

UNIVERSIDADE ESTADUAL PAULISTA “JULIO DE MESQUITA FILHO”
FACULDADE DE ENGENHARIA DE BAURU
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

PAULO SÉRGIO CORRÊA RIBEIRO

**MODELOS DE REGRESSÃO APLICADO NO ABSENTEÍSMO
EM UMA EMPRESA DO SETOR SUCROALCOOLEIRO**

BAURU/SP

2013

Paulo Sérgio Corrêa Ribeiro

**MODELOS DE REGRESSÃO APLICADO NO ABSENTEÍSMO
EM UMA EMPRESA DO SETOR SUCROALCOOLEIRO**

Dissertação apresentada ao curso de Pós-Graduação em Engenharia de Produção, da Faculdade de Engenharia da UNESP, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de Bauru, como requisito parcial para a obtenção do título de mestre.

Área de concentração: Gestão de Operações e Sistemas.

Orientadora:

Prof.^a Dr.^a Gladys D. Cacsire Barriga

BAURU/SP

2013

DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO
UNESP - Campus de Bauru

Ribeiro, Paulo Sérgio Corrêa.

Modelos de regressão aplicado no absenteísmo em uma empresa do setor sucroalcooleiro / Paulo Sérgio Corrêa Ribeiro, 2013.

77 f.

Orientadora: Gladys D. Cacsire Barriga

Dissertação (Mestrado) - Universidade Estadual Paulista. Faculdade de Engenharia, Bauru, 2013

1. Absenteísmo. 2. Distribuição de séries de potência generalizada. 3. Máxima verossimilhança. 4. Modelos inflacionado de zeros. 5. Modelos de regressão. I. Universidade Estadual Paulista. Faculdade de Engenharia. II. Título.

ATA DA DEFESA PÚBLICA DA DISSERTAÇÃO DE Mestrado DE PAULO SERGIO CORREA RIBEIRO, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO, DO(A) FACULDADE DE ENGENHARIA DE BAURU.

Aos 15 dias do mês de agosto do ano de 2013, às 09:00 horas, no(a) ANFITEATRO DA SEÇÃO TÉCNICA DE PÓS-GRADUAÇÃO DA FACULDADE DE ENGENHARIA DE BAURU, reuniu-se a Comissão Examinadora da Defesa Pública, composta pelos seguintes membros: Profa. Dra. GLADYS DOROTEA CACSIRE BARRIGA do(a) Departamento de Engenharia de Produção / Faculdade de Engenharia de Bauru - UNESP, Profa. Dra. LINDA LEE HO do(a) Departamento de Engenharia de Produção / Escola Politécnica - USP, Prof. Dr. DANIEL JUGEND do(a) Departamento de Engenharia de Produção / Faculdade de Engenharia de Bauru - UNESP, sob a presidência do primeiro, a fim de proceder a arguição pública da DISSERTAÇÃO DE Mestrado de PAULO SERGIO CORREA RIBEIRO, intitulado "MODELOS DE REGRESSÃO PARA DADOS DE CONTAGEM INFLACIONADO DE ZEROS: ESTUDO SOBRE O ABSENTEÍSMO DOS TRABALHADORES EM UMA EMPRESA DO SETOR SUCROALCOOLEIRO". Após a exposição, o discente foi arguido oralmente pelos membros da Comissão Examinadora, tendo recebido o conceito final: aprovado. Nada mais havendo, foi lavrada a presente ata, que, após lida e aprovada, foi assinada pelos membros da Comissão Examinadora.


Profa. Dra. GLADYS DOROTEA CACSIRE BARRIGA


Profa. Dra. LINDA LEE HO


Prof. Dr. DANIEL JUGEND

*Dedico este trabalho a meus pais **APARECIDO ANTONIO RIBEIRO** (in memorian) e **CÉLIA CORRÊA RIBEIRO**, pelos valores e ensinamentos transmitidos.*

*Aos meus avós e padrinhos **JOSÉ GOMES CORRÊA** e **ALTELINA ALVES CORRÊA**, pelo incentivo aos estudos, desde as primeiras lições.*

*E, em especial, a minha esposa **ELOÍSE** e a meus filhos (**FELIPE** e **ALICE**), pela compreensão durante os momentos de ausência e pelo apoio nesta jornada.*

AGRADECIMENTOS

*Agradeço, primeiramente, a **DEUS** pela realização deste sonho.*

*De maneira especial, agradeço a oportunidade e a confiança da **Prof^ª. Dr^ª. Gladys D. Cacsire Barriga**, minha querida orientadora. Pela condução exemplar na busca de nossos objetivos. Espero, dedicando-lhe esta conquista, retribuir um pouco do muito que fez por mim nesta caminhada.*

*Agradeço também a **Prof^ª. Dr^ª. Linda Lee Ho** e ao **Prof. Dr. Daniel Jugend** pelas contribuições durante o exame de qualificação.*

*Aos diretores e funcionários da **Compusoftware**, por acreditarem em meu trabalho e por permitir que eu flexibilizasse minha jornada de trabalho para a realização deste sonho.*

*Aos meus **familiares** e **amigos**, pela torcida e pelo carinho.*

“Nenhuma conquista é importante quando não se tem ninguém para dividi-las”.
(Roger Martin)

Essa conquista é nossa!

“Se você quer ser bem-sucedido, precisa ter dedicação total, buscar seu último limite e dar o melhor de si mesmo.

No que diz respeito ao empenho, ao compromisso, ao esforço e à dedicação, não existe meio termo: ou você faz uma coisa bem feita, ou não faz.

Trabalhei muito para chegar ao sucesso. Mas não conseguiria nada se Deus não me ajudasse.”

Ayrton Senna

RESUMO

A crescente produção e utilização de fontes de energia renováveis coloca o Brasil em posição de destaque no cenário global. Este reconhecimento decorre, em grande parte, graças às conquistas alcançadas na substituição do uso do petróleo pelo etanol extraído da cana-de-açúcar e da cogeração de energia a partir do bagaço da cana-de-açúcar. Um grande desafio enfrentado pelas empresas deste setor é o índice de absenteísmo dos trabalhadores da área agrícola, que pode chegar a 11%. O que torna este índice um indicador importante, pois as empresas precisam, cada vez mais, contar com seu quadro de pessoal. Devido aos prejuízos causados pelas ausências dos trabalhadores em seus postos de trabalho, tornam-se oportunos estudos que busquem identificar os motivos que os levam a faltar. Assim, as empresas podem tomar medidas que visem a eliminar ou reduzir estas ocorrências. O objetivo do presente trabalho é utilizar modelos de regressão para a modelagem do número de faltas dos funcionários de uma empresa do setor sucroalcooleiro. Particularmente, é considerado o modelo de regressão inflacionado de zeros baseado na família de distribuição de séries de potência generalizada. Inferências baseadas na teoria de máxima verossimilhança são desenvolvidas para o modelo proposto e o desempenho desta abordagem é ilustrado com um conjunto de dados reais, no qual são estudados os fatores associados à ocorrência de absenteísmo e a não ocorrência de absenteísmo dos trabalhadores da empresa. Os resultados obtidos apontam que o tempo de serviço e o cargo influenciam de forma significativa na proporção de não absenteísmo, essa proporção aumenta conforme o tempo de serviço, e é maior para os trabalhadores da cultura da cana-de-açúcar, sendo menos frequentes os trabalhadores que tem as demais funções dentro da empresa, já o fator salário tem um efeito significativo na média do absenteísmo, observa-se que funcionários com salários altos são mais frequentes ao trabalho. Recomenda-se, então, a adoção de um sistema de recompensas para valorização dos funcionários com melhor desempenho e também uma política de recrutamento interno para promover e motivar os trabalhadores de forma a retê-los na empresa.

Palavras-chave: Absenteísmo, Distribuição de séries de potência generalizada, Máxima verossimilhança, Modelos de regressão, Modelos inflacionado de zeros.

ABSTRACT

The growing production and use of renewable energy sources puts Brazil in a prominent place worldwide. This is largely due to advances in replacing fossil fuels for ethanol extracted from sugarcane and bagasse cogeneration energy. A great challenge faced by companies in the sector is the farm worker absenteeism, that can reach 11%. Absenteeism becomes an important parameter because companies need to rely more and more on their workforce, which makes it necessary to control attendance and productivity of workers. The losses caused by absence from work bring out the need for studies that try to identify the reasons for that. The company, thus, can take actions to eliminate or reduce such occurrences. The aim of the present work is the use of regression models in order to model the frequency of absence of workers of a company of the sugar and alcohol sector. This study focused on the zero inflated regression model based on family of generalized power series distribution . Inferences based on the maximum likelihood theory are developed for the proposed model and the performance of this approach is illustrated by a real data set in which we study factors associated with the occurrence and nonoccurrence of absenteeism in the company. The resulting data show that length of service and position, have a significantly in the proportion influence of non absenteeism. This influence increases conform the length of service and is greater for sugarcane culture workers, it is weaker among workers of other areas within the company. Salary has a significant effect on absenteeism, it was observed that workers with higher salaries are less absent from work. It is recommended, thus, the adoption of a reward system to reward employees with better performances and also an internal recruitment policy to promote and motivate workers , in order to retain them in the company.

Keywords: Absenteeism, Maximum Likelihood, Generalized Power Series Distribution, Regression Models, Zero-Inflated Models.

LISTA DE FIGURAS

Figura 1 - Produção de automóveis no Brasil.....	21
Figura 2 - Mapa de produção do setor sucroenergético	23
Figura 3 - Etapas de produção até o consumidor final.....	25
Figura 4 - Fluxograma do processo produtivo.....	26
Figura 5 - Pirâmide de Maslow – Hierarquia das Necessidades.....	29
Figura 6 – Sequência das necessidades de Maslow	30
Figura 7 – Correlação absenteísmo.....	32
Figura 8 - Número de faltas.....	60
Figura 9 - Proporção de não absenteísmo de funcionários da empresa estratificada por cargo e tempo de serviço.	66

LISTA DE TABELAS

Tabela 1 - Previsão cana-de-açúcar para os próximos anos	24
Tabela 2 - Descrição das variáveis.....	59
Tabela 3 – Caracterização da população de trabalhadores	61
Tabela 4 – Caracterização sócio-demográfica das faltas dos trabalhadores.....	62
Tabela 5 – Caracterização ocupacional das faltas dos trabalhadores	63
Tabela 6 - Ajuste da distribuição de frequência dos conjuntos de dados reais.....	64
Tabela 7 - Algumas estatísticas a partir dos modelos ajustados	64
Tabela 8 – Estimativa de Máxima Verossimilhança (EMV).....	65

LISTA DE SIGLAS

AIC	Akaike information Criterion
BIC	Bayesian information Criterion
BN	Binomial Negativa
CEAM	Comissão de estudos sobre o Álcool-Motor
ECO-92	Conferência das Nações Unidas sobre o Meio Ambiente e o Desenvolvimento
CIPA	Comissão interna de prevenção de acidentes
CLT	Consolidação das leis do Trabalho
EMV	Estimativa de Máxima Verossimilhança
ERP	Entreprise Resouce Planing (Sistema de gestão Integrado)
FLEX-FUEL	Sistema bicomcombustível, capaz de funcionar tanto com gasolina como com etanol
IAA	Instituto de Açúcar e Álcool
FMP	Função Massa de Probabilidade
PROALCOOL	Programa Nacional do Álcool
PSG	Séries de Potência Generalizada
RIO+20	Conferência das Nações Unidas sobre Desenvolvimento Sustentável, para a renovação do compromisso com o desenvolvimento sustentável.
SBC	Schuwartz Bayseian Criterion
SQL	Structured Query Language
ZI	Modelos de Regressão Inflacionado de Zeros
ZIB	Modelo de Regressão de Binomial Inflacionado de Zeros
ZINB	Modelo de Regressão de Binomial Negativa Inflacionado de Zeros
ZIP	Modelo de Regressão de Poisson Inflacionado de Zeros
ZIPSG	Distribuição de Séries de Potência Generalizada Inflacionada de Zeros

SUMÁRIO

1. INTRODUÇÃO.....	12
1.1. FORMULAÇÃO DO PROBLEMA	14
1.2. OBJETIVOS.....	14
1.3. JUSTIFICATIVA.....	15
1.4. DELIMITAÇÃO DA PESQUISA.....	17
1.5. ESTRUTURA DO TRABALHO	18
2. REVISÃO DA LITERATURA	19
2.1. IMPORTÂNCIA DO SETOR SUCROALCOOLEIRO NO BRASIL.....	19
2.2. ABSENTEÍSMO	28
2.2.1. <i>O uso de modelos estatísticos para análise do absenteísmo</i>	35
2.3. MODELOS PARA DADOS DE CONTAGEM.....	38
2.3.1. <i>Modelo de Poisson</i>	39
2.3.2. <i>Modelo Binomial</i>	40
2.3.3. <i>Modelo Binomial Negativa</i>	41
2.4. MODELOS PARA DADOS DE CONTAGEM INFLACIONADO DE ZEROS	41
2.4.1. <i>Modelo Poisson Inflacionado de Zeros (ZIP)</i>	43
2.4.2. <i>Modelo Binomial Inflacionado de Zeros (ZIB)</i>	44
2.4.3. <i>Modelo Binomial Negativa Inflacionado de Zeros (ZINB)</i>	44
2.5. DISTRIBUIÇÃO DE SÉRIES DE POTÊNCIA GENERALIZADA INFLACIONADA DE ZEROS (ZIPSG)	45
2.6. MÉTODO DE ESTIMAÇÃO DOS PARÂMETROS DO MODELO.....	47
2.6.1. <i>Método de Máxima Verossimilhança</i>	47
2.7. CRITÉRIOS PARA COMPARAÇÃO DOS MODELOS	49
2.7.1. <i>AIC – Akaike Information Criterion</i>	50
2.7.2. <i>SBC – Schwarz Bayesian Criterion</i>	52
3. METODOLOGIA DA PESQUISA.....	53
3.1. ORIGEM E PLANEJAMENTO DA PESQUISA	55
3.2. MODELAGEM ESTATÍSTICA	56
3.2.1. <i>Inferência</i>	57
4. APLICAÇÃO EM UMA EMPRESA DO SETOR SUCROALCOOLEIRO.....	59
5. CONSIDERAÇÕES FINAIS	68
5.1. SUGESTÕES PARA TRABALHOS FUTUROS.....	70
REFERÊNCIAS BIBLIOGRÁFICAS	71

1. INTRODUÇÃO

É de grande preocupação mundial a busca por fontes de energia renovável que substitua o uso dos combustíveis fósseis, constantemente o assunto está presente na pauta das reuniões dos chefes de estado. Dois motivos principais são apontados: a questão ambiental devido ao aquecimento global e o risco de escassez dos combustíveis fósseis.

Em 1992, vários países aderiram ao Protocolo de Kyoto durante a ECO-92. As nações que assinaram o tratado assumiram compromisso de redução da emissão de gases que agravam o efeito estufa. Após 20 anos, recentemente, os líderes mundiais voltaram a reforçar seu comprometimento com o tratado durante a RIO+20 (ONU, 2012).

Os estudos para substituição do petróleo pelo etanol extraído a partir da cana-de-açúcar foram iniciados há quase um século no Brasil. E a partir de 2003, o setor sucroalcooleiro ganhou novo impulso com o aumento da demanda de etanol provocada pela fabricação de carros *flex-fuel*. O aumento progressivo da demanda fez o setor sucroalcooleiro atrair investimentos nacionais e internacionais (FICARELLI; RIBEIRO, 2010).

O Brasil é hoje o maior e mais eficiente produtor mundial de açúcar e etanol derivados da cana-de-açúcar, e o maior exportador desses produtos. Isso foi possível devido aos investimentos em pesquisas para o desenvolvimento de variedades mais produtivas; à processos industriais mais eficientes; à imensa disponibilidade de terras aráveis; e ao fato favorável do país estar localizado quase todo em regiões tropicais e chuvosas, o que oferecem excelentes condições à cultura da cana-de-açúcar (SHIKIDA; AZEVEDO; VIAN, 2011).

A soma de todas essas características (territoriais, climáticas e de investimento) gera vantagem comparativa que permite ao país ocupar posição de destaque global na produção da biomassa (energia obtida a partir de recursos naturais, exceto combustíveis fósseis). Relatório do Banco Mundial destacou o país como referência nas ações para alcançar o objetivo de promover energia sustentável para todos proposto pela ONU durante a RIO+20. O país aparece como a terceira nação que mais cresce no consumo de energia renovável ao lado de China e Estados Unidos. A energia renovável ainda

responde por, apenas, 18% da matriz energética mundial. A meta da ONU é que este percentual chegue a 36% até 2030. O Brasil é elogiado no documento por aumentar a geração por meio de fontes alternativas, como a biomassa de cana-de-açúcar e o lixo (YUMKELLA, 2013). No Brasil 45% de toda energia consumida, é produzida por fontes renováveis (BRASIL, 2012b).

Para manter-se nesta posição, as empresas deste setor devem continuar na busca por inovação por meio do aprimoramento tecnológico, do desenvolvimento do capital humano, da melhoria dos processos e melhoria do conceito de gestão.

O presente estudo lança atenção sobre a área agrícola dessas empresas, ou seja, no manejo da cana-de-açúcar, pois esta área representa 70% dos custos totais do negócio segundo dados do Pecege (2011). Deste percentual, 30% são destinados ao pagamento de mão de obra, isto é, 20% dos custos totais das empresas estão direcionados ao pagamento de trabalhadores envolvidos diretamente com a cultura da cana-de-açúcar. Scopinho (2000) destaca que o índice de absenteísmo encontrado na área agrícola das empresas deste setor é alto, em média, 11%. Os prejuízos advindos deste comportamento são de grande impacto às organizações, pode-se afirmar que 5% (cinco por cento) do dinheiro investido em recursos humanos são perdidos com absenteísmo, segundo Pinto Jr. (2005). Portanto, estima-se que, aproximadamente, 1% do orçamento total das empresas se perde com as ausências dos trabalhadores em seus postos de trabalho.

O uso de ferramentas estatísticas no campo da engenharia de produção ampliou nas últimas décadas devido à crescente necessidade de gestão de recursos escassos e também em função dos avanços tecnológicos, que possibilitou às empresas terem acesso a um grande volume de informações. Assim, é possível, por meio da compreensão, simulação de fenômenos e correlações; gerar conhecimentos que auxiliem os gestores em suas tomadas de decisões.

O presente trabalho objetivou responder, por meio de modelos estatísticos, quais são os fatores associados à ocorrência de absenteísmo e a não ocorrência de absenteísmo de trabalhadores de uma empresa do setor sucroalcooleiro.

A assiduidade ao trabalho é o que normalmente se espera dos funcionários, portanto a frequência do valor zero é predominante nesse tipo de conjunto de dados,

assim, torna-se necessária a utilização de modelos de regressão que se ajustem adequadamente aos dados. É possível encontrar na literatura estudos com aplicação de modelos de regressão inflacionado de zeros para modelar o absenteísmo nas áreas da saúde, educação, serviços, entre outras. Percebe-se, porém, uma lacuna de estudos sobre o absenteísmo na área agrícola com a aplicação destes modelos.

Este estudo foi realizado em uma usina localizada no interior do estado de São Paulo. A empresa foi escolhida por possuir uma administração que entende os benefícios à organização que pesquisas como esta podem trazer, outro fator que prevaleceu para a escolha da empresa foi a qualidade das informações disponível em seu sistema de gestão integrado (ERP).

1.1. Formulação do problema

Devido à relevância dos estudos sobre o absenteísmo dos trabalhadores e, em particular, o absenteísmo na área agrícola, o presente trabalho tem como motivação de pesquisa responder as seguintes questões:

- Quais as características que explicam o grau de absenteísmo de trabalhadores da área agrícola de uma empresa do setor sucroalcooleiro?
- Qual modelo de regressão que melhor se ajusta ao conjunto de dados apresentado?

1.2. Objetivos

O objetivo geral deste trabalho é identificar os fatores que influenciam a ocorrência de absenteísmo e a não ocorrência de absenteísmo utilizando modelos de regressão da família de distribuição de séries de potência generalizada inflacionada de zeros.

Além disso, têm-se como objetivos específicos:

- Aplicação dos modelos em um conjunto de dados reais relativo ao absenteísmo de trabalhadores da área agrícola de uma empresa do setor sucroalcooleiro;
- Desenvolvimento de procedimentos inferenciais baseado na teoria de máxima verossimilhança para modelos de regressão da família de distribuição de séries de potência generalizada inflacionada de zeros (CORDEIRO; ANDRADE; DE CASTRO, 2009);
- Comparação dos modelos para identificar qual melhor se ajusta ao conjunto de dados deste estudo.

