

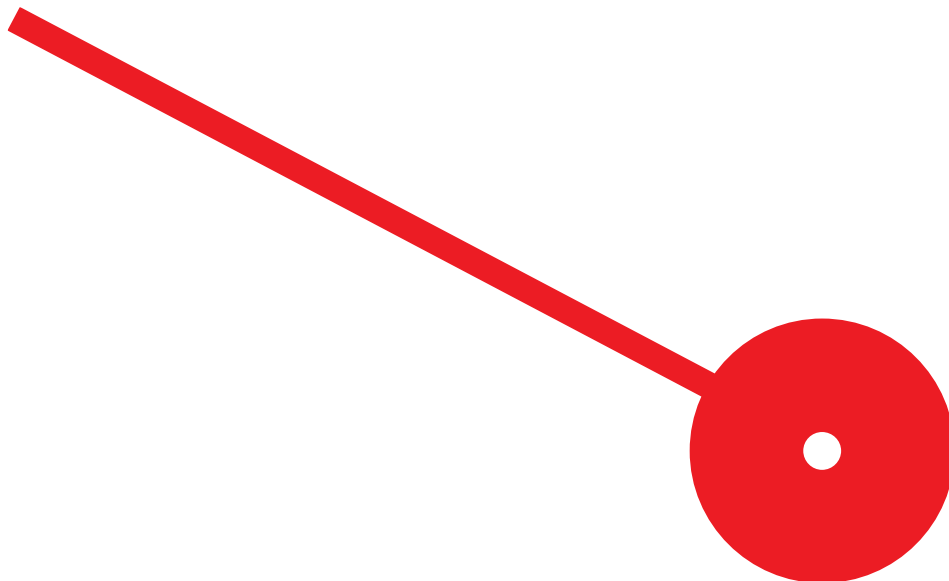
INSTITUTO
SUPERIOR
DE CONTABILIDADE
E ADMINISTRAÇÃO
DO PORTO
POLITÉCNICO
DO PORTO

M

MESTRADO
EM FINANÇAS EMPRESARIAIS

Adequabilidade das
técnicas internas de
validação de modelos de
Probabilidade de
Incumprimento
Hélder Silva Muiambo

09/2021



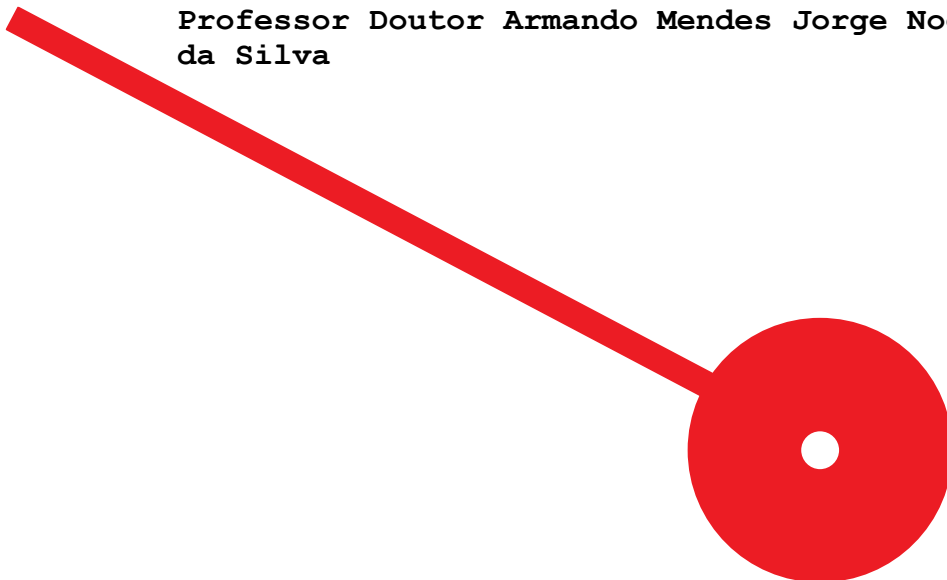
INSTITUTO
SUPERIOR
DE CONTABILIDADE
E ADMINISTRAÇÃO
DO PORTO
POLITÉCNICO
DO PORTO

M MESTRADO
EM FINANÇAS EMPRESARIAIS

Adequabilidade das
técnicas internas de
validação de modelos de
Probabilidade de
Incumprimento
Hélder Silva Muiambo

Dissertação de Mestrado apresentado ao
Instituto Superior de Contabilidade e
Administração do Porto para a obtenção do grau
de Mestre em Finanças Empresariais, sob
orientação da Professora Doutora Isabel
Cristina da Silva Lopes e co-orientação do
Professor Doutor Armando Mendes Jorge Nogueira
da Silva

Hélder Silva Muiambo. Adequabilidade das
técnicas internas de validação de modelos de
Probabilidade de Incumprimento
09/2021



Dedicatória

Aos meus pais Nicéia Domingos Macuácuá Muiambo e João da Silva M. Muiambo.

Agradecimentos

Em primeiro lugar agradeço a Deus que me fortaleceu todo o tempo me dando forças através do Espírito Santo.

Agradeço também aos meus pais Nicéia e Silva, os meus irmãos Morgan, Darwin e Eugénia que apesar da distância, através das redes sociais e telefonemas me prestaram um apoio incondicional e muito encorajamento.

Agradeço igualmente a minha namorada Mariana Juarce, pela compreensão nos momentos que não podíamos nos comunicar, por ter de me dedicar mais aos estudos.

À Professora Doutora Cristina Lopes e ao Professor Doutor Armando Silva, que dedicaram o seu precioso tempo para ler cada versão da presente dissertação tecendo comentários e críticas necessárias para melhorar o trabalho.

À minha Gerente Dra. Anastácia Shatskaia, pela compreensão nos momentos que não podia trabalhar para assistir algumas aulas ou para estar ao computador redigindo a dissertação, agradeço.

Agradeço ao Edson Mahangue, Manuel Macuácuá, Milton Enoque, Ricardo, Baptista Zumba, Eudes Chitsivane, Roman, Natasha, Kayaja, Jorge Falé, Miguel Ramos e Mara, pelo apoio e suporte nos momentos de tristeza e alegria durante os estudos.

Ao meu ex diretor Dr. Rogério Cossa e o Dr. Octávio Manhique, que me incentivaram a realizar o presente mestrado, agradeço.

Ao Banco Internacional de Moçambique (Millenium Bim), que foi a instituição que patrocinou a bolsa de estudos para o curso de mestrado, efectuando todos os pagamentos relacionados a este, agradeço.

