UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
DEPARTAMENTO DE ESTATÍSTICA
Programa de Graduação em Estatística

MARCOS ALVES DE LIMA

ESTUDO DE CASO EM UM VEÍCULO PUBLICITÁRIO COM DADOS DE CONTAGEM LONGITUDINAIS

JUIZ DE FORA
2014

MARCOS ALVES DE LIMA

ESTUDO DE CASO EM UM VEÍCULO PUBLICITÁRIO COM DADOS DE CONTAGEM LONGITUDINAIS

Monografia apresentada ao Curso de Estatística da Universidade Federal de Juiz de Fora, como requisito parcial para à obtenção do título de Bacharel em Estatística.

Orientadores: Prof. Marcel de Toledo Vieira

Prof. Ronaldo Rocha Bastos

JUIZ DE FORA

2014

Lima, Marcos Alves de .

Estudo de caso em um veículo publicitário com dados de contagem longitudinais / Marcos Alves de Lima. -- 2014.

72 p. : il.

Orientadores: Marcel de Toledo Vieira
Ronaldo Rocha Bastos

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas, 2014.

1. Dados de Contagem. 2. Dados Longitudinais. 3. Regressão de Poisson. I. Vieira, Marcel de Toledo, orient. II. Bastos, Ronaldo Rocha, orient. III. Título.

MARCOS ALVES DE LIMA

ESTUDO DE CASO EM UM VEÍCULO PUBLICITÁRIO COM DADOS DE CONTAGEM LONGITUDINAIS

Monografia apresentada ao Curso de Estatística da Universidade Federal de Juiz de Fora, como requisito parcial para à obtenção do título de Bacharel em Estatística.

Aprovada em 18 de dezembro de 2014.

BANCA EXAMINADORA

Marcel de Toledo Vieira, Ph.D.
Universidade Federal de Juiz de Fora

Ronaldo Rocha Bastos, Ph.D.
Universidade Federal de Juiz de Fora

Henrique Steinherz Hippert, D.Sc.
Universidade Federal de Juiz de Fora

Augusto Carvalho Souza, D.Sc.
Universidade Federal de Juiz de Fora

AGRADECIMENTOS

Agradeço muito ao meu pai Mauro e minha mãe Nê, poderia escrever um capítulo só de agradecimentos a esses dois e mesmo assim, não seria o bastante.

Ao meu irmão Marcel, pela grande ajuda em organizar o banco de dados, você sempre será um exemplo a ser seguido. Ao meu irmão Rafael por todo apoio e por mostrar que apesar de toda dificuldade, devemos seguir em frente.

Agradeço aos meus sobrinhos Camila e Thiago e minha afilhada Sophia; seres que detêm o poder de fazer qualquer um esquecer os problemas e o cansaço.

Aos meus amigos, que durante toda essa jornada me incentivaram e tiveram que aceitar minha ausência em diversos momentos. Aos amigos de faculdade, pelo exemplo de dedicação e companheirismo.

Agradeço aos meus orientadores Marcel e Ronaldo. Poucos têm o privilégio e a felicidade de ter dois grandes orientadores ao lado, obrigado pela paciência, críticas e sugestões.

Aos professores do departamento de Estatística, sempre dispostos a tirar dúvidas. Em especial o professor Márcio, pelo bom grado em ajudar nesta monografia. Aos professores Augusto, Luiz Cláudio, Ângela, Clécio, Hippert, Lupércio e Joaquim; importantes incentivadores durante este caminho. Agradeço também ao secretário Bruno pela amizade.

Agradeço a equipe do veículo publicitário pela oportunidade de estágio, pela amizade verdadeira, pelo aprendizado e responsabilidade adquiridos e pelo uso do banco de dados, sem vocês essa monografia não existiria.

Não poderia deixar de agradecer aos amigos do grupo de estudo do CAEd, em especial o professor Fernando Tavares Jr. e também às pessoas que me ajudaram na Hiper-Roll Embalagens.

Por fim, agradeço a todos que contribuíram para minha formação e elaboração deste trabalho.

Grande abraço e muito obrigado.

“Muitas das verdades que temos
dependem de nosso ponto de vista”
Mestre Yoda – Star Wars

RESUMO

Após a Segunda Guerra Mundial, o interesse pelo estudo de populações ao longo do tempo aumentou simultaneamente com os avanços tecnológicos, facilitando o uso de grandes bases de dados e análise das mesmas. Neste tipo de análises, o foco é verificar se determinada característica ou comportamento de um grupo estudado sofre alterações ao longo de determinado período de tempo. A este conjunto de técnicas damos o nome de Análise de Dados Longitudinais.

Juntamente a Análise de Dados Longitudinais, introduziremos conceitos de Modelos Lineares Generalizados (MLGs), sendo estas ferramentas estatísticas muito úteis para a modelagem de uma determinada variável de interesse. Através de MLGs, podemos utilizar diversas informações contidas em um banco de dados definindo uma equação que represente a relação entre as variáveis independentes e variáveis de interesse.

Esta monografia apresenta um estudo de caso com dados de contagem longitudinais, tendo como objetivo verificar a influência de determinadas características pessoais na utilização de promoções contidas em um veículo publicitário e também verificar se o volume de utilizações varia com o tempo. Para isto, foi analisado o comportamento de usuários deste veículo publicitário por um período de 12 edições e empregamos métodos de regressão de Poisson para dados de contagem longitudinais com a finalidade de descrever o perfil dos usuários e auxiliar a tomada de decisão na empresa em que os dados foram obtidos.

Palavras-chave: Dados de Contagem, Dados Longitudinais, Regressão de Poisson.

ABSTRACT

After World War II, the interest in analyzing populations over time simultaneously increased with technological advances, facilitating the use and analysis of large databases. In this type of study, the focus is to see if a certain characteristic or behavior of a study group changes over a given period of time. Such technique is called Longitudinal Data Analysis.

Along with the Longitudinal Data Analysis, we introduce concepts of Generalized Linear Models (GLMs), which are very useful statistical tools for modeling a particular variable of interest. Through GLMs, we can use different information obtained from a database allowing us to define an equation that represents the relationship between independent variables and a variable of interest in a single event.

This monograph presents a case study which considers longitudinal count data, aiming at verifying the influence of a particular trait in the use of promotions contained in an advertising vehicle and also check if the volume of uses varies over time. For this, we analyzed the behavior of users of this advertising vehicle for a period of 12 issues and applied Poisson regression methods for longitudinal count data in order to describe the profile of users and to assist the decision making process in the company where data was obtained.

Keywords: Count Data, Longitudinal Data, Poisson Regression.

LISTA DE QUADROS E TABELAS

LISTA DE GRÁFICOS

LISTA DE ABREVIATURAS E SIGLAS

AEH – Análise de Eventos Históricos.

Emp. – Empresarial.

Entret. – Entretenimento.

MCU – Modelo de Correlação Uniforme.

MLG – Modelos Lineares Generalizados.

MV – Máxima Verossimilhança.

SUMÁRIO

1 INTRODUÇÃO

1.1 Considerações Iniciais e Motivação

Em diversas situações em que se deseja observar mudanças de determinada característica em uma população ao longo do tempo, recomenda-se a utilização de técnicas de pesquisa longitudinal, sendo estas capazes de identificar variações no comportamento de determinado grupo e também a variabilidade entre indivíduos diferentes (VIEIRA, 2012). Juntamente a esta técnica, utilizaremos conceitos de Modelos Lineares Generalizados (MLG), uma vez que este método possui grande aplicabilidade em múltiplas áreas do conhecimento.

A proposta de explicar o comportamento de uma variável resposta através de uma equação capaz de descrever os dados observados teve início com os chamados Modelos Normais Lineares. Porém, em dados que não apresentavam normalidade, tais técnicas não se adequavam bem, gerando resultados muitas vezes imprecisos. Em razão disto, Nelder e Wedderburn (1972) propuseram novas técnicas capazes de gerar resultados confiáveis para todas as variáveis respostas que possuíam distribuição pertencente à família exponencial e a este método, deram o nome de MLG (PAULA, 2013).

A aplicação de tais técnicas a um banco de dados real de um veículo publicitário na região de Juiz de Fora, cidade localizada no interior de Minas Gerais, torna mais fácil e interessante à compreensão da teoria. Conforme mencionado por Miguel (2007) *apud* Silva (2010), o estudo de caso é “um estudo de natureza empírica que investiga um determinado fenômeno, geralmente contemporâneo, dentro de um contexto real de vida”.

O banco de dados utilizado possui informações de características pessoais do usuário deste veiculo publicitário e também informações sobre a utilização do mesmo ao longo do tempo. Com a característica longitudinal das observações somadas a não obrigatoriedade de cadastro para a utilização do veículo publicitário, notamos que a cada período estudado o número de observações se altera e, por conta disto, dizemos que os dados estudados são classificados como desbalanceados (SHAW; MITCHELL-OLDS,1993).

Segundo Shaw e Mitchell-Olds (1993), os chamados dados desbalanceados, que outrora eram de difícil análise, atualmente com o avanço da tecnologia juntamente com *software* capazes de considerar tal fato, vem sendo utilizados com frequência em estudos longitudinais envolvendo seres vivos.

Em suma, podemos dizer que o presente trabalho apresenta uma aplicação de um MLG para dados de contagem observados ao longo do tempo utilizando um banco com dados reais de consumidores de um veículo publicitário.

A utilização da regressão de Poisson é comum quando tratamos de dados do tipo contagem, uma vez que a variável resposta não assume valores negativos (McCULLAGH; NELDER, 1999). Porém, considerações devem ser feitas quando abordamos esta metodologia no contexto longitudinal.

1.2 Objetivo

O objetivo do presente trabalho é a aplicação da metodologia de regressão de Poisson para dados de contagem longitudinais coletados por um veículo publicitário, detectando se o volume de utilizações de tal veículo varia com o tempo e se há influência das características pessoais do usuário na utilização.

1.3 Descrição dos Capítulos

Esta monografia está dividida em 5 capítulos. O presente capítulo apresentou uma introdução ao tema a ser estudado, fazendo uma concisa apresentação do mesmo seguido pelo objetivo do trabalho.

No capítulo 2 é apresentada a metodologia utilizada, onde serão mostradas primeiramente as características de uma distribuição de Poisson, sucedida pela apresentação de Modelos Lineares Generalizados, juntamente com o conceito de família exponencial, na sequência focaremos na utilização do modelo de regressão Poisson e suas particularidades. Este capítulo também introduz conceitos de análise de dados longitudinais e modelagem dos mesmos. Por fim mostraremos como pode ser empregada a análise longitudinal de dados do tipo contagem e a singularidade dos chamados Dados Desbalanceados.

O capítulo 3 é destinado à descrição do banco de dados e do estudo de caso, onde previamente é feita uma breve apresentação da empresa e do produto e logo após é apresentado o banco de dados e análises exploratórias do mesmo. No capítulo 4 serão realizadas aplicações da metodologia descrita e análise dos resultados observados. No último capítulo será apresentada uma conclusão e propostas para possíveis trabalhos futuros.

2 METODOLOGIA

2.1 Distribuição de Poisson

Os dados que utilizaremos para o trabalho são provenientes da contagem em uma população, onde a variável aleatória X pode assumir qualquer valor inteiro não negativo.

Segundo Casella e Berger (2010, p.83), a distribuição de Poisson "é uma distribuição discreta amplamente aplicada e pode servir como modelo para uma série de diferentes tipos de experimentos". Por este motivo, optamos por utilizar a distribuição de Poisson.

Quando nos defrontamos com fenômenos para os quais é esperada pelo menos uma ocorrência em um determinado intervalo de tempo, podemos, algumas vezes, utilizar a distribuição de Poisson para modelar este fenômeno (CASELLA; BERGER, 2010).

A distribuição de Poisson possui apenas um único parâmetro μ. Casella e Berger (2010) definem a função de probabilidade desta distribuição pela seguinte equação:

$$P(X = x|\mu) = \frac{e^{-\mu}\mu^{x}}{x!}$$

Onde X é uma variável aleatória que assume apenas valores nos números inteiros não negativos. Chamaremos de Suporte da Distribuição de X o conjunto dos valores que X assume, denotado por $A(x)$.

Desta forma,

$$A(x) = \{x: 0, 1, 2, \dots\}.$$

O Parâmetro μ, às vezes chamado de parâmetro de intensidade, indica o número esperado de ocorrências de certo evento por unidade de tempo, espaço ou outro indicador de tamanho. É comum encontrar autores que preferem utilizar a letra grega λ para representar o parâmetro.

