



Universidade Estadual da Paraíba
Centro de Ciências e Tecnologia
Departamento de Estatística

Antonio Valmir Paulino

Uso de técnicas de amostragem aplicadas ao tempo de acesso a internet em lan house

Campina Grande
Julho de 2014

Antonio Valmir Paulino

Uso de técnicas de amostragem aplicadas ao tempo de acesso a internet em lan house

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Orientadora:

Prof^a. Dra. Ana Patricia Bastos Peixoto

Campina Grande

Julho de 2014

É expressamente proibida a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano da dissertação.

P328u Paulino, Antonio Valmir.

Uso de técnicas de amostragem aplicadas ao tempo de acesso a internet em lan house [manuscrito] / Antonio Valmir Paulino. - 2014.

35 p. : il. color.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2014.

"Orientação: Profa. Dra. Ana Patricia Bastos Peixoto, Departamento de Estatística".

1. Métodos de amostragem. 2. Inferência estatística. 3. Estratos. 4. Lan house. I. Título.

21. ed. CDD 519

Antonio Valmir Paulino

Uso de técnicas de amostragem aplicadas ao tempo de acesso a internet em lan house

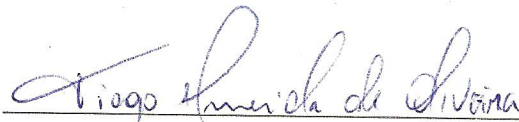
Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Aprovado em: 04 / 07 / 14

Banca Examinadora:



Prof^ª. Dra. Ana Patricia Bastos Peixoto
Orientadora



Prof. Dr. Tiago Almeida de Oliveira
Universidade Estadual da Paraíba



Prof. Dr. Ricardo Alves de Olinda
Universidade Estadual da Paraíba

Dedicatória

Aos meus pais,
Maria das Dores da Silva Paulino e José Paulino (in memória)
A minha irmã Patrícia, dedico com muito amor e carinho.

Agradecimentos

Primeiramente a Deus por me dar a capacidade de alcançar mais este objetivo e força para vencer os obstáculos.

A minha mãe Maria das Dores da Silva Paulino e minha irmã Patrícia por terem acreditado em mim.

Aos meus amigos que conscientes ou inconscientes foram motivadores direto para mais essa vitória.

A Universidade Estadual da Paraíba, especialmente, ao Departamento de Estatística do Centro de Ciências e Tecnologia, e a coordenação do curso.

Meu muito obrigado a minha professora orientadora Ana Patricia pela atenção e colaboração para comigo, sem sua ajuda não teria conseguido.

Aos meus professores pelos ensinamentos.

Resumo

As lan houses são estabelecimentos comerciais que permitem as pessoas a acessar a internet por um preço e tempo controlado, sendo assim perfeitas para elaborar trabalhos escolares, diversões com jogos, imprimir currículos, pesquisar vaga de empregos, comunicar-se periodicamente através de mensagens instantâneas usando as mais variadas redes sociais existentes. A amostragem aleatória estratificada é uma técnica geralmente bastante utilizada para a estimação de parâmetros populacionais, ocasionando uma melhor estimação desses parâmetros populacionais utilizando um número de informações bem menor e um custo pequeno. Neste estudo em particular, foi utilizada a amostragem aleatória estratificada proporcional, com o objetivo de estimar a variável aleatória quantitativa que é o tempo médio e o total populacional do número de horas mensal de acessos a internet pelos usuários da lan house, divididos em três faixas etárias. A primeira faixa etária varia dos 10 anos a 17 anos, a segunda de 18 anos a 30 anos e a terceira de 31 anos ou mais. Na faixa etária de 10 a 17 anos de idade existe um consumo bem significativo chegando a um número de horas de acessos na ordem de 5478 horas representando 65,96% dos acessos na lan house, na faixa etária de 18 a 30 anos de idade o número de horas de acessos foi de 1803 horas representando 21,73% dos acessos e na faixa etária de 31 anos ou mais foi de 1022 horas representando 12,30%. O tempo médio estratificado proporcional dos acessos que foi de 26,79 horas por usuário ao mês, sendo considerada estatisticamente significante, devido ao processo de estratificação da população dos acessos á internet, com isso, conclui-se à importância da utilização dos métodos estatísticos na realização da pesquisas.

Palavras-Chaves: Métodos de amostragem, Inferência estatística, Estratos, Faixa etária

Abstract

The lan houses are commercial establishments that allow people to access Internet for a price and time controlled, so perfect for elaborate studies. school, diversions with games, print curricula vague search jobs, communicate with periodically through instant messages using the most diverse existing social networks. Stratified random sampling is a technique generally much used to estimate population parameters, resulting in a better estimation of all these population parameters utilizing a numbers of info much smaller and a small cost. In this particular study, sampling was utilized stratified random proportional, with the objective of estimating the random variable Quantitative which is the average time and the total population of the number of monthly hrs of Internet users by the lan house, divided into three age groups A first age group ranges from 10 years to 17 years, the second 18 years to 30 years and third of 31 years or more. In the age group 10 to 17 years of age there is a consumption and coming to a significant numbers of hrs of accesses in the order of 5478 hrs accounting for 65.96% of accesses in the lan house in the age group 18 to 30 years of age the number of hrs of accesses was 1803 hours accounting for 21.73% of accesses and aged 31 years or more was 1022 hours accounting for 12.30%. The time medium proportional stratified accesses which was 26.79 hours per user per month was considered statistically significant, due to the process of stratification of population accesses the Internet, it concludes to the importance of using statistical methods in conducting research.

Key Words: Methods of sampling, statistical inference, strata, age

Sumário

Lista de Figuras

Lista de Tabelas

1	Introdução	p. 12
2	Fundamentação Teórica	p. 14
2.1	Marco Histórico	p. 14
2.2	Amostragem aleatória	p. 16
2.2.1	Amostragem aleatória simples	p. 17
2.2.2	Amostragem aleatória estratificada	p. 19
2.2.2.1	Amostragem aleatória estratificada ótima custo variável	p. 23
2.2.2.2	Amostragem aleatória estratificada ótima custo fixo . .	p. 24
2.2.2.3	Amostragem aleatória estratificada proporcional	p. 25
2.2.2.4	Amostragem aleatória estratificada uniforme	p. 25
2.2.3	Amostragem sistemática	p. 25
3	Aplicação	p. 28
3.1	Material e métodos	p. 28
3.2	Resultados	p. 30
4	Conclusão	p. 34
	Referências	p. 35

Lista de Figuras

- 1 Gráfico de barras do tempo médio de acesso a internet em cada faixa etária pelos usuários da lan house p.32
- 2 Setorgrama representando a porcentagem do tempo total de acesso a a internet em cada faixa etária pelos usuários da lan house p.32

Lista de Tabelas

1	Tempo de acesso a internet para cada faixa etária dos usuários da lan house	p. 28
2	Tamanho dos estratos (N_i), fator de proporcionalidade (w_i), tamanho da amostra em cada estrato (n_i) nas três faixas etárias dos usuários da lan house	p. 29
3	Tempos sorteados e selecionados para compor a amostra em cada estrato (n_i) nas três faixa etária dos usuários da lan house, mensurados mensalmente	p. 30
4	Tamanho dos estratos (N_i), tempo médio do estrato (\hat{y}_i), desvio padrão do estrato (σ_i) e tempo total do estrato (\hat{t}_{est}) nas três faixa etária dos usuários da lan house	p. 31

1 Introdução

A utilização da internet ao redor do mundo proliferou-se rapidamente desde a última década, apesar de que a taxa de crescimento parece ter diminuído após 2000. Com a saturação do mercado a fase de crescimento veloz está acabando em países industrializados, mas continua a se espalhar na Ásia, África, América Latina, no Caribe e no Oriente Médio. Por exemplo, o programa PC Conectado (um dos quatro eixos do programa Brasileiro de inclusão digital, que prevê ainda o projeto Casa Brasil, o redirecionamento da infraestrutura tecnológica do governo e a inclusão digital nas escolas) ajudou a indústria a crescer no Brasil.