Resumo:

A presente dissertação apresenta os resultados de uma pesquisa sobre métodos estatísticos mais adequados que os bancos portugueses podem usar para validar os modelos internos de cálculo de probabilidade de incumprimento de clientes de crédito.

Com a implementação do Acordo de Basileia II por volta dos anos 2007 a 2009, os bancos portugueses passaram a ter autonomia para estimar, usando metodologias internas, a probabilidade de incumprimento dos seus clientes. Com essa permissão, veio a obrigação de validar tais metodologias usando algumas técnicas estatísticas como: Curva ROC, índice AUC, Pietra Index, Curva CAP Accuracy Ratio, Matriz de confusão ou BRIER score, sendo que a validação deve ser feita por um departamento interno e independente dentro do banco. Neste sentido, o principal objetivo do presente trabalho é analisar a adequabilidade das técnicas estatísticas usadas pelos bancos para validar os modelos internos de classificação de risco de crédito (cálculo da probabilidade de incumprimento) de modo a aferir, numa base comparativa, sobre a capacidade de validação e qual das técnicas estatísticas acima apresentadas valida de forma mais adequada os modelos desenvolvidos.

Para atingir este objetivo, no presente trabalho, foram desenvolvidos três (3) modelos de regressão logística distintos para calcular a probabilidade de incumprimento na amostra selecionada, composta por empresas portuguesas não financeiras (base de dados disponibilizada pelo Banco de Portugal *Microdata Research Laboratory* - BPLIM). Os resultados mostraram que a regressão logística é um modelo adequado para a classificação dos clientes no que diz respeito ao risco de crédito. Quanto às técnicas estatísticas da Curva ROC, índice AUC, Pietra Index, Curva CAP Accuracy Ratio, Matriz de confusão e BRIER score, usadas para a validação dos três (3) modelos, nenhuma delas se mostrou mais importante em detrimento da outra, pois avaliam aspetos diferentes nos modelos, tendo-se verificado uma complementaridade entre os mesmos.

Palavras chave: Acordos de Basileia, Risco de crédito, Probabilidade de Incumprimento, técnicas internas de validação, Regressão logística.

Abstract:

This dissertation presents the results of a research on the most adequate statistical methods that Portuguese banks can use to validate their internal models for calculating the probability of default of credit customers.

With the implementation of the Basel II Accord around the years 2007 to 2009, Portuguese banks were given autonomy to estimate, using internal methodologies, the probability of default of their customers. With this permission, came the obligation to validate such methodologies using some statistical techniques such as: ROC Curve, AUC index, Pietra Index, CAP Accuracy Ratio Curve, Confusion Matrix or BRIER score, and the validation should be done by an internal and independent department within the bank. In this sense, the main objective of this work is to analyse the suitability of the statistical techniques used by banks to validate the internal models of credit risk classification (probability of default calculation) in order to assess, on a comparative basis, the validation capacity and which of the statistical techniques presented above more adequately validates the models developed.

To achieve this goal, in this work, three (3) distinct logistic regression models were developed to calculate the probability of default in the selected sample, composed of Portuguese non-financial corporations (database provided by the Banco de Portugal Microdata Research Laboratory - BPLIM), the results showed that logistic regression is an adequate model for the classification of customers with respect to credit risk. As for the statistical techniques of ROC Curve, AUC index, Pietra Index, CAP Curve Accuracy Ratio, Confusion Matrix and BRIER score, used for the validation of the three (3) models, none of them proved to be more important in detriment of the other, as they assess different aspects in the models, and a complementarity between them was verified.

Key words: Basel Accords, Credit Risk, Probability of Default, Internal validation techniques, Logistic regression.

Índice

1	Introdução	2
2	Revisão da Literatura	5
2.1	Definição e importância do crédito bancário	5
2.2	Risco de crédito	7
2.3	Acordos de Basileia	9
2.3.1	Acordo de Basileia I	9
2.3.2	Acordo de Basileia II	11
2.3.3	Método IRB em Basileia III	17
2.4	Processo de validação de modelos de risco de crédito	18
2.4.1	Fundamentação e os princípios da validação de modelos.....	18
2.4.2	Principais técnicas estatísticas de validação de modelos de cálculo de probabilidades de incumprimento.....	20
2.5	Estudos anteriores realizados usando o modelo de regressão logística para o cálculo de Probabilidade de <i>Default</i> (PD)	26
3	Metodologia.....	29
3.1	Modelo de Regressão Logística	29
3.2	Dados	30
3.2.1	Manipulação dos dados.....	31
3.2.2	Definição da variável <i>default</i> (incumprimento)	32
3.2.3	Seleção da amostra	33
3.2.4	Apresentação das variáveis	34
4	Resultados	38
4.1	Resultados dos modelos de regressão logística.....	38
4.1.1	Variável dependente	39
4.1.2	Modelo 1	39
4.1.3	Modelo 2	42
4.1.4	Modelo 3	45

4.2	Resultados da aplicação das técnicas de validação aos modelos de incumprimento	48
4.2.1	Matriz de Confusão-MC	48
4.2.2	Curva ROC	49
4.2.3	Curva CAP <i>Accuracy ratio</i>	51
4.3	Discussão global dos resultados da aplicação das técnicas de validação aos modelos de incumprimento.....	52
5	Conclusões.....	56
	Referências bibliográficas.....	59
	ANEXOS	69

Índice de Figuras

Figura 1: Esquema de validação de modelos	20
Figura 2: Representação da Curva ROC	23
Figura 3: Representação da Curva CAP.....	25
Figura 4: Curva ROC do modelo1	49
Figura 5: Curva ROC do modelo2	49
Figura 6: Curva ROC do modelo3	50
Figura 7: Curva CAP do modelo1	51
Figura 8: Curva CAP do modelo2.....	51
Figura 9: Curva CAP do modelo3	51

Índice de Tabelas

Tabela 1: Peso de Risco de Capital por categoria Geral de Ativo de Balanço.....	10
Tabela 2: Lista de diretivas de introdução do Basileia II em Portugal.....	13
Tabela 3: Representação de matriz de confusão.....	21
Tabela 4: Valor de AUC e a capacidade discriminante.....	24
Tabela 5: Total de empresas com default e as respetivas operações.....	33
Tabela 6: Classes de dimensão do crédito.....	33
Tabela 7: Número de empréstimos e de empresas (com e sem default) para amostra.....	33
Tabela 8: Composição da amostra.....	34
Tabela 9: Variáveis que caracterizam o empréstimo.....	34
Tabela 10: Variáveis que caracterizam as empresas.....	35
Tabela 11: Estimação dos parâmetros do modelo1.....	40
Tabela 12: Estimação dos parâmetros do modelo2.....	43
Tabela 13: Estimação dos parâmetros.....	45
Tabela 14: Medidas das probabilidades de incumprimento previstas pelos três modelos.....	47
Tabela 15: Modelo 1.....	48
Tabela 16: Modelo2.....	48
Tabela 17: Modelo 3.....	48
Tabela 18: Resultado das medidas de validação para os modelos de cálculo da PD.....	52