Sendo o Espaço Paramétrico Θ o conjunto de todos os valores que o parâmetro μ pode assumir, temos então:

$$\Theta = \{\mu : 0 \leq \mu < \infty\}$$

Se X tiver distribuição de Poisson com parâmetro μ, então o valor esperado e a variância de X terão o mesmo valor μ (MEYER, 1982).

$$E(X) = \mu \text{ e } Var(X) = \mu$$

Resultado 2.1. Meyer (1982) mostra o valor esperado da variável aleatória X pela seguinte equação:

$$E(X) = \sum_{x} xp(x)$$

Desta forma, temos na Distribuição de Poisson,

$$E(X) = \sum_{x=0}^{\infty} x \frac{e^{-\mu}\mu^{x}}{x!}$$

$$= \sum_{x=0}^{\infty} \frac{e^{-\mu}\mu^{x}}{(x-1)!}$$

$$= \mu e^{-\mu} \sum_{x=0}^{\infty} \frac{\mu^{x-1}}{(x-1)!}$$

Fazendo a seguinte substituição $y = x - 1$, verificamos que a expressão se torna

$$E(X) = \mu\, e^{-\mu} \sum_{y=1}^{\infty} \frac{\mu^{y}}{y!}$$

Empregando a expansão em série de Taylor, onde

$$e^{\mu} = \sum_{y=1}^{\infty} \frac{\mu^{y}}{y!} \qquad (2.1)$$

Temos por fim que

$$E(X) = \mu e^{(-\mu+\mu)} \text{ e}$$

$$E(X) = \mu. \quad \square$$

Resultado 2.2. Cálculo semelhante poderia ser realizado na demonstração da variância, que pode ser definida pela seguinte equação:

$$Var(X) = E(X^2) - [E(X)]^2$$

De modo que

$$E(X^2) = \sum_{x} x^2 p(x)$$

Assim sendo, temos na Distribuição de Poisson,

$$E(X^2) = \sum_{x=0}^{\infty} x^2 \frac{e^{-\mu}\mu^{x}}{x!}$$

$$= \sum_{x=0}^{\infty} x \frac{e^{-\mu}\mu^{x}}{(x-1)!}$$

Novamente, fazendo $y = x - 1$, a equação acima fica

$$E(X^2) = \sum_{y=1}^{\infty} (y+1) \frac{e^{-\mu}\mu^{(y+1)}}{y!}$$

$$= \sum_{y=1}^{\infty} y \frac{e^{-\mu}\mu^{(y+1)}}{y!} + \sum_{y=1}^{\infty} \frac{e^{-\mu}\mu^{(y+1)}}{y!}$$

$$= \left(\mu e^{-\mu} \sum_{y=1}^{\infty} y \frac{\mu^{y}}{y!}\right) + \left(\mu e^{-\mu} \sum_{y=1}^{\infty} \frac{\mu^{y}}{y!}\right) \qquad (2.2)$$

Separaremos a equação (2.2) em duas, de forma que nossa demonstração fique mais clara. Primeiramente iremos resolver a primeira parcela da equação.

$$\left(\mu e^{-\mu} \sum_{y=1}^{\infty} y \frac{\mu^{y}}{y!}\right) \qquad (2.3)$$

Verificamos que

$$\left(\mu e^{-\mu}\sum_{y=1}^{\infty} y\frac{\mu^{y}}{y!}\right) = \left(\mu^{2}e^{-\mu}\sum_{y=1}^{\infty}\frac{\mu^{(y-1)}}{(y-1)!}\right)$$

Deforma análoga ao que fizemos no cálculo do valor esperado e utilizando a expansão em série de Taylor (2.1), chegaremos ao seguinte resultado,

$$\mu e^{-\mu}\sum_{y=1}^{\infty} y\frac{\mu^{y}}{y!} = \mu^{2}$$

Agora, iremos resolver a segunda parcela da equação.

$$\left(\mu e^{-\mu}\sum_{y=1}^{\infty}\frac{\mu^{y}}{y!}\right) \qquad (2.4)$$

Aplicando a expansão em série de Taylor na equação (2.4) iremos obter o seguinte resultado:

$$\mu e^{-\mu}\sum_{y=1}^{\infty}\frac{\mu^{y}}{y!} = \mu$$

Desta forma, se retornarmos a equação (2.2) temos que,

$$E(X^{2}) = \left(\mu e^{-\mu}\sum_{y=1}^{\infty} y\frac{\mu^{y}}{y!}\right) + \left(\mu e^{-\mu}\sum_{y=1}^{\infty}\frac{\mu^{y}}{y!}\right) = \mu^{2} + \mu$$

Por fim, como $Var(X) = E(X^{2}) - [E(X)]^{2}$, observamos que

$$Var(X) = \mu^{2} + \mu - \mu^{2} \text{ e}$$

$$Var(X) = \mu. \quad \square$$

2.2 Modelos Lineares Generalizados (MLG)

A escolha de modelos é de extrema importância em todas as pesquisas em que desejamos encontrar ao menos uma equação que seja mais parcimoniosa possível e que consiga descrever com certo grau de precisão os dados observados (CORDEIRO; DEMÉTRIO, 2008).

Nelder e Wedderburn (1972) mostraram que diferentes técnicas estatísticas podem ser formuladas de uma maneira unificada, como uma classe de modelos de regressão. A essa teoria unificadora, deram o nome de Modelos Lineares Generalizados (MLG).

Através da análise de MLG é possível utilizar as informações contidas em um conjunto de dados produzindo então um modelo que represente a relação existente entre as variáveis de um processo.

Segundo Nelder e Wedderburn (1972), nossa variável resposta, também chamada de variável de interesse ou variável dependente, deve seguir uma distribuição que pertença à família exponencial. Como estamos trabalhando com distribuição de Poisson, mostraremos que a mesma faz parte desta família.

2.2.1 Família Exponencial

O conceito de família exponencial foi inserido na Estatística por Sir Ronald Fisher. Mais tarde, no final do século XIX, os modelos da família exponencial começaram a ser desenvolvidos por Maxwell, Boltzmann e Gibbs.

A família exponencial uniparamétrica pode ser caracterizada pela seguinte função (de probabilidade ou densidade) (CORDEIRO; DEMÉTRIO, 2008)

$$f(x; \mu) = h(x)\exp[\eta(\mu)\, t(x) - b(\mu)]$$

onde as funções $h(x)$, $t(x)$, $\eta(\mu)$ e $b(\mu)$ assumem valores em subconjuntos dos reais.

Uma vez que a Distribuição de Poisson de parâmetro $\mu > 0$, tem a seguinte função de probabilidade.

$$P(X = x|\mu) = \frac{e^{-\mu}\mu^{x}}{x!}$$

A mesma pode ser escrita da seguinte forma

$$P(X = x|\mu) = \frac{1}{x!}\exp(x log\mu - \mu)$$

Por fim, se,

$$h(x) = \frac{1}{x!},$$

$$t(x) = x,$$

$$\eta(\mu) = log\mu, \text{ e}$$

$$b(\mu) = \mu,$$

conseguimos mostrar que a Distribuição de Poisson pertence à Família Exponencial.

Cordeiro e Demétrio (2008) afirmam que a distribuições Normal, Binomial, Binomial Negativa, Gamma, Normal Inversa, Multinomial, Beta e Logaritma também fazem parte da Família Exponencial.

2.2.2 Descrição do Modelo Linear Generalizado

Como estamos interessados em estudar o comportamento de alguma variável em relação a outras, teremos como base o seguinte modelo proposto por Nelder e Wedderburn (1972) para encontrar esse dinamismo.

$$g(\mu) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \dots + \beta_j x_j$$

Onde:

$x_i (i = 1, 2, \dots, j)$ são as variáveis explicativas cujos valores são conhecidos, também chamadas de variáveis regressoras, variáveis auxiliares ou covariáveis;

Os termos β_i $(i = 0, 1, 2, \dots, j)$ são parâmetros desconhecidos, chamados de coeficientes de regressão. Charnet (2008) informa que podemos interpretar β_i como a mudança esperada em Y devido ao aumento de uma unidade em x_i. Alguns autores preferem escrever β_0 como sendo uma constante α. Através de x_i e β_i é possível descrever um modelo que relacione o valor esperado μ, da variável resposta ou dependente (Y), a uma combinação linear das variáveis explicativas (NELDER; WEDDERBURN, 1972).

Um MLG é formado por três componentes: componente aleatório, componente sistemático e função de ligação (NELDER; WEDDERBURN, 1972). O componente aleatório está associado à variável de interesse (Y), sendo esta uma distribuição de probabilidade pertencente à Família Exponencial, caracterizada pelos parâmetros de localização (valor esperado) e dispersão (variância). O componente sistemático é formado pelo preditor linear $\eta = X\beta$ que é uma combinação linear dos parâmetros desconhecidos (β) com as variáveis regressoras independentes (x_i) observadas. O Componente Sistemático está ligado ao valor esperado (μ) da variável de interesse através de uma função de ligação $g(.)$ monótona e diferenciável, que estabelece a relação entre o preditor linear (η) e o valor esperado da função de distribuição do componente aleatório (Y) (CORDEIRO; DEMÉTRIO, 2008).

Devemos ter muito cuidado ao definir os três componentes do MLG. Cordeiro e Demétrio (2008) afirmam que o componente sistemático deve ser estabelecido na fase de planejamento do experimento tendo em vista a natureza dos dados (discreta ou contínua) e seu intervalo de variação (conjunto dos reais, reais positivos, inteiros, etc), o componente aleatório deve ser estabelecido com base nas medidas a serem feitas, podendo estas ser contínuas ou discretas, algumas vezes podendo também representar a presença ou ausência de determinado fator atribuído a alguma categorias, ou também pode ser o valor de uma variável quantitativa. A escolha da função de ligação depende do problema que desejamos estudar, sendo que cada experimento pode ter uma função de ligação diferente, logo, a escolha desta função deve levar em conta a distribuição proposta para os dados.

A função de ligação mais simples encontrada nos modelos lineares clássicos é aquela em que a média μ e o preditor linear η são idênticos, $g(\mu) = \mu$, podendo μ e η assumir qualquer valor no conjunto dos números reais, esta função de ligação é conhecida como ‘identidade’ (McCULLAGH; NELDER, 1999).

Se estivermos trabalhando com dados de contagem e utilizarmos a Distribuição de Poisson, lembrando que $\mu \geqslant 0$, a função identidade não deve ser utilizada, pois η não pode ser negativo. Utilizaremos então $\eta = \ln(\mu)$ (McCULLAGH; NELDER, 1999).

Para a Distribuição Binomial, temos, $0 < \pi < 1$, logo, devemos ter uma função de ligação que tenha correspondência no intervalo (0,1), de acordo com McCullagh e Nelder (1999), que afirmam que cada distribuição de probabilidade tem uma função de ligação que utiliza o valor esperado (parâmetro natural). A esta função de ligação, damos o nome de ligação canônica.

No quadro 2.1, mostraremos a função de ligação canônica de algumas distribuições da Família Exponencial.

Quadro 2.1. Funções de Ligação Canônica de algumas distribuições da Família Exponencial.

Distribuição	**Notação**	**Função de Ligação Canônica g(.)**	**Nome da Ligação Canônica**
Normal	$N(\mu, \sigma^2)$	$\eta = \mu$	Identidade
Poisson	$P(\mu)$	$\eta = \ln(\mu)$	Log
Binomial	$B(n, \pi)$	$\eta = \ln\left\{\frac{\pi}{1-\pi}\right\}$	Logit
Gamma	$G(\mu, r)$	$\eta = \mu^{-1}$	Recíproca
Normal Inversa	$IG(\mu, \sigma^2)$	$\eta = \mu^{-2}$	Quadrática Inversa

Fonte: McCullagh e Nelder (1999); e Paula (2013).

Uma vez que nosso banco de dados é formado por dados não-negativos do tipo contagem, utilizaremos o modelo de Regressão de Poisson nesta monografia, conforme mencionado anteriormente.

2.3 Modelo de Regressão Poisson

Pelas características que um modelo de Poisson possui, este se torna muito importante quando desejamos analisar dados em forma de contagens.

