



Universidade Federal de Uberlândia
Faculdade de Matemática

Bacharelado em Estatística

**PREDIÇÃO DE DESLIGAMENTO DE
FUNCIONÁRIOS POR MEIO DE
MODELOS DE REGRESSÃO LOGÍSTICA**

Georthon Giroldo dos Santos

Uberlândia-MG

2019

Georthon Giroldo dos Santos

**PREDIÇÃO DE DESLIGAMENTO DE
FUNCIONÁRIOS POR MEIO DE
MODELOS DE REGRESSÃO LOGÍSTICA**

Trabalho de conclusão de curso apresentado à Co-
ordenação do Curso de Bacharelado em Estatística
como requisito parcial para obtenção do grau de
Bacharel em Estatística.

Orientador: Prof. Dr. Janser Moura Pereira

Uberlândia-MG

2019



**Universidade Federal de Uberlândia
Faculdade de Matemática**

Coordenação do Curso de Bacharelado em Estatística

A banca examinadora, conforme abaixo assinado, certifica a adequação deste trabalho de conclusão de curso para obtenção do grau de Bacharel em Estatística.

Uberlândia, _____ de _____ de 20_____

BANCA EXAMINADORA

Prof. Dr. Janser Moura Pereira

Prof. Dr. Lúcio Borges de Araújo

Prof. Dr. José Waldemar da Silva

**Uberlândia-MG
2019**

Dedico este trabalho, a minha família e em especial a minha esposa Thalita que, com muito carinho e apoio, não mediram esforços para que eu pudesse completar mais uma etapa de minha vida.

AGRADECIMENTOS

À Universidade Federal de Uberlândia (UFU) pela possibilidade da formação e realização deste trabalho.

A minha mãe Maria Amélia Giroldo, que sempre esteve presente me apoiando e incentivando durante essa etapa da minha vida.

A minha esposa Thalita Gomes dos Santos, que compartilhou comigo esse momento e me incentivou para o desenvolvimento deste trabalho.

Ao Prof. Janser Moura Pereira pela orientação, apoio, incentivo, ensinamentos e paciência durante as aulas e realização deste e de outros trabalhos da universidade.

*"A alegria está na luta, na tentativa, no
sofrimento envolvido e não na vitória
propriamente dita"
- Gandhi, M.*

RESUMO

Este trabalho tem como objetivo verificar através da regressão logística quais fatores podem influenciar para o aumento do índice de turnover dos colaboradores de uma empresa. Os dados são provenientes de uma empresa que faz vendas por telefone em Uberlândia-MG. Foram analisados 264 operadores de televendas de ambos os sexos e idades entre 18 e 48 anos e ainda 18 variáveis comportamentais e de performance dos vendedores que representam possíveis causas que podem impactar para o aumento do turnover. De todas as variáveis avaliadas, cinco foram significativas para a predição do turnover (“Proporção de metas batidas”, “Quantidade de vezes que a meta de clientes foi batida”, “Quantidade de devoluções no período avaliado”, “Quantidade de faltas no período avaliado” e “Sexo”), contudo a variável “PropMetaBatida” possui um impacto significativo para a redução do mesmo, de tal forma que se houver um incremento de ações motivadoras que impactam sobre a meta, a variável correspondente comportará uma expectativa positiva e conseqüentemente redução no turnover. Rever os pesos da remuneração variável de acordo com a real importância para o negócio se apresenta também como medida de baixa rotatividade. Para selecionar o modelo mais parcimonioso foi utilizado o critério de Backward com a estatística de Akaike e para avaliar a qualidade do ajuste, foi realizado o Teste de Hosmer-Lemeshow.

Palavras-chave: Modelo de Regressão Logística, Inferência, Método de Backward, Turnover.

ABSTRACT

The objective of this study is to verify through logistic regression which factors may influence the increase of the turnover rate of a company's employees. The data comes from a company that makes telephone sales in Uberlândia-MG. We analyzed 264 telesales operators of both sexes and ages between 18 and 48 years old and also 18 behavioral and performance variables of the sellers that represent possible causes that can impact for the increase of turnover. Of all the variables evaluated, five were significant for the prediction of turnover ("Proportion of goals hit", "Number of times the customer goal was beaten", "Quantity of returns in the evaluated period", "Number of absences in the period However, the variable "PropMetaBatida" has a significant impact on the reduction of the variable, so that if there is an increase in motivating actions that impact the target, the corresponding variable will have a positive expectation and consequently reduction in turnover. Reviewing the variable remuneration weights according to the real importance for the business is also presented as a measure of low turnover. To select the most parsimonious model, the Backward criterion was used with the Akaike statistic and to evaluate the quality of fit, the Hosmer-Lemeshow test was performed.

Keywords: Logistic Regression Model, Inference, Backward Method, Turnover..