Lista de abreviaturas

AR – *Accuracy ratio*

AUC – *Area Under the Curve*

AIC – *Akaike Information Criterion*

BdP – Banco de Portugal

BPLIM – Banco de Portugal *Microdata Research Laboratory*

CAP – *Cumulative Accuracy Profile*

EAD – *Exposure At Default*

EDF – *Empirical Distribution Function*

EL – *Expected Loss*

EU – União Europeia

FN – *False Negative*

FP – *False Positive*

FAR – *False Alarm Rate*

HR – *Hit Rate*

IC – Instituição de Crédito

IF – Instituição financeira

IRB – *Internal Rating Based*

LGD – *Loss Given Default*

M – *Maturity*

PD – *Probability of Default*

ROC – *Receiver Operating Characteristic*

RWA – *Risk Weighting Assets*

TN – *True Negative*

TP – *True Positive*

TVA – Taxa de Variação anual

UL – *Unexpected Loss*

VaR – *Value At Risk*

1 Introdução

Alcarva (2011, p.21), sintetiza as definições constantes no *site* do Banco de Portugal e no Artigo 2º do Regime Geral das Instituições de Crédito-IC e Sociedades Financeiras (Aprovado pelo Decreto-Lei nº 298/92, de 31 de Dezembro), definindo os bancos “como todas pessoas singulares ou coletivas que, com fins lucrativos, recebem de pessoas, também coletivas ou singulares, depósitos ou outros fundos similares que aplicam por conta própria em operações de crédito ativas e noutros investimentos, complementarmente, prestam aos seus clientes serviços de tesouraria, custódia, mediação, entre outros”.

A concessão de crédito é vista como a principal fonte de lucros dos bancos. Por isso quanto mais crédito o banco fizer, mais lucro terá, esta lógica acompanha os bancos desde a sua criação, tendo sido questionada quando os bancos assumiram altos riscos por deterem taxas de transformação de depósitos em crédito concedido muito altas, gerando muitos prejuízos pela incapacidade de cobrança de parte significativa dos créditos concedidos e descredibilizando o sistema financeiro. Neste âmbito, as dez (10) maiores potências económicas mundiais denominadas G10 (abordadas com detalhe mais à frente), criaram o Comité de Basileia de Supervisão Bancária que por sua vez criou um conjunto de normas para regular a função bancária tendo chamado as tais normas de Acordos de Basileia. O Acordo de Basileia I objetivava apenas regulamentar a concessão de crédito, impondo que os bancos mantivessem um capital mínimo de 8% dos depósitos para “abafar” as perdas pelo risco de crédito, estes acordos sofreram várias adendas com o andar do tempo.

Em Junho de 2004, este Comité lançou uma versão revista da "Convergência Internacional da Medição e Padrões de Capital", designada por Acordo de Basileia II (BCBS, 2005), que para além de cobrir algumas lacunas do primeiro acordo, tinha como principal consequência pela sua implementação a permissão aos bancos para desenvolver os seus próprios modelos de classificação de clientes, o que é chamado de *Internal Rating Based Approach (IRB)*. Com a possibilidade desta abordagem baseada em classificações internas, os bancos podem calcular (usando metodologias internas) as classificações de Probabilidade de Incumprimento (PD), Perda dado o Incumprimento (LGD), Exposição em incumprimento (EAD) e Maturidade (M). Um dos requisitos para o uso do método IRB para as classificações internas é a “validação de estimativas internas” (BCBS, 2005) parágrafo 387 h).

O principal objetivo da presente dissertação é analisar a adequabilidade das técnicas estatísticas usadas pelos bancos para validar os modelos internos de classificação de risco de crédito

(cálculo da probabilidade de incumprimento) de modo a aferir, numa base comparativa entre a Matriz de confusão, a Curva ROC, o índice AUC e *Pietra Index*, e a Curva *CAP-Accuracy Ratio*, sobre a capacidade de validação e qual destas técnicas estatísticas valida de forma mais adequada os modelos desenvolvidos.

Para atingir o objetivo supra citado apoiar-nos-emos em alguns objetivos derivados deste, nomeadamente, fazendo a revisão da literatura sobre o tema em análise, ao longo da qual analisaremos o conceito de crédito, assim como da sua origem. Outro aspeto em que nos iremos focar é a gestão do risco de crédito, devido à sua importância para as Instituições Financeiras (IF). De seguida passaremos para a análise dos Acordos de Basileia, dando ênfase ao Acordo II que apesar de já ter sido revogado pelo atual Acordo III, está na origem das alterações sobre a validação de modelos objeto de estudo. Igualmente será promovida uma comparação entre os três Acordos de Basileia, procurando-se inferir a sua lógica evolutiva. O último ponto desta primeira parte do trabalho será feita a análise das técnicas estatísticas usadas pelos bancos para a validação de modelos.

A primeira parte do trabalho, descrita nos pontos acima, será analisada usando uma metodologia qualitativa caracterizada pela recolha de informação em bibliotecas eletrónicas. Finalmente, e passando para a segunda parte do trabalho, analisaremos e trataremos os dados de empresas portuguesas não financeiras, tendo em vista uma aproximação à quantificação e análise do risco de crédito, com base numa metodologia quantitativa. Serão usados os microdados do Bplim que contém as variáveis necessárias que descrevem os empréstimos das empresas para desenvolver através do *software* Rstudio, três modelos de regressão logística. De seguida serão aplicadas a esses modelos as técnicas estatísticas referidas anteriormente para validar a qualidade dos referidos modelos no cálculo de probabilidades de *default*.

CAPÍTULO II – REVISÃO DA LITERATURA

2 Revisão da Literatura

O crédito é um processo longo e complexo, podendo durar mais de 50 anos (crédito de longo prazo), pois vai desde a análise até ao acompanhamento de modo a garantir que o contraente pague por completo a sua dívida. A presente dissertação tem enfoque na análise preditiva do crédito (análise de crédito), isto é, de entre várias metodologias, ir-se-á analisar e selecionar a técnica mais adequada para validar os modelos internos usados pelos bancos para calcular a probabilidade de incumprimento de crédito das empresas portuguesas não financeiras, em que será usada uma base de dados chamada “microdados” disponibilizada pelo Bplim-Banco de Portugal (BdP).