1.3. Justificativa

Pode-se justificar a importância desta dissertação uma vez que as ausências dos trabalhadores em seus postos de trabalho geram grandes prejuízos às empresas. Além disso, o índice de absenteísmo encontrado na área agrícola de empresas produtoras de cana-de-açúcar (11%) é considerado alto em relação às demais áreas desse tipo de empresa (2,94%), segundo Scopinho (2000). Em empresas de outros setores, o índice médio de absenteísmo encontrado é de 2%, segundo Bachmann & Associados, ABRH-PR e ISAE/FGV (2011).

Para Pinto Jr. (2005), o absenteísmo é uma questão importante ainda não resolvida na literatura. Os estudos são necessários entender os fatores que influenciam os eventos de ausência.

É de grande relevância a administração e a redução dos custos com absenteísmo para otimização dos custos do produto final. Uma vez que, para manter-se neste mercado competitivo, faz-se necessário que as empresas reduzam seus custos, oferecendo produtos com maior qualidade e menor preço.

Sturman (1996) destaca a necessidade de utilizar métodos adequados para a análise do absenteísmo. Segundo o autor, muitas pesquisas são criticadas por utilizar análises inadequadas. Isso ocorre devido à variabilidade dos dados que, normalmente, violam as suposições dos modelos estatísticos. Em seu estudo, o autor realiza a

comparação de vários modelos para explorar a sensibilidade dos modelos com relação ao tamanho da amostra, características da distribuição da variável dependente e características das distribuições das variáveis independentes.

Apoiado na literatura Sturman, (1996) notou que o maior problema decorre do fato de que as taxas de absenteísmo não seguem uma distribuição normal.

Quando se observa o número de faltas de funcionários, tende-se a obter uma alta frequência de valores “zero”, pois, em situações normais, os funcionários não devem faltar ao trabalho. Desta forma, pode-se classificar o absenteísmo como um conjunto de dados de contagem que apresenta uma frequência de valores de “zero” acima da frequência esperada pelos modelos estatísticos.

No Brasil, os estudos sobre o absenteísmo de trabalhadores podem ser encontrados em diferentes contextos, por exemplo, identificar as ausências por motivo de saúde (TEODOZIO, 2010), absenteísmo em Call Center (DIAS, 2005; PINTO JR., 2005), absenteísmo na área da Enfermagem, impacto da ausência de professores na sala da aula (TAVARES; CAMELO; KASMIRSKI, 2009), e absenteísmo em uma empresa de reflorestamento Simões (2010).

O alto nível de competitividade entre as empresas e o crescente destaque do fator humano como elemento de diferenciação demandam a necessidade do equilíbrio entre produtividade e capacidade de produção estimada. Torna-se necessária a prevenção de qualquer tipo de ocorrência que reduza as possibilidades de vantagem competitiva. Neste sentido, a redução do nível de absenteísmo deve ser considerada com um importante indicador. Para tornar isso possível, a empresa deve trabalhar preventivamente na eliminação das causas do absenteísmo dos trabalhadores.

Pinto Jr. (2005) apresenta um crescente interesse de pesquisadores, durante os últimos anos, em relação aos comportamentos contraproducentes dos trabalhadores dentro das organizações. E aponta que as pesquisas estão direcionadas em estudar os antecedentes e consequências do comportamento contraproducente.

O autor ainda diz que é possível, mediante cálculo simples, afirmar que 5% (cinco por cento) do dinheiro investido em recursos humanos são perdidos com absenteísmo, o que pode representar milhares de dólares para uma organização de

faturamento médio e com um número de cerca de mil funcionários. E também motiva os pesquisadores ao destacar que, se os estudos conseguirem identificar melhorias que reduzam em 20% (vinte por cento) o absenteísmo, isso pode trazer um grande resultado para a economia mundial ou de um segmento específico. Esses recursos podem ser investidos na melhoria da qualidade de vida e do bem-estar do trabalhador e na geração de mais empregos.

Cardoso, Cardoso e Santos (2013) apresentam estudo que revela o quanto os altos índices de rotatividade e absenteísmo podem aumentar os custos de uma organização e o quanto podem ser prejudiciais aos negócios. A análise dos fatores rotatividade e absenteísmo pode ser uma fonte significativa para entendimento do clima organizacional, dos resultados sociais, do desempenho produtivo e dos custos de uma empresa.

O estudo com apoio de métodos estatísticos se justifica uma vez que as ausências do passado são um dos melhores indicadores de faltas futuras (STURMAN, 1999).

E apesar dos numerosos estudos sobre o tema (DIONNE; DOSTIE, 2007; DA FONSECA, 2009; SIMÕES, 2010; DOS SANTOS *et al.*, 2011, entre outros), não foi encontrado durante esta pesquisa nenhuma aplicação dos modelos de regressão inflacionado de zeros para modelar o absenteísmo dos trabalhadores da área agrícola de empresas do setor sucroalcooleiro.

1.4. Delimitação da pesquisa

Esta pesquisa trabalhou com um conjunto de dados de contagem discretos. Desta forma, procurou-se delimitar o escopo de análise em relação ao objeto a ser estudado.

Para o estudo de modelos de regressão para dados de contagem discretos, o presente trabalho ficou restrito a avaliar os fatores que influenciam na ocorrência de faltas ao trabalho. Como universo de pesquisa, foram considerados apenas trabalhadores da área agrícola de uma empresa do setor sucroalcooleiro, localizada no interior do estado de São Paulo.

Foram utilizados modelos de regressão da família de distribuição de séries de potência generalizada inflacionada de zeros, são eles:

- Modelo Poisson inflacionado de zeros (ZIP);
- Modelo Binomial inflacionado de zeros (ZIB);
- Modelo Binomial Negativo inflacionado de zeros (ZINB).

1.5. Estrutura do trabalho

A estrutura do trabalho foi elaborada com um primeiro capítulo introdutório do tema de pesquisa, a formulação do problema, a justificativa, os respectivos objetivos e delimitação da pesquisa.

No segundo capítulo, é apresentada a revisão de literatura, na qual é descrita, brevemente, a importância do setor sucroalcooleiro para o Brasil; aborda-se o tema absenteísmo, seus tipos, suas causas e algumas ações preventivas. Neste capítulo, também são apresentados os principais modelos de regressão para dados de contagem inflacionado de zeros.

O capítulo três apresenta a metodologia de pesquisa utilizada neste trabalho desenvolvida a partir da revisão de literatura.

O quarto capítulo propõe a aplicação da abordagem apresentada no capítulo anterior no conjunto de dados selecionados. Neste capítulo, também são realizadas as comparações dos modelos e análise dos resultados.

No capítulo 5, são feitas as considerações finais do trabalho.

2. REVISÃO DA LITERATURA

O referencial teórico, revisão da literatura ou marco teórico tem por objetivo situar o problema no tempo e no espaço. Esse tópico apresenta uma discussão teórica dos problemas, na perspectiva de fundamentá-los nas teorias existentes.

Neste capítulo, é retratada a história e importância do setor sucroalcooleiro no Brasil; o absenteísmo e suas causas; e são apresentados os principais modelos de regressão para dados de contagem inflacionado de zeros.

2.1. Importância do setor sucroalcooleiro no Brasil

A cana-de-açúcar chegou ao Brasil junto com os colonizadores portugueses. A partir do ano de 1532, ganhou destaque como principal cultura do país, principalmente nas regiões de Pernambuco e Bahia. O clima tropical (quente e úmido) e a mão de obra escrava vinda da África ajudaram a espalhar a cultura para outras regiões do país. Após várias dificuldades, depois de 50 anos, o Brasil passou a monopolizar a produção mundial de açúcar, com o custo mais baixo do mundo (UNICA, 2012a).

Segundo Dieese (2006), estudar o setor sucroalcooleiro é o mesmo que estudar a origem da sociedade brasileira tal como ela se configura atualmente, desde a colonização até os movimentos pioneiros de enfrentamento de crises energéticas de dimensão internacional. Trata-se de um marco na história americana, pois foi a primeira grande cultura fora da Europa.

No século XVIII, devido a conflitos externos, portugueses e holandeses se desentenderam, o que desencadeou na colônia brasileira uma crise no mercado da cana-de-açúcar, pois os holandeses diminuíram sua dependência do açúcar brasileiro e iniciaram a produção açucareira no Caribe e, mais tarde, os ingleses e franceses fizeram o mesmo em suas colônias, o que pôs fim ao monopólio do açúcar brasileiro e desencadeou o fim do ciclo da cana-de-açúcar no Brasil colônia (UNICA, 2012a).

O presidente Epitácio Pessoa (1919-1922), preocupado com dependência do petróleo, reclamava da “colossal importação de gasolina no Brasil” e defendia o “uso do álcool em seu lugar”, e previa o “amparo que a solução prestaria à indústria canavieira” (MARCOLIN, 2008 *apud* GATTI JR., 2010).

Em 1923, deu-se início à pesquisa brasileira com veículos movidos a combustíveis alternativos, mais precisamente, o álcool-motor, como era conhecida a mistura de álcool hidratado com derivados do petróleo. Para uma maior estabilidade da mistura, o álcool hidratado era utilizado em maior proporção (DUNHAM, 2007 *apud* GATTI JR., 2010).

Com a crise de 1929, cresceu a pressão para a criação de um mercado para o álcool combustível, uma vez que o mercado de açúcar, interno e externo, se retraiu fortemente com prejuízo para toda a indústria (DUNHAM, 2007 *apud* GATTI JR., 2010). O então presidente, Getúlio Vargas, em defesa da produção açucareira, decretou, no início de 1931, a primeira política governamental efetiva para a utilização do álcool combustível: a adição de 5% de álcool anidro na gasolina para controlar a política de preços e exportação do açúcar. Em agosto do mesmo ano, o governo criou a Comissão de Estudos sobre o Álcool-Motor (CEAM). Vargas também criaria o Instituto do Açúcar e do Álcool (IAA), com o objetivo de regular esse mercado.

Com a crise do petróleo dos anos 70, as economias mundiais foram forçadas a buscar alternativas em relação ao petróleo. No Brasil, foi lançado o Programa Nacional do Álcool (PROALCOOL) em novembro de 1975. Com isso, o Governo Federal passou a controlar toda a produção e comercialização do açúcar e do etanol.

O objetivo desta política era a substituição da gasolina pelo álcool e a redução das importações de petróleo, principal produto que contribuía de forma negativa no saldo da balança comercial.

Em 1979, iniciou-se a produção de veículos a álcool, o que gerou uma demanda pelo álcool hidratado como combustível. A princípio, com o objetivo de garantir a demanda, o preço do etanol foi estabelecido em 55% abaixo do preço da gasolina. A participação nas vendas de veículos leves movidos a etanol chegou a 96% do mercado em 1985 (GUARDABASSI, 2006). E o grau de dependência do Brasil com relação ao petróleo baixou de 72,9% em 1982, para 47,2% em 1987.

Porém, no final desta década, devido aos preços mais atraentes do açúcar no mercado internacional, a produção de etanol foi reduzida drasticamente, e o país passou por uma séria crise de abastecimento. A partir deste momento, os produtores ficaram

desacreditados frente aos consumidores, e a venda destes veículos foi praticamente extinta, chegando ao valor mínimo de 0,07% em 1997 (GUARDABASSI, 2006).

Várias medidas foram tomadas pelo governo para assegurar a venda do excedente de álcool produzido pelas usinas.

Mas a revolução no mercado do etanol ocorreu em 2003 quando a indústria automobilística deu início à produção dos veículos chamados bicombustíveis, capazes de funcionar tanto com gasolina como com etanol, ou qualquer mistura destes combustíveis (GUARDABASSI, 2006). Esse sistema recebeu o nome de *flex-fuel*.

E, novamente devido à disparada dos preços do petróleo no mercado internacional, o etanol voltou a atrair consumidores frente à gasolina. O consumidor atento às incertezas de ambos os setores (petrolífero e sucroalcooleiro) passou a optar por veículos bicombustíveis. Em 2005, as vendas deste tipo de veículos ultrapassaram as vendas de veículos a gasolina, o que representou 52% dos veículos novos vendidos, conforme gráfico apresentado pela Figura 1.

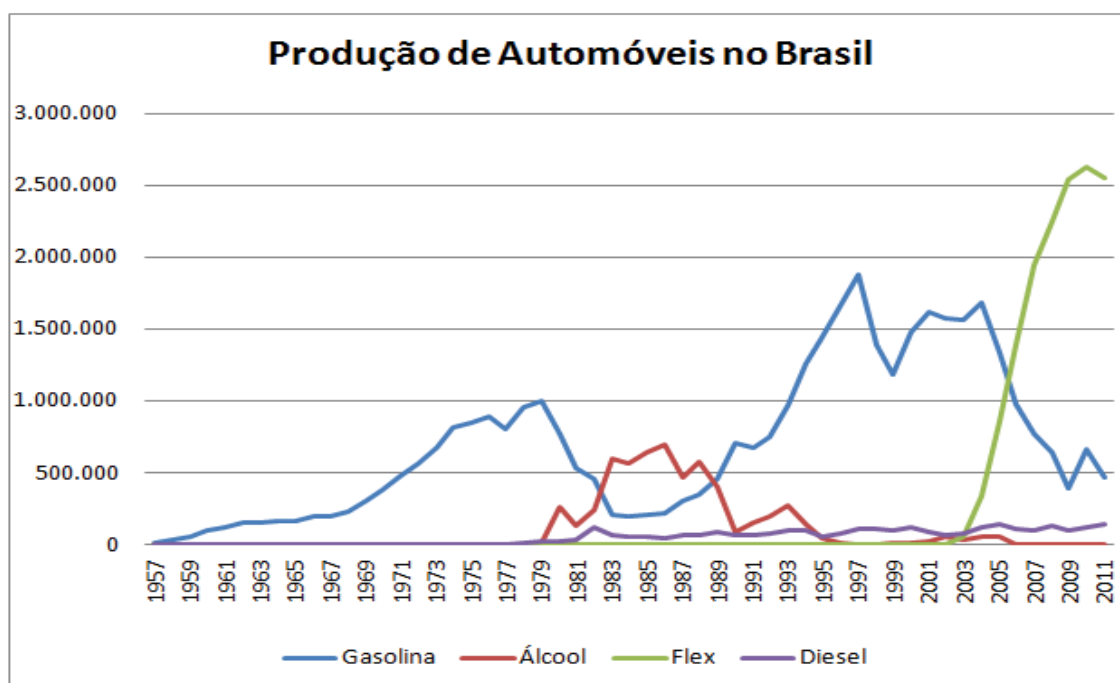


Figura 1 - Produção de automóveis no Brasil

Fonte - Elaborado pelo autor a partir de dados publicados ANFAVEA (2012).

A agroindústria sucroalcooleira absorve intensivamente mão de obra – 1 barril de petróleo emprega 0,06 pessoas – 1 barril de álcool emprega 3,4 pessoas. O setor ainda possui excelente relação de investimento por emprego, segundo levantamento feito pelo BNDES (1998 *apud* AZEVEDO, 2002), para cada R\$ 1 milhão de investimento no setor são gerados 182 empregos, muito maior que os setores da agricultura, bens de consumo, turismo, telecomunicações, indústria em geral e automobilística, bens de capital, pecuária, metalurgia e química.

No mundo todo, para cada 100 quilos de sacarose extraídos da cana, apenas 83 quilos são remunerados pelo preço vigente do açúcar em seus mercados, os outros 17 quilos de sacarose (mel residual) obtêm uma remuneração muito menor. No Brasil, com a fabricação em larga escala do etanol, que usa esse mel residual pouco remunerado nos demais países, há uma equivalência de preços, remunera-se melhor o produtor com um valor próximo ao preço vigente de açúcar.

Outro avanço importante se deu a partir da desregulamentação do setor elétrico brasileiro no final da década de 1990. Esta ação trouxe a possibilidade para que produtores independentes de energia pudessem vender seus excedentes na malha interligada.

Essa ação do governo proporcionou a adição de mais dois produtos na cadeia produtiva: a energia cogerada com a queima do bagaço da cana-de-açúcar e os créditos de carbono remunerados. O impacto tem sido o de elevar a remuneração obtida com a mesma cana-de-açúcar para 113,83%, ou mais, sobre o valor base de referência da sacarose remunerada ao preço vigente do mercado de açúcar.

Para elevar ainda mais a remuneração da cana-de-açúcar, busca-se a incorporação de renda adicional pelo aproveitamento da energia contida na palha da cana-de-açúcar. Já existe tecnologia desenvolvida para a coleta de parte dos resíduos da cana para serem levados até as usinas para seu posterior aproveitamento. Com 50% da palha recolhida, haveria energia suficiente para a geração de um excedente adicional de 23 MW para cada um milhão de toneladas de cana, o que permite a geração um excedente marginal de 126.500 MWh, e uma renda adicional equivalente a 21,5% da receita base obtida com a sacarose remunerada ao preço do açúcar. Isso elevaria a remuneração do setor para 135,33% da renda base obtida pela sacarose ao preço vigente

do açúcar, o que deixaria o Brasil ainda mais confortável na liderança do setor sucroalcooleiro mundial.

De acordo com a UNICA (2012b), no ano de 2011, o setor sucroalcooleiro representou 1,5% do Produto Interno Bruto nacional, movimentou R\$ 48 bilhões, moeu 561 milhões de toneladas de cana, produziu 36 milhões de toneladas de açúcar e 23 bilhões de litros de álcool.

A capacidade para expansão da produção é muito grande, uma vez que somente 2% (7,5 milhões de hectares) das terras aráveis são utilizadas para a produção de cana-de-açúcar. A maior parte da produção encontra-se no estado de São Paulo, conforme Figura 2.

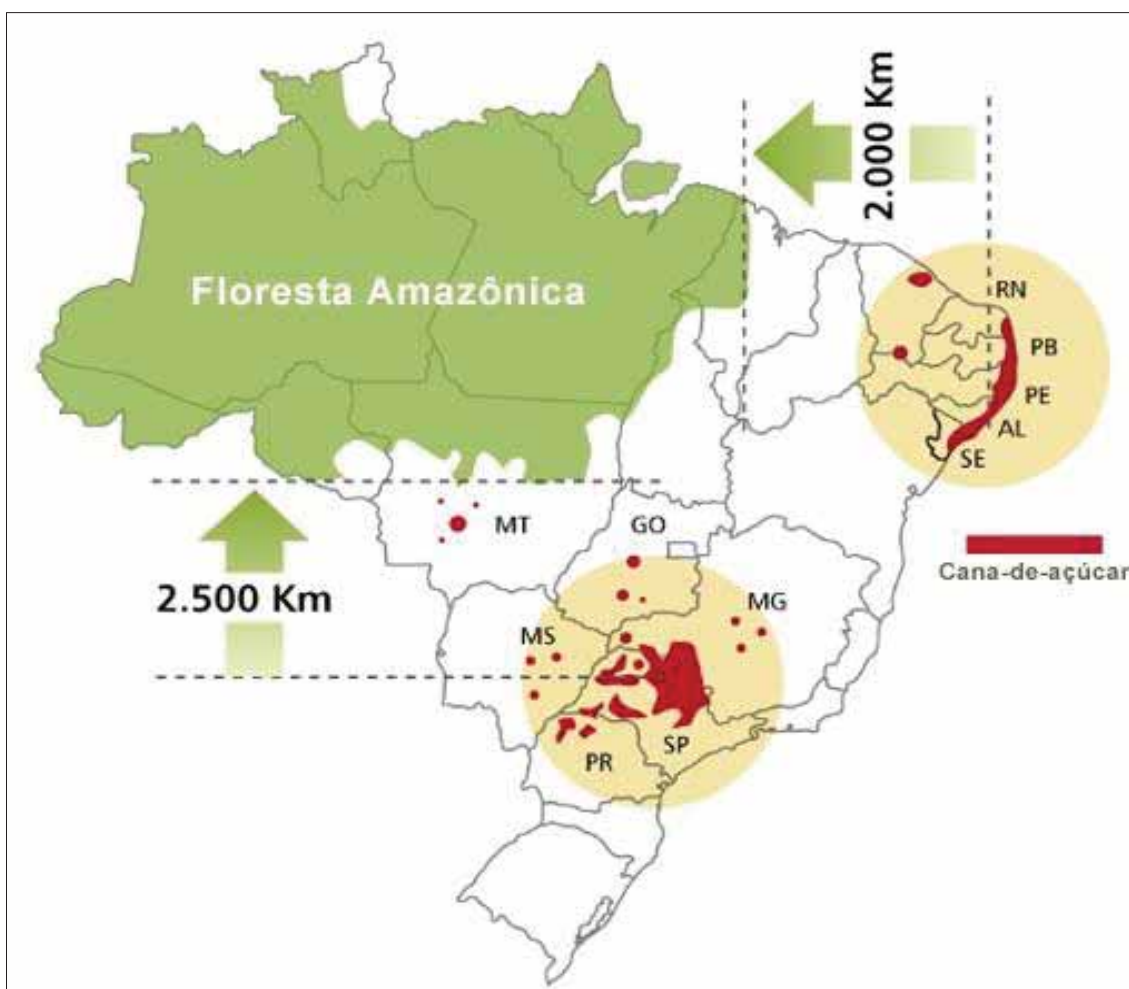


Figura 2 - Mapa de produção do setor sucroenergético
Fonte - UNICA (2012b)

De acordo com as estimativas realizadas pelo Ministério da Agricultura, a soja e cana-de-açúcar, conjuntamente, devem apresentar, nos próximos anos, uma expansão

de área plantada de 6,7 milhões de hectares, sendo 4,8 milhões de hectares de soja e 1,9 milhão de hectares de cana-de-açúcar (BRASIL, 2012a).