SUMÁRIO

Lista de Tabelas	15
1 Introdução	1
2 Metodologia	3
2.1 FORMULAÇÃO MATEMÁTICA	4
2.1.1 Regressão Logística Simples	4
2.1.2 Regressão Logística Múltipla	7
2.1.3 Estimação dos parâmetros	7
2.1.4 Adequação do modelo	8
2.1.5 Multicolinearidade	9
2.1.6 Seleção de variáveis	10
3 Resultados	11
4 Conclusões	15
Referências Bibliográficas	17

LISTA DE TABELAS

3.1	Valores do VIF das variáveis.	11
3.2	Resultados das medidas descritivas em relação ao gênero.	12
3.3	Estatísticas do ajuste do modelo final.	12
3.4	Valores do VIF das variáveis do modelo final.	12

1. INTRODUÇÃO

Desenvolver uma boa equipe de trabalho é um grande desafio e exige muita atenção das empresas. Contratar uma pessoa que não tem o perfil para determinada posição pode causar impactos aos negócios. Por outro lado, perder um bom colaborador é prejudicial para a empresa, em especial quando se trata de bons funcionários [22].

Um dos setores que merece atenção é o comércio varejista brasileiro, pois além de trabalhar com a margem de lucro no limite, enfrenta um novo desafio, a rotatividade de colaboradores. A oferta no setor é ampla, porém devido às características deste ramo (trabalho por escala com finais de semana comprometidos, baixos salários, dentre outros) a “dança” de funcionários é constante nesse setor. O índice de turnover é o termo utilizado para definir o fluxo de admissões e demissões dentro de uma empresa. No entanto, no presente trabalho não será considerado o índice de turnover como variável resposta (variável dependente), ou seja, será considerado como resposta a variável binária, sendo 1 para desligamento e 0 caso contrário [21].

Para [10, p. 87] a definição de rotatividade é “o fluxo de entrada e saída de pessoas em uma organização, ou seja, as entradas para compensar as saídas das pessoas nas organizações.” De tal forma que compreenda que as entradas correspondem às admissões e as saídas às demissões. O turnover pode ser explicado por fatores internos e externos. Destaca-se como fatores internos: política de cargos e salários, benefícios, oportunidades de crescimento, liderança, dentre outros. Já os fatores externos podem-se destacar: conjuntura econômica, nível de oferta de emprego no mercado, cultura e clima organizacional, dentre outros [9].

Além de retratar a rotatividade no quadro de pessoal da empresa, o turnover compromete também seus resultados, ou seja, o retorno do investimento de capital na organização, levando em consideração as oscilações no nível de produtividade, os desperdícios, os altos custos nos processos de demissões e admissões [19].

O turnover se apresenta como vantajoso e desvantajoso para a organização. É positivo ou vantajoso quando as pessoas que saem possuem baixo desempenho; quando a contratação de novos funcionários com ideias inovadoras; se for pela contratação de novos funcionários com menores salários e benefícios e quando representam oportunidade de promoção para outros funcionários. Ao contrário, se apresenta como negativo ou desvantajoso quando: apresentam altos custos com recrutamento e seleção; gastam tempo e dinheiro com a formação de novos funcionários e outros funcionários podem levar informações sobre o negócio para outras empresas [16].

A elevada rotatividade pode prejudicar qualquer atividade, de qualquer organização, mesmo

aquelas de grande porte e consideradas sólidas. Outro fator a ressaltar é que a produtividade cai devido à curva de aprendizado, ou seja, a produção de um novo funcionário não é a mesma pelo fato da compreensão do trabalho e da organização. Alguns problemas podem ser ressaltados como consequências ao turnover: perda de talentos para a concorrência; sobrecarga de trabalho para aqueles que ficam; clima organizacional (pode gerar conflitos); “envilecimento” da marca; perda de competitividade no mercado. Contudo faz-se necessário trabalhar para reduzir o turnover, pois é imprescindível para obter melhores resultados nos empreendimentos e preservar a imagem no mercado [17].

Assim, nos últimos anos houve grande transformação no modo de gerir pessoas, saindo de uma supervisão extremada para um desenvolvimento recíproco entre empresa e indivíduos. As pessoas passaram a ser gerenciadoras do seu crescimento aprimorando o capital intelectual como respostas às necessidades impostas pelo mercado à organização [12].

A necessidade de diferenciar as empresas, visando proporcionar atendimento à diversos mercados, conduz as organizações a um processo de criação e inovação, investindo cada vez mais em processos de aprendizagem e ferramentas, transformando-as em empresas inteligentes. Assim as organizações se beneficiam incorporando novos serviços e grandes ideias, produzindo novos conhecimentos que aliados à estratégia da organização, objetivam maior competitividade no mercado, maximizando resultados [20]

Nesse sentido, prever o desligamento de funcionários por meio da regressão logística representa um estudo de alta relevância para as organizações como um todo. Segundo [14, p. 539] a técnica de Regressão Logística foi desenvolvida na década de 1960, devido à necessidade de se realizar previsões ou explicar alguns acontecimentos em que a variável dependente fosse binária. No entanto, para fazer a previsão de desligamento dos funcionários é necessário identificar quais fatores que influenciam para o aumento do Turnover. Logo, este trabalho tem como objetivo específico realizar uma análise de funcionários em um setor de tele vendas por meio de modelos de regressão logística com a finalidade de encontrar, descrever e prever o risco de desligamento dos colaboradores na empresa (variável dependente) em relação a outras variáveis independentes como idade, escolaridade, quantidade de faltas, devoluções de vendas, entre outros indicadores sociais e de performance com a finalidade de auxiliar os gestores na tomada de decisão em contratações e investimento.