As lan houses são estabelecimentos comerciais que permitem as pessoas a acessar a internet por um preço e tempo controlado, sendo assim perfeitas para elaborar trabalhos escolares, diversões com jogos, imprimir currículos, pesquisar vaga de empregos, comunicar-se periodicamente através de mensagens instantâneas usando as mais variadas redes sociais existentes. Além disso, tais estabelecimentos pagos possuem uma infraestrutura inegavelmente melhor do que a de centros de acesso gratuito (como os tele centros ofertados pelo Ministério das Comunicações), tornando-se bem mais atraentes para o consumidor final. Computadores mais robustos, programas pré-instalados e conexão de maior velocidade são as principais vantagens das lan houses em relação a outros métodos de acesso gratuito disponíveis para os cidadãos brasileiros (WIKIPÉDIA, 2009).

A amostragem aleatória estratificada é uma técnica geralmente bastante utilizada para a estimação de parâmetros populacionais por muitos pesquisadores, devido a ela consiste em dividir população em estratos (sub populações), onde normalmente existe uma homogeneidade dentro desses estratos e uma heterogeneidade entre eles. Isto ocasiona uma melhor estimação desses parâmetros populacionais utilizando um número de informações bem menor e um custo pequeno, para isto leva-se em consideração alguns fatores como: variabilidade dos dados e custo na obtenção da informação dos dados.

Neste estudo em particular, foi utilizada a amostragem aleatória estratificada proporcional, com o objetivo de estimar a variável aleatória quantitativa que é o tempo médio e o total populacional do número de horas mensal de acessos a internet pelos usuários da lan house, divididos em três faixas etárias. Sendo com isso possível permitir que a

administração da lan house tenha informações com relação ao número desses acessos e tome conhecimento da demanda dos mesmos e também com relação à arrecadação da receita. De posse desse estudo pode-se criar mecanismos para melhorar e ampliar ações em relação ao atendimento a esses usuários e das finanças da lan house.

2 Fundamentação Teórica

O conteúdo desta seção relata os principais aspectos da utilização de técnicas de amostragem estratificada, utilizando-se de artigos práticos e teóricos relacionados ao objetivo da pesquisa.

2.1 Marco Histórico

A primeira descrição de registros de interações sociais que poderiam ser realizadas através de redes foi uma série de memorandos escritos por JCR Licklider do MIT, em agosto de 1962, discutindo o “Galactic Network” conceito. Ele imaginou um mundo interconectado conjunto de computadores através dos quais todos poderiam acessar rapidamente os dados e programas de qualquer local. Em essência, o conceito foi parecido com a internet de hoje. Licklider foi o primeiro chefe do programa de pesquisa de computador do DARPA, ele convenceu seus sucessores na DARPA, Ivan Sutherland, Bob Taylor e Lawrence G. Roberts, da importância do conceito de redes (WORDPRESS, 2010)

Leonard Kleinrock do MIT publicou o primeiro trabalho sobre teoria de comutação de pacotes em julho de 1961 e o primeiro livro sobre o assunto em 1964. Kleinrock convenceu Roberts da possibilidade teórica das comunicações usando pacotes ao invés de circuitos, que foi um passo importante no caminho para redes de computadores. O outro grande passo foi fazer os computadores se conversarem. Para explorar isto, em 1965, trabalhando com Thomas Merrill, Roberts conectou o computador TX-2 em Massachusetts com um Q-32 na Califórnia, com uma baixa velocidade de linha telefônica discada, criando à primeira (embora pequena) rede de computadores em toda a zona já construída. O resultado deste trabalho foi à constatação de que os computadores de tempo compartilhado poderiam trabalhar bem juntos, rodando programas e recuperando dados quando necessários em máquinas remotas, mas que o circuito comutado sistema telefônico era totalmente inadequado para o trabalho. Convicção de Kleinrock sobre a necessidade de comutação de pacotes foi confirmada (WORDPRESS, 2010).

A Internet revolucionou o mundo do computador e das comunicações como nada antes. A invenção do telégrafo, telefone, rádio e computador prepararam o terreno para o

surgimento da internet. A internet é ao mesmo tempo uma capacidade de difusão a nível mundial, um mecanismo de difusão de informações e um meio para colaboração e interação entre indivíduos e seus computadores, independentemente de localização geográfica.

A internet nasceu em 1969, nos Estados Unidos. Interligava originalmente laboratórios de pesquisa e se chamava ARPAnet (ARPA: Advanced Research Projects Agency). Era uma rede do departamento de defesa norte-americano. Era o auge da Guerra Fria, e os cientistas queriam uma rede que continuasse de pé em caso de um bombardeio ou ataque nuclear. Surgiu então o conceito central de internet: é uma rede em que todos os pontos se equivalem e não há um comando central. Assim, se B deixa de funcionar, A e C continuam a poder se comunicar. A internet nada mais é do que milhões de computadores em todo o mundo que se comunicam uns com os outros, todos os minutos através de uma rede inacreditavelmente complicada de cabos, filamentos de fibra ótica e fibras de satélite. O nome internet propriamente dita surgiu bem mais tarde, quando a tecnologia da ARPAnet passou a ser usada para conectar universidades e laboratórios, primeiro nos Estados Unidos da América (EUA) e depois em outros países (WORDPRESS, 2010).

A internet é uma inovação, que permite a comunicação e transmissão de dados entre computadores em locais diferentes. É uma evolução extremamente nova, mas isso não significa que não pode-se analisá-la historicamente. Como em outros países, também no Brasil a Internet foi implantada e se desenvolveu junto ao meio acadêmico e científico, tendo 1989 como ano inicial no país. Professores e pesquisadores que houvessem visitado universidades no exterior já conheciam as promissoras redes internacionais de comunicação e, a cada nova descoberta, o Brasil logo tratava de acompanhar. Seguiu-se a primeira onda da internet, lado a lado com os Estados Unidos. Desde então, o crescimento da rede por aqui apresentou uma curva ascendente.

Em meados de 2001 com o surgimento das duas primeiras Lan Houses no estado do Rio de Janeiro (Bunker 22 que futuramente afiliou-se a rede de franquias Monkey e a Challenger, ambas extintas na data presente) a equipe da OptiLab viu neste mercado um nicho promissor e decidiu retomar o projeto que fora desenvolvido para a Saraiva Mega Store adequando seu funcionamento para atender além de Cyber Cafés esse novo conceito de negocio denominado Lan House.

