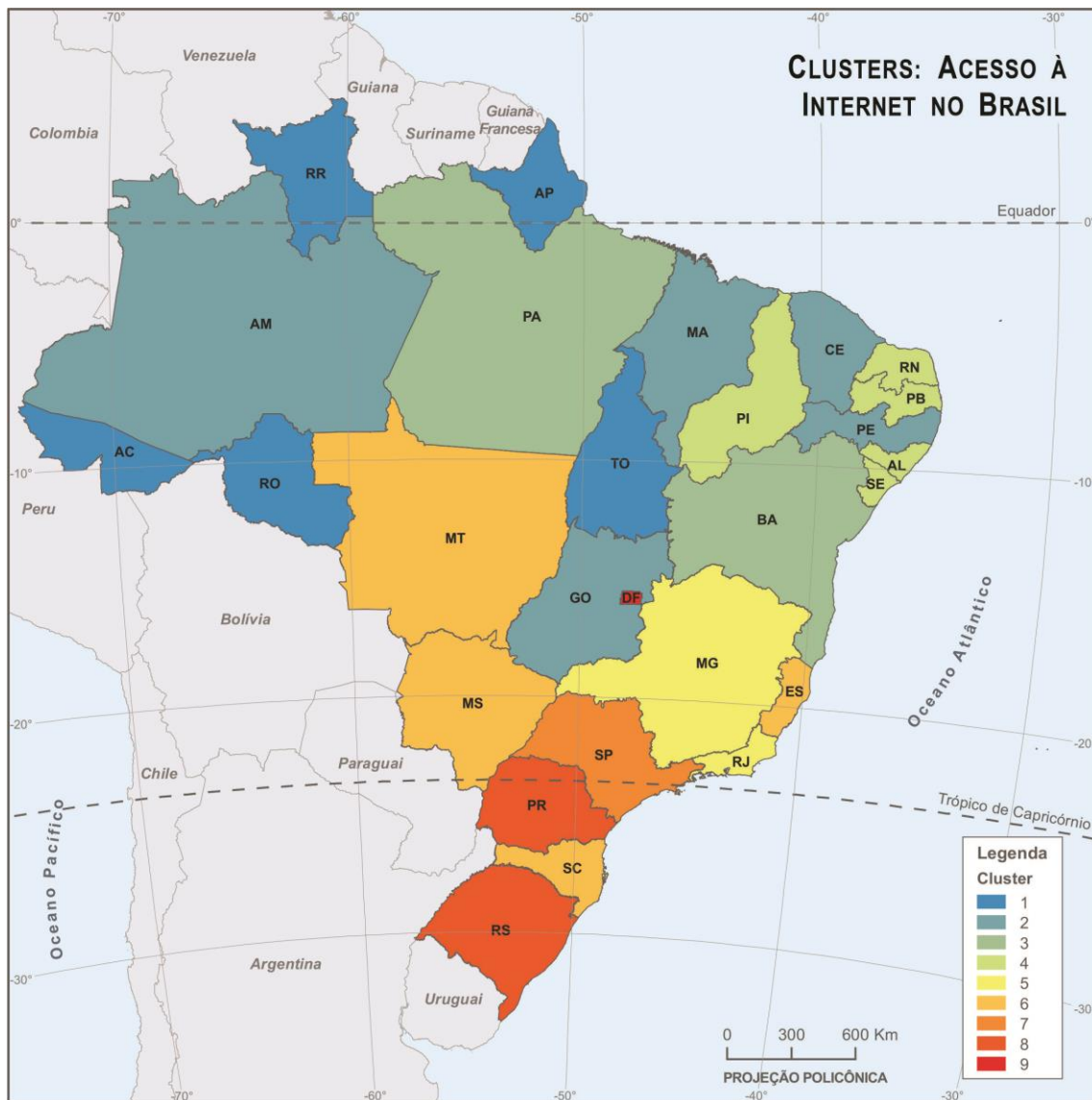


Figura 16- Clusters: Acesso a internet no Brasil

A Tabela 8 apresenta as descritivas dos clusters. A partir dos resultados gerados, é possível traçar os seguintes perfis para os grupos:

Cluster 1: Acre, Rondônia, Roraima, Amapá e Tocantins.

O primeiro agrupamento é formado por um dos menores quantitativos de pessoas que utilizaram a internet no período de referência e os domicílios utilizam pouco acesso à banda larga para acessar a internet. Grande parte do acesso ocorre com o auxílio de telefone celular ou tablet. Para esse grupo, é necessário elevar o número de pessoas com acesso à internet, melhorar o acesso à banda larga, pois significa uma melhora na condição de conexão.