2. METODOLOGIA

No presente trabalho foram utilizados dados de uma empresa de telemarketing da cidade de Uberlândia – MG no período de 2014 a 2016 considerando apenas operadores com no mínimo 3 meses de empresa, totalizando em 264 observações. As variáveis analisadas foram:

y: Representa o desligamento do funcionário (variável dependente - variável binária), se desligado = 1, se não desligado = 0;

x1: PropMetaBatida, é a quantidade de vezes em que o funcionário bateu meta de faturamento, dividido pela quantidade de meses trabalhados;

x2: PropCli, é a quantidade de vezes em que o funcionário bateu meta de clientes, dividido pela quantidade de meses trabalhados;

x3: PropSku, é a quantidade de vezes em que o funcionário bateu a meta de sku(itens), dividido pela quantidade de meses trabalhados;

x4: PropFaltas, é a quantidade de meses em que o funcionário teve ao menos uma falta sem justificativa, dividido pela quantidade de meses trabalhados;

x5: PropDevolucoes, é a quantidade de meses em que o funcionário teve ao menos uma devolução, dividido pela quantidade de meses trabalhados;

x6: PlanZerada, é a soma dos meses em que o funcionário teve planilha zerada pela equipe de qualidade, dividido pela quantidade de meses trabalhados;

x7: TempoCasa, é o tempo que o funcionário tem de empresa em meses;

x8: PropElegivel, é a quantidade de vezes em que o funcionário foi considerado elegível para receber remuneração variável, dividido pela quantidade de meses trabalhados;

x9: AvaliaçõesFB, é a média de notas das avaliações feedbacks realizados no período avaliado;

x10: Sexo, se Masculino = 0 e se Feminino = 1;

x11: Idade, em anos de completos;

x12: EnsinoSuperior, se completo = 1, se incompleto = 0;

x13:MesesTrabalhados, é a quantidade de períodos em que o funcionário teve atuação dentre os 25 meses de avaliação;

x14: BateuFat, é a soma da quantidade de vezes em que o vendedor bateu meta de faturamento;

x15: ElegivelVariavel, é a quantidade de vezes em que o operador bateu o índice que garante que ele receba variável, esse é composto pelo percentual de faturamento, cliente, sku e qualitativo;

x16: DevoluçõesContador, é a soma da quantidade de meses em que o vendedor teve ao menos uma devolução que foi erro dele;

x17: FaltasContador, é a soma da quantidade de meses em que o operador teve ao menos uma falta sem justificativa;

x18: BateuCli, é a soma da quantidade de meses em que o operador bateu meta de clientes;

x19: BateuSku, é a soma da quantidade de meses em que o operador bateu meta de itens.

Adotou-se como metodologia de pesquisa a Análise de Regressão Logística para identificar possíveis fatores que possuem relação com a variável resposta.

2.1 FORMULAÇÃO MATEMÁTICA

2.1.1 REGRESSÃO LOGÍSTICA SIMPLES

Considerando o modelo de regressão linear simples, temos:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad (2.1)$$

em que: Y_i é o fenômeno em estudo (variável dependente); β_0 representa o intercepto (constante); β_1 é o coeficiente associado a variável independente (coeficiente angular); x_i é a variável explicativa (independente) com $i = 0, 1, 2, \dots, n$; ε_i é o termo do erro. O erro ε_i , também chamado de resíduo, representa possíveis variáveis que não foram inseridas no modelo, mas que também contribuiriam para a explicação de Y_i em que [4]:

$$\varepsilon_i \sim^{iid} N(0, \sigma^2) \quad (2.2)$$

ε_i segue distribuição Normal com média 0 e variância σ^2 .

O modelo de regressão logística é semelhante ao modelo de regressão linear. Porém, no logístico a variável dependente Y_i é binária, e pode assumir os seguintes resultados possíveis: $Y_i = 0$ e $Y_i = 1$, conhecidos normalmente como “fracasso” e “sucesso”, respectivamente. Conforme a expressão 2.2, tem-se que $E(\varepsilon_i) = 0$. Além disso em problemas de regressão modela-se a média condicional, que é o valor médio da variável resposta Y dados os valores da variável independente, x_i , designada por $E(Y_i | X = x_i)$, cujo valor será:

$$E(Y_i | x_i) = \beta_0 + \beta_1 x_i \quad (2.3)$$

Contudo, na regressão logística, pode-se afirmar que a variável aleatória Y , tem distribuição Bernoulli $(1, \pi_i)$ pois, assume dois resultados possíveis (sucesso e fracasso) $P[Y_i = 1 | x_i] = \pi_i$, com probabilidade de sucesso e probabilidade de fracasso. Assim, por meio da definição de Esperança Matemática, tem-se que:

$$E(Y_i | x_i) = 0 \cdot (1 - \pi_i) + 1 \cdot \pi_i = \pi_i \quad (2.4)$$

Igualando as expressões 2.3 e 2.4:

$$E(Y_i | x_i) = \beta_0 + \beta_1 x_i = \pi_i \quad (2.5)$$

No entanto, a expressão 2.5 viola as suposições do modelo linear:

(i) Violação do pressuposto de normalidade dos resíduos, pois:

- se $y_i = 1$, então, $\varepsilon_i = 1 - \beta_0 + \beta_1 x_i$
- se $y_i = 0$, então, $\varepsilon_i = 0 - \beta_0 + \beta_1 x_i$

Portanto, não faz sentido assumir que os resíduos seguem distribuição normal.

(ii) Violação do pressuposto de variância constante dos resíduos, também conhecido como pressuposto de homogeneidade de variância dos resíduos.

Pode-se afirmar que $Y_i = \pi_i + \varepsilon_i$, em que a quantidade ε_i admite o valor $1 - \pi_i$ para $Y_i = 1$, ou $-\pi_i$ para $Y_i = 0$. Assim ε_i segue uma distribuição Bernoulli com média zero e variância igual a $\pi_i(1 - \pi_i)$, que podem ser verificadas. Sendo $P(\varepsilon_i = -\pi_i) = 1 - \pi_i$ e $P(\varepsilon_i = 1 - \pi_i) = \pi_i$, então o valor esperado de ε_i é [13]:

$$E(\varepsilon_i) = -\pi_i(1 - \pi_i) + (1 - \pi_i)\pi_i = -\pi_i + \pi_i^2 + \pi_i - \pi_i^2 = 0. \quad (2.6)$$

E a variância de ε_i é:

$$\begin{aligned} Var(\varepsilon_i) &= E(\varepsilon_i^2) - [E(\varepsilon_i)]^2 = E(\varepsilon_i^2) = (-\pi_i)^2(1 - \pi_i) + (1 - \pi_i)^2\pi_i \\ &= (-\pi_i)^2(1 - \pi_i) + (1 - \pi_i)^2\pi_i = \pi_i^2 - \pi_i^3 + (1 - 2\pi_i + \pi_i^2)\pi_i \\ &= \pi_i^2 - \pi_i^3 + \pi_i - 2\pi_i^2 + \pi_i^3 = \pi_i - \pi_i^2 \\ &= \pi_i(1 - \pi_i). \end{aligned} \quad (2.7)$$

No entanto, com base na expressão 2.5 tem-se:

$$Var(\varepsilon_i) = \pi_i(1 - \pi_i) = (\beta_0 + \beta_1 x_i)(1 - \beta_0 + \beta_1 x_i) \quad (2.8)$$

Portanto, observa-se que a variância dos resíduos depende de x_i . Logo, a não é constante.

(iii) Restrição para a resposta média $E(Y_i | x_i)$.

Conforme expressão 2.5 temos:

$$E(Y_i | x_i) = \beta_0 + \beta_1 x_i = \pi_i \quad (2.9)$$

Porém, como π_i é uma probabilidade, seus valores variam entre 0 e 1. Logo, $0 \leq \beta_0 + \beta_1 x_i \leq 1$. No entanto, é uma restrição inapropriada para um modelo linear, que assume valores reais, isto é, no intervalo $(-\infty, \infty)$. Uma alternativa seria [15]:

$$\pi_i = \pi(x_i) = e^{\beta_0 + \beta_1 x_i} \quad (2.10)$$

o que assegura que π_i seja positiva, mas por outro lado, o termo $e^{\beta_0 + \beta_1 x_i}$ pode gerar um valor maior que 1. Assim, para satisfazer essa restrição, ajusta-se um modelo do tipo:

$$\pi_i = \pi(x_i) = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + (e^{\beta_0 + \beta_1 x_i})} \quad (2.11)$$

Em que a equação 2.11, conhecida com função logística, não assume valores negativos ou maiores do que 1 satisfazendo as restrições requeridas [15].

2.1.2 REGRESSÃO LOGÍSTICA MÚLTIPLA

Quando há interesse e/ou necessidade em incorporar mais de uma variável independente no estudo, o modelo de regressão logístico pode ser generalizado.