O presente capítulo divide-se em 3 grandes partes que são, a definição do crédito e do risco de crédito, de seguida a descrição dos acordos de Basileia e por fim a descrição do processo de validação de modelos que inclui a descrição das principais técnicas estatísticas de validação de modelos de avaliação das probabilidades de incumprimento.

2.1 Definição e importância do crédito bancário

A utilização de recursos de terceiros é prática comum nas empresas como alternativa ou complemento aos seus capitais próprios para alavancar os negócios (Lopes, 2013).

De acordo com Caiado e Caiado (2006) os bancos são as instituições financeiras com maior cota no mercado financeiro quando o assunto é crédito concedido, sendo que só os bancos têm autorização para efetuar todas as operações de uma instituição de crédito. O crédito concedido a clientes tem destaque quando se trata de operações ativas, pois além de ser uma das principais e mais antigas formas de financiamento externo a que uma entidade pode recorrer, é também a principal fonte de rendimento dos bancos.

Chrickel, Narayanan e Machado (citados por Sousa, Dos Anjos, Dos Santos, e Murilo 2019), definem o crédito como uma relação de credibilidade e confiança. De acordo com Arriaga e Miranda (2009) tal confiança resolve problemas complexos resultantes da relação entre as contrapartes e reduz, até certo ponto, o clima de incerteza na qual a relação é desenvolvida.

Amaral, et al. (citados por Dias, 2015, p.7) definem o crédito como sendo:

“...o ato de troca económica em que alguém (o credor) realiza uma prestação em dinheiro ou em espécie, em determinado momento, a favor de outrem (o devedor) e aceita o risco de a respetiva contraprestação ser diferida para um momento ou

momentos posteriores, confiando no cumprimento pontual da promessa de reembolso feita pelo devedor, e recebendo um juro como compensação.”

Cabido (citado por Pestana, 2016) afirma que o Crédito é um contrato bilateral, onde uma parte empresta e a outra pede emprestado e promete pagar num prazo acordado.

De forma resumida, o crédito pode ser definido como um ato de cedência de algo, em que a sua devolução pode ser dividida e/ou diferida por um período de tempo acordado e sustentado na esperança da devolução por parte do cedido, em que à tal devolução é acrescido um valor chamado “juro” como recompensação pela cedência.

De acordo com Souza e Bruni (2008) o crédito incita o crescimento da economia, pois aumenta a capacidade de compra e de investimento das empresas e particulares, elevando assim o número de transações comerciais e reduz as dificuldades quando se trata de implementação de grandes projetos.

Desde a sua origem, o crédito tem ganho muito espaço devido à sua importância e à dependência que tem criado aos agentes económicos, pois facilita a realização de projetos para alavancar o sucesso das empresas e dos particulares. Dias (2015), refere-se à importância do crédito para as empresas afirmando que a necessidade de aceder ao crédito cresce a cada dia para o alcance mais rápido dos objetivos, pois a necessidade das empresas de responder de forma rápida à sociedade e à concorrência leva a que haja mais necessidade de recorrer ao crédito de forma a colmatar e alcançar metas.

Em Portugal o crédito financia mais de 50% da atividade económica, de acordo com o Boletim Económico do Banco de Portugal de Setembro de (2020), embora o crédito ao setor não financeiro tenha demonstrado um crescimento lento nos últimos 10 anos, a tendência seria de continuar a crescer, mas este crescimento foi interrompido devido à situação pandémica que o mundo se encontra a enfrentar. A atual pandemia (COVID-19) gerou um choque negativo na saúde pública mundial, com consequências adversas na situação económica de vários países. Apesar do risco de haver uma crise causada pela atual pandemia já estivesse previsto, nenhum país se encontrava preparado para enfrentá-la. As dimensões dos impactos económicos da atual crise ainda não são fáceis de antecipar.

Apesar destes acontecimentos, nos últimos 10 anos o crescimento do crédito exagerado já vinha sendo visto como uma situação de alerta principalmente após a última crise financeira de 2008, que esteve associada a níveis elevados de alavancagem, devido ao risco que as operações

creditícias comportam (o risco de crédito). De acordo com Dias, (2015), este risco tem de ser bem gerido pela instituição financeira de modo a conseguir responder à competitividade do mercado e fazer as respetivas cobranças ao devedor.

2.2 Risco de crédito

As instituições financeiras correm vários riscos na atividade de intermediação financeira, nomeadamente:

- Risco de Crédito;
- Risco Operacional;
- Risco de Liquidez;
- Risco de Mercado;
- Risco de Sistemas;
- Risco de *Compliance*;
- Risco de Reputação; e
- Outros riscos.

Dos riscos acima discriminados, o principal e que merece maior atenção, é o risco de crédito, porque está associado a principal atividade dos bancos, e por ser o enfoque da presente dissertação.

De acordo com Suresh, Anil Kumar e Gowda (2010) o risco de crédito é um dos riscos mais antigos, sendo comum à generalidade dos negócios, e desempenha um papel extremamente importante na performance financeira das Instituições de Crédito (IC), devido a própria natureza do seu negócio. Neste sentido, este risco merece maior atenção, não só por parte das IC, mas de todo o sistema financeiro nacional e internacional incluindo os órgãos de supervisão como o Comité de Supervisão Bancária da Basileia.

O Comité de Supervisão Bancária criou os Acordos de Basileia, que já contam com 3 três versões estando em discussão o 4º acordo, a atualização destes acordos acompanha a evolução e exigências do mercado financeiro no momento em que são atualizados. De acordo o BCBS (2005), o principal objetivo das atualizações dos Acordos é melhorar a capacidade do setor financeiro para absorver choques decorrentes de *stress* financeiro e económico, de qualquer fonte, reduzindo o risco sistémico.

Conforme definido anteriormente, o crédito é representado pela esperança de receber um determinado valor num período futuro acordado. Assim, pode-se definir o risco de crédito

como a possibilidade da contraparte deixar de cumprir com as suas obrigações patentes no contrato (BCBS, 2010). Já Martins (2012), define o risco de crédito como sendo a probabilidade de ocorrerem perdas associadas ao não cumprimento ou à deterioração de garantias de um contrato financeiro por parte de um mutuário.

