

4

Aplicação: Modelo LOGIT para Avaliar o Risco de Crédito

4.1

Preparação de Dados

Foi utilizada uma base de dados de 156 clientes que tiveram o vencimento de seus títulos compreendidos no período de abril a outubro de 2005.

Os dados históricos destes clientes tiveram como fonte tanto o próprio sistema interno da empresa fornecedora, quanto relatórios de uma empresa de cadastro, a Serasa, especializada na elaboração de *rating* de empresas.

Como a fonte de dados externa (Serasa) não possui dados históricos, somente dados no presente, para a amostra foram selecionados clientes que tiveram transações recentes com a empresa, para que houvesse o menor espaço de tempo possível entre as informações históricas de pagamento e as do Serasa.

Baseado nos dados de pagamento destes clientes, foi definida a variável dependente *dummy*, Situação do Cliente (0 – adimplente e 1 inadimplente), de acordo com a seguinte classificação:

- Para clientes adimplentes – clientes com histórico de pagamentos em dia ou com o máximo de atraso em até 30 dias;
- Para clientes inadimplentes – clientes com histórico de pagamentos com atraso superior a 30 dias.

Foi realizado o corte em 30 dias em virtude de que os títulos após 30 dias vencidos, já foram cobrados internamente pela empresa e não tiveram êxito. Após este prazo, eles são enviados para uma empresa de cobrança externa, passando a ter ônus adicional referente ao valor de comissão cobrado por título recebido, ao gerenciamento desta carteira em cobrança externa, além de terem um maior risco de não recebimento, em virtude de já terem sido cobrados internamente.

Adicionalmente, nota-se que quanto maior o tempo de vencimento de um título, maior é a dificuldade em se obter o seu pagamento.

Por fim, foram levantadas as variáveis independentes que tinham possibilidade de explicar a correlação com a inadimplência. Foram listadas 13 variáveis independentes, conforme a seguir:

- i. Idade da empresa;
- ii. Quantidade de consultas por empresa de factoring/fomento mercantil;
- iii. Concentração das consultas de factoring/fomento mercantil;
- iv. Prazo médio de atraso;
- v. Período da última compra;
- vi. Quantidade de pendências vencidas em bancos e financeiras (REFIN);
- vii. Quantidade de pendências vencidas com empresas não-financeiras (PEFIN);
- viii. Quantidade de protestos;
- ix. Período dos protestos;
- x. Quantidade de cheques sem fundos e sustados;
- xi. Quantidade de ações judiciais;
- xii. Quantidade de parcelas por pedido do cliente;
- xiii. Prazo médio de pagamento do cliente;

As empresas de *factoring* e de fomento mercantil, utilizadas nas variáveis “ii” e “iii”, são empresas que prestam serviços principalmente relativos ao adiantamento do valor referente a títulos a receber. Normalmente apresentam taxas financeiras superiores aos bancos de varejo.

Para cada variável independente foram definidas faixas de valores (Tabela 3). Estas faixas foram criadas em virtude da grande dispersão dos dados e a partir

da suposição de que as empresas apresentam o mesmo comportamento em relação à pontualidade no pagamento ao apresentar informações compreendidas na mesma faixa.