Dado um conjunto com p variáveis independentes, em que o vetor da i -ésima linha da matriz (X) das variáveis explicativas é designado por: $(x_{i0}, x_{i1}, \dots, x_{ip})$, , onde cada elemento da matriz corresponde ao ij -ésimo componente , sendo que $i = 1, \dots, n$ e $j = 1, \dots, p$, com $(x_{i0}) = 1$. Tem-se que o vetor de parâmetros desconhecidos é dado por $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$ e β_j é o j -ésimo parâmetro associado à variável explicativa x_j . Assim, a probabilidade de sucesso no modelo de regressão múltipla passa a ser expressa por [18]:

$$\pi_i = \pi(x_i) = P(Y_i = 1|X = x_i) = \frac{e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}}}{1 + (e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}})} = \frac{\exp(x_i^T \beta)}{1 + \exp(x_i^T \beta)} \quad (2.12)$$

E a probabilidade de fracasso por

$$1 - \pi_i = 1 - \pi(x_i) = P(Y_i = 0|X = x_i) = \frac{1}{1 + (e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}})} = \frac{1}{1 + \exp(x_i^T \beta)} \quad (2.13)$$

2.1.3 ESTIMAÇÃO DOS PARÂMETROS

Para ajustar o modelo de regressão logística deve-se estimar os parâmetros β_0 e β_1 por meio do método da máxima verossimilhança. Os estimadores de máxima verossimilhança de β_0 e β_1 são os valores de $\hat{\beta}_0$ e $\hat{\beta}_1$ que maximizam o logaritmo da função de verossimilhança, Conforme expressão 2.17.

Considere uma amostra independente (x_i, y_i) de tamanho n , em que y_i é o valor da variável resposta dicotômica. Sendo que $Y_i = Ber(\pi_i)$, a distribuição de probabilidade é dada por [2]:

$$f(y_i, \pi_i) = \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \quad (2.14)$$

A transformação logit, cujo objetivo é linearizar o modelo aplicando o logaritmo, resulta em [8]:

$$g(x_i) = \ln \left[\frac{\pi(x_i)}{1 - \pi(x_i)} \right] = x_i^T \beta = \beta_0 + \sum_{j=1}^p \beta_j \pi_{ij} \quad (2.15)$$

Esta transformação assume especial importância pois o modelo com esta transformação atende a condição de que a variável resposta pertence à $\mathbb{R}(-\infty, \infty)$ na relação linear.

Admitindo-se que as observações y_i são independentes, a função de máxima verossimilhança é dada por:

$$L(\beta) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \quad (2.16)$$

em que $\beta = (\beta_0, \beta_1)^T$ é o vetor de parâmetros.

Tendo como logaritmo da função de verossimilhança:

$$l(\beta) = \sum_{i=1}^n [y_i x_i^T \beta - \ln(1 + \exp(x_i^T \beta))] \quad (2.17)$$

2.1.4 ADEQUAÇÃO DO MODELO

Após o ajuste do modelo é necessário verificar se de fato representa o que se deseja medir, isto é, o modelo expressa a realidade e, conseqüentemente, proporciona uma melhoria nas decisões [14].

No presente trabalho foram adotados os seguintes critérios para avaliação do modelo:

- (i) o teste de Hosmer e Lemeshow para verificar a adequação do modelo;
- (ii) Critério de Informação de Akaike (AIC) para selecionar o melhor modelo ajustado;
- (iii) o teste de Wald para determinar a significância dos parâmetros.

O teste de Hosmer e Lemeshow consiste em dividir as observações em grupos e compará-las em relação as frequências preditas com as observadas. Sendo a hipótese H_0 de que não existem diferenças entre os valores preditos e observados, a um nível de significância α , ou seja, o modelo proposto ajusta-se bem aos dados [18].

O AIC é utilizado para selecionar ou definir qual é o modelo mais adequado dentre os outros. Classifica-se como o melhor modelo ajustado aquele que obtiver o menor valor de AIC. [6] define AIC como:

$$AIC = -2l(\beta) + 2k \quad (2.18)$$

em que $l(\beta)$ é o logaritmo da função de verossimilhança do modelo e k é a quantidade de parâmetros.

A estatística de Wald permite medir a significância dos parâmetros estimados no modelo ajustado, ou seja, verifica se cada estimativa é significativamente diferente de zero o que indica a permanência ou não de determinada variável no modelo. A estatística de Wald segue uma

distribuição Qui-quadrado e quando a variável dependente tem um único grau de liberdade pode ser calculada elevando-se ao quadrado a razão entre coeficiente que está sendo testado e o respectivo erro-padrão, conforma segue:

$$Wald = \left(\frac{\beta_j}{SE_{\beta_j}} \right)^2 \quad (2.19)$$

Em que β_j é o coeficiente de uma variável independente incluída no modelo; SE é o erro padrão [14].

2.1.5 MULTICOLINEARIDADE

Justifica-se investigar se há multicolinearidade entre as variáveis explicativas, visto que a forte correlação entre elas acarreta vários efeitos negativos no ajuste do modelo de regressão. O problema de multicolinearidade torna a estimativa dos parâmetros imprecisa, por conta de um alto valor do erro padrão, o que não é conveniente estatisticamente [3] [11].

A presença de multicolinearidade pode ser observada quando o coeficiente de determinação R_i^2 apresenta um alto valor, mas nenhum dos coeficientes da regressão é estatisticamente significativo. Portanto, uma das formas de detecção é avaliar o Fator de Inflação da Variância (VIF). Esse fator mede o grau de associação entre as variáveis explicativas, a partir do coeficiente de determinação do modelo de regressão ajustado apenas entre as variáveis independentes. O Fator de Inflação da Variância é definido como [7]:

$$VIF_i = \frac{1}{1 - R_i^2}, \quad (2.20)$$

em que: R_i^2 é o coeficiente de determinação da regressão da variável explicativa X_i sobre as outras variáveis explicativas com $i = 1, 2, \dots, k$, sendo k a quantidade de variáveis explicativas no modelo.