Mas, como as lan houses surgiram? Para entender melhor o assunto, é importante conhecer mais a fundo as origens das lan houses e seu público-alvo original. Tudo começou em março de 1988, na Coreia do Sul, quando foi inaugurado oficialmente o primeiro cyber café do mundo. O estabelecimento (que jamais teve um nome oficial) continha apenas

dois computadores de 16 bits conectados a uma rede telefônica e era voltado para os estudantes da Universidade Hong Ik, que podiam se encontrar no local para trabalhar e comer alguns petiscos.

Alguns anos depois, nascia na mesma região os Nets café, primeiro a oferecer seus serviços de forma comercial para qualquer cidadão coreano. Quem pagasse 5 mil won (cerca de R\$ 11 na cotação atual) podia permanecer durante uma hora em uma das dezessete máquinas disponíveis no local (dez PCs Pentium, cinco Macs e duas estações de trabalhos da Silicon Graphics). Em 1991, o fenômeno começou a contagiar os Estados Unidos e, posteriormente, o resto do mundo. Mas, foi só no final da década de 90 - com o lançamento de games focados no modo multiplayer, como Star Craft e Counter-Strike - que o cyber cafés aos poucos foi mudando seus focos e adotando o nome “lan houses”. O foco mudou da pura internet para a jogatina multiplayer, e não demorou muito para surgirem os primeiros “corujões” (sessões noturnas ininterruptas) e as “lan parties” (festas fechadas nas quais cada um dos participantes levava o seu próprio computador) (BARAN, 1964).

No Brasil, a primeira grande rede de lan houses foi a Monkey Paulista, que teve sua primeira unidade inaugurada em 1998 pelo empresário Sunami Chun (brasileiro, mas com descendência asiática). Disposto a trazer o conceito para as terras tupiniquins, o executivo adotou jovens de classe média alta como público-alvo e conseguiu abrir cerca de 50 filiais ao redor de todo o território nacional. Mas esse tempo de glória não durou tanto, e em 2010 a Monkey já sentia os efeitos da popularização dos desktops, que ficavam cada vez mais baratos e acessíveis. As várias unidades da rede (especialmente as localizadas em regiões nobres) começaram a fechar as portas por conta da pouca procura pelos seus serviços (WORDPRESS, 2010).

Contudo, três anos após o declínio da rede Monkey, alguns centros de inclusão digital (CIDs) ainda conseguem sobreviver nas regiões mais isoladas e carentes. Estima-se que ainda existam pelo menos 100 mil estabelecimentos espalhados em pequenas comunidades ao redor do país, e eles não parecem se abalar frente ao crescente número de aparelhos eletrônicos dentro dos lares das famílias brasileiras.

2.2 Amostragem aleatória

Frequentemente não é possível ou viável a observação de todos os elementos de uma população (censos), e nesses casos tem-se que fazer uso de amostras. A finalidade de

uma amostra é a de descrever, indiretamente, a população. Portanto, é necessário que as amostras coletadas guardem características as mais próximas possíveis da população. A representatividade de uma amostra é a qualidade que as amostras devem ter, e que corresponde a elas terem, ou reproduzirem, as mesmas propriedades da população. Assim, é necessária a observação de alguns cuidados no momento da coleta de amostras, caso contrário problemas quanto à representatividade podem ocorrer (OLIVEIRA et al., 2009).

O procedimento básico para garantir representatividade nas amostras é o sorteio. O sorteio é o procedimento pelo qual é conferida a mesma probabilidade, a todos os elementos de um conjunto, de serem tomados. O sorteio também é chamado de aleatorização ou casualização. Quando a obtenção de uma amostra sofreu algum mecanismo de sorteio, ela é chamada de amostra aleatória. A coleta de amostras aleatórias é chamada de amostragem aleatória. O sorteio garante representatividade porque evita tendenciosidades no momento da coleta. Segundo Oliveira et al. (2009), a amostragem pode ser classificada em amostragem aleatória simples (AAS), amostragem aleatória estratificada (AAE), amostragem aleatória por conglomerado (AAC), amostragem aleatória sistemática (AS).

Independente da natureza da amostragem (AAS, AAE, AAC ou AS), ela pode ainda ser com reposição ou sem reposição. Na amostragem com reposição, os elementos da população, à medida que são sorteados, são devolvidos à população, e podem eventualmente ser sorteados novamente. Quando a amostragem é sem reposição, os elementos são sorteados apenas uma vez.

2.2.1 Amostragem aleatória simples

A amostragem aleatória simples é o processo de amostragem probabilístico em que as combinações de n diferentes elementos, dos N que compoem a população, possuem igual probabilidade de vir a ser a amostra efetivamente sorteada (COCHRAN, 1977). Cada elemento é sorteado sem reposição e, em cada etapa do sorteio, todos os elementos remanescentes têm igual chance de seleção. O número de possíveis amostras é calculado pela combinação de N elementos n a n ($C_{N,n}$) e a probabilidade de cada amostra ser sorteada é definida por $1/(C_{N,n})$

A amostragem aleatória simples (AAS) é o método mais simples e mais importante para a seleção de uma amostra. Além de servir como um plano próprio o seu procedimento é usado de modo repetido em procedimentos de múltiplos estágios. Na prática, a realização do sorteio no processo de AAS é feita mediante várias possibilidades. Pode-se, por exemplo, escrever em “papezinhos” os N elementos da população, colocá-los em uma

caixa, e sorteá-los. Pode-se ainda identificar os elementos com um número, e sorteá-los mediante tabelas de números aleatórios, ou funções randômicas na calculadora ou computador (OLIVEIRA et al. 2009).

O plano de amostragem é descrito do seguinte modo:

- i) Utilizando-se um procedimento aleatório, sorteia-se com igual probabilidade um elemento da população (N).
- ii) Repete-se o processo anterior até que sejam sorteados n unidades, tendo sido este número prefixado anteriormente.
- iii) Caso seja permitido o sorteio de uma unidade mais de uma vez, tem-se o processo AAS com reposição, que será indicado por ASSc. Quando o elemento sorteado é removido de N elementos do sorteio do próximo têm-se o plano ASS sem reposição, que será indicado por ASSs. O primeiro procedimento é também conhecido como ASS irrestrito e o segundo é também conhecido como ASS restrito.

Do ponto de vista prático, o plano ASSs é muito mais interessante, pois satisfaz o princípio intuitivo que “não se ganha mais informação se uma mesma unidade aparece mais de uma vez na amostra”. Por outro lado, o plano ASSc, introduz vantagens matemáticas e estatísticas, como a independência entre as unidades sorteadas, que facilita em muito a determinação das propriedades dos estimadores das quantidades populacionais de interesse.