Segundo Skoglund e Chen (2015, p.30), "...o risco de crédito pode ser definido como o risco subjacente ao risco de incumprimento das contrapartes, desde os clientes de retalho até à negociação das contrapartes".

O risco de crédito é definido como sendo o possível impacto negativo gerado por um devedor no incumprimento do pagamento das suas obrigações, ou pela deterioração da sua qualidade de crédito, perante a instituição financeira, incluindo possíveis restrições à transferência de pagamentos do exterior (Bessis, 2011; Jobst & Zenios, 2001).

Pestana (2016), refere-se ao risco de crédito como sendo o tipo de risco com maior impacto nas instituições bancárias, na medida em que existe a probabilidade de o cliente (mutuário) não liquidar o capital do seu empréstimo ou crédito e/ou não pague os juros de acordo com o estipulado no contrato entre as partes. Usando as palavras de Dias (2015) o risco de crédito inclui todos os riscos em que uma instituição incorre "devido à possibilidade de incumprimento ou um não pagamento no período contratado do valor da dívida", causado por vários motivos ligados ao cliente ou ao banco.

De acordo com Alcarva (2011), em termos organizacionais, a exploração da atividade bancária divide-se em três partes principais: *front office*, *middle office* e *back office*. É no *front office* onde existe o contato direto com o cliente, por isso este setor recebe as propostas de crédito e faz a análise económico-financeira para apoiar a decisão. O *middle office* é responsável por prevenir a degradação da carteira de crédito apoiando as cobranças, nos incumprimentos e também na gestão de clientes com a evolução creditícia negativa. O *back office* é o setor que não tem contato com o cliente, mas a sua responsabilidade é de acolher os clientes em incumprimento. O sucesso da atividade creditícia é o garante da liquidez do banco e depende do bom funcionamento destes setores, daí que Alcarva (2011) afirma que o conceito de liquidez tem constantemente presente a obrigatoriedade de existir um equilíbrio entre a disponibilidade dos recursos investidos e a exigência de mobilização dos depósitos. Por sua vez, a rentabilidade só se conseguirá atingir quando a taxa de juro dos financiamentos (ou operações ativas) for superior à taxa de juro dos depósitos (ou operações passivas).

Encarnação (2005, p.55) afirma que “o risco de crédito é tanto mais relevante para a solidez das IC, quanto maior for a importância dos resultados obtidos via atividade creditícia maior será o volume dos ativos inerentes a essa mesma atividade. Quanto maior é o risco de crédito associado a uma operação, maior deverá ser a sua remuneração”.

Com o acima exposto, verificamos que a gestão do risco de crédito é um importante desafio para as instituições bancárias, sendo que o insucesso na referida gestão, conduz inevitavelmente à sua falência “*too big to fail*”, como foi o caso dos Bancos da Escandinávia e Japão, na década de 90 e a crise do *subprime*, da qual resultou a falência, por exemplo, do *Lehman Brothers*. (Alcarva, 2011)

Por isso, o desenvolvimento de estudos com vista a criar e/ou aperfeiçoar os métodos e técnicas de gestão deste risco, são desenvolvidas desde a criação dos bancos, tendo-se intensificado após a crise de 2008, desses estudos podemos destacar Louhichi & Boujelbene, 2020; Pereira, 2012; Dias, 2015 e o Comité de Basileia, a seguir falaremos deste último que criou normas para a regulamentação bancária internacionais.

2.3 Acordos de Basileia

2.3.1 Acordo de Basileia I

Em 1974, os países integrantes do G10 (Japão, Países Baixos, Suíça, Canadá, Itália, Alemanha, Estados Unidos, Suécia, França, Bélgica e Reino Unido), desafiados pelas mudanças no sistema financeiro, como o aumento de transações no mercado de capitais, arbitragem e especulação, decidiram criar, com apoio do *Bank for International Settlements* (BIS), o Comité de Basileia para regular o sistema financeiro internacional. (Filipe, 2012)

O primeiro documento completo de normas de supervisão bancária internacional intitulado *International Convergence of Capital Measurement and Capital Standards* ou simplesmente, Acordo de Basileia I, foi estabelecido em 1988, com o principal objetivo de reduzir o risco no sector bancário, (Crouhy, Galai & Mark, 2005, p.42) por:

- “Apresentar princípios básicos que através de uma metodologia de avaliação de risco de crédito procuram conciliar liquidez e estabilidade financeira aos que seguem suas regras;
- Contribuir para a solidez, fortalecimento e estabilidade do sistema bancário internacional;

- Estipular níveis mínimos de solvabilidade aos bancos para fazer face ao risco de crédito em que incorrem, isto é, os fundos próprios deviam assumir um valor mínimo capaz de responder aos riscos derivados da sua atividade creditícia.”

Este acordo foi assinado por quase todos países, tendo os bancos centrais transformado os acordos em normas de modo a tornar obrigatória a sua implementação. O acordo propôs a manutenção de um capital mínimo de 8%, para cobrir o risco de crédito reduzindo a possibilidade de falência dos bancos. O primeiro passo para alcançar este objetivo, foi a divisão dos fundos próprios em capital principal ou capital do nível I “*Tier I*” e Capital suplementar ou capital do nível II “*Tier II*”. (BCBS, 1998)

Além da definição dos capitais, o acordo deliberou medidas comuns de ponderação de risco que consistia em atribuir um peso às diferentes categorias de risco a cada ativo para o cálculo dos requisitos de fundos próprios, isto é, cada *Risk Weighting Assets* (RWA), poderia ser ponderado a: 0%, 20%, 50% e 100%, sendo aplicado 8% a 100% dos ativos ponderados para o rácio de solvabilidade, conforme a tabela abaixo:

Tabela 1: Peso de Risco de Capital por categoria Geral de Ativo de Balanço

0%	Caixa e ouro em barras, direitos sobre governos da OECD como títulos de dívida do Tesouro ou hipotecas residenciais
20%	Direitos sobre bancos da OECD e entidades da OECD do setor público como títulos emitidos por órgãos do governo dos EUA ou direitos sobre municípios
50%	Hipotecas residenciais não-seguras
100%	Todos demais direitos como títulos de dívida corporativa e endividamento de países menos desenvolvidos, direitos sobre bancos não da OECD, capital. Imóveis, dependências, planta e equipamentos

Fonte: BIS (BCBS, 1988)

De acordo com Santin e Alegre (2010), o acordo também incluía normas para ponderar o risco de exposição fora do balanço, convertendo essas posições em crédito equivalente e serem incluídas nos cálculos do capital, estando distribuídas como se segue:

- Garantias de empréstimos, incluindo *securities*, têm ponderação de risco de 100%;
- Transações contingenciais têm fator de conversão de 50%;

¹ É comum a todas as instituições que compõem o sistema bancário, estando por isso refletido nas demonstrações financeiras, sendo a base para o cálculo das margens de rentabilidade, é composto por Capital Social e Reservas.