Tabela 3 – Relação das Variáveis Independentes

Variável Independente	Fonte	Faixas por Variável
Idade da empresa	Serasa	1 - até 5 anos // 2- de 6 a 10 anos // 3- de 11 a 20 anos // 4- de 21 a 40 anos // 5 - mais de 41 anos
Quantidade consultas empresas factoring ou fomento mercantil	Serasa	1 - 0 consultas // 2- 1 consulta // 3- 2 consultas // 4- 3 consultas // 5 - 4 consultas // 6 - 5 consultas
Concentração das consultas de empresas de factoring ou fomento mercantil	Serasa	1- a mais recente no último mês // 2- a mais recente de 1 a 2 meses atrás // 3 - a mais recente de 3 a 6 meses atrás // 4 - a mais recente de 7 a 12 meses atrás // 5 - a mais recente com mais de 12 meses // 6 - não há
Prazo Médio de atraso	Serasa	1 - de 0 a 5 dias // 2- de 6 a 10 dias // 3- de 11 a 20 dias // 4- de 21 a 40 dias // 5 - mais de 41 dias
Período da última compra a qualquer fornecedor	Serasa	1- no último mês // 2- de 1 a 2 meses atrás // 3 - de 3 a 6 meses atrás // 4 - de 7 a 12 meses atrás // 5 - mais de 12
Quantidade de pendências vencidas em bancos e financeiras (REFIN)	Serasa	1 - zero // 2- de 1 a 2 // 3- de 3 a 5 // 4- de 6 a 10 // 5 - de 11 a 20 // 6 - mais de 21
Quantidade de pendências vencidas informadas por empresas não financeiras (PEFIN)	Serasa	1 - zero // 2- de 1 a 2 // 3- de 3 a 5 // 4- de 6 a 10 // 5 - de 11 a 20 // 6 - mais de 21
Quantidade de protestos	Serasa	1 - zero // 2- de 1 a 2 // 3- de 3 a 5 // 4- de 6 a 15 // 5 - de 16 a 30 // 6 - de 31 a 50 // 7 - mais de 51
Período em que abrangem os protestos	Serasa	1- no último mês // 2- de 1 a 2 meses atrás // 3 - de 3 a 6 meses atrás // 4 - de 7 a 12 meses atrás // 5 - mais de 12 meses atrás // 6 - não há
Quantidade de cheques sem fundos/sustados	Serasa	1 - zero // 2- de 1 a 2 // 3- de 3 a 5 // 4- de 6 a 15 // 5 - de 16 a 30 // 6 - de 31 a 50 // 7 - mais de 51
Quantidade de ações judiciais	Serasa	1 - zero // 2- 1 ação // 3- 2 ações // 4- de 3 a 5 ações // 5 - mais de 6 ações
Quantidade de Parcelas por Pedido do Cliente	Empresa	1 - 1 parcela // 2 - 2 parcelas // 3 - 3 parcelas // 4 - 4 parcelas // 5 - 5 parcelas // 6 - mais de 6 parcelas
Prazo Médio de Pagamento do Cliente	Empresa	1 - até 20 dias // 2 - 21 a 30 dias // 3 - 31 a 45 dias // 4 - 46 a 60 dias // 5 - 61 a 80 dias // 6 - mais de 81 dias

Para cada cliente da amostra, foi feito o levantamento das variáveis independentes, com a sua respectiva faixa, e a variável dependente associada (0 ou 1). A relação das observações encontram-se no Anexo 1.

De acordo com Hair (p.219-220, 2005), o tamanho da amostra deve seguir os seguintes padrões:

- Proporção Mínima: 5 observações para cada variável independente. Como no modelo foram utilizadas 13 variáveis independentes, deveria-se ter no mínimo 65 observações (5 x 13 variáveis independentes) ;
- Mínimo de 20 observações por grupo.

Utilizou-se o *software* SPSS Versão 11.5 para a criação do modelo. Após a colocação dos dados no software foram definidas aleatoriamente as observações que seriam utilizadas para a criação do modelo e para a sua validação.

Para a obtenção aleatória dos dois grupos foi criada uma variável de seleção, “SELEC 70”, dividindo em grupos com 70% e 30% das observações.

A amostra foi distribuída conforme a Tabela 4.

Tabela 4 – Distribuição da Amostra

		N	Percentual
Casos Selecionados (Modelo)	Incluso na Análise	91	58,3
	Casos sem Dados	12	7,7
	Total	103	66,0
Casos Não-Selecionados (Validação)		53	34,0
Total		156	100,0

Vale ressaltar que, a amostra de 156 observações, quando particionada em amostras de análise e de teste (validação), atende a proporção mínima (5 observações para 1 variável), fornecendo uma razão de aproximadamente 8 para 1 (103 observações para 13 variáveis independentes potenciais) na amostra de análise.

Embora esta proporção cresça para 12 para 1 se a amostra não for dividida, considera-se mais importante validar os resultados com dados não utilizados para a criação do modelo, do que aumentar o número de observações na amostra de análise.