Segundo Silva (2004) ao sortear uma amostra aleatória simples de n elementos de uma população N , tem-se como objetivo obter estimativas para valores populacionais desconhecidos, tais como média ou a proporção. Também é possível descrever o conjunto de todas as diferentes amostras determinadas pelo procedimento de sorteio e construir, portanto, a distribuição amostral associada ao procedimento. Como trabalha-se como uma única amostra e conseqüentemente dispõe-se apenas de estimativas únicas calculadas nessa amostra. Torna-se necessário, portanto estabelecer mecanismos probabilísticos a fim de que se possam construir inferências, ou seja, levantar informações sobre os valores populacionais a partir dessas estimativas. As bases conceituais desses mecanismos residem, na distribuição amostral de cada estimador, conforme se apresenta a seguir,

A média amostral,

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad (2.1)$$

estima a média populacional quando a característica de estudo é quantitativa. Nesse caso a variância populacional σ^2 também será estimada por meio do estimador não-viciado (HANSEN et al., 1953),

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1}. \quad (2.2)$$

Adotando-se importante resultado, segundo o qual a variância da média amostral é calculada em função da variância por elemento σ_X^2

$$V(\bar{y}) = (1 - f) \frac{\sigma^2}{n}, \quad (2.3)$$

a variância da média amostral será calculada, na amostra sorteada pela expressão

$$v(\bar{y}) = (1 - f) \frac{\hat{\sigma}^2}{n}, \quad (2.4)$$

que é um estimador não viciado da variância da distribuição amostral (KUZMA, 1984).

2.2.2 Amostragem aleatória estratificada

A técnica de obtenção de amostras em que a população de N elementos ou unidades amostrais é previamente dividida em grupos mutuamente exclusivos (chamados estratos ou subpopulações), e dentro dos quais são sorteadas amostras casuais simples de tamanho n_h , chama-se amostra aleatória estratificada ou, simplesmente amostragem estratificada (COCHRAN, 1977; KISH, 1965).

A amostragem pode ser utilizada em várias áreas distintas do conhecimento humano, as mais comuns são nas áreas comerciais e empresariais. É uma técnica de suma importância no desenvolvimento de uma pesquisa, pois ela permite obter informações a respeito de parâmetros populacionais até então desconhecidos, por meio da observação de apenas uma parte (amostra) do seu universo de estudo (população). As técnicas de amostragem geralmente mais utilizadas são: amostragem aleatória simples, sistemática, estratificada e por conglomerados.

A amostragem estratificada é uma técnica que consiste em subdividir uma população finita ou infinita de tamanho N unidades, em estratos de tamanhos N_1, N_2, \dots, N_k , de tal forma que haja a maior homogeneidade possível dentro e a maior heterogeneidade possível entre tais estratos, com relação as variáveis (características) em estudo. Esses estratos de tamanhos N_1, N_2, \dots, N_k não são pressupostas e, juntos abrangem a totalidade da população de tamanho N de tal modo que $N_1 + N_2 + \dots + N_k = N$. A amostragem

estratificada consiste em selecionar quantos elementos dos estratos de tamanhos $N_1 + N_2 + \dots + N_k$, serão selecionados em cada estrato, para compor a amostra tamanho n_1, n_2, \dots, n_k . Essas amostras selecionadas juntas abrangem a totalidade de tal modo que $n_1 + n_2 + \dots + n_k = n$.

As variáveis de estratificação de população mais comuns são: limites geográficos, classes sociais, sexo, renda, faixa etária, profissão, etc (COSTA NETO, 2002). Na amostragem estratificada é possível calcular estimativas separadas dos parâmetros populacionais como: média, variância, proporção e total, em cada estrato de tamanho $N_1 + N_2 + \dots + N_k$ da variável em estudo, aumenta a precisão da estimativa global, partindo-se do conhecimento de que a variabilidade da variável em estudo é grande o custo da coleta e análise dos dados são frequentemente menores nesse tipo de amostragem devido as conveniências administrativas e operacionais (MARTINS, 2001).

Suponha-se que se deseja estimar a média, \bar{Y} de um conjunto de valores Y_1, \dots, Y_N , numa população finita. Assumindo-se que a população é estratificada, isto é, constituída por K grupos disjuntos ou estratos de dimensões

$$N_1, \dots, N_k \quad \left(\sum_{i=1}^k N_i = N \right),$$

com membros,

$$Y_{ij} \quad (i = 1, \dots, k; \quad j = 1, \dots, N_i).$$

As médias e variâncias dos estratos (ou subpopulações) são dadas por

$$\bar{Y}_1, \dots, \bar{Y}_k$$

e

$$\sigma_1^2, \dots, \sigma_K^2$$

respectivamente.

A média \bar{Y} e a variância σ^2 da população podem ser escritas da seguinte forma

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^k N_i \bar{Y}_i = \sum_{i=1}^k w_i \bar{Y}_i, \quad (2.5)$$

em que, $w_i = \frac{N_i}{N}$ é o fator de proporcionalidade do estrato i , $i = 1, \dots, k$, e

$$\begin{aligned}\sigma^2 &= \frac{1}{N-1} \sum_{i=1}^k \sum_{j=1}^{N_i} (Y_{ij} - \bar{Y})^2 \\ &= \frac{1}{N-1} \sum_{i=1}^k \sum_{j=1}^{N_i} (Y_{ij} - \hat{Y}_i + \hat{Y}_i - \bar{Y})^2 \\ &= \frac{1}{N-1} \left\{ \sum_{i=1}^k (N_i - 1) \sigma_i^2 + \sum_{j=1}^{N_i} (\hat{Y}_i - \bar{Y})^2 \right\}.\end{aligned}\tag{2.6}$$

Assume-se que uma amostra de dimensão n é escolhida por obtenção de uma amostra aleatória simples de cada estrato. As dimensões de cada estrato serão denotadas por n_1, n_2, \dots, n_k $\left(\sum_{i=1}^k n_i = n \right)$. A amostra aleatória simples proveniente do i -ésimo estrato tem como membros

$$y_{i1}, \dots, y_{in_i} \quad (i = 1, \dots, k),$$

e a média e variância amostrais do i -ésimo estrato são dadas por

$$\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}\tag{2.7}$$

e

$$\hat{\sigma}_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2\tag{2.8}$$

para cada estrato tem-se uma fração de amostragem $f_i = \frac{n_i}{N_i}$, $i = 1, \dots, k$.

Este esquema de amostragem para a obtenção de uma amostra de dimensão total n do conjunto da população é chamado amostragem aleatória (simples) estratificada. O estimador de \bar{Y} usualmente utilizado é a média amostral estratificada

$$\bar{y}_{est} = \sum_{i=1}^k w_i \bar{y}_i.\tag{2.9}$$

Note-se que conhecida as dimensões dos estratos, N_i , e, tem-se, os pesos dos estratos, $w_i = \frac{N_i}{N}$, $i = 1, \dots, k$. A média amostral estratificada \bar{y}_{est} não é, em geral, igual a média amostral da amostra aleatória estratificada. A igualdade apenas se verifica quando,

$$\frac{n_i}{n} = \frac{N_i}{N}, \quad i = 1, \dots, k.$$

Isto implica que as frações de amostragem $f_i = \frac{n_i}{N_i}$ são iguais em todos os estratos. Neste caso diz-se que as dimensões dos estratos, n_i , são escolhidas por amostragem proporcional, já que os n_i são escolhidos por serem proporcionais à dimensão dos estratos.

Este procedimento pode simplificar a escolha dos dados e tem a vantagem do ponto de vista estatística, mas pressupõe que as dimensões dos estratos, N_i , são conhecidas. Se tal não acontecer, os pesos, w_i , têm que ser estimados e o estimador \bar{y}_{est} passará a ser enviesado e perderá a eficiência.