Pode-se observar que, quanto maior o R_i^2 , maior é o valor de VIF, indicando alta colinearidade.

Valores de VIF_i maiores que 10 correspondem a um coeficiente de determinação $R_i^2 > 0,90$ [3][11]. Porém, conforme recomendado pela NBR 14653-2 item A.2.1.5.2 vigente, onde a correlação múltipla superior a 0,80 deve ser verificada especialmente. Logo, aconselha-se limitar a aceitação da variável a um VIF até 5. Logo, $VIF = 2$ significa a não existência de colinearidade entre as variáveis [5]. Dessa forma, justifica-se descartar as variáveis com valores de VIF_i maiores que 5.

2.1.6 SELEÇÃO DE VARIÁVEIS

No presente trabalho ajustou-se um modelo de regressão logística múltipla selecionando as variáveis que irão compor o modelo por meio do critério de *Backward*. Este procedimento caracteriza-se por incorporar, inicialmente, todas as variáveis auxiliares no modelo e percorrer etapas, nas quais uma variável por vez pode vir a ser eliminada. Se em uma dada etapa não houver eliminação de alguma variável, o processo é então interrompido e as variáveis restantes definem o modelo final [23]. A estatística adotada pelo critério de *Backward* para remover uma variável foi o critério de informação de Akaike (AIC).

Para avaliar se o modelo proposto se ajusta bem aos dados utilizou-se o teste de Hosmer e Lemeshow e para aferir o grau de significância de cada coeficiente da equação de regressão logística utilizou-se a estatística de Wald [14].

Os resultados dos testes são apresentados na seção seguinte. Cabe ressaltar que todas as análises estatísticas foram realizadas no software R [1].

3. RESULTADOS

Inicialmente, para selecionar as variáveis a compor o modelo de regressão logística múltipla avaliou-se a multicolinearidade por meio do VIF.

Na Tabela 3.1 são apresentados valores de VIF para cada variável.

Tabela 3.1: Valores do VIF das variáveis.

Variável	VIF
PropDevolucoes(x_5)	868,4360
BateuFat(x_{14})	158,8238
PropCli(x_2)	42,3646
PropSku(x_3)	32,8228
PropFaltas(x_4)	28,2939
PlanZerada(x_6)	16,3475
TempoCasa(x_7)	10,2315
MesesTrabalhados(x_{13})	9,9541
ElegivelVariavel(x_{15})	9,6842
PropElegivel(x_8)	9,4537
AvaliacoesFB(x_9)	4,8742
EnsinoSuperior(x_{12})	3,7639
Idade(x_{11})	3,4571
BateuSku(x_{19})	3,0928
PropMetaBatida(x_1)	2,0376
BateuCli(x_{18})	1,9762
DevolucoesContador(x_{16})	1,2516
FaltasContador(x_{17})	1,0312
Sexo(x_{10})	1,0532

Conforme resultados apresentados na Tabela 3.1, apenas as variáveis com $VIF < 5$, foram consideradas para o ajuste do modelo, sendo elas: AvaliacoesFB, EnsinoSuperior, Idade, BateuSku, PropMetaBatida, BateuCli, DevolucoesContador, FaltasContador e Sexo.

Na Tabela 3.2, são apresentadas medidas descritivas das variáveis selecionadas no modelo em relação ao gênero.

Tabela 3.2: Resultados das medidas descritivas em relação ao gênero.

Gênero	Variável	Mínimo	Máximo	Média	Mediana	Desvio Padrão
Masculino	PropMetaBatida(x_1)	0	0,917	0,2320	0,1810	0,2345
	BateuCli(x_{18})	0	11	2,4821	1,5	2,8603
	DevolucoesContador(x_{16})	0	5	0,9642	1	1,1749
	FaltasContador(x_{17})	0	9	2,4821	2	2,1657
Feminino	PropMetaBatida(x_1)	0	1	0,2830	0,2500	0,2762
	BateuCli(x_{18})	0	20	3,7871	3	4,0665
	DevolucoesContador(x_{16})	0	7	1,1386	1	1,4491
	FaltasContador(x_{17})	0	9	2,0940	2	1,9434

Com base nos resultados apresentados acima, o que mais chama atenção é a diferença na mediana da variável “PropMetaBatida” e “BateuCli”, onde o gênero feminino se destaca em relação ao masculino.

Na Tabela 3.3 são apresentadas as estatísticas referentes ao ajuste do modelo final para a predição da probabilidade de desligamento dos funcionários.

Tabela 3.3: Estatísticas do ajuste do modelo final.