Além disso, os grupos excedem ao tamanho mínimo de 20 observações por grupo.

Na etapa de análise de dados observa-se a ocorrência de 12 *missing cases*, correspondendo a 7,7% da amostra.

Adicionalmente, foi feita uma análise dos dados baseada na estatística descritiva (Anexo 2).

Então, foram geradas informações para os dados referentes a cada variável, como: média, moda, mediana, desvio padrão, variância, assimetria e curtose.

Observou-se que para algumas variáveis, como: quantidade consultas por empresas de factoring, prazo médio de atraso, período da última compra, quantidade REFIN, quantidade PEFIN, quantidade de protestos, quantidade de cheques sem fundos, quantidade de ações judiciais e quantidade de parcelas, a média se encontrava superior a mediana, apresentando uma assimetria positiva.

Isto implica em dizer que nestes casos, mais de 50% dos dados presentes em cada variável se posicionam abaixo da média, conferindo-lhe uma certa tendência de superestimação dos dados.

Pelo coeficiente de curtose pode-se ter uma indicação da intensidade das frequências das observações nas vizinhanças dos valores centrais, da mediana, que divide os dados em 50% a esquerda e 50% a direita, e da média aritmética, caso esta última seja representativa dos valores centrais. Em termos gráficos, a curtose indica o grau de achatamento da curva da amostra.

Analisando os dados para a amostra, pôde ser identificado que a maioria das variáveis independentes apresentou dados de curtose muito distantes de 3 (característica de uma distribuição normal). A variável “Prazo Médio de Atraso” foi a que apresentou dado de curtose o mais próximo de uma distribuição normal, com 2,1, classificada como moderadamente platicúrtica. Esse distanciamento da distribuição normal confere uma menor previsibilidade, em termos de certeza probabilística, em relação ao comportamento dos dados no futuro.

Em virtude da característica das variáveis independentes não serem normalmente distribuídas foi descartado para a criação do modelo o emprego da ferramenta estatística análise discriminante.

4.2

Análise dos Dados

Nesta etapa, foi feita uma análise de correlação entre as variáveis independentes e uma análise entre cada variável independente e a variável dependente ($Y = \text{Situação do Cliente}$). (Anexo 3)

Essa análise de correlação permitiu identificar as variáveis independentes que apresentam uma grande correlação com a variável dependente, pois quanto maior a correlação entre a variável dependente e a independente, maior o poder discriminante entre os grupos, resultando em uma melhor explicação do modelo.

Foram identificadas como variáveis altamente correlacionadas: “Idade da empresa”, “Prazo médio de atraso”, “Período da Última Compra”, “Quant. REFIN”, “Quant. PEFIN”, “Quant. Protestos”, “Período dos Protestos”, “Quant. Cheques sem fundos” e “Prazo Médio Pagamento”. (Anexo 3)

Conclui-se que para esta amostra as demais variáveis não explicam a variável dependente “Situação do Cliente”, o que não significa necessariamente que na população elas não a expliquem.

Adicionalmente, foi realizada uma análise de correlação entre as variáveis independentes (Anexo 3). Através da análise de correlação entre as variáveis independentes, verificou-se aquelas que tinham maior correlação entre si.

Verifica-se que um modelo que apresenta variáveis independentes com alta correlação, ou seja a presença de multicolinearidade, tende a apresentar os betas superestimados.

Segundo Hair (p. 142, 2005),

“... a multicolinearidade refere-se à correlação entre entre três ou mais variáveis independentes. O impacto da multicolinearidade é reduzir o poder preditivo de qualquer variável independente na medida em que ela é associada com as outras variáveis independentes. Quando a colinearidade aumenta, a variância única explicada por cada variável independente diminui e o percentual da previsão compartilhada aumenta.”

Adicionalmente, de acordo com Hair (p. 165, 2005), a multicolinearidade resulta em partes maiores de variância compartilhada e níveis mais baixos de variância única, da qual os efeitos das variáveis independentes individuais podem ser determinados.

Concluindo, a existência da multicolinearidade deve ser observada, pois na sua existência não se deve interpretar os betas da equação isoladamente, pois há a tendência desses coeficientes estarem superestimados, em virtude de uma sobreposição do poder preditivo.