Cordeiro e Demétrio (2008) mostram que dentre as qualidades de um modelo Poisson está a capacidade de proporcionar uma boa descrição de dados experimentais cuja variância é proporcional à média; além disso, se os eventos estudados ocorrerem de forma independente e aleatória no tempo, o modelo de Poisson é capaz de determinar o número de ocorrências deste fenômeno durante um período de tempo determinado.

É pertinente mencionar que, ao utilizarmos a regressão de Poisson para dados de contagem, frequentemente nos deparamos com o problema da superdispersão. Agresti (2007) afirma que, ao empregarmos o modelo Poisson, podemos não estar prevendo toda a variabilidade contida nos dados estudados. Isto acontece porque a distribuição de Poisson possui apenas um único parâmetro μ, utilizado tanto para descrever a média, quanto a variância. Segundo Agresti (2007, p.80): “O fenômeno em que os dados apresentam uma variabilidade maior que o esperado para um MLG é chamado de superdispersão”.

Se a variância da variável resposta apresentar valor superior ao que o modelo de regressão de Poisson para dados de contagem prevê, visto que há heterogeneidade entre indivíduos não captada pelas covariáveis, uma alternativa para observar de forma mais precisa a variação destes dados é a utilização de um MLG com função de ligação logarítmica considerando um preditor linear para o componente aleatório com distribuição Binomial Negativa, já que esta também é de natureza discreta e seu suporte está contido nos números inteiros positivos. Além dessas características, a distribuição Binomial Negativa possui dois parâmetros diferentes, um para a média e outro para a variância, fazendo com que o modelo apresente uma maior variabilidade nas estimativas (AGRESTI, 2007).

Diante destas alegações, é normal questionar o uso de um modelo de regressão de Poisson, visto que o modelo de regressão Binomial Negativo apresenta características semelhantes e minimiza os problemas de superdispersão. Porém, a interpretação dos coeficientes em um MLG Binomial Negativo pode ser trabalhosa.

O modelo de regressão de Poisson desempenha um papel muito importante na análise de dados categóricos tanto quanto o modelo Normal desempenha na análise de dados contínuos. Sendo assim, o modelo log-linear definido pela distribuição de Poisson, $P(\mu)$, é um dos casos especiais do MLG onde se encontra grande aplicabilidade (CORDEIRO; DEMÉTRIO, 2008).

Como dito anteriormente, um MLG é obrigatoriamente formado pelo componente aleatório, componente sistemático e função de ligação. No caso de um Modelo de Poisson temos:

i. Componente Aleatório: Distribuição Poisson, $P(\mu)$.

ii. Componente Sistemático: Combinação linear dos parâmetros desconhecidos com as variáveis regressoras, $\eta = X\beta$.

iii. Função de Ligação: Log-linear, $g(\mu) = \ln(\mu)$.

Ao utilizarmos a função de ligação log-linear, observamos:

$$\ln(\mu) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots + \beta_j x_j$$

assim,

$$\mu = e^{\beta_0}(e^{\beta_1})^{x_1}(e^{\beta_2})^{x_2} \ldots (e^{\beta_j})^{x_j}$$

Note que se utilizarmos apenas uma variável regressora, a equação ficaria da seguinte forma:

$$\mu = e^{\beta_0}(e^{\beta_1})^{x_1}$$

Observamos então, que se encontrarmos o valor para o coeficiente de regressão $\beta_1 = 0$, isto significa que o valor da média μ não se altera se houver alguma mudança na variável explicativa x_1, mas se encontrarmos um valor para $\beta_1 > 0$, isto significa que qualquer alteração na variável regressora implicaria em um aumento na média μ, por outro lado, se for encontrado um valor para o parâmetro $\beta_1 < 0$, temos então que uma alteração em x_1 acarretaria num apoucamento de μ. Porém, dificilmente as observações serão classificadas por apenas um fator (CORDEIRO; DEMÉTRIO, 2008).

Os valores dos parâmetros $\beta_1 \ldots \beta_j$ da regressão Log-linear podem ser estimados pelo método de máxima verossimilhança (MV) em função de β, mostrados em Turkman e Silva (2000) como sendo,

$$L(\beta) = \prod_{i=1}^{n} f(y_i).$$

Ao aplicarmos o logaritmo na função $L(\beta)$, passamos a ter a chamada log verossimilhança,

$$ln\, L(\beta) = l(\beta).$$

Após aplicar o logaritmo, devemos derivar a log-verossimilhança em função de β e igualar o resultado a zero para encontrar uma estimativa para β_i, ou seja,

$$\frac{\partial l(\beta)}{\partial \beta_i} = 0.$$

Por fim, devemos fazer a derivada segunda com o intuito de verificar se o valor encontrado representa um ponto de máximo,

$$\frac{\partial^2 l(\beta)}{\partial \beta_i^2} < 0.$$

Como estamos trabalhando com uma distribuição de Poisson, temos (TURKMAN; SILVA, 2000):

$$Y_i \sim P(\mu_i) \qquad\qquad i = 1, \dots n$$

$$f(y_i|\mu_i) = \exp\{y_i \ln(\mu_i) - \mu_i - \ln(y_i!)\},$$

Uma vez que utilizamos a função de ligação

$$g(\mu) = \ln(\mu),$$

devemos lembrar que

$$\ln(\mu_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \dots + \beta_j x_j,$$

Que também pode ser escrita da forma

$$\ln(\mu_i) = \underline{X_i^T}\, \underline{\beta}$$

Sendo $\underline{X_i} = (x_{i1} \dots x_{ip})$ um vetor de dimensão p associado às covariáveis x_i e $\underline{\beta}$ um vetor dos parâmetros desconhecidos β_i.

A função de verossimilhança será dada então por (TURKMAN; SILVA, 2000)

$$L(\beta) = \exp\left\{\sum_{i=1}^{n} y_i\, \underline{X_i^T \beta} - \sum_{i=1}^{n} e^{\underline{(X_i^T \beta)}} - \sum_{i=1}^{n} \ln(y_i!)\right\}$$

Geralmente, não encontramos equações lineares e soluções analíticas. Nestes casos podemos optar pela utilização de métodos numéricos tais como, por exemplo, o método de Newton-Raphson ou o método escore de Fisher (Anexo A1).

Atualmente há vários *software* que estimam o valor do coeficiente de regressão β_i $(i = 0, 1, 2, \dots, j)$.

2.4 Pesquisas longitudinais, dados longitudinais e questões a serem consideradas.

Os dados observados neste trabalho apresentam característica longitudinal, uma vez que foram observadas as mesmas variáveis para unidades da população durante diversos momentos em um dado período de tempo. Estes momentos, no caso estudado, estão ligados às edições do veiculo publicitário, considerando o prazo de sua validade.

2.4.1 Introdução à Análise de Dados Longitudinais

Estudos envolvendo dados longitudinais ganharam grande notoriedade nas últimas décadas, pois com o rápido avanço da tecnologia de informação, foi possível armazenar e analisar de modo mais rápido e preciso grandes bases de dados (VIEIRA, 2012). Porém, o grande interesse em realizar estudos com o mesmo grupo ao longo do tempo surgiu principalmente após a Segunda Guerra Mundial (MENARD, 1991).

Segundo Vieira (2012), levantamentos de dados que são realizados com os mesmos indivíduos em momentos temporais distintos, podendo ou não conter as mesmas variáveis, são denominadas pesquisas longitudinais. Pesquisas deste tipo são importantes quando deseja-se observar mudanças na característica ou comportamento de certa população ou indivíduo ao longo do tempo. Além disso, estudos envolvendo dados longitudinais também são capazes de verificar variações na variável resposta entre indivíduos diferentes. Conforme Vieira (2012, p.13), sobre estudos longitudinais, “normalmente pressupõe-se que a população alvo seja fixa. Entretanto, no contexto longitudinal, poderíamos conceder que a população apresentasse mudanças ao longo do tempo, o que seria mais realista”.

Uma preocupação presente nos estudos longitudinais é o alto custo que este tipo de pesquisa pode requerer, contudo, isto não foi empecilho em nosso estudo pela forma que obtivemos os dados (ver o capítulo 03).

Como a ideia básica desta metodologia é realizar estudos envolvendo um mesmo grupo através da coleta de informações ao longo do tempo, alguns problemas surgem: é comum em pesquisas envolvendo pessoas, o que é o nosso caso, que um indivíduo participe num momento da pesquisa e não volte a participar novamente, ou que participe em determinada ocasião, omite-se em outra e volte a participar futuramente. Estes fenômenos caracterizam diferentes tipos de não resposta longitudinal.

Dados longitudinais são empregados em diversos ramos de pesquisa, como por exemplo, na epidemiologia, sociologia e economia. No Brasil, um exemplo de estudo longitudinal de grande porte é o Estudo Longitudinal de Crianças Nascidas Vivas em 1982 em Pelotas, RS, Brasil. Diferentemente das modelagens convencionais, quando o interesse na modelagem de um fenômeno envolve dados longitudinais, deve-se levar em conta a variável tempo (t), além disso, a correlação entre observações de um mesmo indivíduo deve ser considerada (VIEIRA, 2012).

2.4.2 Modelagem de Dados Longitudinais com Dados Quantitativos

Com o fácil acesso à tecnologia e diversos *software* funcionando como ferramentas no ajuste de modelos que levam em conta a característica longitudinal dos dados, o interesse por este tipo de estudo têm aumentado consideravelmente desde a década de 1990 (VIEIRA, 2012).

Como dito anteriormente, devemos levar em conta a variável tempo, no caso estudado, temos o tempo (t) como uma medida discreta tendo espaçamento igual, ou não, em todas as ocasiões.

$$t = 1, 2, \dots, T$$

Uma forma de modelar este tipo de dado é pressupor que os valores peculiares associados à variável resposta possam ser ajustados em função das covariáveis, podendo estas variar ou não com o tempo. Ao adotar este método, pode-se modelar a média e covariância separadamente, considerando o fato de que observações repetidas para uma mesma variável são geralmente correlacionadas. Alternativamente surge a modelagem de efeitos aleatórios, levando em conta que a correlação entre observações repetidas é fruto da variação que os coeficientes do modelo apresentam entre os indivíduos. Há também uma variante que leva em consideração os modelos de efeito fixos, com frequência adotados pelos econometristas. Tais métodos são utilizados quando deseja-se realizar inferências sobre os indivíduos (VIEIRA, 2012).

Algumas outras metodologias podem ser consideradas quando almeja-se trabalhar com dados longitudinais, são elas: Modelos de Transição e Análise de Eventos Históricos (AEH). Modelos de Transição são comumente utilizados em modelos de regressão logística e regressão logística multinomial. AEH é também empregada no estudo de análise de transições, sendo o modelo de análise de sobrevivência um caso particular desta metodologia (VIEIRA, 2012).

Não nos aprofundaremos em todos estes métodos citados anteriormente, mas há uma vasta bibliografia tratando desta temática. Diggle *et al.* (2002) é um dos nomes mais citados quando tratamos deste assunto.

Ao modelar dados longitudinais, podemos definir (VIEIRA, 2012):

i. y_{it} , a variável de interesse para o indivíduo i no tempo t.
ii. $\underline{Y_i} = (Y_{i1}, \dots Y_{iT})$, um vetor com T observações repetidas da variável de interesse para as unidades $i = 1, 2, \dots, N$ para T ocasiões da pesquisa.

Modelos adotados com frequência no contexto longitudinal são do tipo,

$$E\left(\underline{Y_i}\right) = \underline{\mu_i}\ (\underline{\beta}).$$

Sendo,

$$\underline{\mu_i}\left(\underline{\beta}\right) = [\mu\left(\underline{x_{i1}}, \underline{\beta}\right), \dots, \mu\left(\underline{x_{iT}}, \underline{\beta}\right)]$$

Dado que $\underline{x_{it}}$ é um vetor das covariáveis fixas e $\underline{\beta}$ é um vetor de parâmetros desconhecidos.

Ao considerar a variável tempo no modelo de regressão, permitindo que as covariáveis (x) possam variar através deste, podemos então fazer comparações entre as respostas de cada indivíduo, relacionando assim, comportamentos antes e depois da ocorrência de algum fenômeno. Uma observação importante sobre este modelo é a propriedade de considerar como controle cada indivíduo na pesquisa, lembrando que pode ocorrer grande variabilidade entre indivíduos (VIEIRA, 2012).