Coefficientes	Estimativa	Erro Padrão	Wald	ODDS	valor p	Valor-p Hosmer& Lemeshow
Intercepto	-1,0722	0,4103	6,8289		0,0089	0,8598
PropMetaBatida(x_1)	-3,9213	1,0180	14,8376	0,0198	0,0001	
BateuCli(x_{18})	0,6045	0,0961	39,5682	1,8303	<0,0000	
DevolucoesContador(x_{16})	0,4845	0,1796	7,2774	1,6234	0,0069	
FaltasContador(x_{17})	0,2454	0,0970	6,4004	1,2782	0,0113	
Sexo(x_{10})	-0,8870	0,3990	4,9420	0,4118	0,0262	

Na Tabela 3.4 são apresentados valores de VIF para cada variável explicativa que foi significativa para o ajuste do modelo.

Tabela 3.4: Valores do VIF das variáveis do modelo final.

Variável	VIF
PropMetaBatida(x_1)	2,0376
BateuCli(x_{18})	1,9762
DevolucoesContador(x_{16})	1,2516
FaltasContador(x_{17})	1,0312
Sexo(x_{10})	1,0532

Podemos observar que os valores do VIF apresentados na Tabela 3.4, estão abaixo do valor limite que é 5, logo as estimativas do modelo não foram afetadas pela multicolinearidade.

Com base na Tabela 3.3, ao nível de significância de 5%, todos os parâmetros do modelo estimado são significativos. A variável “PropMetaBatida” exerce um negativo no modelo, ou seja, quanto mais o funcionário bate metas, maior será a probabilidade de continuar trabalhando na empresa. A variável “Sexo” também possui a estimativa do parâmetro negativa, porém por se tratar de uma variável dicotômica, conclui-se que o indivíduo do sexo feminino, comparado ao sexo masculino, possui maior probabilidade de continuar trabalhando. As variáveis “BateuCli”, “DevolucoesContador” e “FaltasContador” exercem efeitos positivos no modelo, ou seja, quanto maior a quantidade planilhas zeradas, devoluções e Faltas, maior é a probabilidade do funcionário ser desligado. Um ponto importante foi a variável “BateuCli”, com um peso negativo, o que é coerente com a prática, pois quando novos operadores entram para a operação, eles não possuem carteira e trabalham um mailing genérico, que é uma lista de clientes que estão sem comprar há algum tempo e por isso estão disponíveis para serem acionados por qualquer operador, no entanto são clientes que possuem baixo volume de compra e não possuem relacionamento forte com a empresa, assim os vendedores conseguem vender para vários clientes, conseguem bater a meta de clientes, porém essa tem um peso de apenas 10% no índice geral que é composto por faturamento(70%), cliente(10%), Sku(10%) e qualitativo(10%), assim por mais que tenham conseguido performar em clientes, eles não conseguem bater a meta geral e assim não recebem remuneração variável, por isso justifica a variável ter peso negativo.

Pelo teste de Hosmer e Lemeshow, não rejeita a hipótese de que não existem diferenças entre os valores preditos e observados, caracterizando dessa forma o bom ajuste do modelo nos dados.

Substituindo os coeficientes da equação do modelo de regressão logística, pelos coeficientes estimados, tem-se que a probabilidade de o funcionário continuar trabalhando é dada por:

$$\hat{\pi}_i = \frac{-1,0722 - 3,92134x_1 + 0,6045x_{18} + 0,4845x_{16} + 0,2454x_{17} - 0,8870x_{10}}{1 + \exp(-1,0722 - 3,92134x_1 + 0,6045x_{18} + 0,4845x_{16} + 0,2454x_{17} - 0,8870x_{10})} \quad (3.1)$$

Já a probabilidade de ele ser desligado é:

$$1 - \hat{\pi}_i = \frac{1}{1 + \exp(-1,0722 - 3,92134x_1 + 0,6045x_{18} + 0,4845x_{16} + 0,2454x_{17} - 0,8870x_{10})} \quad (3.2)$$

4. CONCLUSÕES

Neste trabalho com o uso da regressão logística foi possível evidenciar as principais variáveis que impactam no turnover de funcionários no setor de televendas na cidade de Uberlândia-MG. As variáveis “PropMetaBatida”, “BateuCli”, “DevolucaoContador”, “FaltasContador” e “Sexo”, foram significativas para a predição do turnover. Os resultados nos mostram que as variáveis “PropMetaBatida” e “Sexo” possuem impacto positivo para a redução do turnover, enquanto “BateuCli”, “DevolucaoContador” e “FaltasContador” afetam esse indicador de forma negativa, vale destacar uma atenção especial para “BateuCli”, uma vez que levando em consideração o conhecimento de negócio, era esperado que seu peso fosse positivo.

Evidenciou-se que o peso atribuído ao cliente (“BateuCli”) no setor de televendas analisado, é irrelevante se comparado por exemplo com o peso atribuído ao faturamento. O que contradiz a política de vendas, uma vez que sendo o cliente um fator relevante para o processo, deveria ser fundamental que o peso atribuído ao mesmo motivasse o crescimento da carteira, e conseqüentemente um maior faturamento.