O valor médio e a variância de \bar{y}_{est} são dados por

$$E(\bar{y}_{est}) = \sum_{i=1}^k w_i E(\bar{y}_i) = \sum_{i=1}^k w_i \bar{Y}_i = \bar{Y},$$

e

$$Var(\bar{y}_{est}) = \sum_{i=1}^k w_i^2 Var(\bar{y}_i) = \sum_{i=1}^k w_i^2 (1 - f_i) \frac{\sigma_i^2}{n_i}, \quad (2.10)$$

já que $cov(\bar{y}_i, \bar{y}_j) = 0$ para $i \neq j$, isto é, as médias amostrais de estratos diferentes não são correlacionadas.

Os resultados para a estimação do total da população, Y_T são análogos.

$$\bar{y}_{T_{est}} = N \times \bar{y}_{est} = \sum_{i=1}^k N_i \bar{y}_i$$

é um estimador de Y_T com

$$Var(\bar{y}_{T_{est}}) = \sum_{i=1}^k N_i^2 (1 - f_i) \frac{\sigma_i^2}{n_i}$$

Assumindo, como habitualmente, uma distribuição aproximadamente normal para \bar{y}_{est} , pode-se construir intervalos de confiança aproximados para \bar{Y} ou Y_T . Assim, para um grau de confiança $(1 - \alpha)100\%$, estes intervalos são dados por

$$\left[\bar{y}_{est} - \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \sqrt{Var(\bar{y}_{est})}; \quad \bar{y}_{est} + \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \sqrt{Var(\bar{y}_{est})} \right] \quad (2.11)$$

e que, $\Phi(\cdot)$ é a distribuição acumulada da normal padrão. E,

$$\left[N \left(\bar{y}_{est} - \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \sqrt{Var(\bar{y}_{est})} \right); \quad N \left(\bar{y}_{est} + \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \sqrt{Var(\bar{y}_{est})} \right) \right] \quad (2.12)$$

É de costume considerar três tipos de amostragens estratificadas: ótima (custo variável e fixo ou Neyman), proporcional e uniforme. Deve-se notar porém, que o uso da amostragem estratificada exige um cuidado adicional nos cálculos dos dimensionamentos dos tamanhos das amostras n_1, n_2, \dots, n_l , em cada estrato para compor a amostra global de tamanho n . O procedimento do dimensionamento depende do critério utilizado para definir os tamanhos das amostras n_1, n_2, \dots, n_l em cada estrato a partir do tamanho da

amostra global n , seja para dados quantitativos ou qualitativos.

2.2.2.1 Amostragem aleatória estratificada ótima custo variável

Na amostragem estratificada ótima custo variável para dimensionar uma amostra de tamanho n_i , num estrato de tamanho N_i . Leva-se em consideração a variabilidade dos dados, ou seja, da variável em estudo, o custo da obtenção da informação sobre essa característica e um número proporcional de elementos num determinado estrato de tamanho N_i .

A variabilidade da característica de interesse na sub população é mensurada pelo seu desvio padrão. Pretendo-se assim, otimizar o tamanho da amostra n_i em relação ao tamanho N_i do estrato, ou seja, obtendo uma quantidade menor de informação sobre a variável do estrato, com base no princípio de que, onde a variabilidade da variável é menor, menos elementos serão selecionados para bem caracterizar o comportamento da variável em estudo. Dessa forma, com um menor número de elementos selecionados na amostra de tamanho n_i , consegue-se uma quantidade de informações equivalentes a obtida nos demais casos com um custo bem menor (OLIVEIRA, 2004).

As principais dificuldades para utilizar esse tipo de amostragem residem nas complicações teóricas relacionadas com respeito a análise dos dados e em que muitas vezes, não se pode mensurar a priori o desvio padrão da variável em estudo nos diversos estratos. Desta forma, o melhor dimensionamento de uma amostra de tamanho n_i , com custo variável e dados quantitativos nos diversos estratos é dado por

$$w_i = \frac{N_i \times (\sigma_i / \sqrt{c_i})}{\sum_{l=1}^l N_l \times (\sigma_l / \sqrt{c_l})}, \quad \text{com } i = 1, \dots, k \quad (2.13)$$

em que, N_i é o tamanho do estrato i , $\sqrt{c_i}$ é a raiz quadrada do custo unitário de obtenção da informação no estrato i , σ_i é o desvio padrão do estrato i , $\sum_{l=1}^l N_l \times (\sigma_l / \sqrt{c_l})$ é o somatório do produto dos estratos com seus respectivos desvios padrões amostrais e dividido pela raiz quadrada de custos unitários. Logo,

$$n_i \cong w_i \times n$$

em que, w_i é o fator de proporcionalidade, n_i é o tamanho da amostra selecionada no estrato i e n é o tamanho da amostra global.

Para efeitos práticos σ_i pode ser substituído por $\hat{\sigma}_i$ na fórmula anterior,

$$w_i = \frac{N_i \times (\hat{\sigma}_i / \sqrt{c_i})}{\sum_{i=1}^l N_i \times (\hat{\sigma}_i / \sqrt{c_i})}, \quad \text{com } i = 1, \dots, k \quad (2.14)$$

em que, $\hat{\sigma}_i$ é o desvio padrão amostral do estrato i .

Quando se trata de dados qualitativos o dimensionamento da amostra é dado por

$$w_i = \frac{N_i \times \left(\sqrt{P_i(1-P_i)} / c_i \right)}{\sum_{i=1}^l N_i \times \left(\sqrt{P_i(1-P_i)} / c_i \right)}, \quad \text{com } i = 1, \dots, k \quad (2.15)$$

em que, N_i é o tamanho do estrato i , $\sum_{i=1}^l N_i \times \left(\sqrt{P_i(1-P_i)} / c_i \right)$ é o somatório do produto dos estratos com seus respectivos desvios padrões amostrais e dividido pela raiz quadrada de custos unitários, $\sqrt{c_i}$ é a raiz quadrada do custo unitário de obtenção da informação no estrato i , $P_i(1-P_i)$ é a variância do estrato i . Logo,

$$n_i \cong w_i \times n$$

em que, w_i é o fator de proporcionalidade, n_i é o tamanho da amostra selecionada no estrato i e n é o tamanho da amostra global.

Para efeitos práticos σ_i pode ser substituído por $\hat{\sigma}_i$ na fórmula anterior,

$$w_i = \frac{N_i \times \left(\sqrt{\hat{p}_i(1-\hat{p}_i)} / c_i \right)}{\sum_{i=1}^l N_i \times \left(\sqrt{\hat{p}_i(1-\hat{p}_i)} / c_i \right)}, \quad \text{com } i = 1, \dots, k \quad (2.16)$$

em que, $\hat{p}_i(1-\hat{p}_i)$ é a variância amostral do estrato i ,

2.2.2.2 Amostragem aleatória estratificada ótima custo fixo

Na amostragem estratificada ótima custo fixo para dimensionar uma amostra de tamanho n_i , num estrato de tamanho N_i . Leva-se em consideração apenas a variabilidade dos dados, ou seja, variação da característica em estudo e um número proporcional de elementos num determinado estrato de tamanho N_i . Contudo, em alguma situação prática não há diferença no custo da obtenção da informação entre os estratos ou as vezes não conhecemos esse custo. Nesse sentido, $c_1 = c_2 = \dots = c_l = c$, ou seja, os custos são iguais. Dessa forma, o melhor dimensionamento de uma amostra de tamanho n_i nos diversos estratos, é simplificando a fórmula da estratificação ótima custo variável pelo dimensionamento de Neyman (OLIVEIRA, 2004). O dimensionamento do tamanho da amostra

com custo fixo, tanto para dados quantitativos, quanto para dados qualitativos é análogo ao dimensionamento da amostra com custo variável, como visto na seção anterior.