Conforme Vieira (2012), ao conjecturar que o modelo se trata de um Modelo de Correlação Uniforme (MCU), também denominado Modelo de Efeitos Aleatórios, tal pode ser expresso pela seguinte equação:

$$Y_{it} = \underline{x_{it}\beta} + u_i + v_{it} \qquad (2.5)$$

Onde u_i, $(i = 1, \ldots, j)$ representa os efeitos aleatórios permanentes, que são variáveis não observadas, também chamadas de fatores individuais específicos e v_{it}, $(t = 1, \ldots, T)$ denominados efeitos aleatórios transitórios.

É comum acharmos em artigos a expressão (2.5) documentada de modo mais geral, podendo ser escrita da seguinte forma:

$$Y_{it} = \underline{x_{it}\beta} + \varepsilon_{it}$$

Onde o termo ε_{it} é o termo de erro do modelo, tendo como característica, valor esperado igual à zero:

$$E(\varepsilon_{it}) = 0$$

Em geral, supõe-se que os componentes u_i e v_{it} do modelo de regressão são mutuamente independentes, com $E(u_i) = E(v_{it}) = 0$. Além disso, $Cov(u_i, v_{it}) = 0$.

Se $E\left(\underline{Y_i}\right) = \underline{\mu_i}\ (\underline{\beta})$, como demonstrado anteriormente, temos então que,

$$E(Y_{it}) = \mu\left(\underline{x_{i1}}, \underline{\beta}\right)$$

E a variância será:

$$Var(Y_{it}) = Var(u_i) + Var(v_{it}).$$

Vieira (2012) informa que para períodos diferentes $(t \neq t')$ a variância deverá ser modelada através de duas formas:

i. Variação para o mesmo indivíduo, $Var(v_{it})$ e
ii. Variação entre indivíduos, $Var(u_i)$.

Desta forma, podemos introduzir o conceito da correlação dada pela fórmula (VIEIRA, 2012):

$$Corr(Y_{it}, Y_{it'}) = \rho = \frac{Var(u_i)}{Var(u_i) + Var(v_i)}$$

Onde ρ assume valores entre zero e um $(0 \leq \rho \leq 1)$,$Var(u_i) = Cov(Y_{it}, Y_{it'})$ e $Var(Y_{it}) = Var(u_i) + Var(v_i)$.

Através do valor encontrado para ρ, podemos fazer alegações sobre a correlação intra-individual, já que a mesma é fruto do efeito aleatório individual específico (v_{it}).

i. Se observarmos $\rho = 0$ quando $Var(u_i) = 0$, isto significa que não há variação entre os indivíduos.

ii. Quando nos depararmos com $\rho = 1$ (que não ocorre na prática) temos que não há nenhum erro de medida no modelo, ou seja, não há nenhuma variação entre as observações feitas para o mesmo indivíduo.

Como dito anteriormente na seção 2.3, há algumas maneiras de estimar o parâmetro desconhecido $\underline{\beta}$. Entre as formas mais comuns estão à função de Máxima Verossimilhança e os Mínimos Quadrados (Anexo A2). Porém, no caso de encontrarmos equações não-lineares, recorremos aos métodos de Newton-Raphson ou o método escore de Fisher.

2.4.3 Análise Longitudinal de Dados do Tipo Contagem

Trabalhar com dados longitudinais, conjuntamente com variáveis categóricas pode ser mais árduo do que trabalhar com variáveis contínuas ou mesmo binárias no contexto longitudinal. Por vezes, encontramos na literatura, ocasiões em que a variável categórica pode ser considerada como contínua, se a mesma puder ser ordenada dentro de um número suficiente de categorias (há autores que sugerem no mínimo cinco). Em outras circunstancias é possível reduzir a quantidade de categorias da variável resposta através de combinação de categorias semelhantes, deixando por fim, a variável de interesse com apenas duas possíveis categorias (dicotômica), porém, tal modificação resulta em perda de informação e, por isto, é pouco recomendada (TWISK, 2003).

Segundo Twisk (2003), a forma mais simples de um estudo longitudinal de dados categóricos é aquele em que uma única variável resposta é medida em apenas dois tempos distintos, tal conjuntura pode ser exemplificada pela Tabela 2.1 a seguir, onde uma variável apresentando três categorias ($cat_1, cat_2\ e\ cat_3$) foi medida em dois tempos dessemelhantes ($t\ e\ t'$).

Tabela 2.1. Indicadores de mudança em dois tempos distintos.

		t'			
		cat_1	cat_2	cat_3	Total
t	cat_1	n_{11}	n_{12}	n_{13}	$n_{1(t)}$
	cat_2	n_{21}	n_{22}	n_{23}	$n_{2(t)}$
	cat_3	n_{31}	n_{32}	n_{33}	$n_{3(t)}$
	Total	$n_{1(t')}$	$n_{2(t')}$	$n_{3(t')}$	N

Fonte: Adaptada de Twisk (2003).

Na Tabela 2.1, n representa o número de indivíduos na categoria e N é o número total de indivíduos. Através da tabela 2.1 podemos verificar mudanças no comportamento dos indivíduos. Sendo assim, observamos sua possível passagem de uma categoria para outra. Por exemplo:

O termo n_{11} representa os indivíduos que estavam na categoria c_1 no tempo t e permaneceram nesta mesma categoria no tempo t'. Já o termo n_{12} indica a quantidade de indivíduos que estavam na categoria c_1 no tempo t, passaram a integrar a categoria c_2 no tempo t'. Através do teste de McNemar (Anexo A3) é possível analisar alterações na classificação dos indivíduos.

Segundo Twisk (2003), ao analisar dados que foram medidos em três ou mais tempos distintos, podemos utilizar a mesma ideia de quando temos apenas dois períodos diferentes. Porém, esta forma se torna um tanto trabalhosa, já que é preciso criar $T-1$ tabelas de dupla entrada de dimensão $r\,x\,c$, onde T é a quantidade de vezes que aquela variável de interesse foi medida e r e c são a quantidade de categorias analisadas sendo r apresentada na forma de linha e c na forma de coluna na tabela. Por exemplo, se desejamos observar mudanças em um grupo em T tempos diferentes, deveremos criar $T-1$ cruzamentos de dados através de tabelas de dupla entrada, como mostraremos no Quadro 2.2:

Quadro 2.2. Indicadores de mudanças em dois momentos distintos do tempo para T vezes medidos.

		$t=2$		
		cat_1	...	cat_r
$t=1$	cat_1	n_{11}	...	n_{1r}
	...	...	...	...
	cat_c	n_{c1}	...	n_{cr}

		$t=3$		
		cat_1	...	cat_r
$t=2$	cat_1	n_{11}	...	n_{1r}
	...	...	...	...
	cat_c	n_{c1}	...	n_{cr}

		$t=4$		
		cat_1	...	cat_r
$t=3$	cat_1	n_{11}	...	n_{1r}
	...	...	...	...
	cat_c	n_{c1}	...	n_{cr}

....

		$t=T$		
		cat_1	...	cat_r
$t=T-1$	cat_1	n_{11}	...	n_{1r}
	...	...	...	...
	cat_c	n_{c1}	...	n_{cr}

Fonte: Adaptada de Twisk (2003).

Este método é utilizado quando desejamos estudar apenas a variação de uma variável resposta, sem levar em consideração as variáveis explicativas. Quando o objetivo é estudar a relação causada por variáveis explicativas é necessário utilizar métodos de regressão (TWISK, 2003).

Twisk (2003) afirma que ao trabalhar com análise longitudinal, é necessário verificar a existência de relação entre a variável resposta com as variáveis explicativas e o tempo. Em se tratando de uma variável do tipo contagem, podemos pressupor que a mesma tenha distribuição de Poisson e aplicar a mesma metodologia da regressão log-linear. Porém, para dados longitudinais, deve-se verificar as correlações intra-sujeitos. Sendo assim, as técnicas utilizadas na análise de regressão de Poisson para dados longitudinais, também conhecida como regressão log-linear longitudinal nada mais é do que uma extensão da regressão discutida na seção 2.3 levando em conta a correlação intra-sujeitos para dados que foram obtidos em pelo menos dois tempos distintos. Ao considerar que utilizaremos o método de regressão de Poisson, temos que nossa variável resposta pode estar relacionada com covariáveis tanto dependentes quanto independentes de uma outra variável tempo (t). Estas covariáveis podem ser do tipo contínua, contagem ou fator.

Ao tratarmos então de um modelo de contagem, lembrando que o preditor linear para um modelo de regressão de Poisson é dado por $\eta = \ln(\mu)$, segundo Twisk (2003), podemos adicionar a este modelo mais um indicador de tempo, sugerindo que o modelo admita mais um termo em sua equação. Contudo, Fitzmaurice *et al* (2011) informam que ao levar em consideração possíveis alterações que o valor esperado pode sofrer ao longo do tempo, o preditor linear poderá ser apresentado da seguinte forma:

$$\eta = \ln(\mu_i/t).$$

Uma vez que t é conhecido, temos que:

$$\ln(\mu_i/t) = \ln(\mu_i) - \ln(t).$$

Empregando o preditor linear sugerido ao modelo de regressão de Poisson, observamos que a equação de regressão assumirá a seguinte forma (FITZMAURICE *et al*, 2011):

$$\ln(\mu_i/t) = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} \ldots + \beta_j x_{jt}$$

Onde *j* é o número de covariáveis consideradas no modelo.

A equação acima também pode ser escrita da seguinte forma:

$$\ln(\mu_i) = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} \ldots + \beta_j x_{jt} + \ln(t)$$

Note que mesmo aparecendo ao lado das covariáveis, a expressão $\ln(t)$ não é precedida por nenhum coeficiente de regressão, assim sendo, não se torna necessária a estimação de mais um parâmetro. Por vezes, encontra-se $\ln(t)$ sendo denominada por *"offset".*

Para verificar as correlações intra-individual, como dito na seção 2.4.2 é necessário criar uma tabela capaz de verificar a correlação no decorrer do tempo, como por exemplo, na Tabela 2.2 (TWISK, 2003).

Tabela 2.2. Estrutura de correlação (considerando T=6).

	t_1	t_2	t_3	t_4	t_5	t_6
t_1		ρ_1	ρ_2	ρ_3	ρ_4	ρ_5
t_2	ρ_1		ρ_6	ρ_7	ρ_8	ρ_9
t_3	ρ_2	ρ_6		ρ_{10}	ρ_{11}	ρ_{12}
t_4	ρ_3	ρ_7	ρ_{10}		ρ_{13}	ρ_{14}
t_5	ρ_4	ρ_8	ρ_{11}	ρ_{13}		ρ_{15}
t_6	ρ_5	ρ_9	ρ_{12}	ρ_{14}	ρ_{15}	

Fonte: Twisk (2003).

Modelos adicionais que levam em consideração observações de caráter não linear de um indivíduo podem ser encontradas na literatura (SHEK; MA, 2011), porém tais métodos não serão abordados neste trabalho.

2.4.4 Dados Desbalanceados

Por fim, ressaltaremos que em consequência da não obrigatoriedade de cadastro no *site* da empresa de publicidade, torna-se comum o fato dos usuários não realizem cadastro em todas as edições que adquiriram o cartão. A ocorrência deste não cadastro pode ter alguns motivos como, por exemplo: o esquecimento de cadastro, o desinteresse por realizar cadastro durante alguma edição, dentre outros. Por estes motivos, é corrente encontrarmos usuários que realizaram cadastro em uma edição e em outras não, e usuários que se cadastram em algumas edições, não realizam cadastro em outras e voltam a se cadastrar posteriormente.

Segundo Shaw e Mitchell-Olds (1993), quando o número de observações é diferente em cada período analisado, dizemos que trata-se de um caso de Dados Longitudinais Desbalanceados, e por esta causa, devemos ter cuidado com as análises e interpretações feitas.

O caso mais comum de desbalanceamento, segundo Shaw e Mitchell-Olds (1993), ocorre na biologia quando a unidade experimental é um ser vivo, desta forma, amostras de tamanhos desiguais podem ser ocasionadas por mortalidade, emigração, dificuldade de localizar indivíduos em um censo específico ou mesmo a incapacidade de realocar os indivíduos dentro de determinado tratamento.

É comum o desejo por forçar um equilíbrio nos dados a serem analisados, já que isto pode ser feito de maneira simples, basta excluir os dados de indivíduos que não apareçam em todos os momentos que foram realizadas as pesquisas. Porém, ao descartar algumas informações já coletadas, além de muitas vezes causar prejuízo (perda de tempo e dinheiro), esta exclusão, mesmo que seja feita de forma aleatória, pode diminuir a precisão das estimativas, além disso, pode-se observar diferença nos valores estimados dependendo de quais dados foram deletados (SHAW; MITCHELL-OLDS,1993).