² Refere-se às componentes não especificadas no capital básico, nomeadamente, Reservas Ocultas, Provisões Gerais/Provisões para Riscos de Crédito, Instrumentos de Capital de Natureza Híbrida e Dívida Subordinada.

- Obrigações de curto prazo, relacionadas as operações comerciais, com fator de conversão de 20%;
- Posições com prazo de maturação superior a 1 ano possuem fator de conversão de risco de crédito de 50%.

De acordo com Filipe (2012), em 1996, o acordo de Basileia I sofreu uma adenda, tendo esta adenda adicionado para o cálculo do capital as exigências de medição de risco de mercado, passando o cálculo do capital a obedecer a seguinte fórmula:

$$\text{Rácio de Solvabilidade} = \frac{\text{Fundos Próprios Elegíveis (Tier I+Tier II)}}{\text{Risco de Crédito+Risco de Mercado}} > 8\% \quad (1)$$

Este foi o início de um conjunto de adendas que o acordo de Basileia I foi sofrendo, sendo que em 2004 a adenda foi chamada de Acordo de Basileia II.

2.3.2 Acordo de Basileia II

O acordo de Basileia II é mais abrangente que o Basileia I, pois além dos riscos de crédito³ e de mercado⁴, o mesmo adiciona o risco operacional⁵. Enquanto no Basileia I os bancos tinham de seguir os mesmos princípios, independentemente da dimensão e da carteira de clientes, o acordo Basileia II procurava ser mais sensível aos riscos permitindo que os bancos possam escolher entre modelos padrão ou modelos internos. (Filipe, 2012)

A definição de capital elegível é crucial (não tendo alterado com a adenda ao 1º acordo), para que os bancos mantenham o capital total equivalente ao exigido, pelo menos 8% do total dos ativos ponderados pelo risco, sendo que para o novo acordo a medição da solvabilidade passou a ser de acordo com a seguinte fórmula:

$$\text{Rácio de Solvabilidade} = \frac{\text{Fundos Próprios Elegíveis (Tier I+Tier II)}}{\text{Risco de Crédito+Risco de Mercado+Risco Operacional}} > 8\% \quad (2)$$

De acordo com BCBS (2006), o acordo de Basileia II está dividido em 3 pilares:

2.3.2.1 Pilar 1: Requisitos de fundos próprios

O primeiro pilar do Basileia II tem em vista responder à dimensão de capital necessário para cobrir os riscos de crédito, de mercado e operacional, estabelecendo a forma de cálculo dos requisitos mínimos de capital e as várias formas de cálculo destes riscos.

³ O risco de crédito engloba toda a possibilidade de incumprimento presente numa transação financeira.

⁴ O risco de mercado corresponde às oscilações nos preços causado por eventos que atingem sistematicamente todo o mercado, também é conhecido como risco sistémico.

⁵ O risco operacional engloba todas as falhas internas à instituição seja, nos sistemas, pessoas, produtos ou processos que possam acontecer e acarretar prejuízos de vários tipos, desde financeiras até mesmo reputacionais.

A este pilar será dada maior importância na presente dissertação, por tratar do cálculo dos requisitos de capital para o risco de crédito, que está diretamente relacionado com o cálculo de probabilidades de incumprimento.

a) Mensuração do risco de crédito

O cálculo de requisitos mínimos para cobrir o risco de crédito pressupõe a estimativa dos fatores de risco, ou seja, a probabilidade de incumprimento (*Probability of Default – PD*), a perda dada o incumprimento (*Loss Given Default – LGD*), a exposição em risco (*Exposure At Default – EAD*) e a maturidade efetiva (*Maturity – M*), através de métodos disponibilizados pelos acordos de Basileia. (BCBS, 2011)

Conforme indicado anteriormente, a inovação do Basel II face ao anterior acordo reside no uso de notações de crédito (internas e/ou externas), isto é, os bancos passariam a calcular o risco de crédito através de dois métodos, método padrão – *standardised approach* e o método *Internal Ratings-Based Approach* (IRB) que constitui o ênfase deste trabalho. Estes métodos divergem em dois aspetos essenciais, enquanto o método Padrão se baseia nas avaliações externas do risco produzidas por agências de notação, o método IRB baseia-se em sistemas de avaliação do risco de crédito desenvolvidos pelos próprios bancos. (Antão & Lacerda, 2008)

De acordo com BCBS (2005), sujeito a certas condições mínimas e requisitos de divulgação, os bancos que tenham recebido a aprovação do órgão supervisor (normalmente o banco central) para utilizar o método IRB podem basear-se nas suas próprias estimativas internas de risco para determinar a exigência de capital para uma dada exposição.

Para cada classe de ativos cobertos no âmbito do método IRB, o BCBS (2010), define três elementos principais, nomeadamente:

- *Fatores de risco* — estimativas de parâmetros de risco fornecidas pelos bancos, algumas das quais são fornecidas pelo órgão supervisor.
- *Funções de peso do risco* — os meios pelos quais os fatores de risco são transformados em ativos ponderados pelo risco e, portanto, requisitos de capital.
- *Requisitos mínimos* — as normas mínimas que devem ser cumpridas por um banco para utilizar o método IRB para uma determinada classe de ativos.

Para as diferentes classes de ativos, o Basel II dividiu o método IRB em dois, o IRB básico e o IRB avançado. No IRB básico, os bancos fornecem as suas próprias estimativas de probabilidades de incumprimento e confiam nas estimativas do supervisor para outros fatores

de risco de crédito; enquanto no IRB avançado além das PD, os bancos fornecem também as suas próprias estimativas de LGD, EAD, e os seus próprios cálculos de M. Tanto para um ou para o outro método, os bancos devem sempre considerar o peso de risco para efeitos de cálculo dos requisitos de capital. (BCBS, 2010)

b) Funcionamento do método IRB em Portugal

O processo de implementação do acordo de Basileia II em Portugal iniciou-se em 2007, com a emissão de um conjunto de diretivas e regulamentos (descritos na tabela2). Estes regulamentos e diretivas, serão usados para descrever a presente sessão do trabalho.