2.2.2.3 Amostragem aleatória estratificada proporcional

Na amostragem estratificada proporcional para dimensionar uma amostra de tamanho n_i , num estrato de tamanho N_i , o número de elementos selecionados para compor a amostra é proporcional ao número de elementos existentes no estrato, isso porque em outras situações práticas, além dos custos de obtenção da informação sobre a variável em estudo serem iguais em cada estrato as variâncias também o são, ou seja, usa-se isso como pressuposição. Nestas ocasiões, o dimensionamento é feito de forma proporcional aos tamanhos dos estratos.

$$w_i = \frac{N_i}{\sum_{i=1}^l N_i} = \frac{N_i}{N}, \quad \text{com } i = 1, \dots, l \quad (2.17)$$

em que, w_i é o fator de proporcionalidade, N_i é o tamanho do estrato i , N é o tamanho da população.

Logo,

$$n_i \cong w_i \times n$$

em que, w_i é o fator de proporcionalidade, n_i é o tamanho da amostra selecionada no estrato i e n é o tamanho da amostra global.

2.2.2.4 Amostragem aleatória estratificada uniforme

É um procedimento que não leva em consideração as características dos estratos. Em cada estrato a amostra tem o mesmo tamanho, logo

$$n_1 = n_2 = \dots = n_k = n$$

e conseqüentemente o tamanho da amostra aleatória estratificada é $n \times k$ sendo k o número de estratos.

2.2.3 Amostragem sistemática

A amostragem sistemática é um processo de amostragem probabilístico não aleatório, onde o critério de probabilidade se estabelece através da aleatorizada da primeira unidade amostral. Em um processo sistemático, as unidades amostrais são selecionadas a partir

de um esquema rígidos e pré-estabelecidos de sistematização, com o propósito de cobrir a população em toda sua extensão, a fim de obter um modelo sistemático simples e uniforme. Algumas vantagens da amostragem sistemática, de acordo com Husch, Miller e Beers (1972), são

- i) A sistematização proporciona uma boa estimativa da média e do total, devido à distribuição uniforme da amostra em toda a população;
- ii) Uma amostra sistemática é executada com maior e menor custo que uma aleatória, desde que a escolha das unidades amostrais seja mecânica e uniforme;
- iii) O deslocamento entre as unidades é mais fácil pelo fato de seguir direção fixa e preestabelecida, resultando em tempo menor e, por consequência, um menor custo de amostragem;
- iv) O tamanho da população não precisa ser conhecido, uma vez que cada unidade que ocorre dentro do intervalo de amostragem fixado, é selecionada sequencialmente, após ser definida a unidade inicial.

A amostragem sistemática é uma variação da amostragem aleatória simples. Sua aplicação requer que a população seja ordenada de modo tal que cada um de seus elementos possam ser unicamente identificado pela posição. Apresentam condições para satisfação desse requisito uma população identificada a partir de uma lista que englobe todos os seus elementos, uma fila de pessoas ou um conjunto de candidatos a um concurso, identificados pela ficha de inscrição (GIL, 2010).

Uma amostra sistemática é obtida determinando-se inicialmente o intervalo de amostragem k . Como ponto de partida, escolhe-se aleatoriamente um número j , entre 1 e k , em seguida seleciona-se os elementos classificados (de ordem) $j + k, j + 2k, j + 3k, \dots, j + (n - 1)k$, até completar o tamanho da amostra desejado n . Observe que, neste plano de amostragem, a fração amostral é de $1/k$. Neste caso, dizemos ter uma amostra sistemática de 1 em k , de tamanho n .

Para efetuar a seleção da amostra, procede-se ao sorteio de um ponto de partida entre 1 e o fator de expansão, k , definido pela razão entre o número de elementos da população e o número de elementos da amostra, isto é:

Uma vez sorteado o primeiro elemento da amostra, os demais serão encontrados, somando-se $k, 2k, 3k, \dots$ até completar a amostra. Portanto se a população tem

$$k = \frac{N}{n}$$

elementos, existem k possíveis amostras sistemáticas de tamanho n .

3 Aplicação

Encontram-se nesta seção as principais metodologias que serviram de base para este trabalho, tanto na parte da descrição dos dados, quanto na utilização das técnicas de amostragem.

3.1 Material e métodos

Os dados utilizados para a elaboração desse trabalho foram obtidos em uma lan house, localizada na cidade de Remígio, no ano de 2011. O estudo pretende mensurar a variável tempo em horas mensal de acesso a internet comum, aos usuários da lan house. A mensuração é feita por meio do sistema de cadastro e controle dos tempos de acesso dos usuários, ou seja, quando os usuários acessam a internet o sistema computa esse tempo de acesso em horas e no final do mês ele faz o somatório total desses acessos em horas.

A população dos usuários cadastrados na lan house é num total de 310 e foram divididos e listados em três faixas etárias formando assim três estratos. No primeiro são os usuários de 10 anos a 17 anos e listados de 1 até 160, no segundo de 18 anos a 30 anos e listados de 1 até 95 e no terceiro de 31 anos ou mais e listados de 1 até 55 (Tabela 1).

Tabela 1: Tempo de acesso a internet para cada faixa etária dos usuários da lan house

Faixa etária	Número de usuários
10 a 17	160
18 a 30	95
> 31	55
Total	310

Para estimar o tempo médio e o total populacional de horas mensal estratificados de acesso a internet das três faixas etárias foi feito o dimensionamento do tamanho da amostra global n para população finita com $N = 310$ usuários e um erro tolerável de 15%, para a variável horas de acesso a internet mensal foi obtido pelas expressões

$$n_0 = \frac{1}{\varepsilon_0^2}, \quad (3.1)$$

em que n_0 é a primeira aproximação da amostra global e ε_0 é o erro amostral tolerável.

Logo,

$$n = \frac{N \times n_0}{N + n_0} \quad (3.2)$$

em que N é o tamanho da população; n é o tamanho da amostra global.

Após obtenção da amostra global, utilizou-se da amostragem estratificada proporcional, para encontrar os valores das amostras em cada estrato, representando as três faixas etárias do estudo, como pode ser observado na Tabela 2.

Tabela 2: Tamanho dos estratos (N_i), fator de proporcionalidade (w_i), tamanho da amostra em cada estrato (n_i) nas três faixas etárias dos usuários da lan house

Faixa etária	N_i	w_i	n_i
10 a 17	160	0,52	21
18 a 30	95	0,30	12
> 31	55	0,18	7
Total	310	1,00	40

Para dimensionar as amostras de tamanho n_i em cada faixa etária, ou seja, em cada estrato, utilizou-se a amostragem estratificada proporcional. Feito isso, dentro de cada estrato, procedeu-se a amostragem sistemática para selecionar os usuários que acessaram a internet, e com isso, obter o tempo de acesso em horas mensurados mensalmente em cada faixa etária (3). A amostra sistemática se deu pelo uso do fator de sistematização dado por,

$$s = \left[\frac{N_i}{n_i} \right],$$

em que, N_i é o tamanho do estrato i e n_i é o tamanho da amostra do estrato i .