Visto que o problema de não considerar a diferença existente entre o tamanho das amostras pode levar a conclusões equivocadas, se torna necessário o uso de métodos corretos juntamente com um *software* capaz de conjecturar tal fato. Atentando-se a isto, consideraremos o ajuste de modelos de efeitos aleatórios a partir do *software* Stata para realizar as estimativas dos parâmetros, uma vez que tais modelos permitem a consideração de dados desbalanceados.

3 BANCO DE DADOS

O Capítulo 3, intitulado Banco de Dados, será omitido desta versão da monografia com o propósito de garantir ao veiculo publicitário preservação de seus dados e informações dos clientes.

Páginas omitidas: 35 a 48.

4 APLICAÇÃO E RESULTADOS

Neste capítulo, iremos aplicar os métodos de regressão de Poisson no banco de dados longitudinal estudado a fim de observar se de fato há alguma diferença na quantidade de promoções utilizadas pelo usuário, dado as características do mesmo e se também houve diferença na quantidade de utilizações ao longo do período analisado. Ajustaremos o modelo tendo como variável de interesse a quantidade de utilizações (qtd_utiliz). Utilizaremos as seguintes variáveis explicativas para realizar nossa análise:

sexo_usu,
etaria_usu,
zona_usu,
prof_usu e
t,

descritas no capítulo 3.

Primeiramente iremos considerar o Efeito Aleatório tendo uma distribuição Normal e analisaremos os resultados gerados. Posteriormente iremos ajustar o modelo supondo que o efeito aleatório tenha distribuição Gamma e verificaremos se os resultados encontrados são diferentes para os dois casos.

4.1 Modelo de regressão de Poisson considerando o efeito aleatório com distribuição Normal.

O comando utilizado no *software* Stata para produzir as análises foi:

```
xtpoisson qtdutliz i.sexo_usu i. etaria_usu i.zona_usu i.prof_usu t, normal irr
```

E os resultados obtidos podem ser observados na Tabela 4.1 a seguir:

Tabela 4.1. Modelo de regressão de Poisson considerando o efeito aleatório com distribuição Normal.

Random–effects Poisson regression		Number of obs	=	9839
Group Variable: numero_usu		Number of groups	=	6425
Random effects u_i ~ Gaussian		Obs per group: min	=	1
		avg	=	1.5
		max	=	12
		Wald chi2 (19)	=	555.45
Log likelihood = -22851.495		Prob > chi2	=	0.0000

qtdutliz	IRR	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Sexo						
Masculino	1.019215	.0204445	0.95	0.343	.9799219	1.060083
Faixa-etária						
20 a 34 anos	1.110027	.0298636	3.88	0.000	1.053012	1.170129
35 a 49 anos	1.110534	.0420379	2.77	0.006	1.031123	1.19606
50 anos ou mais	1.174141	.0632403	2.98	0.003	1.05651	1.304868
Zona						
Leste	.8300681	.0262494	-5.89	0.000	.7801821	.8831438
Nordeste	.7779232	.0320119	-6.10	0.000	.7176446	.8432649
Norte	.7584578	.0274647	-7.63	0.000	.7064938	.814244
Oeste	.839497	.0325633	-4.51	0.000	.77804	.9058086
Sudeste	.8875276	.0365464	-2.90	0.004	.8187123	.9621271
Sul	.8860076	.0309298	-3.47	0.001	.8274137	.9487508
Outras cidades	.6348009	.1016135	-2.84	0.005	.4638573	.8687416
Profissão						
Emp. e autônomos	.9554807	.024393	-1.78	0.074	.9088478	1.004506
Área tecnológica	.8795636	.0348071	-3.24	0.001	.8139215	.9504997
Saúde	.9429621	.0357726	-1.55	0.122	.8753922	1.015747
Serviço público	.9535714	.0467284	-0.97	0.332	.8662462	1.0497
Educ. e docência	.9827258	.0426113	-0.40	0.688	.9026596	1.069894
Artes, entret. e bem estar	.9357641	.0475431	-1.31	0.191	.8470705	1.033744
Outros	1.083697	.0540289	1.61	0.107	.9828114	1.194938
Tempo	-------	---------	---------	-------	----------	------------
/lnsig2u	-1.186078	.0298134	-39.78	0.000	-1.244512	-1.127645
sigma_u	.5526452	.0082381			.5367323	.5690298

Likelihood-ratio test of sigma_u=0: chibar2(01) = 8164.52 Pr>=chibar2 = 0.000

Na tabela apresentada, observamos que a estatística de Wald possuindo distribuição Qui-quadrado com 19 graus de liberdade indica que há significância em ao menos uma variável categórica estudada. As estatísticas de Wald e Log Likelihood indicam que o modelo estudado, considerando as variáveis citadas, é significativo, e para este caso, mostram também que há redução na soma dos quadrados dos resíduos (*deviance*).

O IRR representa a razão de incidência, indicando o quanto uma categoria influi na quantidade de promoções utilizadas em relação à categoria de referência. A contribuição pode ser calculada através da exponencial do coeficiente aleatório (e^{β_i}), para cada estimativa dos coeficientes de regressão.

Observamos a significância de cada categoria de uma variável através do valor de $P > |z|$. Considerando um nível de significância de 5%, dizemos que a categoria é significativa se $P > |z|$ apresentar um valor menor do que 0,05. Caso contrário, dizemos que a mesma não é significativa ao nível de 5%.

No final da tabela temos o valor para sigma_u, sendo este um indicador para a variância marginal. Uma vez que observarmos sigma maior que zero, significa que a variância é maior que a média do intercepto aleatório μ. No caso estudado temos que sigma_u > 0, indicando a presença de uma superdispersão. Analogamente temos o valor de /lnsig2u, que nada mais é do que o logaritmo do quadrado de sigma_u.

Com relação às variáveis independentes analisadas, observamos através dos resultados obtidos que a variável sexo não é significativa ao nível de 5%, ou seja, os homens e as mulheres cadastradas utilizam em média a mesma quantidade de promoções ao longo do tempo.

Quanto à faixa-etária, notamos que na medida em que a idade aumenta, a quantidade de utilizações também se eleva. Em comparação com a primeira faixa-etária (0 a 19 anos), observamos que pessoas que tem idades entre 20 a 49 anos utilizam cerca de 11% a mais que os mais jovens. Essa diferença amplia-se um pouco mais quando observamos os mais velhos (50 ou mais anos), sendo que estes consomem em média, 17% a mais de promoções que os mais jovens.

Observando a variável referente á região onde o usuário reside, nota-se que a região central é a que mais utiliza as promoções. As estimativas sugerem que, apesar de serem significativas ao nível de 5%, todas as outras regiões têm consumo inferior ao da região central, sendo a zona norte a que menos utiliza, cerca de 25% a menos que o centro. E como era de se esperar, moradores de outras cidades utilizam menos ainda.

Com relação à ocupação do indivíduo, temos que estudantes e as demais profissões possuem características semelhantes em relação à quantidade de utilizações. Apenas usuários que se enquadram na categoria “Área Tecnológica” apresentam uma utilização mais baixa, cerca de 13% a menos que os estudantes.

Os resultados obtidos através do modelo de regressão de Poisson reafirmam de forma mais veemente as análises exploratórias feitas no capítulo 3.

As estimativas feitas para os coeficientes de regressão serão apresentados na Tabela 4.2 a seguir:

Tabela 4.2. Coeficientes de regressão encontrados, considerando o efeito aleatório com distribuição Normal.

Qtdutliz	Coef.	P>|z|	[95% Conf. Interval]	
Sexo				
Masculino	.0190325	0.343	-.0202824	.0583475
Faixa-etária				
20 a 34 anos	.1043845	0.000	.0516546	.1571144
35 a 49 anos	.1048407	0.006	.0306487	.1790328
50 anos ou mais	.1605366	0.003	.0549711	.2661021
Zona				
Leste	-.1862476	0.000	-.2482279	-.1242672
Nordeste	-.2511274	0.000	-.3317808	-.1704741
Norte	-.2764681	0.000	-.3474409	-.2054953
Oeste	-.1749523	0.000	-.2509774	-.0989273
Sudeste	-.1193156	0.004	-.2000225	-.0386087
Sul	-.1210298	0.001	-.1894505	-.0526091
Outras cidades	-.4544439	0.005	-.7681783	-.1407096
Profissão				
Empresarial e autônomos	-.0455407	0.074	-.0955776	.0044962
Área tecnológica	-.1283294	0.001	-.2058914	-.0507675
Saúde	-.0587292	0.122	-.1330832	.0156247
Serviço público	-.0475409	0.332	-.1435861	.0485043
Educ. e docência	-.0174251	0.688	-.1024097	.0675595
Artes, entret. e bem estar	-.0663919	0.191	-.1659713	.0331875
Outros	.0803781	0.107	-.017338	.1780942
Tempo	---------	---------	----------	----------
Constante	1.34654	0.000	1.292153	1.400927

Note que agora também há um valor para a constante β_0.

Uma vez que a variável sexo é a única não significativa no modelo, podemos retira-la do mesmo. Contudo, para este trabalho, a retirada da variável sexo não afetou a significância encontrada nas categorias e alterou muito pouco a os valores estimados para os coeficientes de regressão. Sendo assim, não iremos expor os valores encontrados, e abriremos a oportunidade de utilizar estes resultados em trabalhos futuros.

4.2 Modelo de regressão de Poisson considerando o efeito aleatório com distribuição Gamma.

Com a finalidade de observar se há alguma diferença ao considerar o efeito aleatório com distribuição Gamma, já que este é o *default* do Stata para modelos de Poisson, iremos apresentar um Modelo de regressão de Poisson considerando o efeito aleatório com distribuição Gamma para os mesmo dados longitudinais do veículo publicitário.

O comando utilizado no *software* Stata para produzir as análises foi:

```
xtpoisson qtdutliz i.sexo_usu i. etaria_usu i.zona_usu i.prof_usu t, irr
```

A partir dos resultados obtidos na Tabela 4.3 a seguir, observa-se que as mesmas categorias são significativas ao nível de 5% tanto para modelos que consideram o efeito aleatório com distribuição Normal, quanto para modelos que consideram o efeito aleatório com distribuição Gamma.

Nota-se também que os valores encontrados no IRR são bem parecidos em ambos os casos. Portanto, há motivos para supor que ambos os modelos podem ser utilizados neste caso sem que haja grande modificação nos resultados observados. Inclusive, quando retira-se a variável referente ao sexo do usuário, novamente observa-se pouca alteração nos resultados encontrados.

Tabela 4.3. Modelo de regressão de Poisson considerando o efeito aleatório com distribuição Gamma.

Random–effects Poisson regression	Number of obs	=	9839
Group Variable: numero_usu	Number of groups	=	6425
Random effects u_i ~ Gamma	Obs per group: min	=	1
	avg	=	1.5
	max	=	12
	Wald chi2 (19)	=	632.36
Log likelihood = -23092.662	Prob > chi2	=	0.0000

Qtdutliz	IRR	Std. Err.	z	P>|z|	[95% Conf. Interval]	
Sexo						
Masculino	1.025966	.0200091	1.31	0.189	.9874885	1.065942
Faixa-etária						
20 a 34 anos	1.109456	.0291766	3.95	0.000	1.053719	1.16814
35 a 49 anos	1.111398	.041088	2.86	0.004	1.033716	1.194918
50 anos ou mais	1.203834	.0628494	3.55	0.000	1.086744	1.333539
Zona						
Leste	.8325208	.025594	-5.96	0.000	.7838388	.8842262
Nordeste	.7768653	.0311067	-6.31	0.000	.7182283	.8402895
Norte	.7473771	.026518	-8.21	0.000	.6971689	.8012013
Oeste	.8136653	.0309949	-5.41	0.000	.7551288	.8767395
Sudeste	.8997572	.0360069	-2.64	0.008	.8318817	.9731709
Sul	.8668982	.029667	-4.17	0.000	.810659	.9270389
Outras cidades	.6045244	.0966031	-3.15	0.002	.441969	.8268676
Profissão						
Emp. e autônomos	.9603344	.0239851	-1.62	0.105	.9144565	1.008514
Área tecnológica	.8913144	.03431	-2.99	0.003	.8265421	.9611625
Saúde	.9519447	.0352001	-1.33	0.183	.8853946	1.023497
Serviço público	.9428895	.0451577	-1.23	0.219	.858409	1.035684
Educ. e docência	1.001298	.042243	0.03	0.975	.921834	1.087612
Artes, entret. e bem estar	.9247766	.0458921	-1.58	0.115	.8390656	1.019243
Outros	1.07251	.052206	1.44	0.150	.9749175	1.179872
Tempo	---------	---------	----------	---------	----------	----------
/lnalpha	-1.144665	.0279714			-1.199488	-1.089842
Alpha	.3183304	.0089041			.3013484	.3362695
Likelihood-ratio test of alpha=0:		Chibar2(01)= 7682.19			Prob>=chibar2 = 0.000	

E os valores encontrados para os coeficientes de regressão serão mostrados na Tabela 4.4 a seguir:

Tabela 4.4. Coeficientes de regressão encontrados, considerando o efeito aleatório com distribuição Gamma.