Observa-se na Tabela 5, um resumo das correlações entre variáveis independentes, com significância no nível de 0,01, ou seja, de 1% (bi-caudal).

Tabela 5 – Quadro Resumo de Correlação entre Variáveis

Variável Independente	Alta Correlação com:
Idade da empresa	Prazo médio de atraso , período da última compra, quant. Protestos , período protestos e quant. Cheques sem fundos.
Quantidade consultas empresas factoring ou fomento mercantil	Conc. Consultas de factoring, quant. REFIN e quant. Ações judiciais.
Concentração das consultas de empresas de factoring ou fomento mercantil	Quant. Cons. Factoring.
Prazo Médio de atraso	Idade empresa , per. Última compra, quant. REFIN, quant. Protestos , per. Protestos e quant. cheques sem fundos.
Período da última compra a qualquer fornecedor	Idade empresa, prazo medio de atraso, quant. REFIN, quant. Protestos, quant. Cheques sem fundos e prazo médio de pagamento.
Quantidade de pendencias vencidas em bancos e financeiras (REFIN)	Quant. Cons. Factoring, prazo médio de atraso, per. Última compra, quant. PEFIN, quant. Protestos, per. Protestos, quant. Cheques sem fundos e quant. Ações judiciais.
Quantidade de pendencias vencidas informadas por empresas não financeiras (PEFIN)	Quant. REFIN, quant. Protestos , per. Protestos, quant. Cheques sem fundos e quant. Ações judiciais.
Quantidade de protestos	Idade da empresa , prazo médio de atraso , período da última compra, quant. REFIN, quant. PEFIN , per. Protestos, quant. Cheques sem fundos e prazo médio de pagamento.
Período em que abrangem os protestos	Idade da empresa, prazo médio de atraso, período da última compra, quant. REFIN, quant. PEFIN, quant. Protestos e quant. Cheques sem fundos.
Quantidade de cheques sem fundos/sustados	Idade da empresa, prazo médio de atraso, período da última compra, quant. REFIN, quant. PEFIN, quant. Protestos e per. protestos.
Quantidade de ações judiciais	Quant. Cons. factoring, quant. REFIN e quant. PEFIN.
Quantidade de Parcelas por Pedido do Cliente	Prazo médio de pagamento.
Prazo Médio de Pagamento do Cliente	Per. Última compra, quant. Protestos e quant. Parcelas.

Após a análise de correlação, foi aplicada a ferramenta regressão logística.

A escolha do modelo de Regressão Logística (LOGIT) deveu-se à não normalidade da curva de distribuição do risco de crédito e a característica dicotômica da variável dependente desejada ($Y = 0$ ou $Y = 1$). Não foi utilizado o

modelo probit, modelo semelhante ao logit, por este pressupor que haja normalidade na curva de distribuição.

De acordo com Hair (p. 245, 2005), a logit é útil para situações nas quais se deseja prever a presença ou ausência de uma característica, ou resultado, baseado em valores das variáveis independentes. Pode ser utilizada, por exemplo, para se mensurar a probabilidade do risco de crédito em situações de operação de vendas a prazo, empréstimos ou financiamentos.

A probabilidade máxima pode ser estimada pela logit, após a transformação da variável dependente em variável de base logarítmica, permitindo que seja calculada a probabilidade de um certo evento acontecer.

Para os dados do estudo, foram gerados resultados para os seguintes métodos *Forward Stepwise: Conditional, Wald e Likelihood Ratio*. Esses métodos diferenciam entre si em função do critério empregado para orientar a entrada das variáveis independentes.

Para o método *Conditional* busca-se a maior probabilidade condicional, no método *Wald*, busca-se um maior coeficiente de *Wald* e no *Likelihood Ratio*, objetiva-se uma maior redução no valor $-2LL$ (“ $- 2 \log$ verossimilhança”).

Inicialmente foram gerados resultados para os três critérios acima para orientar a entrada das variáveis, sendo que os três métodos apresentaram o mesmo resultado final. As tabelas seguintes referentes ao *output* do SPSS, foram geradas pelo método *Forward Stepwise Likelihood Ratio*, que tem como objetivo reduzir o valor de $- 2LL$.