Tabela 1 demonstra a previsão de área plantada/colhida e produção de cana-de-açúcar para os próximos anos.

Tabela 1 - Previsão cana-de-açúcar para os próximos anos

Safra	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18	2018/19	2019/20	2020/21	2021/22
área (mil ha)	9.060	9.211	9.388	9.574	9.763	9.954	10.145	10.337	10.528	1.0720	10.911
cana-de-açúcar (mil t)	607.852	589.432	632.772	694.144	750.765	771.434	764.288	745.788	741.144	758.562	793.206

Fonte - Elaborado pelo autor a partir de dados publicados Brasil (2012a).

As maiores expansões de produção devem ocorrer nos estados de Goiás, 40,5%; São Paulo, 39,7%; e Minas Gerais, 32,6%. Em São Paulo, a produção deve aumentar em 135,8 milhões de toneladas. Para atender a esse crescimento, a área no estado deve aumentar em 29,7% no final do período das projeções. Pelas previsões realizadas, o estado de Goiás é o que deve apresentar, nos próximos anos, maiores aumentos da produção (40,5%) e da área de cana-de-açúcar (41,3%) (BRASIL, 2012a).

A cadeia produtiva sucroalcooleira é bastante estruturada no Brasil, único país do mundo que domina todos os estágios da sua tecnologia de produção. Representa um conjunto amplo e articulado de atividades que vão desde a produção da cana-de-açúcar, até a colocação do açúcar e do álcool no mercado para consumo final, passando por todos os elos de processamento. O agronegócio da cana-de-açúcar compõe-se de elos geradores de várias oportunidades de negócios: produção da cana-de-açúcar, processamento de produtos derivados, serviços de pesquisa, capacitação, assistência técnica e creditícia, transporte, comercialização, exportação, serviços portuários, entre outras (SEBRAE, 2008).

Trata-se de uma cadeia complexa, em função do número de elos que a compõe. Especialmente no conjunto de atividades centrais, apresenta uma diversidade considerável de estruturas produtivas desempenhadas por grandes e pequenas empresas, assim como de produtos e subprodutos colocados à disposição do mercado e de outras atividades industriais. Por outro lado, o setor vem passando por grandes mudanças tecnológicas, tanto de mercado quanto de processos produtivos, sobretudo de processos de aproveitamento dos subprodutos (bagaço e palha de cana), dados do SEBRAE (2008).

A Figura 3 apresenta um modelo proposto por Waack e Neves (1998 *apud* AZEVEDO, 2002), no qual mostra a cadeia produtiva de uma indústria sucroalcooleira.

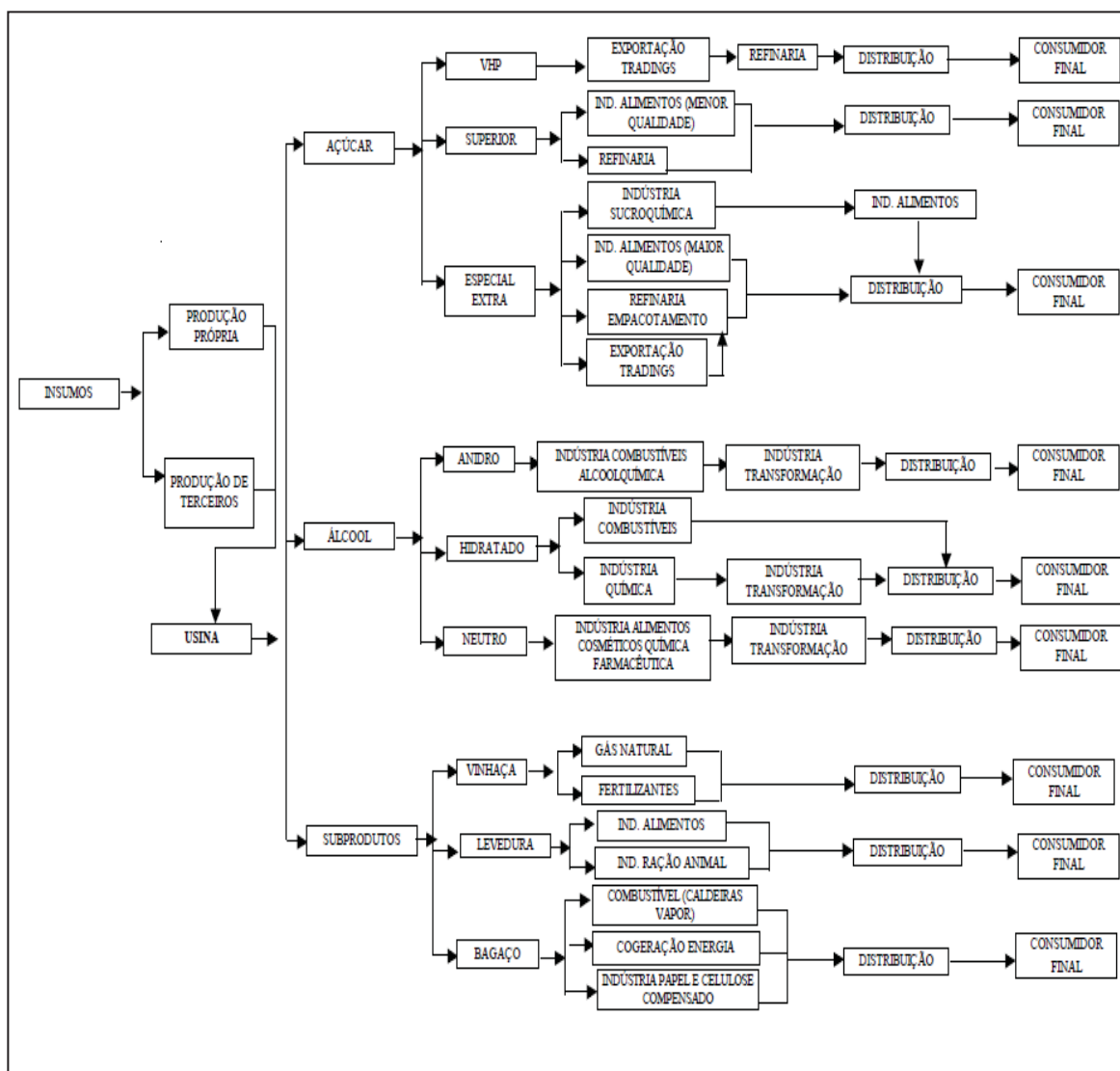


Figura 3 - Etapas de produção até o consumidor final
Fonte - Waack e Neves (1998 *apud* AZEVEDO, 2002).

Os processos de produção do açúcar e do etanol possuem etapas idênticas até a fase de separação do caldo, a partir daí, ocorrem distinções entre os processos. A Figura 4 ilustra todo o processo industrial de uma usina:

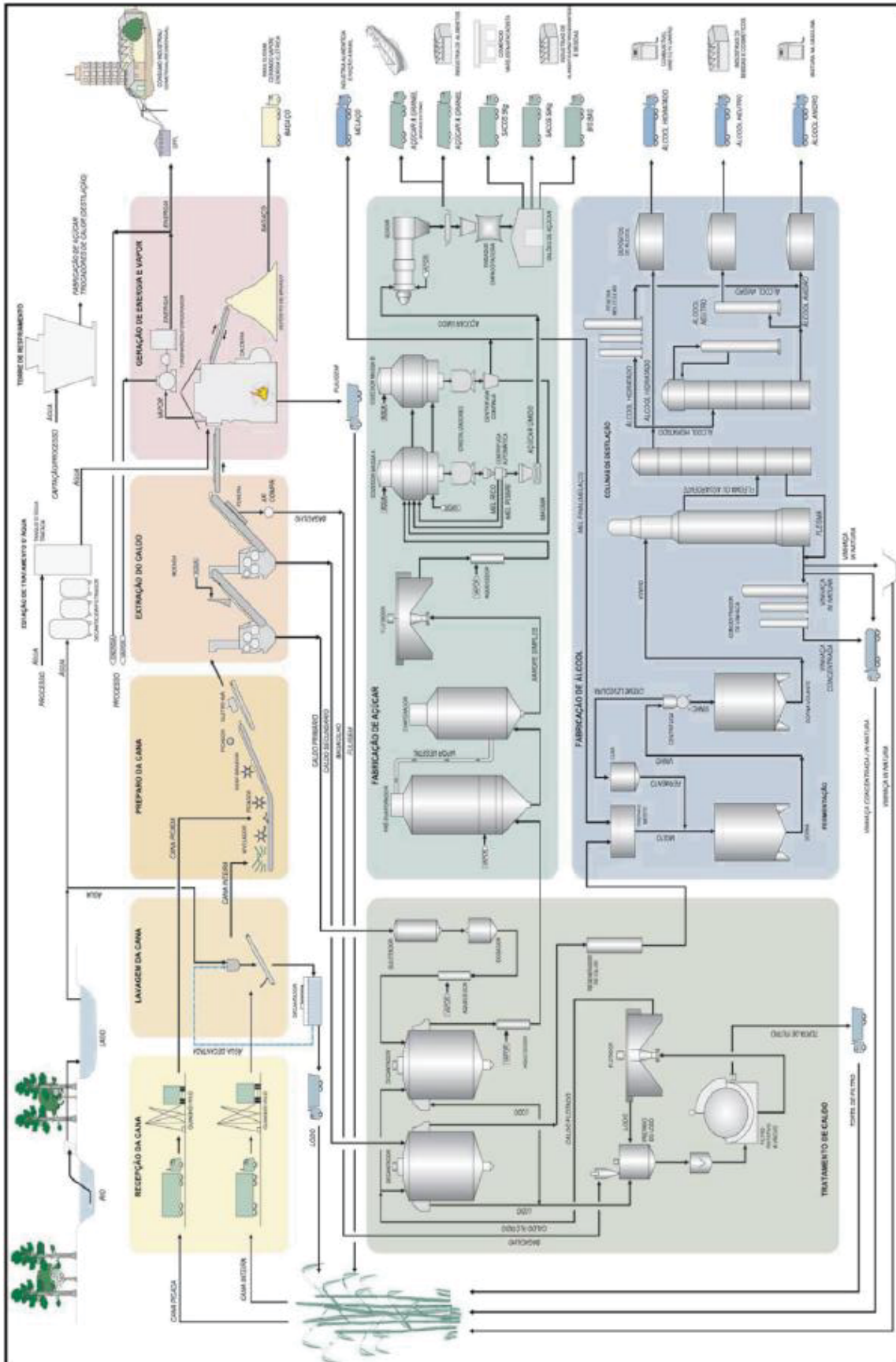


Figura 4 - Fluxograma do processo produtivo
 Fonte – Santa Elisa (2005 *apud* PAIVA, 2009).

Apresentado o fluxograma dos processos internos de uma indústria sucroalcooleira, passa-se à descrição sucinta de cada um:

1. Plantio: preparo e reforma do solo, plantio e crescimento do canavial;
2. Corte e carregamento: queima, movimentação de matéria-prima dentro do canavial, corte e carregamento de matéria-prima nas unidades de transporte;
3. Transporte, pesagem e amostragem: transporte de matéria-prima até a usina, pesagem e amostragem dos veículos de transporte na entrada da usina;
4. Recepção de matéria-prima: descarga, movimentação de matéria-prima dentro da usina e alimentação de matéria-prima para a moagem;
5. Preparo da cana: preparo da cana para a moagem;
6. Extração: extração do caldo e embebição do bagaço;
7. Clarificação: sulfitação, aquecimento, decantação e filtração do caldo para a fabricação de açúcar e de álcool;
8. Evaporação: evaporação do caldo para a fabricação de açúcar;
9. Fabricação de açúcar: cozimento da massa, cristalização e centrifugação do açúcar;
10. Fermentação: fermentação do caldo para fabricação de álcool;
11. Tratamento de levedura: tratamento, centrifugação e secagem da levedura utilizada na fermentação;
12. Destilação: destilação, retificação e desidratação do álcool;
13. Armazenamento: secagem, ensaque e armazenamento do açúcar;
14. Geração de vapor: geração de vapor a partir do bagaço.

Devido ao grande volume de mão de obra empregada principalmente nas atividades de manejo, corte e carregamento, o índice de absenteísmo influencia diretamente as atividades seguintes do processo de fabricação (a partir do item 4, conforme descrito anteriormente).

A próxima seção aborda o conceito do absenteísmo, suas causas, consequências e quais medidas podem ser adotadas para combater o absenteísmo nas empresas.

2.2. Absenteísmo

A Gestão de Recursos Humanos assume cada vez mais um papel importante para o desenvolvimento das empresas. Sendo o capital humano o principal recurso de diferenciação, o que o torna um fator-chave de sucesso na luta por diferencial competitivo sustentável. As práticas e políticas de Recursos Humanos evoluíram, quer do ponto de vista teórico, quer da aplicação ao longo do Século XX ancoradas nas teorias organizacionais e teorias comportamentais (NEVES; GONÇALVES, 2009).

O trabalho realizado pelas organizações tem como finalidade transformar todos os seus recursos disponíveis (recursos humanos, tecnológicos e materiais) em produtos, visando a gerar retorno financeiro (TAMAYO, 2004).

As instalações físicas, os equipamentos e os recursos financeiros são necessários para a organização, mas pessoas são particularmente mais importantes, pois são elas que planejam e fabricam os produtos, executam serviços, controlam a qualidade, vendem os produtos, alocam recursos, estabelecem estratégias e objetivos para a organização (CHIAVENATO, 2008).

No trabalho das pessoas, um dos fatores que vem se destacando é a motivação. Esta sofre constante modificação e sua variação está diretamente ligada a cada indivíduo, que, por sua vez, tem oscilações em seu estado de ânimo.

Para que a organização possa manter seus colaboradores motivados, é preciso que se identifiquem quais são as necessidades pessoais dos seus colaboradores, podendo, assim, elaborar metas que busquem o comprometimento desse para alcançar as suas necessidades pessoais, respeitando, também, as necessidades da empresa em que está inserido.

A saúde e o bem-estar no trabalho têm despertado cada vez mais o interesse das organizações, o que se deve, em grande parte, à crescente conscientização dos inúmeros prejuízos que estes causam para as organizações, como perda de receita, queda de produtividade, rotatividade e desaceleração do ritmo de trabalho (TAMAYO, 2004).

A qualidade de vida no trabalho pode ser representada pelo grau de satisfação das necessidades pessoais por meio das atividades exercidas na organização. A satisfação dos colaboradores pode estar relacionada diretamente com a função exercida,

com o crescimento dentro da organização, com o reconhecimento das tarefas exercidas, pelo relacionamento com os colegas, com o ambiente de trabalho, pelo poder de tomar e participar de decisões, pelo salário justo e pelos benefícios oferecidos; fatores estes que afetam as atitudes e o comportamento dos colaboradores, influenciando diretamente na produtividade de cada um (CHIAVENATO, 2008).

Há uma diversidade de estudos que analisam as causas e o processo dos comportamentos voluntários humanos, dentre elas, temos a hierarquia das necessidades postulada por Maslow em 1943 e que, por vez, é a mais conhecida. Para o autor, o motivo do comportamento humano reside no próprio indivíduo, tais motivos são denominados como necessidades e podem ser conscientes ou inconscientes. Estão organizadas de forma hierárquica, arranjadas em uma pirâmide de importância no comportamento humano, em que, sua base, localizam-se as necessidades primárias, enquanto no topo estão as necessidades secundárias. Dessa forma, o indivíduo se impulsiona a satisfazê-las seguindo tal ordem, e, à medida que uma necessidade é atendida, outra surge (NOLTEMEYER *et al.*, 2012).

A hipótese central da teoria de Maslow é a existência de uma hierarquia das necessidades humanas, constituída pelas necessidades biológicas, psicológicas e sociais. Somente à medida que as necessidades inferiores da hierarquia são satisfeitas, pelo menos em parte, é que surgirão as necessidades superiores da hierarquia (NOLTEMEYER *et al.*, 2012). A Figura 5 apresenta graficamente a pirâmide das necessidades.



Figura 5 - Pirâmide de Maslow – Hierarquia das Necessidades
Fonte: Adaptada de Maslow (1987).

Bergamini (2005) ressalta como uma ideia revolucionária encontrada por Maslow, hierarquizar os motivos humanos para compreender melhor o comportamento humano em sua diversidade. A hierarquia, neste caso, facilita melhor a compreensão sobre o tipo de objetivo que está sendo perseguido pelo indivíduo, ou seja, pelas necessidades que estão energizando as suas ações comportamentais. O autor propõe uma visão sequencial das necessidades, conforme apresentado na Figura 6.

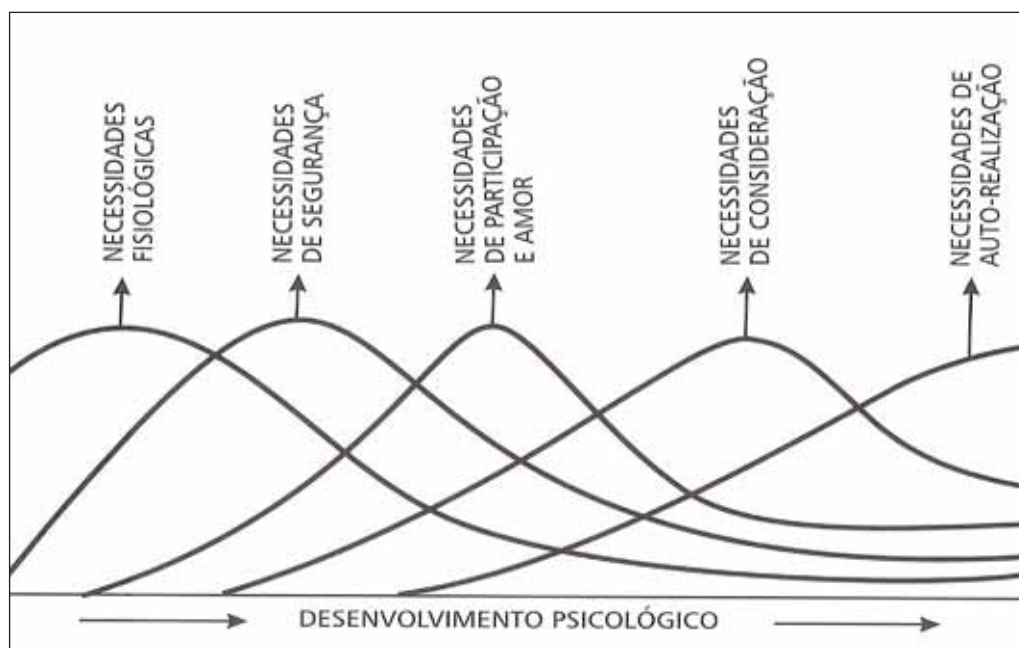


Figura 6 – Sequência das necessidades de Maslow
Fonte: Adaptada de Maslow (1987).

Um dos problemas associados aos baixos níveis de desempenho e de produtividade é o absenteísmo. Este comportamento, ligado ao capital humano, “retira competitividade das empresas, uma vez que, aumenta significativamente os custos e não permite que estas respondam às solicitações do mercado” (NEVES, 2010 *apud* DA SILVA, 2011).

O termo absenteísmo originou-se da palavra “absentismo” aplicado aos proprietários rurais que abandonavam o campo para viver na cidade. No período industrial, esse termo foi aplicado aos trabalhadores que faltavam ao serviço (QUICK e LAPERLOSA, 1982 *apud* PENATTI; ZAGO; QUELHAS, 2006).