Qtdutliz	Coef.	P>|z|	[95% Conf. Interval]	
Sexo				
Masculino	.0256341	0.189	-.0125904	.0638587
Faixa-etária				
20 a 34 anos	.1038697	0.000	.0523263	.1554131
35 a 49 anos	.1056188	0.004	.0331597	.1780779
50 anos ou mais	.1855112	0.000	.083186	.2878365
Zona				
Leste	-.1832971	0.000	-.2435519	-.1230424
Nordeste	-.2524883	0.000	-.3309678	-.1740089
Norte	-.2911853	0.000	-.3607276	-.2216431
Oeste	-.2062061	0.000	-.2808669	-.1315453
Sudeste	-.1056303	0.008	-.1840651	-.0271956
Sul	-.1428337	0.000	-.2099077	-.0757597
Outras cidades	-.5033132	0.002	-.8165156	-.1901107
Profissão				
Empresarial e autônomos	-.0404737	0.105	-.0894253	.0084778
Área tecnológica	-.1150581	0.003	-.1905045	-.0396117
Saúde	-.0492483	0.183	-.1217219	.0232253
Serviço público	-.0588062	0.219	-.1526746	.0350622
Educ. e docência	.0012974	0.975	-.0813901	.0839848
Artes, entret. e bem estar	-.0782031	0.115	-.1754664	.0190603
Outros	.0700018	0.150	-.0254024	.1654059
Tempo	---------	-----	-----------	----------
Constante	1.521783	0.000	1.468919	1.574647

Como podemos observar, os coeficientes estimados para o modelo de regressão de Poisson que considerou o efeito aleatório tendo distribuição Gamma apresentou estimativas bem parecidas em relação ao modelo que considerou o efeito aleatório tendo distribuição Normal.

5 CONCLUSÃO E CONSIDERAÇÕES FINAIS.

Esta monografia conseguiu demonstrar como uma análise de regressão pode ser importante para o reconhecimento de quanto uma determinada covariável pode influenciar no valor da variável independente. No presente estudo de caso, a aplicação da metodologia de regressão de Poisson para dados de contagem longitudinais possibilitou o conhecimento das características pessoais dos usuários que se beneficiaram do veiculo publicitário, conforme a quantidade de promoções utilizadas pelos mesmos.

Através dos resultados apresentados no capítulo 4, podemos perceber que os usuários cadastrados que mais se beneficiam com as promoções da revista são pessoas mais velhas, com idade acima dos 49 anos. Com relação à zona onde o usuário reside, notamos que a região central é a que apresenta maior volume de utilizações. E em relação à ocupação do indivíduo, temos que estudantes e profissionais de diversos ramos têm consumo semelhante, exceto os trabalhadores da área tecnológica que apresentam consumo inferior.

O presente trabalho demonstrou de maneira breve como as técnicas de modelagem para dados de contagem longitudinais podem ser importantes na estratégia adotada dentro de qualquer setor. Por se tratar de um estudo de caso contendo dados reais, torna-se mais fácil e interessante para qualquer pessoa compreender o quanto as técnicas aqui empregadas são poderosas na tomada de decisão.

A maior dificuldade encontrada nesta monografia foi a construção do banco de dados utilizado. Uma vez que os dados estavam arquivados no servidor na empresa em outro formato. Sendo assim, foi preciso adaptar o banco a um formato que o *software* utilizado conseguisse realizar as operações e estatísticas desejadas.

Hoje, como sabemos o formato que devemos utilizar, pode-se fazer diversas propostas para trabalhos futuros. Uma vez que o atual servidor possui dados desde a primeira edição até a edição atual, sendo que este banco possui muitas informações, é possível observar se há sazonalidade na quantidade de vendas o que poderia permitir a otimização da tiragem a ser feita. Trabalhos que consideram não apenas os usuários cadastrados, mas também os não cadastrados podem ser feitos a fim de observar a característica de consumo do usuário juiz-forano em relação ao ramo do anunciante, período da semana e dia em que houve mais promoções utilizadas. Lembrando que também o veículo publicitário está presente no mercado em outras cidades, é possível comparar as características de consumo de juiz-foranos com os demais. Por fim, temos que diversos outros trabalhos podem ser realizados utilizando o banco de dados da empresa.

BIBLIOGRAFIA

AGRESTI, A. **An Introduction to Categorical Data Analysis**. 2ed. New Jersey, EUA: Wiley, 2007.

BENAZZI, João Renato de Souza Coelho; PEDRA, Bruno Yagelovic. **Compras Coletivas: uma análise exploratória de sua utilidade para as empresas anunciantes.** In: Simpósio em Tecnologias Digitais e Sociabilidade, 2011 Salvador.

CASELLA, George; BERGER, Roger L. **Inferência Estatística**. 2ed. São Paulo: Cengage Learning, 2010.

CHARNET, Reinaldo et al. **Análise de Modelos de Regressão Linear: com aplicações**. 2ed. São Paulo: Editora da Unicamp, 2008.

CORDEIRO, Gauss M.; DEMÉTRIO, Clarice G.B. **Modelos Lineares Generalizados e Extensões.** Piracicaba, 2008.

COSTA, Silvano Cesar da. **Modelos Lineares Generalizados Mistos Para Dados Longitudinais**. 2003. Tese (Doutorado em Agronomia) – Escola Superior de Agricultura Luiz Queiroz, Universidade de São Paulo, São Paulo.

DIGGLE, P.J. et al. **Analysis of Longitudinal Data**. 2 ed. Oxford, EUA: Oxford University Press, 2002.

DRUKKER, David. **Testing for Serial Correlation in Linear Panel-data Models.** The Stata Journal, 2003. p. 168-177.

FITZMAURICE, Garrett et al. **Applied Longitudinal Analysis**. Boston, Massachusetts, EUA: Wiley, 2004.

McCULLAGH, P.; NELDER, J. A. **Generalized Linear Models**. 2 ed. Flórida, EUA: Chapman & Hall, 1999.

MENARD, S. **Longitudinal Research**. Newbury Park: Sage, 1991.

MEYER, Paul L. **Probabilidade. Aplicações à Estatística**. 2ed. Rio de Janeiro: Livros Técnicos e Científicos Editora S.A., 1982

MIGUEL, Paulo Augusto Cauchick. **Estudo de caso na engenharia de produção: estruturação e recomendações para sua condução**. Revista Produção, São Paulo, v. 17, n. 1, p. 216-229, 2007.

NELDER, J. A.; WEDDERBURN, R. W. M. **Generalized linear models**. Journal of the Royal Statistical Society. Serie A (General), Vol 135, No. 3, (1972), p. 370-384.

PAULA, Gilberto. **Modelos de Regressão: com apoio computacional.** São Paulo: Universidade de São Paulo, 2013.

PORTAL ACTION, 2014. Em **<http://www.portalaction.com.br>** Acesso em 10 nov. 2014

RABE-HESKETH, Sophia; SKRONDAL, Anders. **Multilevel and Longitudinal Modeling Using Stata - volume II: categorical responses, counts, and survival**. 3ed. Texas, Eua: Stata Press, 2012.

SHAW, R. G.; MITCHELL-OLDS, T. **Anova for Unbalanced Data: an overview. Ecology**, Vol 74, No. 6, (1993), p. 1638-1645.

SHEK, Daniel; MA, Cecilia. **Longitudinal Data Analyses Using Linear Mixed Models in SPSS: concepts, procedures and illustrations**. The Scientific World Journal, 2011. Em <http://www.hindawi.com/journals/tswj/2011/246739/abs/> Acesso em 04 out 2014.

TURKMAN, Antónia; SILVA, Giovani. **Modelos Lineares Generalizados: da teoria à prática.** Lisboa, Portugal, 2000.

TWISK, Jos W.R. **Applied Longitudinal Data Analysis for Epidemiology: a practical guide.** Cambridge, EUA: Cambridge University Press, 2003.

VIEIRA, Marcel de Toledo. **A Consideração da Amostragem Complexa na Análise de Dados Longitudinais.** In: 57ª Reunião Anual da RBras, Universidade de São Paulo, Piracicaba, 2012.

WERKEMA, M. C. C.; AGUIAR, S. **Análise de regressão: como entender o relacionamento entre as variáveis de um processo.** Belo Horizonte: Fundação Christiano Ottoni da Escola de Engenharia da UFMG, 1996.

ANEXOS

Anexo A1 - Método de Newton-Raphson e Método Escore de Fisher

Método de Newton-Raphson

Em diversas situações em que se deseja estimar os parâmetros de um modelo de regressão, por exemplo, β_0 e β_1, através de métodos de máxima verossimilhança, podemos encontrar equações não-lineares. Uma solução para resolver este tipo de problema é recorrer a métodos numéricos interativos como o Método de Newton-Raphson.

A prática deste método é feita expandindo uma determinada função $U(\beta)$ em torno de um ponto inicial $\beta^{(0)}$, uma vez que:

- $U(\beta)$ são derivadas de primeira ordem do logaritmo da função de verossimilhança em relação aos parâmetros do modelo.
- $U'(\beta)$ são derivadas segundas do logaritmo da função de verossimilhança.

Temos então que:

$$U(\beta) \approx U\left(\beta^{(0)}\right) + U'(\beta^{(0)})(\beta - \beta^{(0)}).$$

Ao repetirmos o processo acima, conseguimos chegar ao processo interativo, tal que:

$$\beta^{(m+1)} = \beta^{(m)} + [-U'\left(\beta^{(m)}\right)]^{-1} U'\left(\beta^{(m)}\right),$$

sendo $m = (0, 1, \dots, k)$ e $-U'$ uma matriz positiva definida.

Caso $-U'$ não for positiva definida, pode-se substitui-la pela matriz de informação de Fisher (PORTAL ACTION, 2014).

Método Escore de Fisher

O Método Escore de Fisher surge como opção quando o Método de Newton-Raphson (descrito na página anterior) apresenta demora em convergir a um valor, ou mesmo oscilar muito.

Resumidamente, substitui-se $U'(\beta)$ pelo seu valor esperado, $E[U'(\beta)]$.

Por vezes, é mais fácil calcular $E[U'(\beta)]$ do que $U'(\beta)$. O valor esperado de $U'(\beta)$ é conhecido como Informação de Fisher (PORTAL ACTION, 2014).

Anexo A2 - Mínimos Quadrados

A técnica dos Mínimos Quadrados é utilizada para encontrar estimadores de parâmetros. A fim de tornar fácil a demonstração do método de Mínimos Quadrados, vamos utilizar a equação de regressão escrita da forma mais simples:

$$y_i = \beta_0 + \beta_1 x_i$$

Deseja-se então, encontrar o mínimo da seguinte função, em β_0 e β_1:

$$\sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)]^2$$

Em seguida, devemos fazer a derivada em função de β_0 e β_1.

Primeiramente, derivando em relação à β_0, chegamos ao seguinte resultado:

$$\frac{\partial}{\partial \beta_0} \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)]^2 \; = \; -2 \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)],$$

Igualando o resultado a zero temos:

$$-2 \sum_{i=1}^{n} [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)] = 0 \,,$$

Notamos então que:

$$\hat{\beta}_0 = \frac{1}{n} \sum_{i=1}^{n} y_i - \hat{\beta}_1 \frac{1}{n} \sum_{i=1}^{n} x_i$$

Agora, iremos realizar cálculos semelhantes, porém, derivando em relação a β_1.

$$\frac{\partial}{\partial \beta_1} \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)]^2 \; = \; -2 \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_i)] \, x_i$$

Igualando o resultado a zero, temos:

$$-2\sum_{i=1}^{n}[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]\, x_i = 0$$

Como já obtivemos o valor de $\hat{\beta}_0$, iremos substitui-lo na equação anterior e por fim, chegaremos a seguinte equação:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2}$$

Assim, obtemos então os estimadores de β_0 e β_1 (CHARNET *et al*, 2008).