Tabela 8- Descritivas dos Clusters

Grupos	Estados	Descritivas	Var.1	Var.2	Var.3	Var.4	Var.5	Var.6	Var.7	Var.8	Var.9	Var.10
1	Acre, Rondônia, Roraima, Amapá e Tocantins	Média	437,8	1106,2	139,4	484	90,8	309,4	54,2	194,2	113,6	136,8
		Desvio Padrão	214	75	75	228	63	196,8	17,58	52,2	40,6	61,9
		Coeficiente de variação	48,88	6,78	53,83	47,19	69,34	63,59	32,44	26,87	35,77	45,25
2	Amazonas, Maranhão, Ceará, Goiás e Pernambuco	Média	2621	1028	909	3207	616	2134	294,6	1081,6	688,8	861
		Desvio Padrão	844	144,3	337	1065	313	1030	47,6	187,3	44,3	247
		Coeficiente de variação	32,2	14,04	37,11	33,22	50,79	48,27	16,16	17,32	6,43	28,74
3	Pará Bahia	Média	4192	969,5	1521	5243	955	3168	559	2055	1204,5	1455
		Desvio Padrão	2261	119,5	858	2478	812	2508	31,1	76,4	75,7	827
		Coeficiente de variação	53,93	12,33	56,41	47,28	85,12	79,18	5,57	3,72	6,28	56,8
4	Piauí, Rio Grande do Norte, Paraíba, Alagoas e Sergipe	Média	1199	889,4	436,6	1558	273,2	946	156,2	583,2	332,2	356,5
		Desvio Padrão	302	81,8	113,5	374	102,5	346	23,2	72,9	47,1	106,1
		Coeficiente de variação	25,22	9,2	25,99	24,02	37,53	36,55	14,85	12,51	14,17	29,76
5	Minas Gerais Rio de Janeiro	Média	9442	1465	3618	11343	2989,5	9250	652	2146	1556	3075
		Desvio Padrão	445	204	191	1349	71,4	345	224	893	238	14,4
		Coeficiente de variação	4,71	13,9	5,28	11,89	2,39	3,73	34,41	41,63	15,32	0,47
6	Espírito Santo, Santa Catarina Mato Grosso e Mato Grosso do Sul	Média	2052	1471,8	783	2467	637	1999	157	504	359	619
		Desvio Padrão	947	135,4	405	1226	391	1187	14,45	46,3	55,8	300
		Coeficiente de variação	46,17	9,2	51,75	49,71	61,4	59,38	9,2	9,19	15,55	48,43
7	São Paulo	Média	25310	1636	9787	31434	8509	27165	1341	4479	3562	8727,6
		Desvio Padrão	*	*	*	*	*	*	*	*	*	*
		Coeficiente de variação	*	*	*	*	*	*	*	*	*	*
8	Rio Grande do Sul Paraná	Média	5706,5	1546	2226	6894	1887	5795,5	363,5	1184	840	1728,9
		Desvio Padrão	2,12	73,5	180	379	42,4	61,5	113,8	359	127,3	42,7
		Coeficiente de variação	0,04	4,76	8,1	5,5	2,25	1,06	31,32	30,29	15,15	2,47
9	Distrito Federal	Média	1875	2518	698	2218	600	1907	92	294	209	533,3
		Desvio Padrão	*	*	*	*	*	*	*	*	*	*
		Coeficiente de variação	*	*	*	*	*	*	*	*	*	*

Cluster 2: Amazonas, Maranhão, Ceará, Pernambuco e Goiás.

Esse grupo apresenta posições inferiores a média relativa ao quantitativo de pessoas e domicílios com acesso a internet. O acesso à internet ocorre, sobretudo, por telefone móvel celular. Deve se elevar o acesso por microcomputador. O baixo uso do microcomputador aqui está relacionado ao custo do equipamento (MATIAS, 2011).

Cluster 3: Pará e Bahia.

Os estados apresentam pessoas que utilizaram a internet acima da média, contudo o número de domicílios particulares permanentes com acesso está abaixo da média. Ampliar o rendimento médio mensal per capita é uma das saídas para ampliar o acesso. Para esses estados o acesso ocorre com o uso do telefone celular na sua maioria.

Cluster 4: Piauí, Rio Grande do Norte, Paraíba, Alagoas e Sergipe.

Esse agrupamento apresenta valores inferiores a médias para todas as variáveis. Observa-se um baixo acesso a internet pelas pessoas e pelos domicílios. O rendimento médio mensal domiciliar nesse grupo precisa ser melhorado para que ocorra uma universalização do acesso à internet que facilitaria, por exemplo, uma maior participação política no planejamento territorial. O nível de instrução da população deve ser melhorado, pois o acesso à internet apresenta relação com essa variável.