Milkovich e Boudreau (2008) definem o absenteísmo como faltas ou ausências dos empregados ao trabalho. Em um sentido mais amplo, é a soma dos períodos em que os empregados da organização se encontram ausentes do trabalho, seja por falta, seja por atraso, devido a algum motivo interveniente.

Trata-se de um problema complexo, pois vem causando grandes preocupações às empresas. O mercado atual é altamente competitivo e é de grande importância que se tenha um controle bem delineado das despesas da empresa em todos os seus segmentos. As empresas trabalham com uma capacidade, cada vez mais, enxuta de empregados e com prazos reduzidos de entrega.

O absenteísmo torna-se um fator importante, pois tem grande impacto econômico devido a sua interferência na produção, aumento do custo operacional e redução da eficiência no trabalho. Acarreta, ainda, uma sobrecarga àqueles que permanecem no trabalho tendo que executar também as atividades dos colegas ausentes, prejudicando o grupo de trabalho e a própria rotatividade de pessoal, entre outros.

O absenteísmo tem exigido muito das organizações e de seus administradores, pois suas causas estão ligadas a múltiplos fatores, que vão desde as questões sociais, saúde, gestão de pessoas, entre outros problemas, o que torna este tema complexo e difícil de ser gerenciado (PENATTI; ZAGO; QUELHAS, 2006). Essas causas precisam ser diagnosticadas para que medidas preventivas e cuidados possam ser tomados, pois os efeitos das ausências podem ser irreversíveis.

Para o empregado, o absenteísmo pode expressar sentimentos de hostilidade e de falta de equidade ou de injustiça, referentes à sua situação concreta no trabalho. Também pode ser uma forma de fuga motivada pelo baixo salário, clima laboral ruim ou outras circunstâncias geradoras de insatisfação. Faltas significam, muitas vezes, que algo não está bem e que precisa de ações saneadoras para o problema. É um indicador fiel do nível de satisfação com o trabalho. É considerado um inimigo oculto da produtividade (POMI, 2013).

O absenteísmo é um fenômeno social vinculado diretamente à atitude do indivíduo e da sociedade ante o trabalho.

A análise dos indicadores de absenteísmo pode ser uma fonte significativa de entendimento dos custos, do clima organizacional e de ganhos efetivos de resultados financeiros e sociais. Muitas vezes, esses indicadores são controlados de forma simplista e operacional, perdendo-se o enfoque estratégico na gestão de valor dos ativos humanos (POMI, 2013).

Para Chiavenato (1997), o absenteísmo está relacionado com doença efetivamente comprovada ou não, com razões diversas de caráter familiar, com atrasos involuntários por força maior, com faltas voluntárias por motivos pessoais, com dificuldades e problemas financeiros, com problemas de transporte, baixa motivação para trabalhar, supervisão precária da chefia e políticas inadequadas da organização.

Da Silva (2011) correlaciona o absenteísmo com a satisfação dos funcionários, bem como com os indicadores associados às condições de trabalho, às práticas e políticas da empresa e com a produtividade. Para o autor as condições de trabalho e as políticas da empresa afetam a satisfação e a motivação dos funcionários, que conseqüentemente, reflete no absenteísmo e na produtividade. Esta correlação pode ser observada na figura 7.

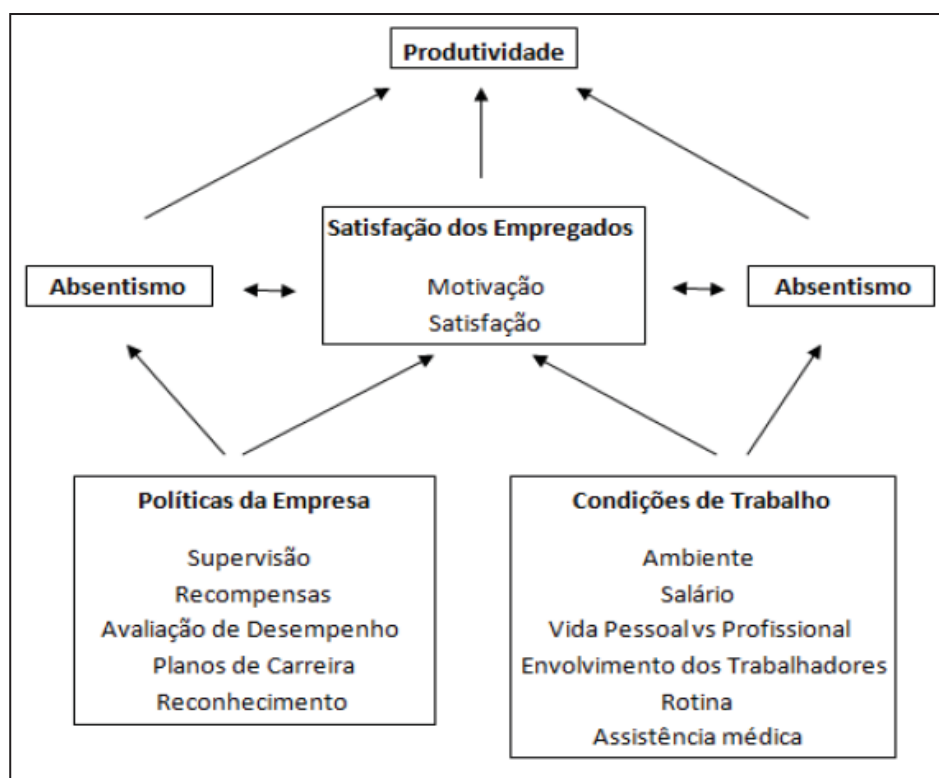


Figura 7 – Correlação absenteísmo
Fonte: Da Silva (2011).

Peretz e Fried (2012) realizaram estudo com 21 países para avaliar a correlação dos aspectos culturais da sociedade e práticas organizacionais de avaliação de desempenho com os índices de absenteísmo e rotatividade. Os resultados chamam atenção da área de Recursos Humanos sobre a importância de se considerar as diferenças culturais para a aplicação das práticas organizacionais de avaliação de desempenho. Os autores sugerem a adoção de sistemas de avaliação de desempenho que

utilizem múltiplas fontes de avaliação (por exemplo, avaliação 360 graus). No entanto, as empresas devem ter cautela na implementação de qualquer tipo de sistema, pois se os critérios não forem claros, isso pode provocar efeito inverso e desta forma pode ocorrer desmotivação dos funcionários e conseqüentemente um aumento do absenteísmo e da rotatividade.

Para Hausknecht, Hiller e Vance (2008), os níveis mais elevados de absenteísmo em um grupo de trabalho estão associados a níveis mais baixos de satisfação no trabalho e de comprometimento organizacional nos mercados de trabalho com desemprego baixo, e vice-versa. Comprometimento organizacional é o fator crucial: o absenteísmo é maior em unidades de trabalho com baixos níveis de compromisso, independentemente do nível de satisfação.

Stockmeier (2005) classifica três tipos de absenteísmo:

- **Absenteísmo por motivo de saúde ocupacional:** acidentes de trabalho, doenças ocupacionais e do trabalho, por exemplo;
- **Absenteísmo por motivo de saúde assistencial:** doenças comuns, sazonais, doenças cirúrgicas não ocupacionais, parto, acidentes domésticos e de trânsito, por exemplo;
- **Absenteísmo por motivos que não são de saúde:** atrasos, baixa motivação, problemas familiares, por exemplo.

Dentre as causas mais frequentes do absenteísmo, pode-se destacar:

- Insatisfação no trabalho;
- Falta de comprometimento com resultados;
- Falta de identificação com a tarefa;
- Ausência de supervisão eficiente/tratamento injusto;
- Falta de motivação para o trabalho;
- Ausência de feedback;
- Más relações interpessoais e intersetoriais;
- Absenteísmo-doença.

Em qualquer um dos motivos anteriores, é dever do gerente e do pessoal de recursos humanos tomarem as medidas cabíveis previstas na CLT, dissídio coletivo ou acordo entre empresa e sindicato para inibir a falta ao trabalho, seja justificada ou não; palestras de orientação aos funcionários devem ser ministradas ou folders devem ser fornecidos para total esclarecimento ou lembrança das regras de combate ao absenteísmo.

No combate aos motivos que não são por saúde, os encarregados são de extrema importância ao fiscalizar a ação de seus subordinados no que tange a produtividade no trabalho. Devem também, em qualquer caso de suspeita de doença ocupacional ou não, enviar o seu funcionário para avaliação médica.

Para Penatti, Zago e Quelhas (2006):

Existem medidas auxiliares no combate ao absenteísmo, diminuição de acidentes/doenças do trabalho, e conseqüentemente proteção da saúde ocupacional dos trabalhadores, além de proteger empresa e trabalhadores de prejuízos financeiros, o que aumenta o lucro de ambas as partes:

Algumas medidas de combate ao absenteísmo por motivo de saúde (ocupacional ou assistencial) que podem ser aplicadas:

- **Ginástica laboral;**
- **Programa de ergonomia;**
- **Programas de segurança do trabalho (CIPA).**

Algumas medidas de combate ao absenteísmo por motivos que não são de saúde:

- **Problemas sócio-econômicos:** incentivar a direção da empresa a implantar, desde que viável e conforme apresente lucros, um incentivo salarial, participação em lucros e resultados e implantação de plano de carreira aos funcionários;
- **Problemas pessoais no trabalho:** realizar reuniões com funcionários que se desentendem no trabalho, se possível com a assistência de um profissional de psicologia, para contornar o problema no ambiente de trabalho;
- **Problemas pessoais no lar:** assistência social e psicológica para contornar desavenças com cônjuge e familiares, desavenças com vizinhos, desavenças com amigos, falta de amizades;
- **Problemas de adaptação ao trabalho:** realizar rodízios no trabalho em turnos/noturno, trabalho insalubre, trabalho perigoso, realizar análise de perfil profissional e ou realizar programas de autoestima e valorização da própria função para evitar aversão pessoal à tarefa a ser realizada;

- **Problemas de abuso de drogas:** identificar os usuários de drogas/álcool e dar uma oportunidade de ingressarem em programas de recuperação para drogados, alcoólicos anônimos ou de combate ao fumo.

Para Chiavenato (2000), com a melhoria na saúde do colaborador e o fortalecimento das relações sociais entre a empresa e os colaboradores, podem ocorrer resultados positivos para ambas as partes. A empresa pode ter aumento de sua produtividade, melhoria na disposição dos trabalhadores, aumento do recrutamento e retenção do trabalhador e diminuição dos índices de absenteísmo. Já os colaboradores têm mais motivação, conseqüentemente, uma melhor qualidade de vida no trabalho, havendo, então, uma redução do tempo de enfermidade ou incapacitação e diminuição dos custos de cuidados com a saúde. Lembrando, também, que o absenteísmo é visto como resultado de um processo no qual o trabalhador apresenta autonomia frente a suas condições físicas de saúde, suas necessidades e sua motivação, sendo que o estado motivacional para trabalhar é um fator significativo no resultado final do absenteísmo.

2.2.1. O uso de modelos estatísticos para análise do absenteísmo

Vários modelos estatísticos são encontrados na literatura para modelar o absenteísmo.

Dos Reis *et al.* (2003) utilizaram o modelo ZINB para análise dos fatores do absenteísmo-doença em profissionais de enfermagem. O estudo apontou que os grupos com maior índice de absenteísmo foram os técnicos de enfermagem, mulheres e estatutários (OR=1,61; 1,47; 1,53 respectivamente). Os diagnósticos mais frequentes foram os relacionados ao aparelho respiratório. A análise multivariável, ao considerar separadamente os sexos, mostrou efeito apenas do vínculo empregatício (RR=1,45 e RR=2,43) para mulheres e homens. Os autores concluem que existe relação entre afastamento e vínculo empregatício. Os autores também recomendam estudos futuros com a inclusão de outras variáveis, como tempo na empresa, turno e carga reprodutiva.

Dias (2005) fez uso do modelo de regressão linear para modelar o absenteísmo em Call-Center. Os resultados indicaram, essencialmente, que o absenteísmo geral da amostra (7,1%) e dos operadores receptivos (6,8%) é pouco preditivo pelo modelo, pois não explicou, significativamente, a variância da amostra. As conclusões do estudo

refutam grande parte do modelo proposto, mas foram levantadas contribuições para a organização estudada, bem como para prática dos profissionais de Recursos Humanos.

Pinto Jr. (2005) assim como Dias (2005), realizou estudo sobre o absenteísmo em Call-Center com modelo de regressão linear. Como resultado, observou-se que os fatores recompensa e absenteísmo, justificam grande parte do resultado operacional e de desempenho, junto aos operadores ativos, talvez devido à remuneração variável que existe nesta área. Quanto aos operadores receptivos, o desempenho se vincula mais ao apoio da chefia e coesão entre os colegas; o absenteísmo pode ser gerado por uma insatisfação ou outros fatores que merecem de estudos adicionais.

Dionne e Dostie (2007) trabalharam com o modelo de Poisson com parâmetro adicional, para identificar as determinantes do absenteísmo usando dados empregador-empregado. No estudo os autores encontraram fortes evidências de que o horário de trabalho padrão durante a semana, opções de trabalho em casa e semanas de trabalho reduzidas foram associados com ausência reduzida, ao passo que o trabalho por turnos e semanas de trabalho longas foram associados com o aumento da ausência.

Silva, Pinheiro e Sakurai (2008) utilizaram ANOVA para identificação do perfil do absenteísmo em um banco estatal de Minas Gerais. Segundo os autores, o levantamento do perfil de absenteísmo nessa empresa mostrou-se importante para elucidação de alguns comportamentos referentes ao absenteísmo. Alguns dos itens levantados, como os referentes aos dados sociodemográficos, revelaram o envelhecimento da população trabalhadora, provavelmente decorrente da política restritiva de contratações e dos planos de adequação de quadros. Notou-se, ainda, crescimento no número de trabalhadores do sexo feminino, bem como dos níveis de escolaridade dos trabalhadores. Conclui-se, também, que as duas morbidades que mais afetaram os trabalhadores da empresa são aquelas tidas como ligadas às formas como o trabalho se organiza. Esse fato reforça a necessidade de um detalhamento do estudo atual no sentido de se investigarem as relações de determinação entre a organização do trabalho e as morbidades prevalentes observadas no banco pesquisado.

Da Fonseca (2009) utilizou o modelo de regressão logística univariada e multivariada, para a identificação dos fatores que influenciam o absenteísmo-doença em funcionários de uma indústria automobilística. O objetivo principal do estudo foi verificar a associação do absenteísmo por problema de saúde, presenteísmo e procura

por atendimento médico com o escore de atividade física habitual (AFH) de Baeke e seus três componentes: atividade física ocupacional (AFO), exercício físico de lazer (EFL) e atividade de lazer e locomoção (ALL). O escore AFO foi associado com o absenteísmo e apresentou proximidade à significância estatística com a procura por atendimento médico, ambos na análise univariada e não confirmada no modelo multivariado. A autora concluiu que o escore de atividade física habitual não foi associado estatisticamente ao absenteísmo por problema de saúde, presenteísmo e procura por atendimento médico. Dentre os escores avaliados, apenas o ocupacional se mostrou mais próximo da associação.

Tavares, Camelo e Kasmirski (2009) consideraram modelo de regressão linear simples para identificar os determinantes do absenteísmo dos professores da rede estadual paulista. O estudo foi realizado em duas etapas, na primeira, observa-se que os atributos pessoais dos professores, relacionados à sua carreira, ajudam a explicar sua decisão em faltar. Conclui-se que quanto mais estável é professor, em termos de tempo de carreira e do *status* ocupacional, maiores são os incentivos a faltar. Por outro lado, quanto maior o engajamento do professor na escola, menor a incidência de faltas. Na segunda etapa, estimou-se que 10 dias a mais de falta dos professores estão associados a uma nota em matemática 5% de um desvio-padrão abaixo da média, quando estimado pelo modelo MQO. Para dar conta da eventual endogeneidade da taxa de absenteísmo do professor, utilizou-se a técnica de variáveis instrumentais, cujo instrumento foi a distância da casa do professor até a escola. O instrumento utilizado pareceu ser capaz de isolar boa parte da endogeneidade por trás do absenteísmo docente, mas a partir desta estratégia econométrica o absenteísmo não se mostrou significativo para explicar o desempenho dos alunos.

Simões (2010) considerou o modelo de regressão logística bivariada e multivariada, para determinar os fatores associados ao absenteísmo-doença dos trabalhadores rurais de uma empresa florestal em Minas Gerais. Os resultados indicaram uma prevalência de 54,0% de atestados na população, com a perda de 7655 dias. Os agravos mais prevalentes foram os referentes ao aparelho osteomuscular (23,5%), seguidos do aparelho respiratório (14,0%) e as lesões, envenenamentos e outras consequências de causas externas (9,0%). A análise multivariada revelou que as variáveis função, setor e tempo de trabalho na empresa estiveram associadas a ocorrência do absenteísmo-doença. Os achados fornecem evidências importantes sobre

a relação entre fatores ocupacionais e demográficos e o absenteísmo-doença dos trabalhadores florestais. Conclui-se que o trabalho rural nos ambientes florestais tem relevância o suficiente para que estudos mais aprofundados sejam realizados, considerando a complexidade do tema relacionado a saúde e trabalho.

Dos Santos *et al.* (2011) fizeram uso dos modelos ZIP e ZINB para análise do absenteísmo-doença em trabalhadores de hospitais de Santa Catarina. Como resultados os autores apontam que o baixo suporte do colega aumentou a expectativa de dias de licenças para Tratamento de Saúde (LTS) em 2,04 (IC95%: 1,05-3,93), unido ao trabalho de baixa exigência aumentou 2,68 (IC95%: 1,37-5,27) e ao de alta exigência (*iso-strain*) 78% (IC95%: 1,02-3,12) mais do que o modelo demanda-controle (MDC) com alto suporte do colega. Este estudo mostrou a importância do suporte do colega sobre a duração absenteísmo-doença, e auxiliou conjecturar que a variável tempo nos trabalhadores de hospital leva a adaptação às exigências das tarefas e a monotonia. A conclusão dos autores aponta que intervenções no relacionamento entre os trabalhadores provavelmente diminuirão a duração de LTS.

Conforme Sturman (1996), é necessário uma análise criteriosa para escolha do modelo a ser utilizado para o estudo do absenteísmo, pois a escolha de um modelo inadequado pode trazer um viés sobre o resultado apurado.

Não foram encontrados estudos na literatura com aplicação dos modelos de regressão inflacionado de zeros para modelar o absenteísmo dos trabalhadores da área agrícola de empresas do setor sucroalcooleiro.

2.3. Modelos para dados de Contagem

Os modelos de regressão para dados de contagem assumem lugar de destaque na estatística aplicada.

Para Ramalho (1996), o elemento diferenciador deste tipo de modelo, que motivou o seu aparecimento, é a natureza especial da variável dependente, que assume apenas valores inteiros não negativos, correspondentes à ocorrência de um dado número de acontecimentos num certo intervalo de tempo. Ou seja, contagem de um número de

eventos que ocorrem numa mesma unidade de observação durante um intervalo de tempo definido.

Ramalho (1996) também afirma que o modelo clássico de regressão linear não se ajusta adequadamente para explicar uma variável discreta, que pode tomar somente um pequeno número de valores com probabilidades estritamente positivas, depende de um conjunto de regressores.

2.3.1. Modelo de Poisson

Dados expressos como contagem contabilizam o número de vezes que certo evento ocorre em um determinado período de tempo. Tal evento é uma realização de uma variável aleatória que assume distribuição de Poisson (COSTA, 2003 *apud* DA SILVA; FIACCONE; AMORIM, 2010).

“O modelo de regressão de Poisson tem por característica a análise de dados contados na forma de proporções ou razões de contagem” (McCULLAGH e NELDER, 1989 *apud* TADANO; UGAYA; FRANCO, 2009).

Segundo Jonhson e Kotz (1969 *apud* MONTROYA, 2009), uma variável aleatória Y tem distribuição de Poisson com parâmetro λ , denotada por $P(\lambda)$, se sua função de probabilidade for dada por

$$f(y, \lambda) = P(Y = y) = \frac{e^{-\lambda} \lambda^y}{y!}, \quad y = 0, 1, 2, \dots, \quad (1)$$

com $\lambda > 0$ o único parâmetro especificando a distribuição.

A esperança e a variância da distribuição de Poisson são dadas por $E[Y] = \lambda$ e $Var[Y] = \lambda$.