Tabela 2: Lista de diretivas de introdução do Basileia II em Portugal

Diretiva	Descrição
2006/48/CE	Relativa ao acesso à atividade das instituições de crédito e ao seu exercício
2006/49/CE	Relativa à adequação dos fundos próprios das empresas de investimento e das instituições de crédito
2007/64/CE	Relativa aos serviços de pagamento no mercado interno
648/2012	Relativo aos derivados do mercado de balcão, às contrapartes centrais e aos repositórios de transações
575/2013	Relativo aos requisitos prudenciais para as instituições de crédito e para as empresas de investimento

Fonte: Adaptação própria das diretivas do BdP

Pelo disposto no nº 2 do artigo 143º do (Regulamento UE 575/2013), a autorização prévia para utilizar o Método IRB, incluindo estimativas próprias de LGD e de fatores de conversão, é exigida para cada classe de risco e para cada sistema de notação e método dos modelos internos utilizado nas posições em risco sobre ações, bem como para cada modelo utilizado na estimativa de LGD e de fatores de conversão.

Concretizando, conforme referido anteriormente, o BCBS (2010) define quatro (4) fatores de risco de crédito:

Probabilidade de Incumprimento (PD) – é a incerteza do não reembolso por parte do mutuário para um determinado empréstimo no período, normalmente de um ano, dependendo não só da classificação do mutuário mas também da situação económica geral.

Exposição em risco (EAD) – é o total do valor do empréstimo que está exposto à perda quando um respetivo empréstimo é incumpridor.

Perda dado o Incumprimento (LGD) – corresponde, em percentagem, a parte do empréstimo que o banco perde após o incumprimento por parte do mutuário, isto é, a proporção do empréstimo que não é recuperada.

Maturidade efetiva (M) – é o prazo de vencimento do empréstimo considerando os recebimentos futuros.

O método IRB tem como principal finalidade, através dos parâmetros acima definidos, o cálculo da perda esperada (*Expected Loss - EL*) e a perda não esperada (*Unexpected Loss - UL*). A perda esperada já é conhecida, pois de acordo com Colquitti (2007), o EL representa o custo de extensão do crédito, que o credor precisa de recuperar ao fixar o preço do empréstimo. O montante de risco de crédito é o produto dos três (3) parâmetros apresentados anteriormente pela seguinte fórmula:

$$EL(\text{montante}) = PD * LGD * EA \quad (3)$$

A perda esperada também pode ser escrita em forma de percentagem no momento do incumprimento, pela seguinte fórmula:

$$EL(\%) = PD * LGD \quad (4)$$

Ainda de acordo com Colquitti (2007, p.221), as perdas não esperadas (UL) ocorrem geralmente em condições de *stress* e quando ocorrem têm um impacto significativamente grande na carteira da instituição e na rentabilidade global. O EL corresponde à média de uma função de distribuição de crédito, enquanto o UL é o desvio padrão da distribuição que mede a volatilidade ou potencial da exposição máxima de perda a um determinado nível de confiança (por exemplo, 95%, 99%). As perdas não esperadas incluem também o montante de capital que o banco tem como almofada no caso de ocorrerem tais perdas extremas.

As funções de peso de risco produzem requisitos de capital para a parte do UL de acordo com o descrito na Directiva (2006/48/CE), o montante das posições ponderadas pelo risco em relação às posições sobre as empresas, as instituições e as administrações centrais e os bancos centrais são calculadas usando as fórmulas abaixo:

A correlação será dada por:

$$R = 0.12 * \frac{1-e^{-50*PD}}{1-e^{-50}} + 0.24 * \left[1 - \left(\frac{1-e^{-50*PD}}{1-e^{-50}} \right) \right] \quad (5)$$

O ponderador de risco (RW) será dado por:

$$RW = \left\{ LGD * N \left[(1 - R)^{-0.5} * G(PD) + \left(\frac{R}{1-R} \right)^{0.5} * G(99,9\%) \right] - PD * LGD \right\} * \left[\frac{1-(M-2.5)*b}{1-1.5*b} \right] * 12.5 * 1.06 \quad (6)$$

Onde:

- $N(x)$ indica a função de distribuição cumulativa de uma variável aleatória normal (isto é, representa a probabilidade de uma variável aleatória normal de média 0 e variância 1 ser inferior ou igual a x);
- $G(z)$ indica a função de distribuição cumulativa inversa para uma variável aleatória normal (isto é, o valor x por forma a que $N(x) = z$);
- b corresponde ao fator associado à data de vencimento e é dado por: $b = [0.11852 - 0.05478 * \ln(PD)]^2$. (Diretiva 2006/487CE)

De acordo com Tsukahara (2013), a perda máxima de crédito para o período de um ano com intervalo de confiança de 99% (*VaR-value at risk*) é dada pela seguinte fórmula:

$$VaR = LGD * N \left[(1 - R)^{-0.5} * G(PD) + \left(\frac{R}{1-R} \right)^{0.5} * G(99,9\%) \right] \quad (7)$$

A expressão $PD * LG$ corresponde à perda esperada em percentagem conforme o descrito na equação 4.

Neste sentido a perda não esperada (UL) será dada pela diferença entre a perda máxima de crédito dado o intervalo de confiança e a perda esperada em percentagem, sendo:

$$UL = LGD * N \left[(1 - R)^{-0.5} * G(PD) + \left(\frac{R}{1-R} \right)^{0.5} * G(99,9\%) \right] - PD * LGD \quad (8)$$

De acordo com o BCBS (2005) o multiplicador abaixo é um fator de ajuste que permite adequar o cálculo da UL ao prazo da operação:

$$\frac{1-(M-2.5)*b}{1-1.5*b} \quad (9)$$

2.3.2.2 Pilar 2: Processo de supervisão

Este pilar visa reforçar o processo de supervisão quanto à adequação de capital nas instituições de crédito, o que exige a convergência de políticas e práticas de supervisão.

No âmbito deste pilar o comité de Basileia definiu na versão do Basileia de 2006 quatro princípios fundamentais:

- Os Bancos devem dispor de um processo para avaliar a suficiência de fundos próprios, bem como uma estratégia para manter um nível de capital adequado;

- As autoridades de supervisão devem rever os processos internos e as estratégias das instituições e avaliar a capacidade das instituições para controlar e assegurar o cumprimento dos rácios prudenciais;
- As autoridades de supervisão devem esperar que as instituições operem com rácios de capital superior aos mínimos definidos e podem exigir que mantenham fundos próprios superiores a esse mínimo; e
- As autoridades de supervisão devem atuar para impedir que os fundos próprios se tornem inferiores aos mínimos necessários, tendo em conta o perfil de risco de cada instituição, podendo mesmo adotar medidas corretivas. (Caiado & Caiado, 2008)

De acordo com Caiado e Caiado (2008), espera-se que com este pilar as instituições mobilizem mais recursos para identificar e analisar os riscos que enfrentam, procurem medir com maior precisão o impacto de perdas potenciais subjacentes a estes riscos e recorram a instrumentos de mitigação dos riscos em que incorrem, o que à partida poderá contribuir para a obtenção de melhores performances, designadamente na área económica. É no âmbito deste pilar que é introduzido o processo de autoavaliação, por cada instituição, das necessidades de capital económico, face aos riscos assumidos no decurso da sua atividade.