Em relação aos pesos atribuídos às variáveis ficou perceptível que o trabalho realizado para os determinar necessita ser feito não só pelo setor em questão (televendas), e sim por um comitê que envolva planejamento e análise de outros fatores, como por exemplo (tecnologia utilizada, fator humano e sua satisfação, pesquisa de mercado, dentre outros), observando que o planejamento pode evitar eventuais imprevistos e obstáculos minimizado as variações no fechamento que incorreriam em elevação dos custos para a organização, onde pode ser ressaltado o turnover.

Para estudos futuros em torno da variável “PropMetaBatida”, cabe um plano de ação para tentar identificar formas (Treinamento em Produtos, Treinamento de Sistemas, Argumentos de Vendas, Pesquisa de Microclima) que melhorem a performance dos vendedores, pois batendo meta, existe uma grande chance do turnover ser reduzido. Outro ponto que ficou evidente foi acerca da variável “BateuCli”, que teoricamente deveria ser positivo, então a sugestão seria distribuir melhor os pesos e medir para avaliar a performance dos vendedores e o impacto disso na remuneração variável paga pela empresa.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] (2018), R. D. C. T.: *R: A language and environment for statistical computing*. Em *R Foundation for Statistical Computing*, 2018. <http://www.R-project.org/>.
- [2] ABREU, H. J.: *Aplicação da análise de sobrevivência em um problema de CreditScoring e comparação com a regressão logística*. Dissertação de Mestrado, 2004.
- [3] al., M. H. K. et: *Applied linear models*. New York: McGraw-Hill Irwin, 5^a ed., 2004.
- [4] al., R. C. et: *Análise de modelos de regressão linear: com aplicações*. Editora da Unicamp, 2^a ed., 2008.
- [5] ASSIS, M. L. S. B. M. M. C. M. C.: *Fator de inflação da variância e regressões auxiliares para diagnóstico do problema de multicolinearidade nos modelos de regressão*. , 2017. <https://ibape-nacional.com.br/biblioteca/wp-content/uploads/2017/08/025.pdf> , acessado em 24/01/2019.
- [6] BARRETO, T. E. L. S. G.: *Análise dos critérios de informação para a seleção de ordem em modelos auto regressivos*. 2011.
- [7] BERK, K. N.: *Tolerance and condition in regression computations*. Journal of the American Statistical Association, 360(72):863–866, 1977.
- [8] CABRAL, I. C.: *Aplicação do modelo de regressão logística num estudo de mercado*. Dissertação de Mestrado, 2013.
- [9] CHIAVENATO, I.: *Recursos Humanos: Capítulo 3*. Atlas, 4^a ed., 1997.
- [10] CHIAVENATO, I.: *Gestão de Pessoas: O novo papel dos recursos humanos na organização*. Elsevier, 2^a ed., 2004.
- [11] DUNLOP, A. T.: *Statistics and data analysis: From Elementary to Intermediate*. Prentice Hall, 2000.
- [12] DUTRA, J. S.: *Gestão por Competências: um modelo avançado para o gerenciamento de pessoas*. Editora Gente, p. 120, 2001.
- [13] FIGUEIRA, C. V.: *Modelos de regressão logística*. Dissertação de Mestrado, 2006.

- [14] FILHO(Coord.), S. L. C. E. P. D.: *Análise multivariada para cursos de administração, ciências contábeis e economia*. Atlas, 2007.
- [15] GAUVREAU, M. P. K.: *Princípios de Bioestatística*. Thomson, 2^a ed., 2006.
- [16] HARRIS, M.: *Human Resource Management: a practical approach*. Copyright: Orlando, 1997.
- [17] KENOBY.: *Turnover: Guia completo para reduzir a rotatividade de colaboradores.*, 2019. <http://www.kenoby.com/blog/turnover/>, acessado em 02/07/2019.
- [18] LEMESHOW, D. W. H. S.: *Applied Logistic Regression*. New York: John Wiley, 1989.
- [19] LUCENA, M. D.S.: *Planejamento de recursos humanos*. Atlas, 10^a ed., 2007.
- [20] NISEMBAUM, H.: *Gestão do Conhecimento: Enriquecendo o capital humano*. Em *Manual de Gestão de pessoas e equipes*, p. 688. Editora Gente, 2002.
- [21] SEBRAE: *Rotatividade de colaboradores é recorrente no varejo.*, 2016. <http://www.sebrae.com.br/sites/PortalSebrae/artigos/rotatividade-de-colaboradores-e-recorrente-no-varejo,ceb3438af1c92410VgnVCM100000b272010aRCRD>, acessado em 30/09/2018.
- [22] SEBRAE: *Saiba o que é turnover e entenda o impacto da rotatividade no negócio.*, 2016. <http://www.sebrae.com.br/sites/PortalSebrae/artigos/entenda-o-que-e-turnover-e-o-impacto-da-rotatividade-no-negocio,44e08fa0672f0510VgnVCM1000004c00210aRCRD>, acessado em 30/09/2018.
- [23] SMITH, N. D.: *Applied regression analysis*. John Wiley e Sons, 3^a ed., 1998.