Pela Tabela 6 verifica-se que as variáveis foram selecionadas em 4 etapas e observa-se através da linha “Modelo” que o modelo apresenta adequação estatisticamente significativa, mesmo a um alfa de 1%, rejeitando a hipótese nula de que as variáveis independentes não explicam a variável dependente Y – “Situação do Cliente”.

Tabela 6 – Omnibus Tests of Model Coefficients

		Qui-quadrada	Grau de liberdade	Sig.
Passo 1	Passo	36,523	1	0,000
	Bloco	36,523	1	0,000
	Modelo	36,523	1	0,000
Passo 2	Passo	13,043	1	0,000
	Bloco	49,565	2	0,000
	Modelo	49,565	2	0,000
Passo 3	Passo	6,432	1	0,011
	Bloco	55,997	3	0,000
	Modelo	55,997	3	0,000
Passo 4	Passo	4,006	1	0,045
	Bloco	60,003	4	0,000
	Modelo	60,003	4	0,000

Adicionalmente, pode ser verificado na Tabela 7, que houve uma redução do valor de $-2 \text{ Log Likelihood}$, conforme foram sendo introduzidas as variáveis no modelo. Seu valor foi reduzido de 77,4 para 53,9. Este valor significa a medida geral de como o modelo se ajusta.

De acordo com Hair (p. 264, 2005), o $-2 \text{ Log Likelihood}$ é dado pelo valor de verossimilhança, que compreende -2 vezes o logaritmo do valor da verossimilhança e é chamado de -2 LL ou $-2 \text{ log verossimilhança}$, assim sendo, um modelo bem ajustado terá um valor pequeno para -2 LL , sendo o seu valor mínimo igual a zero.

Ao comparar o valor de -2 LL após a introdução de cada variável independente, verifica-se que a diferença representa a mudança no ajuste preditivo de uma equação para a outra.

Tabela 7 – Resumo do Modelo

Passo	$-2 \text{ Log likelihood}$	Cox & Snell R^2	Nagelkerke R^2
1	77,386	0,331	0,463
2	64,343	0,420	0,588
3	57,912	0,460	0,644
4	53,906	0,483	0,676

Adicionalmente, existem as medidas *Cox & Snell*, *Nagelkerke* e “pseudo” R^2 que também têm o intuito de apresentar o ajuste geral do modelo. Segundo Hair (p. 264, 2005), estas medidas de adequação de ajuste comparam as probabilidades estimadas com as probabilidades observadas, sendo que valores mais altos significam um melhor ajuste do modelo.

A medida R^2 Cox & Snell está limitada, no sentido em que não pode alcançar o valor máximo de 1; e, conseqüentemente, Nagelkerke propôs uma modificação que apresenta o domínio de 0 a 1.

No modelo gerado, obteve-se como valor de Cox & Snell, variando de 0,331 para 0,483 e de Nagelkerke, variando de 0,463 para 0,676. Observa-se que conforme foram incluindo as variáveis, o modelo foi melhor se ajustando, de forma a apresentar valores mais altos.

A terceira medida é o “pseudo” R^2 , baseado na melhora no valor de $-2LL$.

O “Pseudo” R^2 para o Modelo Logit, pode ser calculado como:

$$R^2 \text{ logit} = \frac{2LL_{\text{nulo}} - (-2LL_{\text{modelo}})}{-2LL_{\text{nulo}}}$$

Apresentando o seguinte valor para o modelo em questão:

$$R^2 \text{ logit} = (77,386 - 53,906) / (77,386) = 0,3034$$

Segundo Hair (p. 264, 2005), a medida final de ajuste do modelo é o valor de *Hosmer and Lemeshow Test*, o qual mede a correspondência entre valores reais e os previstos da variável dependente. Um bom ajuste de modelo é indicado por um valor qui-quadrado não-significante.