Anexo A3 - Teste de McNemar

O Teste de McNemar é utilizado quando deseja-se realizar alguma comparação entre amostras pareadas através de tabelas de contingencia. Quando as amostras são correlacionadas, é indicado o uso do Teste de McNemar para constatar se há alguma diferença entre as frequências de duas amostras.

O teste pode ser aplicado em diversos casos, como por exemplo, observar a presença ou ausência de determinado fator em uma amostra ao longo do tempo. Utilizaremos a Tabela A4.1 para exemplificar nosso teste.

Tabela A4.1. Frequências Oriundas de Amostras Pareadas.

		$Amostra\ I$		
		$I\ Ausente$	$II\ Presente$	Total
$Amostra\ II$	$II\ Presente$	*A*	*b*	*a+b*
	$II\ Ausente$	*C*	*d*	*c+d*
	Total	*a+c*	*b+d*	*n=a+b+c+d*

Fonte: Portal Action <http://www.portalaction.com.br>.

Nosso interesse é verificar se houve alguma alteração quanto à classificação do indivíduo entre as duas amostras, ou seja, estamos interessados em estudar as frequências *a* e *d.*

McNemar propôs a seguinte estatística, que têm distribuição Qui-Quadrado com 1 grau de liberdade.

$$Q_{obs}^2 = \frac{(a-d)^2}{(a+d)} \sim \chi_{1GL}^2$$

Observa-se que a hipótese nula, neste caso, admite que não há diferenças entre as duas amostras (PORTAL ACTION, 2014).

Anexo A4 – Número de pessoas que realizaram cadastro em determinada edição do Veículo Publicitário.

Este anexo é uma saída do *software* Stata, indicando a quantidade de pessoas cadastradas em cada uma das edições da revista.

A coluna *Pattern* indica a edição em que houveram cadastros

```
xtdes, patterns(1000)

numero_usu:  1, 2, ..., 6673                                  n =       6669
       t:  1, 2, ..., 12                                      T =         12
           Delta(t) = 1 unit
           Span(t)  = 12 periods
           (numero_usu*t uniquely identifies each observation)

Distribution of T_i:   min      5%     25%       50%       75%     95%     max
                         1       1       1         1         2       4      12

     Freq.  Percent    Cum. |  Pattern
 ---------------------------+--------------
      719     10.78   10.78 |  .....1......
      587      8.80   19.58 |  .1..........
      571      8.56   28.15 |  ......1.....
      542      8.13   36.27 |  1...........
      433      6.49   42.77 |  ..1.........
      411      6.16   48.93 |  .......1....
      370      5.55   54.48 |  ....1.......
      322      4.83   59.30 |  ...1........
      275      4.12   63.43 |  ..........1.
      258      3.87   67.30 |  .........1..
      205      3.07   70.37 |  ...........1
      183      2.74   73.11 |  ........1...
      130      1.95   75.06 |  11..........
       72      1.08   76.14 |  ......11....
       63      0.94   77.09 |  ....11......
       58      0.87   77.96 |  .....11.....
       54      0.81   78.77 |  .11.........
       48      0.72   79.49 |  ..11........
       43      0.64   80.13 |  .1...1......
       38      0.57   80.70 |  ...11.......
       29      0.43   81.14 |  .........11.
       25      0.37   81.51 |  .......11...
       25      0.37   81.89 |  1....1......
       23      0.34   82.23 |  1.....1.....
       22      0.33   82.56 |  ..........11
       22      0.33   82.89 |  ....1.1.....
       22      0.33   83.22 |  .1..1.......
       21      0.31   83.54 |  11...1......
       20      0.30   83.84 |  ..1..1......
       19      0.28   84.12 |  .1.1........
       19      0.28   84.41 |  1.1.........
       18      0.27   84.68 |  .1....1.....
       16      0.24   84.92 |  ...1..1.....
       15      0.22   85.14 |  1..1........
       14      0.21   85.35 |  ........11..
       14      0.21   85.56 |  .......1..1.
       13      0.19   85.75 |  .....1.1....
       13      0.19   85.95 |  ....111.....
       13      0.19   86.14 |  ..1...1.....
       13      0.19   86.34 |  1...1.......
       12      0.18   86.52 |  .......1...1
       11      0.16   86.68 |  ......111...
       11      0.16   86.85 |  ...111......
       11      0.16   87.01 |  ..1.1.......
       11      0.16   87.18 |  ..111.......
       11      0.16   87.34 |  11....1.....
       11      0.16   87.51 |  111.........
       10      0.15   87.66 |  .......1.1..
       10      0.15   87.81 |  ......1....1
```

10	0.15	87.96	......1.1...
10	0.15	88.11	.....1....1.
10	0.15	88.26	....1..1....
9	0.13	88.39	111
9	0.13	88.53	......1...1.
9	0.13	88.66	..1....1....
9	0.13	88.80	.1........1.
9	0.13	88.93	.1...11.....
9	0.13	89.07	.1..11......
9	0.13	89.20	.11...1.....
8	0.12	89.32	.........1.1
8	0.12	89.44	...1.1......
8	0.12	89.56	.1.....1....
8	0.12	89.68	.11..1......
8	0.12	89.80	11...11.....
8	0.12	89.92	11..1.......
7	0.10	90.03	......1..1..
7	0.10	90.13	1....11.....
6	0.09	90.22	........1.1.
6	0.09	90.31	.....1...1..
6	0.09	90.40	.....111....
6	0.09	90.49	.111........
5	0.07	90.57	......1111..
5	0.07	90.64	....11.1....
5	0.07	90.72	....1111....
5	0.07	90.79	...11111....
5	0.07	90.87	..1.11......
5	0.07	90.94	..11.11.....
5	0.07	91.02	.111.1......
5	0.07	91.09	1......1....
5	0.07	91.17	11.1........
4	0.06	91.23	.......11..1
4	0.06	91.29	....1.....1.
4	0.06	91.35	....1....1..
4	0.06	91.41	...1..11....
4	0.06	91.47	...1.11.....
4	0.06	91.53	..1........1
4	0.06	91.59	..1.......1.
4	0.06	91.65	..1......1..
4	0.06	91.71	..1..11.....
4	0.06	91.77	..111.1.....
4	0.06	91.83	.1.1.11.....
4	0.06	91.89	.11....1....
4	0.06	91.95	.11.1.......
4	0.06	92.01	.11.11......
4	0.06	92.07	.1111111....
4	0.06	92.13	11.....1....
4	0.06	92.19	111..1......
4	0.06	92.25	111.1.......
3	0.04	92.29	........1..1
3	0.04	92.34	.......1.11.
3	0.04	92.38	.......11.1.
3	0.04	92.43	.......111..
3	0.04	92.47	.......1111.
3	0.04	92.52	......11.1..
3	0.04	92.56	.....1.....1
3	0.04	92.61	.....1.111..
3	0.04	92.65	.....11...11
3	0.04	92.70	...1.......1
3	0.04	92.74	...1......1.
3	0.04	92.79	...1...1....
3	0.04	92.83	...11.1.....
3	0.04	92.88	..1..1.1....
3	0.04	92.92	..11..1.....
3	0.04	92.97	..1111......
3	0.04	93.01	.1..1.1.....
3	0.04	93.06	.1.1..1.....
3	0.04	93.10	.1.11.......
3	0.04	93.15	1........1..
3	0.04	93.19	1.....11....
3	0.04	93.24	1...11......
3	0.04	93.28	1..1.1......
3	0.04	93.33	1.1.1.......
3	0.04	93.37	1.11........
3	0.04	93.42	11..111.....
3	0.04	93.46	111..1...1..

2	0.03	93.49	........11.1
2	0.03	93.52	.......1.1.1
2	0.03	93.55	11111
2	0.03	93.58	......1..11.
2	0.03	93.61	......1.11..
2	0.03	93.64	......11...1
2	0.03	93.67	......111.1.
2	0.03	93.70	......1111.1
2	0.03	93.73	.....1....11
2	0.03	93.76	.....1..111.
2	0.03	93.79	.....1.11...
2	0.03	93.82	.....1.11..1
2	0.03	93.85	.....11....1
2	0.03	93.88	.....11...1.
2	0.03	93.91	.....11..1..
2	0.03	93.94	.....11..111
2	0.03	93.97	.....11.1...
2	0.03	94.00	.....111..1.
2	0.03	94.03	.....111.11.
2	0.03	94.06	.....1111...
2	0.03	94.09	....1.....11
2	0.03	94.12	....1...1...
2	0.03	94.15	....1..1..1.
2	0.03	94.18	....11..1...
2	0.03	94.21	....1111..1.
2	0.03	94.24	...1.....1..
2	0.03	94.27	...1....1...
2	0.03	94.30	...1.1...1..
2	0.03	94.33	...1.1.1....
2	0.03	94.36	...1.111....
2	0.03	94.39	...11......1
2	0.03	94.42	...11.....1.
2	0.03	94.45	...111.1....
2	0.03	94.48	...111.1..11
2	0.03	94.51	...1111.....
2	0.03	94.54	..1....1.1..
2	0.03	94.57	..1...1...1.
2	0.03	94.60	..1...11....
2	0.03	94.63	..1...11...1
2	0.03	94.66	..11111.....
2	0.03	94.69	.1.......1..
2	0.03	94.72	.1.......11.
2	0.03	94.75	.1......1...
2	0.03	94.78	.1..1....1..
2	0.03	94.81	.1.1...1....
2	0.03	94.84	.1.1.1......
2	0.03	94.87	.1.111......
2	0.03	94.90	.11..1.1....
2	0.03	94.93	.11..111.11.
2	0.03	94.96	.111...1....
2	0.03	94.99	.1111.......
2	0.03	95.02	.111111.....
2	0.03	95.05	1..........1
2	0.03	95.08	1.........1.
2	0.03	95.11	1..1..1.....
2	0.03	95.14	1..1.11.....
2	0.03	95.17	11......1...
2	0.03	95.20	11....11....
2	0.03	95.23	11..11......
2	0.03	95.26	11..11...1..
2	0.03	95.29	11.111......
2	0.03	95.32	111....1..11
2	0.03	95.35	111.111.....
2	0.03	95.38	1111........
2	0.03	95.41	1111...1....
2	0.03	95.44	11111.......
2	0.03	95.47	111111......
2	0.03	95.50	11111111...1
1	0.01	95.52	........1.11
1	0.01	95.53	........111.
1	0.01	95.55	.......1.111
1	0.01	95.56	......1...11
1	0.01	95.58	......1..111
1	0.01	95.59	......1.11.1
1	0.01	95.61	......11..11
1	0.01	95.62	......11.111