A variável resposta de uma regressão de Poisson deve seguir uma distribuição de Poisson e os dados devem possuir dispersão igual, ou seja, a média da variável resposta deve ser igual à variância (TADANO; UGAYA; FRANCO, 2009).

Ramalho (1996) apresenta as razões para utilização do modelo de Poisson:

- A distribuição de Poisson ajusta-se totalmente à principal característica dos dados de contagem, o fato de estes assumirem apenas valores inteiros não negativos;
- Após a sua estimação, é possível inferir, relativamente, a probabilidade de ocorrência futura de um dado acontecimento, admitindo que o modelo esteja bem especificado;
- O modelo de Poisson possui uma estrutura simples, que permite ser facilmente estimado.

A distribuição de Poisson pressupõe igualdade entre sua média e variância, contudo, na prática, a variância dos dados tende a ser muito maior que a média desses, fenômeno conhecido como superdispersão Costa(2003 *apud* DA SILVA; FIACCONE; AMORIM, 2010).

2.3.2. Modelo Binomial

Montoya (2009) apresenta o modelo Binomial e aponta suas aplicações práticas ao citar que, em muitos experimentos, existem situações em que o item analisado retorna respostas de “sucesso” ou “fracasso” (sucesso é o evento de interesse); ou seja, a variável resposta assume apenas dois resultados possíveis.

A distribuição binomial é formada quando o número de ensaios (n) é fixo, o parâmetro π (a probabilidade de sucesso) é a mesma para cada ensaio e os ensaios são todos independentes. Uma variável aleatória Y tem distribuição binomial com parâmetros n e π , denotada por $bin(n; \pi)$, sua função de probabilidade é dada por

$$f(y; n, \pi) = \binom{n}{y} \pi^y (1 - \pi)^{n-y}, \quad \text{para } y = 0, 1, \dots, n, \quad (2)$$

com $0 < \pi < 1$.

A média e a variância da distribuição binomial são dadas por $E[Y] = n\pi$ e $Var[Y] = n\pi(1 - \pi)$.

2.3.3. Modelo Binomial Negativa

A distribuição Binomial Negativa pode ser escrita de diversas maneiras de acordo com a parametrização utilizada. Conforme Nelder e Wedderburn (1972 *apud* MONTOYA, 2009), em que $p = \frac{\mu}{\mu+\phi}$, $0 < p \leq 1$, o parâmetro ϕ^{-1} é o parâmetro de dispersão e é assumido que $\phi \geq 0$. A distribuição binomial negativa como parâmetros μ e ϕ , denotado por $NB(\mu, \phi)$, tem função de probabilidade dada por

$$f(y; \mu, \phi) = \frac{\Gamma(\phi+y)}{\Gamma(y+1)\Gamma(\phi)} \left(\frac{\mu}{\mu+\phi}\right)^y \left(\frac{\phi}{\mu+\phi}\right)^\phi \quad \text{para } y = 0, 1, 2, \dots, \quad (3)$$

Esta distribuição tem valor esperado $E(Y) = \mu$ e a variância $Var(Y) = \frac{\mu(\phi+\mu)}{\phi} = \mu + \frac{\mu^2}{\phi}$. Os autores alertam para o fato de que a média é a mesma encontrada na distribuição de Poisson, porém esta distribuição apresenta uma modificação na função de variância de acordo com o parâmetro de dispersão, que lhe permite acomodar, desse modo, uma variação extra.

Como exemplos, podem-se destacar o estudo da relação entre acidentes de caminhões e o desenho geométrico das seções da estrada (MIAOU, 1994 *apud* MONTOYA, 2009) e o estudo da distribuição Binomial Negativa na análise de fenômenos recorrentes (NAVARRO *et al.*, 2001 *apud* MONTOYA, 2009).

2.4. Modelos para dados de contagem inflacionado de zeros

Recomenda-se um exame cuidadoso antes de aplicar qualquer técnica estatística em um conjunto de dados. Aconselha-se que uma análise exploratória seja realizada para obter informações adicionais e maior sensibilidade sobre a variável em estudo.

Em dados de contagem, durante a análise, pode-se, por exemplo, identificar uma heterogeneidade das variâncias, o que ocasiona uma variabilidade maior do que a esperada. Segundo Borgatto (2004), esse fenômeno é conhecido como superdispersão,

que pode ocorrer devido: à variabilidade da média, a excesso de zeros, à correlação entre indivíduos e/ou à omissão de variáveis não observadas.

Quando as pressuposições do modelo são violadas, a análise resultante pode levar a resultados equivocados. Segundo Borgatto (2004), este tipo de violação do modelo dá origem às chamadas falhas sistemáticas (não linearidade, não normalidade, não homocedasticidade, não independência, entre outros), que acarretam certo viés no modelo para predizer os valores, o que pode levar a super ou subestimação de variáveis importantes.

“O modelo clássico de Regressão de Poisson é frequentemente útil em descrever a média μ_i , contudo, ele subestima a variância dos dados quando há extra-variabilidade nos mesmos. Não levar em consideração o excesso de zeros e/ou a extra-variabilidade nos dados, pode levar à estimação incorreta dos erros-padrão, e conseqüentemente a uma avaliação incorreta da significância dos parâmetros da regressão individual” (LEE *et al.*, 2005).

Tal problema pode ser resolvido pelo modelo clássico de regressão de Poisson, ou seja, é possível, por exemplo, estimar adicionalmente um parâmetro de dispersão. Outra forma seria usar o modelo de regressão binomial negativo. Embora tais modelos sejam capazes de capturar a superdispersão nos dados, eles, muitas vezes, não são suficientes para modelar o excesso de zeros (ZEILEIS *et al.*, 2008 *apud* DA SILVA; FIACCONE; AMORIM, 2010).

Fenômeno bastante comum em dados de contagem é o excesso de zeros, muitas vezes, ocasionado por uma combinação dos chamados zeros estruturais e zeros amostrais, favorece-se, assim, a presença de um número de zeros muito maior do que se esperaria em uma distribuição de Poisson (COSTA, 2003 *apud* DA SILVA; FIACCONE; AMORIM, 2010).

As distribuições inflacionadas de zeros têm-se mostrado úteis para modelar os resultados do processo produtivo industrial (LAMBERT, 1992) e outras aplicações, como nas áreas da Medicina (BOHNING *et al.*, 1999; MAZIN *et al.*, 2008; XIANG *et al.*, 2006), da Saúde Pública (FAMOYE; SINGH, 2006; NAGAMINE; CANDOLO; MOURA, 2008; WANG *et al.*, 2003; ZHOU e TU, 2000), das Ciências Ambientais (AGARWAL; GELFAND; CITRON-POUSTY, 2002; NAYA *et al.*, 2008) e da Agricultura (DAGNE, 2010; HALL, 2000).

Pode-se destacar um crescente interesse pelos modelos inflacionados de zeros (ZI), por exemplo, o modelo Poisson inflacionado por zeros (ZIP) estudado por Lambert (1992), Hall (2000) e outros; e modelo binomial inflacionado de zeros (ZIB), discutido por Hall (2000) e Vieira, Hinde e Demétrio (2000).

2.4.1. Modelo Poisson Inflacionado de Zeros (ZIP)

Para o estudo de variáveis respostas nas quais prevalece a ocorrência de muitos zeros, Lambert (1992) orienta utilizar modelos para dados que se ajustem à superdispersão nos dados causada pelo excesso de zeros.

Esse modelo é dado por:

$$P(Y_i = y_i) = \begin{cases} \pi_i + (1 - \pi_i)e^{-\mu_i}, & y_i = 0 \\ (1 - \pi_i)\frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}, & y_i = 1, 2, \dots, \end{cases} \quad (4)$$

O valor esperado e a variância de y_i no modelo ZIP são, respectivamente, definidos por

$$E(Y_i) = (1 - \pi_i)\mu_i \quad \text{e} \quad \text{Var}(Y_i) = \mu_i(1 - \pi_i)(1 + \pi_i) \quad (5)$$

Conforme Couto (2008), esta metodologia geral para modelar dados de contagem inflacionado de zeros (modelos ZIP – zero-inflated Poisson regression) foi discutida por Ridout, Demétrio e Hinde (1998), em que foi apresentada uma revisão de literatura com ênfase em horticultura. Vieira (1998) e Vieira, Hinde e Demétrio (2000), aplicaram esses modelos em experimento para controle biológico de pragas. Outros estudos sobre excessos de zeros foram também realizados na simulação de dados sobre raízes de maçã, sobre ovos de jacaré, e na indústria em processos de fabricação ao modelar o número de defeitos.

2.4.2. Modelo Binomial Inflacionado de Zeros (ZIB)

O modelo de regressão binomial com excesso de zeros tem a seguinte forma:

$$P(Y_i = y_i) = \begin{cases} \pi + (1 - \pi_i) \left(1 + \frac{\mu_i}{k_i - \mu_i}\right)^{-k_i}, & y_i = 0 \\ (1 - \pi_i) \binom{k}{y} \left(\frac{\mu_i}{k_i - \mu_i}\right)^{y_i} \left(1 + \frac{\mu_i}{k_i - \mu_i}\right)^{-k_i}, & y_i = 1, 2, \dots, n_i \end{cases} \quad (6)$$

onde $0 < \pi < 1$, $0 < k_i < 1$. A média e a variância para o modelo ZIB são, respectivamente, dadas por

$$E(Y_i) = (1 - \pi_i)\mu_i \quad \text{e} \quad Var(Y_i) = \mu_i(1 - \pi_i)(1 - (n_i^{-1} \mu_i - \mu_i \pi_i)) \quad (7)$$

Entre os diversos estudos realizados, tem-se, por exemplo, Hall (2000), que estuda o modelo binomial inflacionado de zeros ao considerar o efeito aleatório para dados de horticultura.

2.4.3. Modelo Binomial Negativa Inflacionado de Zeros (ZINB)

Cheung (2002 *apud* FUMES, 2010) faz analogia ao modelo ZIP, com parâmetro de média μ_i , probabilidade π para os zeros estruturais e $1 - \pi_i$ para os dados amostrais, com $i = 1, 2, \dots, n$ e um parâmetro adicional de dispersão ϕ , esse modelo é dado por:

$$P(Y_i = y_i) = \begin{cases} \pi + (1 - \pi_i) \left(\frac{\phi}{\phi + \mu_i}\right)^\phi & y_i = 0 \\ (1 - \pi_i) \frac{\Gamma(\phi + y_i)}{\Gamma(y_i + 1)\Gamma(\phi)} \left(\frac{\mu_i}{\mu_i + \phi}\right)^{y_i} \left(\frac{\phi}{\mu_i + \phi}\right)^\phi & y_i = 1, 2, \dots, \end{cases} \quad (8)$$

em que o seu valor esperado é definido por $E(Y_i) = (1 - \pi_i)\mu_i$ e sua variância por

$$Var(Y_i) = (1 - \pi_i) \left(1 + \frac{\mu_i}{\phi} + \pi_i \mu_i\right) \mu_i \quad (9)$$

Nota-se que a média é a mesma encontrada na distribuição de Poisson Inflacionada de Zeros, porém esta distribuição apresenta modificação na função de variância que depende do parâmetro de dispersão, o que pode ser uma alternativa para os casos de superdispersão.

Estudos diversos têm sido feitos sobre o modelo ZINB. Por exemplo, Ridout, Hinde e Demétrio (2001) propuseram um teste escore para testar o modelo ZIP contra uma alternativa de que o modelo correto é o ZINB. Yau, Wang e Lee (2003) mostraram um estudo de dados de contagens com sobredispersão e excesso de zeros, com o uso de uma mistura do modelo de regressão binomial negativa inflacionado de zeros. Enquanto Mwalili, Lesaffre e Declerck (2008) ilustram como o modelo de regressão ZINB pode ser corrigido por erros de classificação.

2.5. Distribuição de séries de potência generalizada inflacionada de zeros (ZIPSG)

Uma variável aleatória discreta Y tem distribuição de Séries de Potência Generalizada (PSG) (CORDEIRO; ANDRADE; DE CASTRO, 2009) com média $\mu > 0$ e parâmetro de dispersão $\phi \geq 0$ se a sua função massa de probabilidade (fmp) for expressa como

$$P(Y = y) = \frac{a(y, \phi)g(\mu, \phi)^y}{A(\mu, \phi)}, y \in S \quad (10)$$

onde S é um subconjunto do conjunto dos inteiros não negativos, $a(y, \phi) > 0$ e que $g(\mu, \phi)$ e $A(\mu, \phi)$ (μ média e ϕ parâmetro de dispersão) são funções analíticas positivas, finitas e duas vezes diferenciáveis. Algumas distribuições importantes desta classe são: Binomial, Poisson, Binomial Negativa e a Poisson Generalizada.

Por exemplo, se k é um inteiro positivo, $a(y, \phi) = \binom{k}{y}$ e $A(\mu, \phi) = (1 + \frac{\mu}{k-\mu})^k$, $g(\mu, \phi) = \frac{\mu}{k-\mu}$ e $S = \{0, 1, \dots, k\}$, então (10) define a distribuição binomial. Alguns exemplos desta família que serão utilizados neste trabalho, onde $S = \{0, 1, \dots\}$ são os seguintes:

$$a(y, \phi) = \frac{1}{y!}, \quad g(\mu, \phi) = \mu \quad e \quad A(\mu, \phi) = e^\mu : \text{Poisson},$$

$$a(y, \phi) = \frac{\Gamma(\phi+y)}{y!\Gamma(\phi)}; \quad g(\mu, \phi) = \left(1 - \frac{\mu}{\mu+\phi}\right)^\phi \quad e \quad A(\mu, \phi) = \frac{\mu}{\phi+\mu} : \text{Binomial negativa},$$

$$a(y, \phi) = \frac{(1+\phi y)^{y-1}}{y!}, \quad g(\mu, \phi) = \frac{\mu e^{-\mu\phi(1+\mu\phi)^{-1}}}{1+\mu\phi} \quad e \quad A(\mu, \phi) = e^{\mu(1+\mu\phi)^{-1}} : \text{Poisson generalizada}$$

A distribuição de séries de potência generalizada inflacionada de zeros (ZIPSG) é o resultado da mistura entre distribuição PSG (10) e a distribuição degenerada em zero (JOHNSON; KEMP; KOTZ, 2005). A variável aleatória Y tem distribuição ZIPSG se sua fmp for dada por:

$$P(Y = y) = \begin{cases} \pi + (1-\pi) \frac{a(0, \phi)}{A(\mu_i, \phi)}, & y = 0 \\ (1-\pi) \frac{a(y, \phi)g(\mu, \phi)^y}{A(\mu, \phi)}, & y = 1, 2, \dots, \end{cases} \quad (11)$$

onde $0 \leq \pi < 1$. Quando $\phi = 0$ a variável aleatória Y tem uma distribuição de séries de potência generalizada com média μ e parâmetro de dispersão ϕ .

Em muitas aplicações práticas, é comum assumir que os parâmetros π_i , μ_i e ϕ_i dependem de vetores de variáveis explicativas (ou covariáveis) \mathbf{x}_{1i} , \mathbf{x}_{2i} e \mathbf{x}_{3i} , respectivamente. A proposição de Lord, Guikema e Geedipally (2008) é relacionar o π_i para covariáveis \mathbf{x}_{1i} pela ligação logística e μ_i e ϕ_i para covariáveis \mathbf{x}_{2i} e \mathbf{x}_{3i} , pela ligação logarítmica, respectivamente, isto é,

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \mathbf{x}_{1i}^\top \alpha, \quad \log(\mu_i) = \mathbf{x}_{2i}^\top \beta, \quad e \quad \log(\phi_i) = \mathbf{x}_{3i}^\top \gamma, \quad (12)$$

onde $\alpha = (\alpha_1, \dots, \alpha_{p1})^\top$, $\beta = (\beta_1, \dots, \beta_{p2})^\top$ e $\gamma = (\gamma_1, \dots, \gamma_{p3})^\top$ são parâmetros desconhecidos.

2.6. Método de estimação dos parâmetros do modelo

Estimação significa o ato de produzir estimativas de parâmetros populacionais a partir de uma amostra. Dentre os métodos clássicos da inferência estatística, existem dois caminhos possíveis para se obter estimativas de parâmetros: a *estimação pontual* e a *estimação por intervalos*. A estimação pontual refere-se à atribuição de um único valor numérico a um certo parâmetro populacional a partir de estatísticas amostrais. A estimação por intervalos utiliza as informações contidas na amostra para estabelecer uma afirmação quanto à probabilidade, ou grau de confiança, com que um certo intervalo de valores irá conter o verdadeiro valor do parâmetro populacional (NAGHETTINI; PINTO, 2007).

2.6.1. Método de Máxima Verossimilhança

Existem vários métodos para estimar os parâmetros. Dentre os mais utilizados, destaca-se o método de máxima verossimilhança (EMV). Para Cordeiro e Demétrio (2007), este método fornece estimadores com boas propriedades de consistência e eficiência assintótica.

O método da máxima verossimilhança consiste, basicamente, em maximizar uma função dos parâmetros da distribuição, conhecida como *função de verossimilhança*. O equacionamento para a condição de máximo resulta em um sistema de igual número de equações e incógnitas, cujas soluções produzem os estimadores de máxima verossimilhança” (NAGHETTINI; PINTO, 2007).

A partir de um conjunto de dados (amostra) e um modelo estatístico, a estimativa por máxima verossimilhança busca estimar os valores para os diferentes parâmetros do modelo. O método busca aqueles valores para os parâmetros de maneira a maximizar a probabilidade dos dados da amostra. O método de máxima verossimilhança apresenta-se como um método geral para estimação de parâmetros.

Uma amostra aleatória (y_1, y_2, \dots, y_n) , retirada de uma população com uma função de densidade de probabilidade $f(y, \theta)$, a qual depende do vetor de parâmetros θ , tem uma função de massa de probabilidade (fmp) conjunta dada por

$$\prod_{i=1}^n f(y_i, \theta)$$

A função de densidade de probabilidade conjunta é, simplesmente, o produto das densidades de cada uma das observações,

$$f(y_1, \theta) \times f(y_2, \theta) \times \dots \times f(y_n, \theta)$$

onde θ é um vetor de parâmetros (fixo) e y_i é uma variável aleatória (variável).

Antes da retirada da amostra, cada observação é uma variável aleatória cuja função de densidade de probabilidade é igual à função de densidade de probabilidade da população. A média e a variância de cada observação a ser retirada são iguais à média e variância da população em questão. Na função de densidade conjunta, antes de retirada a amostra, θ é fixo e y_i é variável.

Contudo, uma vez que tenha sido obtida uma amostra específica, y_i torna-se fixo e a função de densidade de probabilidade conjunta, pode, então ser reinterpretada como sendo uma função do vetor de parâmetros θ , que se tornam variáveis. Para uma dada amostra (y_1, y_2, \dots, y_n) , a função de densidade de probabilidade conjunta, vista como função do vetor de parâmetros desconhecidos θ , é denominada de função de verossimilhança.

Em econometria, o problema que se coloca é o de, dada uma amostra, obter-se uma estimativa dos valores dos parâmetros populacionais desconhecidos. Uma possibilidade para a resolução do problema de estimação é escolher o vetor $\hat{\theta}$ que maximize a probabilidade de obtenção da amostra específica (y_1, y_2, \dots, y_n) que se tem em mãos.

Tem-se, portanto, a função de verossimilhança $L(\theta, y)$, onde y é fixo e θ é a variável, e o problema consiste em obter-se o vetor $\hat{\theta}$ que maximiza esta função. O estimador de máxima verossimilhança $\hat{\theta}$ é o vetor que faz

$$L(\hat{\theta}, y) > L(\hat{\theta}, y)$$

onde $\hat{\theta}$ é qualquer outro estimador de θ .

2.7. Critérios para comparação dos modelos

Comparar diferentes modelos é a questão mais importante em modelagem estatística. Com o avanço da capacidade de processamento dos atuais computadores pessoais, desenvolvimento de novos softwares estatísticos e acessibilidade desses softwares à comunidade científica, múltiplos modelos mais complexos podem ser, simultaneamente, ajustados aos dados em análise. Isto implica que a modelagem estatística é um processo iterativo, pois os modelos devem ser construídos gradualmente e comparados entre si para a escolha do mais parcimonioso entre os de melhores ajustes. Além disso, a variância dos dados em torno da média (ou mesmo da mediana) nem sempre pode ser admitida como uniforme ao longo do intervalo da variável explanatória (homocedasticidade) (VAN BUUREN; FREDRICKS, 2001 *apud* PAIVA; FREIRE; CECATTI, 2008).