Cluster 5: Minas Gerais e Rio de Janeiro

Apresenta bons indicadores como acesso da população à internet e número considerável de domicílios nesse acesso. Todas as variáveis estão acima da média e apresentam baixo coeficiente de variação. A população desses estados acesso mais a internet via microcomputador do que nos grupos analisados anteriormente. Para esse grupo é necessário novos estudos para averiguar as diferenças regionais internas.

Cluster 6: Espírito Santo, Santa Catarina, Mato Grosso e Mato Grosso do Sul

Apesar de apresenta o rendimento médio mensal domiciliar acima da média, trata-se de um cluster que agrega estados que precisa desenvolver em relação as outras variáveis o número de pessoas com acesso a internet não se aproxima da média.

Verifica-se também que O acesso à banda larga não é eficiente nesse agrupamento. Conexões lentas desestimulam o uso da rede de computadores, também há falta de acesso a níveis mais elevados de instrução nessa região. (MATIAS, 2011).

Cluster 7: São Paulo

É um cluster que agrega apenas um estado por possuir, algumas particularidades em relação ao seu desenvolvimento socioeconômico. Possui o maior acesso a internet tanto de moradores quanto de domicílios. O acesso ocorre, sobretudo por via de microcomputador. Apresenta a maior taxa de pessoas com ensino médio completo que acessam a internet. Esse destaque deixa São Paulo isolado em grupo, contudo, melhorias devem ser pensadas em relação à disseminação de uma internet mais rápida e eficiente.

Cluster 8: Rio Grande do Sul e Paraná.

Possui um dos melhores rendimentos médios mensais domiciliares per capita e um valor acima da média quanto ao número de pessoas que utilizaram a internet. Temos aqui um caso parecido com o cluster 6, ou seja, indica-se novos estudos para averiguar as diferenças regionais internas quanto ao uso da internet.

Cluster 9: Distrito Federal

Nesse encontramos o melhor rendimento médio mensal domiciliar per capita. Contudo é necessário melhorar o acesso à internet nos domicílios, sobretudo nas chamadas cidades satélites. Existe um número considerável de pessoas que acessa a internet via celular e tablet. É necessário melhorar o grau de instrução da população, o número de pessoas que utilizaram a internet com ensino médio completo foi baixo.

Nessa pesquisa, defende-se o maior acesso à internet de qualidade para promover a participação da população no processo de planejamento territorial. Contudo, sabemos que apenas o acesso à tecnologia não vai garantir esse feito. A qualidade do uso da internet nos processos decisórios só será possível mediante regularidade e quando ela puder influenciar de fato as decisões das políticas públicas. Assim, são necessários mecanismos que possibilitem o engajamento cívico no espaço, para produzir debates, ampliando a cultura política da população.

Espera-se que essa pesquisa explicita algumas evidências nos agrupamentos e que políticas específicas para cada cluster possam ser elaboradas. As políticas públicas do Estado, direta ou indiretamente, afetam profundamente quase todos os aspectos da organização espacial, pois ele é o agente central para o funcionamento da sociedade capitalista, e por isso, também cabe a ele diminuir a pobreza estrutural encontrada no país.

Reforça-se, aqui, que a internet por si só não geram participação no planejamento, mas oferecem elementos pertinentes para tal, uma vez que podem privilegiar elementos caros à democracia, sobretudo aqueles que dizem respeito à participação. Assim, a internet pode permitir a comunicação de muitos, e através deles os cidadãos podem mobilizar, participar mais atividade da gestão da cidade, dos governos a partir de um prisma mais tecnológico. (SANTOS, 2008).

Ressalta-se que a internet não cria nada de novo que não existe na realidade. Ela altera de inúmeras maneiras, as experiências de interação humana na sociedade. Existem várias organizações sociais, novas unificações coletivas novas formas de lutas políticas e novas formas de se fazer política gerando tempos e espaços diferentes, pois a internet aqui é traduzida como um canal de participação, expressão e busca de informações pelos cidadãos, permitindo aos indivíduos ou à coletividade acompanhar políticas públicas e participar da tomada de decisão.

6- Considerações Finais

Planejar é necessário, pois, a sociedade produz seus meios de vida a partir de um intercâmbio. Essa pesquisa buscou analisar, através da Estatística Multivariada o acesso à internet no Brasil. Concebe-se nessa pesquisa a importância da Internet para participação civil, levando em conta participação, planejamento e gestão. A Internet propicia o advento de serviços cívicos, centros de informação, sistemas de informação, pesquisas junto aos cidadãos, etc. Existem mecanismos que podem potencializa a participação como os e-mails, listas de discussão, formas de votos, sondagens, sugestões orçamentárias, transferências, sistemas de apoio a decisões complexas.