1	0.01	95.64	......111..1
1	0.01	95.65	111111
1	0.01	95.67	.....1...1.1
1	0.01	95.68	.....1..1...
1	0.01	95.70	.....1..1.1.
1	0.01	95.71	.....1.1...1
1	0.01	95.73	.....1.11.1.
1	0.01	95.74	.....1.1111.
1	0.01	95.76	.....11.111.
1	0.01	95.77	.....1111.1.
1	0.01	95.79	....1......1
1	0.01	95.80	....1..1...1
1	0.01	95.82	....1..1.1..
1	0.01	95.83	....1..1.111
1	0.01	95.85	....1..11..1
1	0.01	95.86	....1..111..
1	0.01	95.88	....1.1....1
1	0.01	95.89	....1.1...1.
1	0.01	95.91	....1.1...11
1	0.01	95.92	....1.1..1..
1	0.01	95.94	....1.1.1...
1	0.01	95.95	....1.11....
1	0.01	95.97	....1.111...
1	0.01	95.98	....1.111.1.
1	0.01	96.00	....11....1.
1	0.01	96.01	....11.11...
1	0.01	96.03	....11.11..1
1	0.01	96.04	....11.11.11
1	0.01	96.06	....111.11..
1	0.01	96.07	....111.1111
1	0.01	96.09	....1111.11.
1	0.01	96.10	....111111..
1	0.01	96.12	11111111
1	0.01	96.13	...1.....11.
1	0.01	96.15	...1...1..1.
1	0.01	96.16	...1...1..11
1	0.01	96.18	...1...11...
1	0.01	96.19	...1..1....1
1	0.01	96.21	...1..1.1...
1	0.01	96.22	...1..11..1.
1	0.01	96.24	...1.1...1.1
1	0.01	96.25	...1.1..11..
1	0.01	96.27	...1.1.1..1.
1	0.01	96.28	...1.1.1.11.
1	0.01	96.30	...1.111.1..
1	0.01	96.31	...1.1111...
1	0.01	96.33	...1.111111.
1	0.01	96.34	...1.1111111
1	0.01	96.36	...11...1...
1	0.01	96.37	...11...11..
1	0.01	96.39	...11..1....
1	0.01	96.40	...11..1..1.
1	0.01	96.42	...11..11...
1	0.01	96.43	...11..1111.
1	0.01	96.45	...11.1...1.
1	0.01	96.46	...11.1..1..
1	0.01	96.48	...11.1.1...
1	0.01	96.49	...11.1.111.
1	0.01	96.51	...11.11....
1	0.01	96.52	...11.11...1
1	0.01	96.54	...11.111...
1	0.01	96.55	...11.1111..
1	0.01	96.57	...111..1...
1	0.01	96.58	...1111..1..
1	0.01	96.60	...11111.1.1
1	0.01	96.61	...11111.111
1	0.01	96.63	...111111.1.
1	0.01	96.64	...11111111.
1	0.01	96.66	...111111111
1	0.01	96.67	..1......1.1
1	0.01	96.69	..1......11.
1	0.01	96.70	..1.....1...
1	0.01	96.72	..1.....1111
1	0.01	96.73	..1....1..1.
1	0.01	96.75	..1....11...
1	0.01	96.76	..1....11..1

```
1        0.01    96.78 |  ..1...1.1...
1        0.01    96.79 |  ..1..1.....1
1        0.01    96.81 |  ..1..1....1.
1        0.01    96.82 |  ..1..1..1...
1        0.01    96.84 |  ..1..1..11..
1        0.01    96.85 |  ..1..11....1
1        0.01    96.87 |  ..1..11.1...
1        0.01    96.88 |  ..1..11.1111
1        0.01    96.90 |  ..1.1..1....
1        0.01    96.91 |  ..1.1.1.....
1        0.01    96.93 |  ..1.1.1..1..
1        0.01    96.94 |  ..1.1.1.1...
1        0.01    96.96 |  ..1.1.111...
1        0.01    96.97 |  ..1.11...1..
1        0.01    96.99 |  ..1.11.1....
1        0.01    97.00 |  ..1.11.1.1..
1        0.01    97.02 |  ..1.111.....
1        0.01    97.03 |  ..1.111...1.
1        0.01    97.05 |  ..1.111..1..
1        0.01    97.06 |  ..1.1111....
1        0.01    97.08 |  ..1.1111.1..
1        0.01    97.09 |  ..1.11111...
1        0.01    97.11 |  ..1.111111..
1        0.01    97.12 |  ..1.11111111
1        0.01    97.14 |  ..11......11
1        0.01    97.15 |  ..11...1....
1        0.01    97.17 |  ..11...1.1.1
1        0.01    97.18 |  ..11...11...
1        0.01    97.20 |  ..11..11.1..
1        0.01    97.21 |  ..11.1.11..1
1        0.01    97.23 |  ..11.111....
1        0.01    97.24 |  ..11.111...1
1        0.01    97.26 |  ..11.111.1.1
1        0.01    97.27 |  ..11.111.11.
1        0.01    97.29 |  ..11.1111.1.
1        0.01    97.30 |  ..111.1....1
1        0.01    97.32 |  ..111.111...
1        0.01    97.33 |  ..1111.1....
1        0.01    97.35 |  ..1111.111..
1        0.01    97.36 |  ..1111.1111.
1        0.01    97.38 |  ..11111...1.
1        0.01    97.39 |  ..111111.1..
1        0.01    97.41 |  ..111111.11.
1        0.01    97.42 |  ..11111111..
1        0.01    97.44 |  ..11111111.1
1        0.01    97.45 |  ..111111111.
1        0.01    97.47 |  .1.........1
1        0.01    97.48 |  .1......1.1.
1        0.01    97.50 |  .1.....1...1
1        0.01    97.51 |  .1.....1..1.
1        0.01    97.53 |  .1.....1..11
1        0.01    97.54 |  .1.....11...
1        0.01    97.56 |  .1.....11.11
1        0.01    97.57 |  .1.....1111.
1        0.01    97.59 |  .1....1.11..
1        0.01    97.60 |  .1....11....
1        0.01    97.62 |  .1....11..1.
1        0.01    97.63 |  .1....1111..
1        0.01    97.65 |  .1...1...1..
1        0.01    97.66 |  .1...1.1....
1        0.01    97.68 |  .1...11...1.
1        0.01    97.69 |  .1...111111.
1        0.01    97.71 |  .1..1....1.1
1        0.01    97.72 |  .1..1..1....
1        0.01    97.74 |  .1..1.1...1.
1        0.01    97.75 |  .1..1.1.1.1.
1        0.01    97.77 |  .1..11...11.
1        0.01    97.78 |  .1..11..1...
1        0.01    97.80 |  .1..11.1.1..
1        0.01    97.81 |  .1..11.11...
1        0.01    97.83 |  .1..111.....
1        0.01    97.84 |  .1..1111....
1        0.01    97.86 |  .1..1111.11.
1        0.01    97.87 |  .1.1..1....1
1        0.01    97.89 |  .1.1.1...1..
1        0.01    97.90 |  .1.1.1.11...
```

```
1       0.01   97.92 |   .1.11......1
1       0.01   97.93 |   .1.11.1.....
1       0.01   97.95 |   .1.11.11....
1       0.01   97.96 |   .1.111....11
1       0.01   97.98 |   .1.111...1..
1       0.01   97.99 |   .1.111.1....
1       0.01   98.01 |   .1.1111.....
1       0.01   98.02 |   .1.11111....
1       0.01   98.04 |   .1.11111..11
1       0.01   98.05 |   .11.......1.
1       0.01   98.07 |   .11.....1...
1       0.01   98.08 |   .11.....11..
1       0.01   98.10 |   .11....1.1..
1       0.01   98.11 |   .11..1.111..
1       0.01   98.13 |   .11..11.....
1       0.01   98.14 |   .11..11...1.
1       0.01   98.16 |   .11..111....
1       0.01   98.17 |   .11..1111...
1       0.01   98.19 |   .11.1.1.....
1       0.01   98.20 |   .11.11....1.
1       0.01   98.22 |   .11.11...1..
1       0.01   98.23 |   .11.11.1....
1       0.01   98.25 |   .11.111.....
1       0.01   98.26 |   .11.111....1
1       0.01   98.28 |   .11.111...1.
1       0.01   98.29 |   .11.111..11.
1       0.01   98.31 |   .11.111.1...
1       0.01   98.32 |   .11.1111.11.
1       0.01   98.34 |   .11.11111.1.
1       0.01   98.35 |   .111......1.
1       0.01   98.37 |   .111...1..1.
1       0.01   98.38 |   .111..1.....
1       0.01   98.40 |   .111..11..1.
1       0.01   98.41 |   .111.11.....
1       0.01   98.43 |   .111.11....1
1       0.01   98.44 |   .111.11...1.
1       0.01   98.46 |   .111.11.11..
1       0.01   98.47 |   .111.11.11.1
1       0.01   98.49 |   .111.11.111.
1       0.01   98.50 |   .111.111....
1       0.01   98.52 |   .111.1111...
1       0.01   98.53 |   .111.111111.
1       0.01   98.55 |   .1111..111..
1       0.01   98.56 |   .1111.1.....
1       0.01   98.58 |   .1111.111...
1       0.01   98.59 |   .1111.1111..
1       0.01   98.61 |   .1111.111111
1       0.01   98.62 |   .11111.....1
1       0.01   98.64 |   .111111..1..
1       0.01   98.65 |   .111111..11.
1       0.01   98.67 |   .1111111..1.
1       0.01   98.68 |   .11111111.11
1       0.01   98.70 |   .11111111111
1       0.01   98.71 |   1.......1...
1       0.01   98.73 |   1......1..1.
1       0.01   98.74 |   1....1...1..
1       0.01   98.76 |   1....1..1...
1       0.01   98.77 |   1....1.1....
1       0.01   98.79 |   1....1.11...
1       0.01   98.80 |   1....11...11
1       0.01   98.82 |   1....111.11.
1       0.01   98.83 |   1...1..1....
1       0.01   98.85 |   1...1.11....
1       0.01   98.86 |   1...11..1...
1       0.01   98.88 |   1...11.1....
1       0.01   98.89 |   1...1111..1.
1       0.01   98.91 |   1..1.111....
1       0.01   98.92 |   1..11.......
1       0.01   98.94 |   1..1111.....
1       0.01   98.95 |   1.1....1....
1       0.01   98.97 |   1.1...1.....
1       0.01   98.98 |   1.1...11....
1       0.01   99.00 |   1.1..1.1....
1       0.01   99.01 |   1.1..11.....
1       0.01   99.03 |   1.1..111....
1       0.01   99.04 |   1.1.1..11...
```

```
          1       0.01   99.06 |  1.1.1111....
          1       0.01   99.07 |  1.11..1.....
          1       0.01   99.09 |  1.11.1......
          1       0.01   99.10 |  1.11.1.1.1..
          1       0.01   99.12 |  1.11.1.11...
          1       0.01   99.13 |  1.11.111....
          1       0.01   99.15 |  1.111..1..1.
          1       0.01   99.16 |  1.1111....1.
          1       0.01   99.18 |  1.11111.....
          1       0.01   99.19 |  11........1.
          1       0.01   99.21 |  11......1.1.
          1       0.01   99.22 |  11.....1.1..
          1       0.01   99.24 |  11....11.1..
          1       0.01   99.25 |  11...111....
          1       0.01   99.27 |  11...111.1..
          1       0.01   99.28 |  11...1111..1
          1       0.01   99.30 |  11...11111..
          1       0.01   99.31 |  11...111111.
          1       0.01   99.33 |  11..1..1....
          1       0.01   99.34 |  11..1.1.....
          1       0.01   99.36 |  11..1.11....
          1       0.01   99.37 |  11..11.1....
          1       0.01   99.39 |  11..11.111..
          1       0.01   99.40 |  11..1111....
          1       0.01   99.42 |  11..1111111.
          1       0.01   99.43 |  11.1...1111.
          1       0.01   99.45 |  11.1.1......
          1       0.01   99.46 |  11.1.1.1....
          1       0.01   99.48 |  11.1.111....
          1       0.01   99.49 |  11.11.......
          1       0.01   99.51 |  11.11...11..
          1       0.01   99.52 |  11.11..1....
          1       0.01   99.54 |  11.111.11.1.
          1       0.01   99.55 |  11.1111.1...
          1       0.01   99.57 |  11.11111....
          1       0.01   99.58 |  11.11111.11.
          1       0.01   99.60 |  11.111111...
          1       0.01   99.61 |  111...1.....
          1       0.01   99.63 |  111...11....
          1       0.01   99.64 |  111.1.1.....
          1       0.01   99.66 |  111.11......
          1       0.01   99.67 |  111.111...1.
          1       0.01   99.69 |  111.111..111
          1       0.01   99.70 |  111.1111....
          1       0.01   99.72 |  111.1111..11
          1       0.01   99.73 |  1111..11....
          1       0.01   99.75 |  1111..111...
          1       0.01   99.76 |  1111..111.1.
          1       0.01   99.78 |  1111.11.....
          1       0.01   99.79 |  1111.1111..1
          1       0.01   99.81 |  11111....1..
          1       0.01   99.82 |  11111.111...
          1       0.01   99.84 |  11111.111..1
          1       0.01   99.85 |  11111.11111.
          1       0.01   99.87 |  111111...1..
          1       0.01   99.88 |  111111.1.1.1
          1       0.01   99.90 |  1111111.....
          1       0.01   99.91 |  1111111..11.
          1       0.01   99.93 |  1111111.1...
          1       0.01   99.94 |  1111111.11.1
          1       0.01   99.96 |  11111111....
          1       0.01   99.97 |  11111111.111
          1       0.01   99.99 |  11111111111.
          1       0.01  100.00 |  111111111111
-------------------------------+-------------
       6669     100.00         |  XXXXXXXXXXXX
```