Tabela 8 – *Hosmer and Lemeshow Test*

Passo	Qui-quadrada	Grau de liberdade	Sig.
1	0,154	2	0,926
2	2,457	4	0,652
3	5,630	6	0,466
4	8,600	6	0,197

Após o teste de *Hosmer and Lemeshow*, é utilizado o Método de Matrizes de Classificação desenvolvido para a Análise Discriminante para avaliar a precisão preditiva do modelo de regressão logística.

A Matriz de Classificação (Tabela 9), mostra a classificação dos casos preditos pelo modelo comparando com os dados reais referentes à variável Y, Situação do Cliente.

Observa-se que o modelo apresenta melhor explicação para os clientes com a situação de solvente, com o percentual de acerto de 95,2% para a amostra de casos selecionados (amostra de análise), utilizada para gerar o modelo.

Considerando o modelo de uma forma geral, 90,1% dos clientes são classificados corretamente em um dos dois grupos, apresentando, a princípio, um expressivo percentual representativo de convergência.

Tabela 9 – Matriz de Classificação

Observação	Preditos							
	Casos Selecionados (Modelo)			Casos Não-Selecionados (Validação)				
	Situação Cliente		Percentual Correto	Situação Cliente		Percentual Correto		
	0,00	1,00		0,00	1,00			
Passo 1	Situação Cliente	0,00	54	8	87,1	30	1	96,8
		1,00	9	20	69	5	7	58,3
	Percentual Geral				81,3			86,0
Passo 2	Situação Cliente	0,00	58	4	93,5	31	0	100,0
		1,00	10	19	65,5	3	9	75,0
	Percentual Geral				84,6			93,0
Passo 3	Situação Cliente	0,00	59	3	95,2	31	0	100,0
		1,00	7	22	75,9	4	8	66,7
	Percentual Geral				89			90,7
Passo 4	Situação Cliente	0,00	59	3	95,2	31	0	100,0
		1,00	6	23	79,3	3	9	75,0
	Percentual Geral				90,1			93,0

Ao observar na Tabela 9 os dados referentes aos casos não selecionados para a criação do modelo e utilizados como amostra de validação, verifica-se que 100% dos casos foram enquadrados corretamente para a situação de cliente adimplente e 75% para a situação de cliente inadimplente, totalizando uma média de 93% de acerto geral do modelo.

Para identificar o percentual mínimo de classificação aceitável é necessário utilizar o critério de chance proporcional, que será calculado a seguir.

O critério de chance proporcional, conforme Hair (p. 241, 2005) leva em consideração a diferença dos tamanhos dos grupos analisados, sendo para o modelo em análise, os grupos de adimplentes ($Y=0$) e de inadimplentes ($Y = 1$). A fórmula para o critério de chance proporcional é:

$$C_{\text{PRO}} = p^2 + (1 - p^2)$$

Onde:

C_{PRO} = critério de chance proporcional

p = proporção de empresas no grupo 1

$1 - p$ = proporção de empresas no grupo 2

Para o modelo em questão, foram obtidas 65 observações para a situação adimplente e 26 para inadimplente, tendo como valor de chance proporcional $(0,714)^2 + (0,286)^2 = 0,592$.

Conforme Hair (p.242, 2005), o critério de chance máxima é simplesmente o percentual corretamente classificado se todas as observações fossem colocadas no grupo com maior probabilidade de ocorrência.

Como o grupo 1 ocorre 71,4% do tempo, estaria correto 71,4% do tempo se fossem designadas todas as observações a este grupo. No entanto, para a amostra de análise foi obtido na matriz de classificação um percentual de acerto de 95,2%, bem superior aos 71,4%.

A precisão de classificação de 90,1% do modelo é substancialmente maior que o critério de chance proporcional obtido (59,2%) e maior que o critério de chance máxima de 71,4%. Também excede a referência sugerida do valor mais 25%, o que neste caso estabelece o valor de referência em $71,4 \times 1,25 = 89,25\%$.

A tabela a seguir, Tabela 10, apresenta as variáveis presentes na equação, gerada pela ferramenta de regressão logística bem como os seus betas e a significância das variáveis independentes.