A modelagem por análise de regressão linear e ajuste pelo método dos mínimos quadrados ainda é a técnica mais utilizada nas publicações científicas que envolvem modelagem estatística. Isto tem ocorrido não só pelo fato de menor exigência computacional no cálculo dos parâmetros do modelo, mas principalmente, pela parcimônia e pela maior facilidade de entendimento do modelo gerado. No entanto, modelos aditivos generalizados são muito mais flexíveis e podem proporcionar um excelente ajuste na presença de relacionamentos não lineares. Por outro lado, devido a essa flexibilidade, o pesquisador deve ser cauteloso na escolha do modelo para não sobreajustar os dados, ou seja, aplicar um modelo excessivamente complexo (com vários graus de liberdade) de modo a produzir um bom ajuste que, provavelmente, não será replicado em estudos de validação posteriores (PAIVA; FREIRE; CECATTI, 2008).

Existem diversas ferramentas estatísticas para analisar e selecionar o modelo mais adequado. Van Buuren e Fredriks (2001 *apud* PAIVA; FREIRE; CECATTI, 2008) destacaram as seguintes ferramentas:

- Inspeção visual da forma das curvas de referência;
- Percentis plotados contra os valores dos dados individuais;
- Percentis empíricos e percentis ajustados plotados juntos;

- Verificação da proporção dos dados com valores acima e abaixo dos percentis ajustados;
- Testes estatísticos para verificar a normalidade dos resíduos;
- Gráfico quantil-quantil dos z-escores (“*qq-plot*” e “*detrended qq-plot*”);
- O gráfico “*worm plot*”;
- Desvio global (*deviance*);
- Critério de informação de Akaike (AIC);
- Critério bayesiano de Schwarz (SBC);
- Teste Q de Royston e Wright (2000).

Para Paiva, Freire e Cecatti (2008), não existe uma regra para a escolha das ferramentas, nem em que sequência devem ser utilizadas.

A análise visual dos gráficos produzidos depende da subjetividade do pesquisador e nem sempre será suficiente. Por isso, critérios objetivos devem ser utilizados na seleção de modelos.

Os critérios numéricos são bastante usados para discriminar modelos porque ponderam menos parâmetros com melhor adequação (menores desvios). A regra é selecionar entre os modelos candidatos que produzam o menor valor do AIC ou SBC. Menores valores para AIC e SBC indicam melhor ajuste, de modo que ambos os critérios permitem comparação entre modelos não aninhados e penalizam aqueles com maior número de parâmetros (PAIVA; FREIRE; CECATTI, 2008).

2.7.1. AIC – Akaike Information Criterion

Segundo Burnham e Anderson (1998), o AIC fundamenta-se no conceito de entropia por oferecer uma medida relativa da informação perdida quando um determinado modelo é usado para descrever a realidade e pode ser dito para descrever o equilíbrio entre polarização e variância na construção do modelo, ou, então, da precisão e complexidade do modelo. O cálculo geral do AIC pode ser descrito conforme a equação

$$AIC = 2k - 2\ln(L) \quad (12)$$

em que k é a quantidade de parâmetros estimados no modelo e L , o valor maximizado da função de verossimilhança para o modelo estimado.

A lógica do AIC é que não há hipótese sendo testada como no teste F. Ao contrário, o critério permite que se determine qual modelo é o mais correto e quanto. O método pode ser utilizado para comparar qualquer tipo de modelo: lineares, não lineares, aninhados e não aninhados. A base teórica matemática do método de Akaike é bastante complexa; combina a teoria da máxima verossimilhança, a teoria da informação e o conceito de entropia da informação (FLORIANO *et al.*, 2008).

A sequência de procedimentos para escolha do modelo pelo AIC deve respeitar os seis passos a seguir (MOTULSKY; CHRISTOPOULOS, 2003 *apud* FLORIANO *et al.*, 2008):

1. Ajustar as equações;
2. Anotar a soma de quadrados do erro de cada modelo; no caso de fatores ponderados, usar a soma ponderada de quadrados do erro;
3. Determinar o número de observações (n); se foi usada uma variável de frequência, ela deve ser considerada como multiplicador, e o valor total de n deve ser aquele determinado pela soma das frequências de cada classe de valor;
4. Determinar o valor de k para cada modelo; nos modelos não lineares sem um intercepto, deve-se adicionar 1 ao valor do número de parâmetros do modelo, porque a regressão não linear estima o valor da soma de quadrados;
5. Calcular o AIC, ou, se necessário, o AIC_c (corrigido);
6. O modelo com menor AIC ou AIC_c é o mais próximo de ser o correto; somente se pode comparar um dos dois, ou o AIC simples ou o corrigido (AIC_c), nunca o AIC de um com o AIC_c de outro modelo.

Ainda segundo Motulsky e Christopoulos (2003 *apud* FLORIANO *et al.*, 2008), somente devem ser comparados modelos que se ajustam bem aos dados, deve-se, antes, eliminar todos os modelos que não apresentam bons resultados e ficar com os dois ou três melhores para a comparação final por meio de ferramentas estatísticas, como o AIC.

Para Lindsey (1997 *apud* BORGATTO, 2004), o AIC é mais apropriado para dados provenientes de pequenas amostras.

2.7.2. SBC – *Schwarz Bayesian Criterion*

Também conhecido como critério bayesiano de Schwarz ou critério de informação bayesiano (BIC). Foi desenvolvido por Gideon E. Schwarz em 1978.

É uma estatística semelhante ao critério de Akaike com a característica de impor uma penalidade maior pela inclusão de coeficientes adicionais a serem estimados.

O cálculo geral do SBC pode ser descrito conforme a equação:

$$SBC = k \ln(k) - 2\ln(L) \quad (13)$$

em que k é a quantidade de parâmetros estimados no modelo e L , o valor maximizado da função de verossimilhança para o modelo estimado.

Schwarz (1978) conclui que, para um grande número de observações os procedimentos diferem marcadamente entre si. Desta forma, o critério de Akaike não pode ser assintoticamente ótimo.

Portanto, o critério SBC é recomendado para dados provenientes de grandes amostras.

3. METODOLOGIA DA PESQUISA

Gil (1996) define pesquisa como:

procedimento racional e sistemático que tem como objetivo proporcionar respostas aos problemas que são propostos. A pesquisa é requerida quando não se dispõe de informações suficientes para responder ao problema, ou então quando a informação disponível se encontra em tal estado de desordem que não possa ser adequadamente relacionada ao problema.

Ainda segundo Gil (1996), é possível classificar as pesquisas com base em seus objetivos, como exploratórias, descritivas e explicativas. Yin (2005) descreve que, de acordo com esses objetivos, existem cinco principais estratégias (ou procedimentos) de pesquisa: experimentos, levantamentos, análise de arquivos, pesquisas históricas e estudos de caso.

Quanto às abordagens de pesquisa, existe a quantitativa e a qualitativa. A quantitativa é o ato de mensurar variáveis de pesquisa e, uma forma de atingir isso é pelo uso da linguagem matemática. Na pesquisa quantitativa, não há subjetividade na apreensão dos fatos, o pesquisador não interfere ou pouco interfere nas variáveis da pesquisa, elas são definidas antes da realização da observação ou experimentação e tem o caráter da generalização e replicação (MIGUEL, 2010).

A abordagem qualitativa, por sua vez, não é a ausência da quantificação, e sim “a ênfase na perspectiva do indivíduo que está sendo estudado” (BRYMAN, 1989 *apud* MIGUEL, 2010). De acordo com o autor, na pesquisa qualitativa, a preocupação é obter informações do indivíduo e interpretar o ambiente em que a problemática ocorre. “Isso implica que o ambiente natural dos indivíduos é o ambiente da pesquisa”.

Miguel (2010) ressalta que, na Engenharia de Produção, uma pesquisa qualitativa significa que o pesquisador visita a organização, faz observações e coleta evidências. Nessa abordagem, a realidade subjetiva dos indivíduos é considerada relevante e contribui para o desenvolvimento da pesquisa, mas o autor ressalta que a pesquisa que utiliza a abordagem qualitativa tende a ser menos estruturada para poder captar as perspectivas e as interpretações das pessoas pesquisadas, e que isso não significa ser menos rigorosa, mas sim mais crítica.

Os métodos de pesquisa mais apropriados na área da Engenharia de Produção para conduzir uma pesquisa qualitativa, ainda segundo Miguel (2010), são o estudo de caso e a pesquisa-ação, sendo que essa última uma pesquisa concebida e realizada em

estreita associação com uma ação ou com a resolução de um problema coletivo, com participação das pessoas envolvidas no problema investigado.

Assim, pode-se dizer que o estudo de caso é um método qualitativo exploratório que tem caráter empírico e investiga um fenômeno atual, dentro de um contexto real, e é extraído de múltiplas fontes de evidência, em que o contexto em que ocorre é um fator importante (YIN, 2005; MIGUEL, 2010) e, desta forma, está claro se tratar de um método que preenche todos os requisitos para o tipo de problema apresentado nesse trabalho.

No entanto, Yin (2005) ressalta que o estudo de caso ainda é tratado como uma forma de pesquisa menos desejável do que os experimentos e levantamentos, justamente pela ausência ou menor rigor na pesquisa em relação a outros métodos, por fornecerem pouca base para generalização e por levarem tempo para serem desenvolvidos, com resultados em documentos volumosos e ilegíveis. Miguel (2010) concorda que, frequentemente, os trabalhos baseados em estudo de caso estão sujeitos a críticas em função das limitações metodológicas na escolha do caso, na coleta e análise dos dados, na apresentação dos resultados e na geração de conclusões suportadas pelas evidências, mas que, quando bem conduzidos, podem conferir maior validade e confiabilidade no seu uso.

A partir das classificações apresentadas, este trabalho se caracteriza como:

- Uma pesquisa **aplicada**, uma vez que objetiva gerar conhecimento para aplicação de práticas dirigidas à solução de problemas específicos, que envolve verdades e interesses locais;
- Uma pesquisa **quantitativa**, pois requer o uso de técnicas estatísticas na análise das informações obtidas;
- Uma pesquisa **explicativa**, por possuir um objetivo explicativo que visa a identificar características de relação entre as variáveis em estudo para explicar a razão de ocorrência do fenômeno;
- E, do ponto de vista dos procedimentos técnicos, a pesquisa caracteriza-se como **modelagem quantitativa**, em que, com o uso de modelos matemáticos, busca-se a resolução de problemas reais.

3.1. Origem e planejamento da pesquisa

Inicialmente, foi realizada uma revisão da literatura sobre o que já havia sido escrito no tema modelos para dados de contagem inflacionado de zeros. Verificou-se que, embora muita coisa tenha sido escrita sobre o tema, pouco existe sobre a utilização desses modelos para descrever os fatores associados ao absenteísmo de trabalhadores da área agrícola.

Outra percepção que originou o interesse por esse trabalho veio de observações do autor em relação à relevância da mão de obra na formação do custo das empresas do setor sucroalcooleiro. O custo dos trabalhadores da área agrícola representa 20% do custo total do negócio (PECEGE, 2011).

Uma vez decidido o tema, primeiramente, o planejamento consistiu em mapear a literatura em busca dos motivos que justificassem a adoção dos modelos estatísticos para a determinação das causas do absenteísmo dos trabalhadores.

A partir dessa proposição, definiu-se uma estrutura conceitual teórica que sustentasse esses conhecimentos, composta pela importância do setor sucroalcooleiro no Brasil; o absenteísmo e suas implicações; apresentação dos modelos para dados de contagem; e modelos para dados de contagem inflacionado de zeros para embasar o desenvolvimento dessa pesquisa.

Após levantamento do referencial teórico, procurou-se, então, planejar o caso em si. Foi selecionada a empresa de análise, de forma a poder efetuar os contatos necessários para explicar os motivos do estudo, solicitar autorização para entrevistas e coleta de dados a partir de consulta direta ao banco de dados da empresa.

Sobre o critério de escolha do caso, a empresa em estudo é uma usina que nasceu da união de dois grandes grupos nacionais e, desde o início de suas atividades, trabalha com uma administração executiva que aplica as principais ferramentas de gestão disponíveis. A empresa busca, constantemente, aliar tecnologia de ponta e desenvolvimento de seus colaboradores. Devido à qualidade das informações, bom uso de seu sistema integrado de gestão (ERP) e conceito de gestão praticado pela empresa; essa foi escolhida para o estudo. Conforme Miguel (2010), deve-se tomar cuidado de não se fazer generalizações inadequadas. Com relação à utilização de apenas um caso para a pesquisa, Yin (2005) relata que, entre outras possibilidades aceitáveis para essa

escolha, se o caso único for representativo ou típico, com o objetivo de captar as circunstâncias e condições de uma situação diária ou de um lugar comum, então isso pode ser considerado justificável.

A empresa aceitou participar da pesquisa por entender que, quanto mais conhecer sobre os fatores que levam seus colaboradores a faltar no trabalho, melhores podem ser as estratégias para o combate do absenteísmo.

O próximo passo foi definir a abordagem metodológica que determinou os métodos e técnicas, tanto para a coleta quanto para a análise dos dados. Com relação à técnica de coleta de dados, foi realizada uma entrevista não estruturada com o gerente de recursos humanos, o objetivo desta entrevista foi o de levantar, junto ao gestor, quais seriam, na visão dele, os fatores que podem levar um trabalhador da área agrícola a faltar ao trabalho. Na sequência, foi consultado o gerente agrícola com o mesmo objetivo.

A partir das entrevistas, identificou-se como possíveis fatores, tanto na ocorrência de faltas como na não ocorrência de faltas dos funcionários, o cargo, o estado civil, a idade, o tempo de serviço, o salário e o grau de instrução dos trabalhadores.

A coleta de dados foi efetuada, inicialmente, com a consulta ao sistema de gestão de RH da empresa por meio de uma consulta SQL (Structured Query Language) montada com objetivo específico de fornecer uma amostra aleatória dos dados.

Com relação à análise dos dados, conforme já ressaltado por Miguel (2010), nem tudo o que foi coletado deverá ser incluído no relatório de pesquisa, “geralmente é necessário reduzir os dados de forma que seja incluído na análise somente aquilo que é essencial e que tem estreita ligação com os objetivos e constructos da pesquisa” (MIGUEL, 2010).

3.2. Modelagem estatística

Neste trabalho, a distribuição de séries de potência generalizada é modificada em (11), para a modelagem de dados de contagem inflacionados de zeros. Considera-se esta família já que inclui uma ampla classe de modelos discretos utilizados na literatura.

Isto é, tem-se como casos particulares os modelos ZIP, ZIB e ZINB, etc. Para o modelo modificado são desenvolvidos procedimentos inferenciais baseados na teoria de máxima verossimilhança (DURBIN, 1988).

Utiliza-se, também, o critério de Akaike (“Akaike Information Criterion”, AIC) e do critério bayesiano de Schwartz (“Schwartz Bayesian Criterion”, SBC) para comparação dos submodelos ZIP, ZIB e ZINB.

3.2.1. Inferência

Seja $\boldsymbol{\vartheta} = (\alpha^\top, \beta^\top, \gamma^\top)$ o vetor de parâmetros da distribuição de probabilidade da variável aleatória Y dada em (11). A partir de uma amostra de n observações independentes, $(y_1, \delta_1, x_{11}, x_{21}, x_{31}), \dots, (y_n, \delta_n, x_{1n}, x_{2n}, x_{3n})$, a função de log-verossimilhança correspondente é dada por

$$l(\boldsymbol{\vartheta}) = \sum_{i=1}^n \delta_i \log \left(\pi_i + (1 - \pi_i) \frac{a(0, \phi_i)}{A(\mu_i, \phi_i)} \right) + \sum_{i=1}^n (1 - \delta_i) \log \left((1 - \pi_i) \frac{a(y_i, \phi_i) g(\mu_i, \phi_i)^{y_i}}{A(\mu_i, \phi_i)} \right), \quad (15)$$

onde, $\delta_i = 1$ se $y_i = 0$ e $\delta_i = 0$ caso contrário; e π , μ_i e ϕ_i são dadas na Equação (12).

A estimativa de máxima verossimilhança dos parâmetros do modelo é obtida maximizando a função de verossimilhança em (15). A maximização numérica do logaritmo da função de verossimilhança é realizada com a utilização de *softwares* existentes. Neste trabalho foi utilizado o *software* R (ver, R DEVELOPMENT CORE TEAM, 2010).

A inferência assintótica para o vetor de parâmetros $\boldsymbol{\vartheta} = (\alpha^\top, \beta^\top, \gamma^\top)$ pode basear-se na distribuição assintótica do EMV de $\boldsymbol{\vartheta}$, $\hat{\boldsymbol{\vartheta}} = (\hat{\alpha}^\top, \hat{\beta}^\top, \hat{\gamma}^\top)$. Sob condições de regularidade adequadas, pode ser mostrado que a distribuição assintótica do estimador de máxima verossimilhança $\hat{\boldsymbol{\vartheta}}$ é normal multivariada com vetor de média $\boldsymbol{\vartheta}$ e matriz de covariância $\Sigma(\hat{\boldsymbol{\vartheta}})$, que pode ser estimada por

$$\hat{\Sigma}(\hat{\boldsymbol{\vartheta}}) = \left\{ - \frac{\partial^2 \ell(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^\top} \right\}^{-1}, \quad (16)$$

avaliada em $\mathfrak{P} = \widehat{\mathfrak{P}}$. As segundas derivadas requeridas são calculadas numericamente.

Além da estimação, o teste de hipóteses é outra questão importante. Sejam \mathfrak{P}_1 e \mathfrak{P}_2 subconjuntos disjuntos de \mathfrak{P} , o objetivo é testar $H_0 : \mathfrak{P}_1 = \mathfrak{P}_{01}$ contra $H_1 : \mathfrak{P}_1 \neq \mathfrak{P}_{01}$, \mathfrak{P}_2 não especificados. Seja $\widehat{\mathfrak{P}}_0$ o EMV restrito a H_0 . Define-se o logaritmo da estatística da razão de verossimilhança por

$$TRV = 2(l(\widehat{\mathfrak{P}}) - l(\widehat{\mathfrak{P}}_0)), \quad (17)$$

onde $l(\cdot)$ é o logaritmo da função de verossimilhança. Sob H_0 e algumas condições de regularidade, TRV converge em distribuição para uma distribuição qui-quadrado com $\dim(\mathfrak{P}_1)$ graus de liberdade.

4. APLICAÇÃO EM UMA EMPRESA DO SETOR SUCROALCOOLEIRO

Para atingir os objetivos do estudo, considerou-se como variável resposta o número de faltas dos funcionários da área agrícola de uma usina localizada no interior do estado de São Paulo, com orçamento anual de 167,5 milhões de reais, com quadro de pessoal composto por 1301 funcionários, destes 1086 atuam diretamente na área agrícola (86,5%).

A investigação considerou dados de intervalos mensais no período de 01/01/2009 a 31/12/2011, foram observados 2026 funcionários. Foi retirada uma amostra aleatória de tamanho $n = 943$ (funcionários). Para o estudo foram consideradas às seguintes variáveis apresentada na Tabela 2.

Tabela 2 - Descrição das variáveis

Variável	Descrição	Detalhe
Y	número de faltas	Em dias
x ₁	Cargo	0 = Trabalhadores na cultura de cana-de-açúcar 1 = Caso contrário
x ₂	Estado Civil	1 = Solteiro 0 = Caso contrário
x ₃	Idade	Em anos
x ₄	Tempo de serviço	Em anos
x ₅	Salário	Em unidades de salários mínimos
x ₆	Grau de Instrução	0 = Se escolaridade ≥ 5 anos de escolarização 1 = Caso contrário

Fonte - Elaborado pelo autor.

Foram selecionados apenas os funcionários da área agrícola, excluídos da amostra o gerente agrícola e os coordenadores, devido à diferença salarial existente em relação aos demais trabalhadores. Também não foram considerados funcionários afastados por motivo de auxílio maternidade e funcionários de férias.

Ao considerar que o índice de absenteísmo (IA) deve refletir a percentagem do total de dias perdidos (TDP), em decorrência de ausências, em relação ao total de dias disponíveis (TDD), a fórmula utilizada para calcular o absenteísmo neste estudo é dada por:

$$IA = (TDP) / (TDD) \times 100$$

O total de dias perdidos (TDP) retirados da amostra foi de 3064 e o total de dias disponíveis (TDD) é o resultado da multiplicação do número de funcionários da amostra (943) pelo número de dias de cada período de observação (30).