Vislumbra-se uma interatividade entre as TDICs e o exercício da cidadania em modelos de cidades digitais no contexto da chamada sociedade da informação. Nessa perspectiva, informação e comunicação são considerados fatores chaves para o exercício da cidadania e ampliação da participação dos cidadãos em processos decisórios.

A análise de cluster foi essencial para a criação de grupos e definição dos perfis. A separação das Unidades da Federação em conglomerados foi viável para conseguir agrupar estados similares entre si e divergentes em relação aos outros grupos. Assim, é possível ter maior assertividade na formulação de políticas públicas de inclusão digital já que se tratando de planejamento territorial os recursos são escassos.

Assim, a contribuição dessa pesquisa encontra-se na esfera da possibilidade de redefinição de políticas públicas para que mais pessoas possam ter acesso à internet e a participação população nas decisões eletrônicas. O essencial é contribuir para o reconhecimento do papel desempenhado pela tecnologia no processo de planejamento.

São pertinentes e urgentes pesquisas de cunho informacional que avancem na reflexão dessa sociedade contemporânea com redes invisíveis, espaços tradicionais e virtuais, buscando uma nova abordagem do território que deve estar ligado às teorias tradicionais e as do mundo técnico-científico-informacional. Além de auxiliar no debate do aproveitamento das políticas públicas de inclusão enquanto elementos de valorização dos lugares, organização recursos e projetos que favorecem a trocas de saberes e de experiências criando redes de ajuda mútua, desenvolvendo maior participação da população nas decisões políticas.

Espera-se contribuir para novas maneiras de pensar a política, a participação e o planejamento territorial favorecendo a reconstituição de laços sociais, desburocratização das administrações, otimização do tempo real, racionalização dos recursos e equipamentos coletivos das cidades, experimentar novas formas democráticas. Os indivíduos estarão mais propensos a participar do planejamento da sua cidade, do seu estado ou do seu país se possuir acesso à internet e estiverem de posse de motivos e oportunidades necessários para fazê-lo.

Referências

DULCI, TOMÁS e XAVIER. Interfaces dos capitais humanos, cultural e social na situação ocupacional e nos rendimentos dos indivíduos. In: AGUIAR, N. (Org.). **Desigualdades sociais, redes de sociabilidade e participação política**. Belo Horizonte. Editora UFMG. 2007.

FÁVERO, Luiz Paulo et al. **Análise de dados**: modelagem multivariada para a tomada de decisões. Rio de Janeiro: Elsevier, 2009.

FRALEY, C et al. **mclust Version 4 for R**: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. Department of Statistics-University of Washington, 2012. Disponível em <<https://www.stat.washington.edu/research/reports/2012/tr597.pdf>> Acesso 01 Nov 2016.

GUIDI, Leda. Democracia eletrônica em Bolonha: A rede Iperbole e a construção de uma comunidade participativa *on-line*. In: CEPIK, Marco e EISENBERG, José. **Internet e Política**. Belo Horizonte: Editora UFMG, 2002.

HAIR, J, Jr. Et. al. **Análise multivariada de dados**. Porto Alegre: Bookman, 2009.

Instituto Brasileiro de Geografia e Estatística-IBGE. (PNAD) **Acesso à internet e à televisão e posse de telefone móvel celular para uso pessoal**. Rio de Janeiro: IBGE, 2016.

LATTIN, James M. **Análise de dados multivariados**. São Paulo: Cengage Learning, 2011.

MAIA, Rousiley. Redes cívicas e Internet. Do ambiente informativo denso às condições da deliberação pública. In: CEPIK, Marco e EISENBERG, José. **Internet e Política**. Belo Horizonte: Editora UFMG, 2002.

MATIAS, V R S. **O Phármakon da democracia eletrônica na gestão territorial urbana**. Tese (doutorado) - Universidade Federal de Minas Gerais, Instituto de Geociências, 2011.

MATOS, Renata Assis de. **Comparação de metodologias de análise de agrupamentos na presença de variáveis categóricas e contínuas**; Dissertação (Mestrado em Estatística) – Programa de Pós- Graduação em Estatística, UFMG, Belo Horizonte, 2007.

MINGOTI, Sueli A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. 1ª edição, 2ª reimpressão. Editora UFMG. Belo Horizonte, 2013.

ROGERSON, Peter A. **Métodos estatísticos para geografia**: um guia para o estudante. Tradução técnica: Paulo Fernando Braga Carvalho e José Irineu Rangel Rigotti. Porto Alegre: Bookman, 2012.

SANTOS, Boaventura de Sousa. **A gramática do tempo: para uma nova cultura política**. 2ªed. São Paulo: Cortez, 2008.

UHLIR, Paul F. **Diretrizes políticas para o desenvolvimento e promoção da informação governamental de domínio público.** Brasília: UNESCO, 2006.