Tabela 10 – Variáveis na Equação

		B	S.E.	Wald	Grau de Liberdade	Sig.	Exp (B)
Passo 1	PRZ_ATR	1,902	0,431	19,522	1	0,000	6,702
	Constante	-3,721	0,692	28,914	1		0,024
Passo 2	PRZ_ATR	1,569	0,485	10,472	1	0,001	4,800
	QT_PROT	0,527	0,156	11,365	1	0,001	1,693
	Constante	-4,474	0,845	28,014	1	0,000	0,011
Passo 3	ID_EMPR	-0,722	0,297	5,919	1	0,015	0,486
	PRZ_ATR	1,390	0,502	7,678	1	0,006	4,016
	QT_PROT	0,483	0,164	8,708	1	0,003	1,620
	Constante	-2,206	1,117	3,902	1	0,048	0,110
Passo 4	ID_EMPR	-0,734	0,309	5,650	1	0,017	0,480
	PRZ_ATR	1,408	0,500	7,919	1	0,005	4,086
	QT_PEFIN	0,934	0,477	3,841	1	0,050	2,545
	QT_PROT	0,436	0,178	6,046	1	0,014	1,547
	Constante	-3,314	1,262	6,895	1	0,009	0,036

O teste de *Wald* é usado em regressão logística para obter a significância de cada coeficiente logístico (B). A estatística *WALD* é obtida por: $(\text{Coef. B} / \text{Erro Padrão})^2$. Identifica o quanto a variável independente participa individualmente da explicação da variável dependente.

Observa-se que pelo teste de *Wald* a significância de quase todos os coeficientes logísticos foi aceitável a um alfa de 5%, com exceção do coeficiente referente à variável QT_PEFIN que teve um alfa de 10%, que embora superior, também é aceitável para entrar no modelo.

Esses resultados representam uma expressiva adequação do modelo para explicar o comportamento da variável dependente Situação do Cliente.

A partir do resultado do logit, obtêm-se a seguinte equação final:

$$\text{SIT_CLIENTE} = - 3,314 + (- 0,734 \times \text{ID_EMPR}) + (1,408 \times \text{PRZ_ATR}) + (0,934 \times \text{QT_PEFIN}) + (0,436 \times \text{QT_PROT})$$

Onde:

SIT_CLIENTE = Situação do Cliente (adimplente ou inadimplente)

ID_EMPR = Idade da empresa

PRZ_ATR = Prazo médio de atraso no pagamento a fornecedores

QT_PEFIN = Quantidade de pendências vencidas informadas por instituições não-financeiras

QT_PROT = Quantidade de protestos

Após a definição da equação, verifica-se o grau de correlação entre as variáveis independentes contidas na mesma, obtendo a Tabela 11, para avaliar a possibilidade da ocorrência da multicolinearidade, ou seja, correlação entre duas ou mais variáveis independentes.

Tabela 11 – Correlação entre Variáveis Independentes da Equação

Variável Independente definida na Equação	Alta Correlação com:
Idade da empresa	Prazo médio de atraso e quant. Protestos.
Prazo Médio de atraso	Idade empresa e quant. Protestos.
Quantidade de pendencias vencidas informadas por empresas não financeiras (PEFIN)	Quant. Protestos.
Quantidade de protestos	Idade da empresa, prazo médio de atraso e quant. PEFIN.

Analisando a tabela com as correlações entre as variáveis independentes do modelo, observa-se que as mesmas estão altamente correlacionadas entre si.

Assim sendo, foi realizado o teste VIF que identifica o grau de colinearidade entre as variáveis. Para casos em que o valor de VIF for menor ou igual a 5 considera-se que não há presença de colinearidade estatisticamente representativa entre as variáveis.

Pelos resultados (Tabela 12), verifica-se que todos os valores de VIF apresentados foram inferiores a 5, logo, não havendo uma presença significativa de colinearidade entre as variáveis.

Tabela 12 – Estatística VIF

Modelo	Estatística de Colinearidade	
	Tolerância	VIF
1 (Constante)		
Idade Empresa	0,850	1,177
Prazo Médio de Atraso	0,647	1,547
Quant. PEFIN	0,707	1,414
Quant. Protestos	0,523	1,911