O fator de sistematização foi igual a 8 para todos os estratos. Com isso, o primeiro elemento (m) a ser sorteado em cada estrato está ($1 \leq m \leq 8$). Para os usuários da faixa etária 1, o elemento sorteado foi $m = 1$, para faixa etária 2, $m = 7$ e para a faixa etária 3, $m = 5$. A disposição dos demais elementos sorteados nos estratos pode ser observada na Tabela 3.

A partir das amostras selecionadas, todos os procedimentos inferenciais foram tomados a fim de se obter os resultados que justificaram o uso das técnicas amostrais.

Tabela 3: Tempos sorteados e selecionados para compor a amostra em cada estrato (n_i) nas três faixa etária dos usuários da lan house, mensurados mensalmente

Sorteio	Tempo (h)	Sorteio	Tempo (h)	Sorteio	Tempo (h)
1°	41	7°	8	5°	27
9°	34	15°	15	13°	4
17°	35	23°	21	21°	19
25°	27	31°	7	29°	10
33°	28	39°	14	37°	15
41°	45	47°	30	45°	25
49°	40	55°	20	53°	30
57°	38	63°	11		
65°	39	71°	12		
73°	31	79°	32		
81°	29	87°	34		
89°	25	95°	24		
97°	36				
105°	37				
113°	32				
121°	29				
129°	43				
137°	41				
145°	26				
153°	28				
160°	35				

3.2 Resultados

Os resultados apresentam o tempo de acesso nas faixas etárias distintas sendo possível observar na Tabela 4 que após calculados os valores das estimativas tempo médio, desvio padrão e total populacional de acessos à internet feitos pelo os usuários da lan house em cada uma das suas três faixas etárias, ficou evidente que no estrato que corresponde à faixa etária de 10 a 17 anos de idade que esses parâmetros: tempo médio; desvio padrão e total populacional de acessos à internet foram bem maiores que nas outras duas faixas etárias de 18 a 30 anos de idade e 31 anos de idade ou mais. Isso se deve ao fato de que o público de usuários de lan house seja em grande parte constituído de pessoas jovens (crianças e pré-adolescentes) que durante seus acessos à internet passam jogando games. Jogam coletivamente, conectados em rede, conectados nas mais variadas redes sociais como salas de “bate papo”, e-mail, download de músicas são os principais usos, conjugados com a navegação pela internet e games. Esse grupo também, uma vez ou outra, aproveita para fazer pesquisas escolares.

Tabela 4: Tamanho dos estratos (N_i), tempo médio do estrato (\hat{y}_i), desvio padrão do estrato (σ_i) e tempo total do estrato (\hat{t}_{est}) nas três faixa etária dos usuários da lan house

Faixa etária	N_i	\hat{y}_i	σ_i	t_{est}
10 a 17	160	34,24	5,99	5.478
18 a 30	95	19,00	9,36	1.805
> 31	55	18,57	9,50	1.022
Total	310			8.305

Os adolescentes até 17 anos formam o principal público das lan houses. São eles quem vão à lan house com maior frequência, passam maior parte do tempo e também apresentam o maior consumo. As lan houses têm um estilo essencialmente adolescente/jovem, seja porque são eles os principais frequentadores e acabam influenciando no estilo do estabelecimento, seja como recurso mercadológico para atrair o segmento mais lucrativo para o negócio. Isso caracteriza que nesse estrato que corresponde à faixa etária de 10 a 17 anos de idade que existe uma homogeneidade dentro desse estrato com relação ao tempo de acessos à internet devido ao valor da estimativa de seu desvio padrão que foi de 5,99 revelando uma pequena variabilidade dos acessos à internet no estrato (Tabela 4).

Já nos outros dois estratos que corresponde as duas faixas etárias de 18 a 30 anos de idade e 31 anos de idade ou mais os valores de seus desvios-padrão respectivamente foram 9,36 e 9,50 revelando uma significativa variabilidade dos acessos à internet. Isso se deve ao fato de que são pessoas adultas com responsabilidades de trabalho, estudos e outras responsabilidades (Tabela 4). A Figura 1 apresenta o comportamento do usuários e alguns fatores podem contribuir consideravelmente para esses resultados, tais como, maior disponibilidade de tempo, facilidade em aprender a manusear aparelhos eletrônicos e curiosidade própria dos adolescentes. Há também uma maior porcentagem de acesso na faixa etária de 10 a 17 anos como pode ser visualizada na Figura 2 em relação ao total de horas de acesso por mês.

Calculada a estimativa do tempo médio de acessos à internet em cada uma de suas três faixas etárias. Foi calculada a estimativa estratificada do tempo médio desses acessos à internet que foi de 26,79 horas por usuário no mês e um desvio padrão 1,16 horas. Calcula-se o valor da estimativa do total populacional estratificado que foi de 8305 horas por mês. Os valores das estimativas estratificadas do tempo médio e total populacional apresentaram uma melhor estatística com relação às estimativas em cada estrato.

Na faixa etária de 10 a 17 anos que em números absoluto representa 5.478 horas de acessos à internet por mês, ou seja, uma porcentagem de 65,96% dos acessos isso

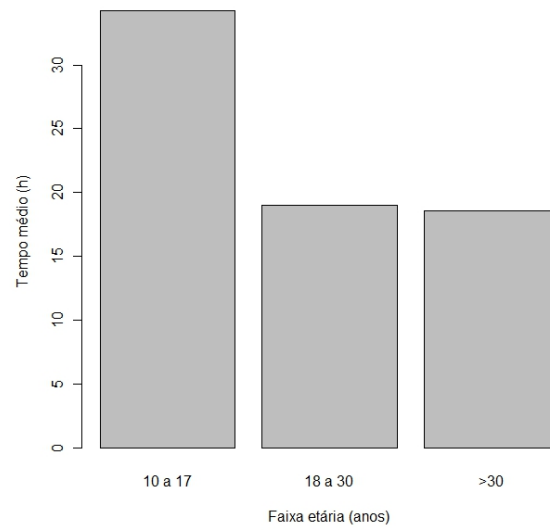


Figura 1: Gráfico de barras do tempo médio de acesso a internet em cada faixa etária pelos usuários da lan house

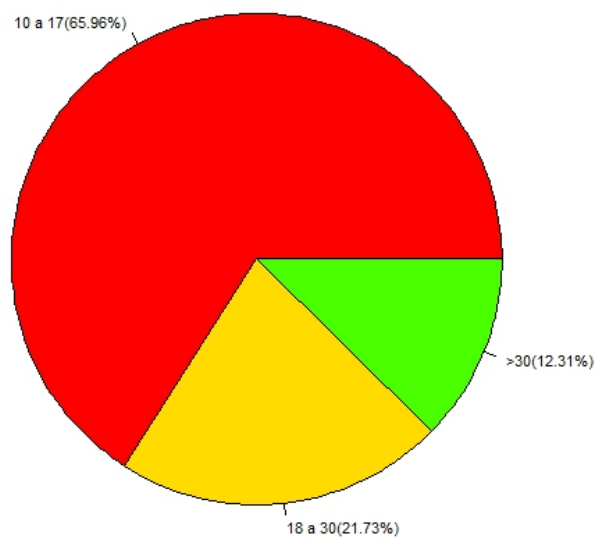


Figura 2: Setorgrama representando a porcentagem do tempo total de acesso a a internet em cada faixa etária pelos usuários da lan house

influencia significativamente na arrecadação da lan house, nas outras faixas etárias os valores das estimativas absolutas e relativas foram na faixa etária de 18 a 30 anos 1.805 horas e a porcentagem foi de 21,73% e na faixa de 31 anos ou mais foi de 1.022 horas e a porcentagem foi de 12,30%.