Sendo assim:

$$IA = 3064 / 28290 \times 100$$

$$IA = 10,83 \%$$

Com base na fórmula utilizada para realização do índice de absenteísmo, observou-se que, o índice na empresa, corresponde a 10,83% do total de dias disponíveis.

Conforme apresentado no gráfico da Figura 8, foram observados 354 funcionários com zero falta, isso representa 37,5% do total da amostra.

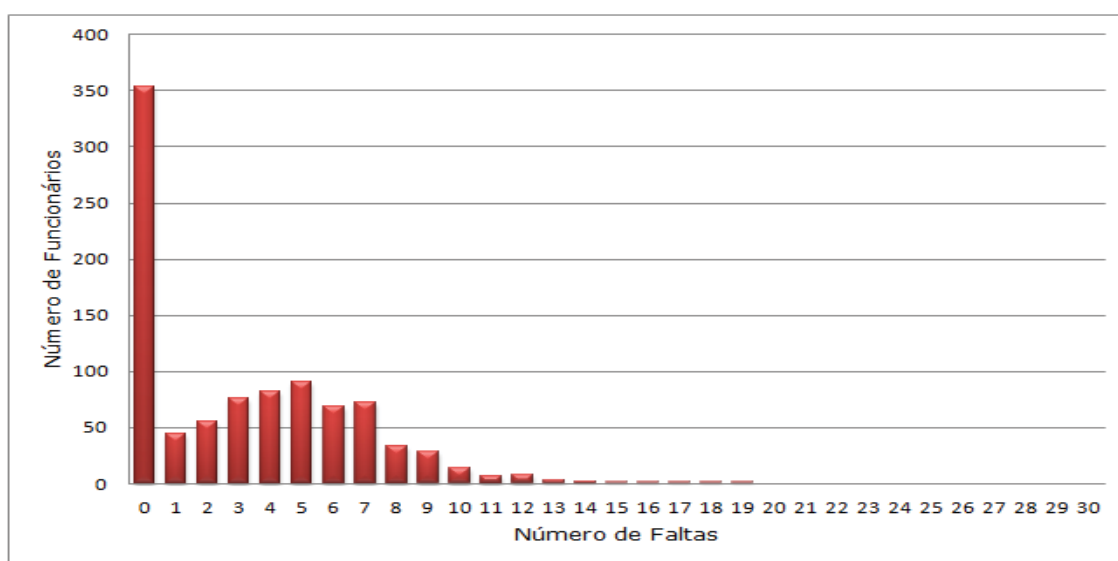


Figura 8 - Número de faltas
Fonte - Elaborado pelo autor.

Nota-se na Tabela 3 que no grupo de funcionários do estudo, houve predomínio de trabalhadores do sexo masculino (95,3%), com média de idade de 33,9 anos, sendo a faixa etária predominante de 20 a 29 anos (39,8%). Também predominaram os funcionários solteiros (59,30%) e grande concentração de trabalhadores com baixo índice de escolaridade, Ensino Fundamental (68,7%).

Em relação ao tempo de empresa, preponderou aquele grupo com tempo de serviço inferior a um ano (68%), com média de 1,5 anos de empresa. Com relação ao salário, predominou os funcionários com remuneração equivalente a 1 salário mínimo (50,6%). Este grupo é representado em sua maioria, pelos trabalhadores na cultura da cana-de-açúcar (37,8%) e os auxiliares de serviços rurais (11,2%). Entre as funções, destacaram-se os trabalhadores na cultura da cana-de-açúcar (37,8%), seguidos pelos motoristas (14,7%) e auxiliares de serviços rurais (11,2%).

Tabela 3 – Caracterização da população de trabalhadores

População	Número	Percentual
Sexo		
Masculino	899	95,3%
Feminino	44	4,7%
Faixa etária (em anos)		
18 – 19	15	1,6%
20 – 29	375	39,8%
30 – 39	294	31,2%
40 – 49	181	19,2%
50 – 59	64	6,8%
60 e +	14	1,5%
Estado Civil		
Solteiro	559	59,3%
Casado/Amasiado	376	39,9%
Separado/Divorciado	7	0,7%
Viúvo	1	0,1%
Grau de instrução		
Analfabeto	7	0,7%
Fundamental	648	68,7%
Médio	274	29,1%
Superior Incompleto	3	0,3%
Superior Completo	10	1,1%
Pós-Graduado	1	0,1%
Tempo de empresa (em anos)		
1	641	68,0%
2	197	20,9%
3	45	4,8%
4	53	5,6%
5 e +	7	0,7%
Salário (em Salários Mín.)		
1	477	50,6%
2	298	31,6%
3	129	13,7%
4	27	2,9%
5 e +	12	1,3%
Funções		
Ajudante Prev. Incêndio	6	0,6%
Aux. Administrativo	9	1,0%
Aux. Servs. Rurais	106	11,2%
Borracheiro	8	0,8%
Controlador Tráfego	6	0,6%
Coordenador Agrícola	11	1,2%
Eletricista Automotivo	9	1,0%
Engatador Julieta	35	3,7%
Fiscal de Turma	16	1,7%
Frentista	8	0,8%
Líder de Campo	36	3,8%
Mecânico	46	4,9%
Motorista	139	14,7%
Oper. Máq. Agrícola	20	2,1%
Oper. Colheitadeira Cana	72	7,6%
Trab. Cultura Cana	356	37,8%
Tratorista	60	6,4%

Fonte - Elaborado pelo autor.

Conforme Tabela 4, o absenteísmo por faixa etária predominou nos funcionários com idade entre 18 e 19 anos (14,0%). Em relação ao estado civil, o grupo dos casados foi o que mais se ausentou do trabalho, com índice de absenteísmo de 14,6%. Considerando-se o grau de instrução, o grupo que mais se destacou foi o de indivíduos com Ensino Médio (11,9%), seguido dos que tem apenas o Ensino Fundamental (10,5%).

Tabela 4 – Caracterização sócio-demográfica das faltas dos trabalhadores

População	Número	Percentual	Absenteísmo
Sexo			
Masculino	2986	97,5%	11,1%
Feminino	78	2,5%	5,9%
Faixa etária (em anos)			
18 – 19	63	2,1%	14,0%
20 – 29	1178	38,4%	10,5%
30 – 39	1007	32,9%	11,4%
40 – 49	596	19,5%	11,0%
50 – 59	189	6,2%	9,8%
60 e +	31	1,0%	7,4%
Estado Civil			
Solteiro	1395	45,5%	8,3%
Casado/Amasiado	1642	53,6%	14,6%
Separado/Divorciado	27	0,9%	12,9%
Viúvo	0	0,0%	0,0%
Grau de instrução			
Analfabeto	7	0,2%	3,3%
Fundamental	2041	66,6%	10,5%
Médio	976	31,9%	11,9%
Superior Incompleto	9	0,3%	10,0%
Superior Completo	31	1,0%	10,3%
Pos-Graduado	0	0,0%	0,0%

Fonte - Elaborado pelo autor.

Pode-se observar na Tabela 5, com relação a variável tempo de empresa, que o maior índice de absenteísmo foi encontrado no grupo de trabalhadores com um ano ou menos de empresa (11,1%), isso representa 69,5% de todas as faltas da amostra.

Na análise de faltas conforme remuneração predominou os trabalhadores com renda equivalente a um salário mínimo (11,3%), responsável por 52,6% do total de faltas da amostra.

Os engatadores de julieta (20,2%), os tratoristas (19,2%), os operadores de colheitadeira de cana (17,5%) e os auxiliares de serviços rurais (17,4%) foram, nessa ordem, as funções com maior índice de absenteísmo.

Percebe-se também que o menor índice de absenteísmo foi encontrado nos trabalhadores na cultura de cana-de-açúcar (2,4%), seguidos pelos fiscais de turma (7,3%). O baixo número de faltas nestes cargos ocorre porque estes trabalhadores recebem por produtividade, ou seja, se o funcionário não trabalhar também não ganha. As demais funções dentro da empresa não tem esse tipo de remuneração variável.

Tabela 5 – Caracterização ocupacional das faltas dos trabalhadores

População	Número	Percentual	Absenteísmo
Tempo de empresa (em anos)			
1	2129	69,5%	11,1%
2	615	20,1%	10,4%
3	140	4,6%	10,4%
4	162	5,3%	10,2%
5 e +	18	0,6%	8,6%
Salário (em Salários Mín.)			
1	1611	52,6%	11,3%
2	949	31,0%	10,6%
3	403	13,2%	10,4%
4	74	2,4%	9,1%
5 e +	27	0,9%	7,5%
Funções			
Ajudante Prev. Incêndio	24	0,8%	13,3%
Aux. Administrativo	41	1,3%	15,2%
Aux. Servs. Rurais	554	18,1%	17,4%
Borracheiro	34	1,1%	14,2%
Controlador Tráfego	26	0,8%	14,4%
Coordenador Agrícola	40	1,3%	12,1%
Eletricista Automotivo	33	1,1%	12,2%
Engatador Julieta	212	6,9%	20,2%
Fiscal de Turma	35	1,1%	7,3%
Frentista	30	1,0%	12,5%
Líder de Campo	150	4,9%	13,9%
Mecânico	208	6,8%	15,1%
Motorista	626	20,4%	15,0%
Oper. Máq. Agrícola	67	2,2%	11,2%
Oper. Colheitadeira Cana	379	12,4%	17,5%
Trab. Cultura Cana	260	8,5%	2,4%
Tratorista	345	11,3%	19,2%

Fonte - Elaborado pelo autor.

A Tabela 6 apresenta a distribuição de frequência do número de faltas dos 943 funcionários, conjuntamente com as frequências esperadas dos modelos ZINB, ZIB e ZIP. Pode-se observar que o ajuste por ZINB é melhor do que ZIP e ZIB em termos da estatística qui-quadrado (χ^2) de Pearson.

Tabela 6 - Ajuste da distribuição de frequência dos conjuntos de dados reais

Faltas	Nº de Funcionários (n)	Frequência esperada		
		ZINB	ZIB	ZIP
0	354	328,9	367,4	321,4
1	44	32,8	9,9	16,7
2	55	62,5	31,4	44,2
3	76	85,9	64,4	77,9
4	82	95,3	95,4	102,9
5	91	90,6	108,8	108,8
6	69	76,5	99,5	95,8
7	72	58,8	74,9	72,3
8	33	41,9	47,3	47,8
9	28	28,0	25,4	28,1
10	14	17,7	11,7	14,8
11	7	10,7	4,7	7,1
12	8	6,2	1,6	3,1
≥13	10	7,3	0,6	2,1
χ^2		18,8	331,1	104,4
p-value		0,17	0,00	0,00

Fonte - Elaborado pelo autor.

No caso dos modelos ZIP e ZIB, somente foram consideradas a ligação logística para proporção de não absentismo e ligação logarítmica para o número médio de faltas, pois ambos os modelos não incluem o parâmetro de dispersão. Além disso, as covariáveis consideradas foram estatisticamente significativas em todos os modelos.

Também foram comparados os membros do modelo de regressão ZIPSG (ZINB, ZIP e ZIB) por meio da inspeção dos critérios AIC e SBC. O modelo com o menor valor de qualquer um destes critérios (entre todos os modelos concorrentes) é, normalmente, tomado como o modelo preferido para descrever esses dados. A Tabela 7 mostra estas estatísticas em ordem crescente das estatísticas.

Tabela 7 - Algumas estatísticas a partir dos modelos ajustados

Modelo	Estatísticas		
	$-\max l(\vartheta)$	AIC	SBC
ZINB	1730,561	3465,121	3474,819
ZIB	1834,538	3673,076	3682,775
ZIP	1980,104	3964,208	3973,906

Fonte - Elaborado pelo autor.

O modelo ZINB destacou-se como o melhor modelo sendo escolhido como o modelo de trabalho.

O modelo descrito na Seção 2.5 foi ajustado com todas as combinações possíveis das covariáveis descritas na Tabela 2. Após a aplicação do teste da razão de

verossimilhança para a seleção de covariáveis (ou fatores), foi obtido o seguinte modelo de regressão ZIPSG:

$$\pi_i = \frac{\exp(\alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{4i})}{1 + \exp(\alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{4i})}, \log(\mu_i) = \beta_0 + \beta_1 x_{5i}, \text{ e } \log(\phi_i) = \gamma_0 + \gamma_1 x_{4i} \quad (18)$$

A Tabela 8 apresenta a estimativa máxima verossimilhança (EMV) dos parâmetros para o modelo de regressão ZINB.

Tabela 8 – Estimativa de Máxima Verossimilhança (EMV)

Parâmetros	Estimativa (est)	Erro Padrão (ep)	est / ep	p-value
α_0	0,651	0,211	3,10	
α_1 Cargo	-4,523	0,320	14,10	0,000001
α_2 Tempo de serviço	0,630	0,142	4,40	0,000001
β_0	1,705	0,054	31,70	
β_1 Salário	-0,051	0,028	1,80	0,070000
γ_0	10,969	0,649	16,90	
γ_1 Tempo de serviço	-12,993	0,652	19,90	0,000000

Fonte - Elaborado pelo autor.

A média e variância do modelo ZINB são, respectivamente,

$$E(Y_i) = (1 - \pi_i)\mu_i \text{ e } \text{Var}(Y_i) = (1 - \pi_i)\mu_i (1 + \pi_i\mu_i + \mu_i\phi_i), i = 1, \dots, 943.$$

A estimativa de máxima verossimilhança (EMV) da média do absentismo é dada por

$$\log(\hat{\mu}_i) = 1,705 - 0,051\text{Salário}_i, i = 1, \dots, 943, \quad (19)$$

(0,054) (0,028)

os valores expressos entre parênteses são o erro padrão dos coeficientes de regressão da EMV, o p-valor correspondente resultou em 0,07, isto indica que o salário tem um efeito significativo no número de médio de absentismo ao nível de significância de 10%. Além disso, da Equação (19), observa-se que funcionários com salários altos são

mais frequentes ao trabalho. A estimativa indica que para cada 1 salário de aumento, o absenteísmo é reduzido em 5,1%.

A estimativa da proporção de não absenteísmo é dada por:

$$\log\left(\frac{\hat{\pi}_i}{1-\hat{\pi}_i}\right) = 0,651 - 4,523Carg o_i + 0,63Tempo\ de\ serviço_i, \quad i = 1, \dots, 943, \quad (20)$$

(0,211) (0,32) (0,142)

os p-valores para avaliar o efeito das covariáveis cargo e tempo da equação (20) resultaram em valores menores que 0,000001, o que indica que essas covariáveis influenciam de forma significativa na proporção de não absenteísmo.

A Figura 9, representa a proporção de não absenteísmo estratificada por cargo e tempo de serviço, em que se observa que a proporção de não absenteísmo aumenta conforme o tempo de serviço e que essa proporção é maior para os trabalhadores da cultura da cana-de-açúcar.

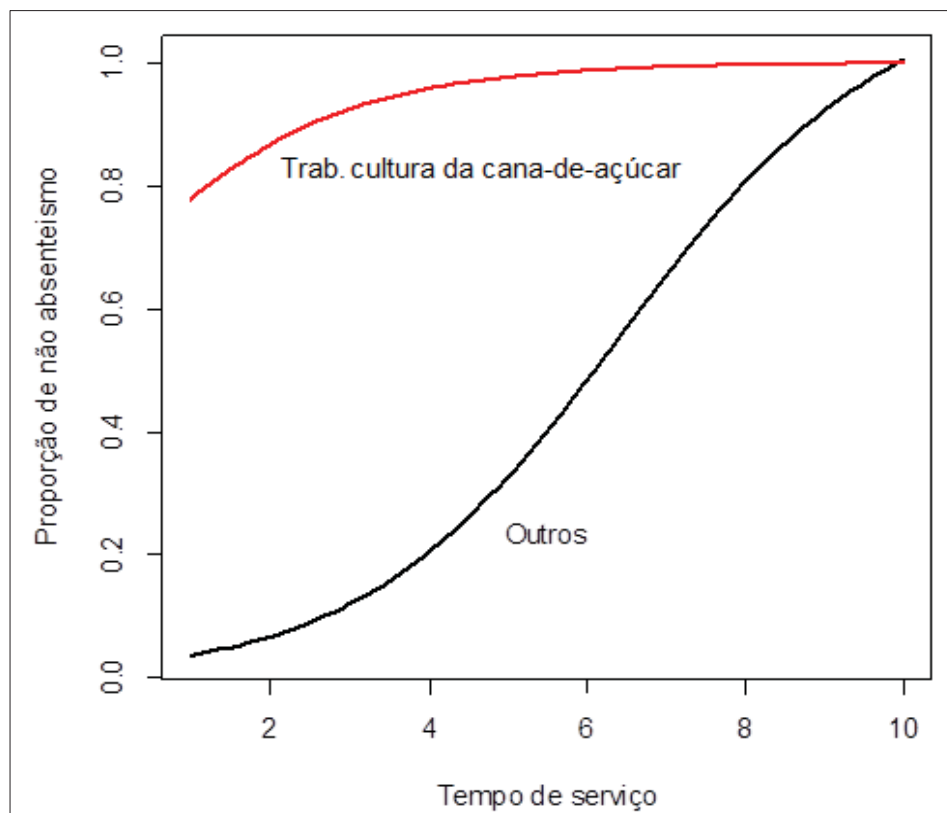


Figura 9 - Proporção de não absenteísmo de funcionários da empresa estratificada por cargo e tempo de serviço.

Fonte - Elaborado pelo autor.

A EMV do parâmetro de dispersão do modelo é dada por:

$$\log(\hat{\phi}_i) = 10,969 - 12,993 \text{Tempo de servicio}_i, \quad i = 1, \dots, 943. \quad (21)$$

(0,649) (0,652)

o p-valor correspondente à equação (21) resultou em 0,00, isto indica que o tempo de serviço é um fator que influencia de forma significativa na dispersão do modelo. Além disso, nota-se que a dispersão diminui com o tempo de serviço.

As Equações (19) e (20) podem ser utilizadas para previsão da média do absenteísmo e da proporção de não absenteísmo, respectivamente. Além disso, a partir das equações (19)-(21), pode-se estimar a média do modelo ZINB.

5. CONSIDERAÇÕES FINAIS

Este estudo teve como objetivo principal identificar os fatores que influenciam a ocorrência de absenteísmo e a não ocorrência de absenteísmo em trabalhadores da área agrícola de uma empresa do setor sucroalcooleiro.

Para atingir este objetivo, bem como os objetivos específicos, a investigação conduzida utilizou a família de distribuição de séries de potência generalizada inflacionada de zeros, como modelo alternativo para modelar dados de contagem em presença de muitos zeros no contexto dos modelos de regressão, com finalidade de estudar o efeito das covariáveis (fatores) nos parâmetros do modelo ZIPSG.

Os critérios de comparação AIC e SBC indicaram que o modelo ZINB é o que melhor se ajustou ao conjunto de dados estudado.

Para o modelo de regressão ZINB, pode-se observar que as covariáveis cargo dos trabalhadores e o tempo de serviço são estatisticamente significativas para o não absenteísmo.

Verificou-se também que o salário influencia significativamente na média de incidências de absenteísmo e somente a covariável tempo de serviço influencia significativamente na dispersão do modelo ZINB.

A partir do modelo desenvolvido e das análises efetuadas, foi possível chegar às seguintes conclusões específicas da empresa estudada:

- O índice de absenteísmo apresentado pela empresa (10,83%) está na média apresentada pelo setor no Brasil (11%), mas é necessário tomar medidas para combater este problema, pois se comparado a outros setores da indústria brasileira (2%), este nível de absenteísmo é muito elevado;
- A satisfação dos funcionários é influenciada pelas condições de trabalho oferecidas pela empresa e pelas políticas e práticas de recursos humanos implementadas;
- Ao nível das condições de trabalho, as variáveis cargo e tempo de serviço influenciam de forma significativa em proporção de não absenteísmo. Isto é, funcionários com tempo maior de empresa e com cargos que lhes permitam,

maior autonomia em suas rotinas de trabalho, realizam suas tarefas com maior satisfação e motivação. O que proporciona um índice de absenteísmo menor;

- Quanto às políticas e práticas de recursos humanos implementadas na empresa concluí-se que, o salário e o reconhecimento do trabalho prestado estão positivamente correlacionados com o absenteísmo;
- Verifica-se que o absenteísmo está estritamente influenciado pelo salário dos funcionários, ou seja, quanto maior é o salário dos funcionários menor é o absenteísmo na empresa.

O modelo analisado neste trabalho pode corroborar para as práticas dos profissionais de recursos humanos, uma vez que as análises estatísticas possibilitaram o reconhecimento de fatores importantes que influenciam a ocorrência e não ocorrência do absenteísmo.