Após obtenção dos resultados para cada estrato, estimou-se o tempo médio estratificado de acesso a internet mensurados mensalmente em horas dos usuários da lan house que foi de 26,79 horas de acesso por usuário, com um desvio padrão de 1,16 horas. Com esses resultados pôde-se obter um intervalo de confiança (85%) para o tempo médio estratificado mensurados mensalmente foi

$$IC(\bar{y}_{est}; 85\%) = [25,13; 28,46]$$

É possível observar que o intervalo de confiança não contém a constante zero, sendo possível obter conclusões a respeito do mesmo. Isso significa que depois de calculada a estimativa do tempo médio estratificado do número de horas de acessos à internet pelo os usuários da lan house que foi de 26,79 horas/mês, foi construído o intervalo de confiança com um nível de significância de 85%. Portanto, tem-se 85% de confiança que o verdadeiro valor da média populacional se encontra entre os limites [25,13; 28,46] ou então se corre um risco de 15% de que o verdadeiro valor da média populacional seja menor que 25,13 ou maior que 28,46 horas/mês.

De acordo com o tempo total de 8.305 horas de acesso mensal, pode-se construir o intervalo de confiança para o total populacional estratificado de acesso a internet mensurados mensalmente em horas dos usuários da lan house, sendo o mesmo

$$IC(T; 85\%) = [7.786; 8.824]$$

Pode-se dizer que os significantes resultados obtidos na faixa etária de 10 a 17, deve-se ao fato de que os usuários dessa faixa etária procuram frequentar o estabelecimento com regularidade num determinado horário, não porque seja mais tranquilo, mas para se encontrar com os demais frequentadores da mesma faixa etária. Evitam ir a horários diferentes, pois nesse caso, a presença na lan house praticamente se resumiria ao uso do equipamento. A lan house é mais do que um ponto de encontro entre os seus jovens usuários, é um local onde se conhecem, podem conhecer pessoas novas, conversam, consolidam vínculos e amizades e passam boa parte do dia.

4 Conclusão

Neste estudo, foram utilizados alguns tipos de amostragens aleatórias como, amostragem aleatória simples, sistemática e a estratificada proporcional. A amostragem aleatória simples que é o método mais simples e mais importante para selecionar uma amostra foi utilizada suas metodologias para calcular as estimativas dos parâmetros tempo médio e total populacional em cada um dos estratos. A amostragem aleatória sistemática é processo sistemático onde as unidades amostrais são selecionadas a partir de um critério pré-estabelecido de sistematização, foi utilizada suas metodologias no processo de escolha e seleção dos indivíduos e seu respectivo tempo de acessos à internet em cada um dos estratos. A amostragem estratificada proporcional leva em consideração para a escolha do tamanho de uma amostra um número proporcional de elementos existentes em cada estrato e suas metodologias foram utilizadas para calcular o dimensionamento do tamanho da amostra em cada estrato consequentemente utilizou-se suas metodologias para estimar a variável aleatória quantitativa que é o tempo médio dos acessos à internet dos usuários de uma lan house localizada na cidade de Remígio PB, com a finalidade de generalizar os resultados para toda a população de usuários.

Calculadas as estimativas do tempo médio e do total populacional dos acessos à internet pelos usuários da lan house em cada uma das suas três faixas etárias ficou evidente que na faixa etária de 10 a 17 anos de idade existe um consumo bem significativo chegando a um número de horas de acessos na ordem de 5.478 horas representando 65,96% dos acessos na lan house, na faixa etária de 18 a 30 anos de idade o número de horas de acessos foi de 1.803 horas representando 21,73% dos acessos e na faixa etária de 31 anos ou mais foi de 1.022 horas representando 12,30%. O tempo médio estratificado proporcional dos acessos que foi de 26,79 horas por usuário ao mês, sendo considerada estatisticamente significativa, devido ao processo de estratificação da população dos acessos à internet, com isso, conclui-se à importância da utilização dos métodos estatística na realização de pesquisas.

Referências

- COCHRAN, W.G. **Sampling techniques**. New York: John Wiley & Sons, 3 ed. 1977.
- COSTA NETO, P. L. O. **Estatística**. 2ª ed. São Paulo: Edgard Blucher, 2002.
- GIL, A. C. **Métodos e Técnicas de Pesquisa Social**. 6ª ed., São Paulo: Editora Atlas, 2010.
- HANSEN, M. H.; HURWITZ, W. N.; MADOW, W. G. **Sample survey methods and theory**. New York: John Wiley & Sons, 1953.
- HUSCH, B.; MILLER, C. I.; BEERS, T. W. **Forest Mensuration**. 2ª ed. The Ronald Press Company, New York, USA. 1972.
- OLIVEIRA, P. H. F. C. **Amostragem Básica - Aplicada em Auditoria**. 2ª ed. Rio de Janeiro: Editora Científica Moderna Ltda, 2004.
- OLIVEIRA, M. S. de; BEARZOTI, E.; BOAS, F. L. V.; NOGUEIRA, D. A.; NICOLAU, L. A. **Introdução à Estatística**. Lavras : UFLA, 2000, 334p.
- MARTINS, G.A. **Estatística geral e aplicada**. 2ª ed. São Paulo: Atlas; 2001.
- MILLER, G.A (1991) "Licklider, psicólogo," jornal da sociedade Acústica da América 89, não.4B, pp 1887. Disponível em: <<http://www.en.wikipedia.org/wiki/J-C-R-Licklider>>. Acesso em: 23 Mai 2014.
- SILVA, N. N. da. **Amostragem probabilística**. 2ª ed. São Paulo: EDUSP, 2004.
- KUZMA, I. W. **Basic statistic for the Health Sciences**. Mountain View, Mayfield Publishing Company, 1984.
- KISH, L. **Survey sample**. New York: John Wiley & Sons. 1965.
- WIKIPÉDIA. Desenvolvido pela Wikimedia Foundation. Apresenta conteúdo enciclopédico. Disponível em: <http://www.pt.wikipedia.org/wiki/lan_house>. Acesso em: 23 Mai 2014.
- <<http://www.hadouken.wordpress.com/2010/03/11/>>, o-fim -da -monkey-paulista-e-da-era-das-lan houses. Acesso em: 23 Mai 2014.
- <<http://www.softwarelivre.gov.br/noticias/New-item.2005-03-29.2402>>. Acesso em: 10 Abr 2014.
- < <http://www.P.Baran>, "on Distributed communications Networks," IEEE.trans.gomm.Sistems, March. 1964<. Acesso em: 23 jun 2014