Os trabalhadores possuem expectativas com relação ao seu trabalho. Dificilmente esses aceitam trabalhar em condições desmotivadoras. A maioria trabalha para garantir com seu esforço o sustento de si, e de seus familiares, mas isso não significa que o dinheiro seja tudo; além de salário, busca-se reconhecimento.

A gestão de recursos humanos deve tomar medidas com vista a aumentar os níveis de satisfação dos seus funcionários para diminuir o nível de absenteísmo. Conforme discutido anteriormente, os prejuízos advindos deste tipo de comportamento são enormes e envolvem muito mais do que aspectos financeiros, atingem a produção, o grupo de trabalho e a própria rotatividade de pessoal, entre outros.

Recomenda-se à empresa, como contribuição de ordem prática, aplicar um sistema de recompensas na sua estrutura, de forma a distinguir os funcionários com melhor desempenho, e a estimular os restantes funcionários. Por outro lado, quando for necessária uma pessoa para uma função mais responsável, a gestão da empresa deve efetuar um recrutamento interno a fim de possibilitar as promoções (plano de carreira).

Também é importante destacar a relevância de estudos que busquem cruzar dados de diferentes níveis da organização, como este, para expandir as fronteiras dos modelos de gestão existentes.

5.1. Sugestões para trabalhos futuros

A aplicação deste estudo em outras empresas do setor sucroalcooleiro pode complementar os resultados alcançados até o momento.

Assim como foi feito para o modelo ZINB, pode-se fazer o ajuste para o modelo de Poisson Generalizado.

Os modelos de regressão utilizados neste trabalho são aplicados em estudos para dados discretos. A utilização de modelos para dados contínuos também pode ajudar as empresas do setor sucroalcooleiro a maximizarem seus resultados, por exemplo, o estudo sobre os fatores ligados ao volume de perdas no processo de colheita mecanizado da cana-de-açúcar.

REFERÊNCIAS BIBLIOGRÁFICAS

AGARWAL, D.; GELFAND, A.; CITRON-POUSTY, S. Zero-inflated models with application to spatial count data. **Environmental and Ecological Statistics**, v.9, n.4, p. 341-355, 2002.

ANFAVEA. **Anuário da indústria Automobilística Brasileira**. São Paulo. 2012.

AZEVEDO, H. J. Uma análise da cadeia produtiva de cana-de-açúcar na região norte fluminense. Boletim técnico nº 6. **Observatório Socioeconômico da Região Norte Fluminense**, abril de 2002.

BACHMANN & ASSOCIADOS; ABRH-PR; ISAE/FGV. **3º Benchmarking Paranaense de Recursos Humanos: Dados de 2010**. Curitiba. 2011.

BERGAMINI, C. W. **Psicologia Aplicada à Administração de Empresas: Psicologia do Comportamento Organizacional**. 4. ed. São Paulo: Atlas, 2005.

BOHNING, D. *et al.* The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. **Journal of the Royal Statistical Society. Series A (Statistics in Society)**, v. 162, n. 2, p. 195-209, 1999.

BORGATTO, A. F. **Modelos para proporções com superdispersão e excesso de zeros - um procedimento bayesiano**. 2004. 104f. Tese (Doutorado). ESALQ-USP, Piracicaba, São Paulo. 2004.

BRASIL, Ministério da Agricultura Pecuária e Abastecimento. **Brasil projeções do agronegócio 2011/2012 a 2021/2022**. Disponível em <<http://www.agricultura.gov.br>>. Acesso em: Out. de 2012a.

BRASIL, Ministério de Minas e Energia. **Balanco Energético Nacional 2012**. Disponível em <<https://ben.epe.gov.br>>. Acesso em: Out. de 2012b.

BURNHAM, K. P.; ANDERSON, D. R. **Model Selection and Inference - A practical information-theoretic approach**. Ed. Springer 1998, 488p.

CARDOSO, M. F.; CARDOSO, J. de F.; SANTOS, S. R. O impacto da rotatividade e do absenteísmo de pessoal sobre o custo do produto: um estudo em uma indústria gaúcha. **Rev. Gestão, Finanças e Contabilidade**, v. 3, n. 1, p. 107-121, 2013.

CHIAVENATO, I. **Recursos Humanos**. São Paulo. Editora Atlas. 1997.

CHIAVENATO, I. **Administração: Teoria, Processo e Prática**. 3. ed. São Paulo: Mcgraw-Hill, 2000.

CHIAVENATO, I. **Gestão de Pessoas: o novo papel dos recursos humanos nas organizações**. 3. ed. Rio de Janeiro: Elsevier, 2008.

CORDEIRO, G. M.; ANDRADE, M.; DE CASTRO, M. Power series generalized nonlinear models. **Computational Statistics & Data Analysis (Print)**, v. 53, p. 1155-1166, 2009.

CORDEIRO, G. M.; DEMÉTRIO C. G. B. **Modelos Lineares Generalizados**. Mini-curso para o 12º SEAGRO e a 52ª Reunião Anual da RBRAS, UFSM, Santa Maria. 2007.

COUTO, M. R. M. **Transformação de dados com excesso de zero em experimentos com culturas Olerícolas**. 2008. 72f. Dissertação (Mestrado). Centro de Ciências Rurais, Universidade Federal de Santa Maria, Santa Maria, 2008.

DA FONSECA, V. R. **Atividade física, absenteísmo, e demanda por atendimento à saúde de funcionários de indústria automobilística de São Caetano do Sul**. 2009. Tese (Doutorado). Faculdade de Medicina. Universidade de São Paulo, São Paulo, 2009.

DA SILVA, N. B.; FIACCONE, R. L.; AMORIM, L. D. A. F. Uma aplicação dos Modelos Lineares Generalizados Hierárquicos duplos em dados longitudinais de contagem com excesso de zeros. **In: SINAPE**, 19., 2010. São Pedro.

DA SILVA, C. A. F. **Estudo do contributo da contabilidade de gestão na aplicação do modelo Scorecard RH em duas PME têxteis: o caso particular do absenteísmo**. 2011. Dissertação (Mestrado). Escola de Economia e Gestão. Universidade do Minho, Braga, 2011.

DAGNE, G. A. Bayesian semiparametric zero-inflated Poisson model for longitudinal count data. **Mathematical Biosciences**. Florida, v. 224, p. 126-130, 2010.

DIAS, M. **Absenteísmo em contact center** - Estudo de caso com preditores micro e macroorganizacionais. 2005. 109f. Dissertação (Mestrado). Instituto de Psicologia, Universidade Federal de Uberlândia, Uberlândia, 2005.

DIEESE. **Estudo de caso 4** – Setor Sucroalcooleiro, 2006.

DIONNE, G.; DOSTIE, B. New Evidence on the Determinants of Absenteeism Using Linked Employer-Employee Data. **Industrial and Labor Relations Review**. v.61, n.1, p. 106-120, 2007.

DOS REIS, R. J. *et al.* Fatores relacionados ao absenteísmo por doença em profissionais de enfermagem. **Revista de Saúde Pública**. v. 37, n. 5, p. 616-623, 2003.

DOS SANTOS, K. *et al.* Absenteísmo-doença, modelo demanda-controle e suporte social: um estudo caso-controle aninhado em uma coorte de trabalhadores de hospitais, Santa Catarina, Brasil. **Revista Brasileira de Epidemiologia**. v. 14, n. 4, p.609-619, 2011.

DURBIN, J. Maximum likelihood estimation of the parameters of a system of simultaneous regression equations. **Econometric Theory**. USA. v. 4, p. 159-170, 1988.

FAMOYE, F.; SINGH, K. P. Zero-Inflated Generalized Poisson Regression Model with an Application to Domestic Violence Data. **Journal of Data Science**. v. 4, p. 117-130, 2006.

FICARELLI, T. R.; RIBEIRO, H. Dinâmica do arrendamento de terras para o setor sucroalcooleiro: Estudo de caso no Estado de São Paulo. **Informações Econômicas**. v. 40, n. 1, p. 44-54, 2010.

FLORIANO, E. P. *et al.* Ajuste e seleção de modelos tradicionais para série temporal de dados de Altura de árvores. **Revista Ciência Florestal**. Santa Maria, v. 16, n. 2, p. 177-199, 2008.

FUMES, G.; CORRENTE, J. E. Modelos Inflacionados de zeros: Aplicações na Análise de um questionário de frequência alimentar. **Revista Brasileira de Biomedicina**. São Paulo, v.28, n.1, p.24-38, 2010.

GATTI JR., W. 35 anos da criação do pró-álcool: do álcool-motor ao veículo *flex-fuel*. In: SEMEAD – SEMINÁRIOS DE ADMINISTRAÇÃO, 13., 2010, São Paulo.

GIL, A. C. **Métodos e Técnicas de Pesquisa Social**. São Paulo: Atlas, 1996.

GUARDABASSI, P. M. **Sustentabilidade da Biomassa como Fonte de Energia: Perspectivas para Países em Desenvolvimento**. 2006. 123p. Dissertação de mestrado–Programa Interunidades de Pós-Graduação em Energia. Universidade de São Paulo.

HALL, D. B. Zero-inflated Poisson and binomial regression with random effects: A case study. **Biometrics**. v. 56, p.1030-1039, 2000.

HAUSKNECHT, J. P.; HILLER, N. J.; VANCE, R. J. **Deconstructing absenteeism: Satisfaction, commitment, and unemployment**. Ithaca, NY: School of Industrial and Labor Relations, Cornell University. 2008. Disponível em <<http://digitalcommons.ilr.cornell.edu/briefs/45>>. Acesso em: Nov. de 2012.

JOHNSON, N. L.; KEMP, A. W.; KOTZ, S. **Univariate Discrete Distributions**. 3 ed. Hoboken: John Wiley & Sons, Inc. 2005.

LAMBERT, D. Zero-Inflated Poisson Regression, with an Application to defects in Manufacturing. **American Society for Quality**. v.34, n.1, p.1-14, 1992.

LEE, A. H. *et al.* Modelling bivariate count series with excess zeros. **Mathematical Biosciences**, v. 196, p. 226 – 237, 2005.

LORD, D.; GUIKEMA, S.; GEEDIPALLY, S. Application of the Conway-Maxwell-poisson generalized linear model for analyzing motor vehicle crashes. **Accident Analysis & Prevention**, v. 40, n. 3, p. 1123-1134, 2008.

MASLOW, A.M **Motivation and Personality**. New York : Harper e Row, 1987.

MAZIN, S. C. *et al.* Uso de um modelo Bayesiano de Poisson com excesso de zeros na análise de dados de lesões miocárdicas em recém-nascidos com cardiopatias congênitas Complexas. **Revista Brasileira de Biomedicina**. São Paulo, v.26, n.4, p.113-125, 2008.

MIGUEL, P. A. C. (Coord.). **Metodologia de pesquisa em engenharia de produção e gestão de operações**. Rio de Janeiro: Elsevier, 2010.

MILKOVICH, G. T.; BOUDREAU, J. W. **Administração de Recursos Humanos**. São Paulo: Atlas, 2008.

MONTOYA, A. G. M. **Inferência e diagnóstico em modelos para dados de contagem com excesso de zeros**. 2009. 95 p. Dissertação (Mestrado) – Pós-Graduação em Estatística – Instituto de Matemática, Estatística e Computação Científica, UNICAMP, Campinas, 2009.

MWALILI, S. M.; LESAFFRE, E.; DECLERCK, D. The zero-inflated negative binomial regression model with correction for misclassification: an example in caries research. **Statistical Methods in Medical Research**. v.17, n.4, p.123, 2008.

NAGAMINE, C. M. L.; CANDOLO, C.; MOURA, M. S. de A. Uma aplicação de modelos para dados de contagem inflacionados de zeros na modelagem do número de ovos do mosquito *Aedes Aegypti*. **Revista Brasileira Biomedicina**. São Paulo, v.26, n.1, p.99-114, 2008.

NAGHETTINI, M.; PINTO, E. J. A. **Hidrologia estatística**. Belo Horizonte: CPRM, 2007.

NAYA, H. *et al.* A comparison between Poisson and zero-inflated Poisson regression models with an application to number of Black spots in Curriedale sheep. **EDP Sciences**. Les Ulis, v.40, p.379-394, 2008.

NEVES, J.; GONÇALVES, S. P. A investigação em Gestão de Recursos Humanos em Portugal: Resultados e tendências. **Revista Portuguesa e Brasileira de Gestão**, v.8, n.4, p.66-83, 2009.

NOLTEMEYER, A. *et al.* The relationship among deficiency needs and growth needs: An empirical investigation of Maslow's theory. **Children and Youth Services Review**. Miami, v.34, n.9, p.1862-1867, 2012.

ONU, **Resolution adopted by the General Assembly: The future we want**. General Assembly. (A/RES/66/288), 2012.

PAIVA, C. S. M.; FREIRE, D. M. C.; CECATTI, J. G. Modelos Aditivos Generalizados para Posição, Escala e Forma (GAMLSS) na Modelagem de Curvas de Referência. **Revista Brasileira de Ciências da Saúde**. João Pessoa, v.12, n.3, p.289-310, 2008.

PAIVA, R. P. O. de. **Modelagem do planejamento agregado da produção em usinas cooperadas do setor sucroenergético utilizando programação matemática e otimização robusta**. 2009. 241 p. Tese (Doutorado). Centro de ciências exatas e tecnologia, Universidade Federal de São Carlos, São Carlos, 2009.

PECEGE. **Custos de produção de cana-de-açúcar, açúcar e etanol no Brasil: Fechamento da safra 2010/2011**. Piracicaba: Universidade de São Paulo, Escola Superior de Agricultura “Luiz de Queiroz”, Programa de Educação Continuada em Economia e Gestão de Empresas/Departamento de Economia, Administração e Sociologia. 2011. 141 p. Relatório apresentado à Confederação da Agricultura e Pecuária do Brasil – CNA.

PENATTI, I.; ZAGO, J. S.; QUELHAS, O. Absenteísmo: As consequências na gestão de pessoas. **In:** SEGET, 3., 2006. Resende-RJ.

PERETZ, H.; FRIED, Y. National Cultures, Performance Appraisal Practices, and Organizational Absenteeism and Turnover: A Study Across 21 Countries. **Journal of Applied Psychology**, v. 97, n. 2, p. 448–459, 2012.

PINTO JR., H. **Impacto de características pessoais, comprometimento e clima organizacional no desempenho**: estudo de caso em uma Empresa de tele-atendimento. 2005. Dissertação (Mestrado). Instituto de Psicologia, Universidade Federal de Uberlândia, Uberlândia, 2005.

POMI, R. **Absenteísmo e Rotatividade de Pessoal**. Disponível em: <www.codigorh.com.br/absentesmo-e-rotatividade-de-pessoal>. Acesso em: Abr. de 2013.

R DEVELOPMENT CORE TEAM. **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing, Vienna, Austria. 2010.

RAMALHO, J. J. S. **Modelos de regressão para dados de contagem**. 1996. Dissertação (Mestrado). Instituto Supervisor de Economia e Gestão, Universidade Técnica de Lisboa, Lisboa, 1996.

RIDOUT, M.; DEMÉTRIO, C. G. B.; HINDE, J. Models for count data with many zeros. **In: INTERNATIONAL BIOMETRIC CONFERENCE**, Cape Town, 1998. Proceedings, Cape Town: IBC, p.1 – 13, 1998.

RIDOUT, M.; HINDE, J.; DEMÉTRIO, C. A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. **Biometrics**, v. 57, n. 1, p. 219-223, 2001.

SCHWARZ, G. Estimating the dimension of a model. **The Annals of Statistics**, v. 6, n. 2, p. 461-464, 1978.

SCOPINHO, R. A. Qualidade total, saúde e trabalho: uma análise em empresas sucroalcooleiras paulistas. **Revista de Administração Contemporânea**, v. 4, n. 1, P. 93-112, 2000.

SEBRAE. **Cadeia produtiva da indústria sucroalcooleira**: Cenários econômicos e estudos setoriais. Recife, 2008.

SHIKIDA, P. F. A.; AZEVEDO, P. F. de; VIAN, C. E. de F. Desafios da Agroindústria no Brasil Pós-desregulamentação: uma análise das capacidades tecnológicas. **Revista de Economia e Sociologia Rural**, v. 40, n. 3, p. 599-628, 2011.

SILVA, L. S.; PINHEIRO, T. M. M.; SAKURAI, E. Perfil do absenteísmo em um banco estatal em Minas Gerais: análise no período de 1998 a 2003. **Ciência & Saúde Coletiva**, v. 13, n.2, p. 2049-2058, 2008.

SIMÕES, M. R. L. **Análise do absenteísmo-doença dos trabalhadores rurais de uma empresa florestal em Minas Gerais**. 2010. Dissertação (Mestrado). Escola de Enfermagem, Universidade Federal de Minas Gerais, Belo Horizonte, 2010.

STOCKMEIER, T. E. **Programa de combate ao absenteísmo**: Baseado no documento elaborado pelo Médico do Trabalho Dr. Thomas Eduard Stockmeier, PhD em Medicina Ocupacional pela ANAMT, 2005. Disponível em: <<http://www.aa.med.br/upload/biblioteca/Programa%20de%20combate%20ao%20abse nte%C3%ADsmo.doc>>. Acesso em: Out. de 2012.

STURMAN, M. C. **Multiple approaches to absenteeism analysis**. Ithaca, NY: Cornell University, School of Industrial and Labor Relations, Center for Advanced Human Resource Studies. 1996. Disponível em: <<http://digitalcommons.ilr.cornell.edu/cahrswp/178>>. Acesso em: Nov. de 2012.

STURMAN, M. C. Multiple approaches to analyzing count data in studies of individual differences: The propensity for type I errors, illustrated with the case of absenteeism prediction. **Educational and Psychology Measurement**, v.59, p.414-430, 1999.

TADANO, Y. de S.; UGAYA, C. M.; FRANCO, A. T. Método de Regressão de Poisson: Metodologia para avaliação do impacto da poluição atmosférica na saúde populacional. **Ambiente & Saúde**. v.12, n. 2, p. 241-255, jul-dez 2009.

TAMAYO, A. **Cultura e Saúde nas Organizações**. Porto Alegre: Artmed, p. 12 – 116, 2004.

TAVARES, P. A.; CAMELO, R. de S.; KASMIRSKI, P. R. **A falta faz falta? Um estudo sobre o absenteísmo dos professores da Rede estadual paulista de ensino e seus efeitos sobre o desempenho escolar**. Texto para discussão. Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto. 2009.

TEODOZIO, J. M. O absenteísmo no trabalho por doença: estratégias de prevenção. **Revista Eficaz**. v. 2, n. 2, 2010. Disponível em: <<http://www.institutoeficaz.com.br/revistacientifica>>. Acesso em: Nov. de 2012.

UNICA. **Dados e Cotações**. Disponível em: <<http://www.unica.com.br>>. Acesso em: Out. 2012a.

UNICA. **Setor Sucroenergético**. Disponível em: <<http://www.unica.com.br>>. Acesso em: Out. 2012b.

VIEIRA, A. M. C. **Modelos para dados de proporções com superdispersão aplicados ao controle biológico**. Piracicaba, 1998. 61p. Dissertação (mestrado) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo.

VIEIRA, A. M. C.; HINDE, J.; DEMÉTRIO, C. G. B. Zero-inflated proportion data models applied to a biological control assay. **Journal of Applied Statistics**, v.27, p 373-389, 2000.

WANG, K. *et al.* A bivariate zero-inflated Poisson regression model to analyze occupational injuries. **Accident Analysis & Prevention**. v. 35, p. 625-629, 2003.

XIANG, L. *et al.* A score test for zero-inflation in correlated count data. **Statistics in Medicine**. v. 25, p. 1660-1671, 2006.

YAU, K.; WANG, K.; LEE, A. Zero-Inflated Negative Binomial Mixed Regression Modeling of Over-Dispersed Count Data with Extra Zeros. **Biometrical Journal**, v. 45, p. 437-452, 2003.

YIN, R. K. **Estudo de Caso**: planejamento e métodos. Tradução: Daniel Grassi. 3. ed. Porto Alegre: Bookman, 2005.

YUMKELLA, K. **Sustainable Energy for all**: The SE4ALL Global Tracking Framework full report. Disponível em < www.worldbank.org/se4all >. Acesso em: Jul. de 2013.

ZHOU, X.; TU, W. Confidence intervals for the mean of diagnostic test charge data containing zeros. **Biometrics**, v. 56, n. 4, p. 1118-1125, 2000.