2.3.2.3 Pilar 3: Disciplina de mercado

O objetivo deste pilar é estimular uma maior disciplina do mercado através do aumento da transparência das instituições sujeita ao Basileia II, a fim de que os agentes de mercado possam estar bem informados e possam perceber melhor o perfil de risco dos bancos, ou seja, através deste pilar o Comité de Basileia recomenda a transparência e disponibilização da informação financeira ao mercado através da divulgação da descrição de metodologias de gestão de risco por linha de negócio. (Caiado & Caiado, 2008)

Uma vez que, os vários países apresentam diferenças significativas em termos de autoridades de supervisão, o Comité de Basileia definiu que as recomendações propostas neste pilar deveriam fazer parte do processo de gestão bancária, nomeadamente serem aprovadas pelos próprios conselhos executivos das instituições bancárias. (BCBS, 2006)

Caiado e Caiado (2008), acrescentam que com este pilar, que preconiza a suficiência e a consistência na divulgação da informação nos diferentes mercados, de modo a promover a confiança e a credibilidade do sistema financeiro, espera-se que os participantes no mercado passem a dispor de informação que lhes permita ajuizar das práticas de gestão das instituições

em função da respetiva solidez, através da influência que podem exercer, mormente no domínio da capacidade de endividamento e dos respetivos custos e da valorização de seu capital.

2.3.3 Método IRB em Basileia III

Após a crise *subprime* de 2008, o acordo de Basileia II foi criticado pela incapacidade de prevenir crises financeiras produzidas pelo aumento descontrolado da liquidez no mercado, pela alavancagem e assunção excessiva de risco; por isso foi apresentada mais uma reforma às medidas de supervisão, o Acordo de Basileia III, tendo a versão final, sido emitida em 2017.

Almeida (citado por Fernandes, 2014), de forma resumida aponta os principais desequilíbrios que estiveram na base da elaboração do novo acordo como sendo os seguintes:

- Excesso de alavancagem;
- Capital de qualidade insuficiente; e
- Inexistência de reservas para fazer face a períodos de escassez de liquidez.

Neste contexto, as ferramentas introduzidas no acordo foram o *Liquidity Coverage Ratio (LCR)*, que define que a qualidade da carteira de investimentos em termos de liquidez deve ser proporcional aos fluxos de caixa líquidos a 30 dias e o *Net Stable Funding Ratio (NSFR)*, que impõe que os fundos estáveis disponíveis devem ser superiores, ou iguais, aos necessários, sendo quer o LCR como NSFR devem alcançar os níveis mínimos impostos pelo regulador. (Mendes, citado por Rodrigues, 2018)

No que diz respeito aos fundos de médio/longo prazo, dever-se-á criar uma “esponja de capital” ou *buffers* contra-cíclica; assim, deverão ser reforçadas as provisões para as perdas inesperadas em cerca de 2,5% dos Ativos ponderados pelo risco (*Unexpected Losses*).

Na UE, a implementação do Basileia III foi, de forma progressiva entre 2013 e 2019, promovendo a adoção pelas entidades bancárias da abordagem IRB para o cálculo das perdas esperadas e inesperadas (e.g. modelos do tipo *Value at Risk (VaR)*).

No que diz respeito ao método IRB, não sofreu alterações significativas de um acordo para o outro sendo que a instituição que implementá-lo deverá ter as próprias metodologias para estimar os riscos. Schlickemaier (2012), sustenta esse pensamento afirmando que o Basileia III é semelhante ao Basileia II, diferindo nalguns aspetos importantes; embora Basileia III permita a referência a notações de risco de agências, altera o método IRB padrão para uma abordagem "IRB modificada" porque aplicando este método o banco avalia as exposições e determina se as estimativas de risco baseadas nas agências são apropriadas, além disso, Basileia

III também exige que as notações de crédito externas sejam disponibilizadas ao público, numa base não seletiva e gratuita.

2.4 Processo de validação de modelos de risco de crédito

2.4.1 Fundamentação e os princípios da validação de modelos

Conforme indicado anteriormente o Comité de Basileia tem incentivado os bancos a utilizarem métodos internos para estimar os riscos, mas este processo vem acompanhado da respetiva validação, que é imposta como condição para o uso destes métodos.

Neste sentido, as alíneas a-e do artigos 185º do (Regulamento 573/2013) descrevem os requisitos sob os quais uma instituição deve validar as estimativas internas, destacando-se para a presente dissertação o que refere que as instituições devem comparar regularmente as taxas de incumprimento observadas com as estimativas de PD por cada grau de notação e, quando estas taxas se situarem fora do intervalo previsto para esse grau, devem analisar os motivos específicos na origem de tais desvios.

O BCBS (2007) indica 6 princípios de validação a saber:

1º Princípio – A validação avalia as estimativas de risco e o uso de *ratings* na concessão de crédito, isto é, mesmo que as estimativas de risco sejam baseadas no passado, devem ser capazes de prever o futuro de alguma medida.

2º Princípio – A responsabilidade da validação é do próprio banco e não do órgão supervisor.

3º Princípio – A validação é um processo iterativo. A definição do período e das ferramentas de validação para responder às mudanças do mercado ou de operações deve ser entre os bancos e os órgãos supervisores, devendo estes valorizar o diálogo.

4º Princípio – Não há um único método de validação. Não existe uma única prática de validação de modelos que possa ser utilizada em todas as carteiras dos bancos.

5º Princípio – A validação não deve abordar apenas os aspetos quantitativos, é necessário verificar também os elementos qualitativos presentes nele.

6º Princípio – O processo e os resultados da validação devem ser analisados de forma independente. A auditoria interna constitui a área competente para tal revisão.

Mesmo que não se encontrem apontados de forma direta, podemos adicionar aos princípios acima indicados os seguintes:

- A validação deve ser construtiva e não ameaçadora;
- A validação não fornece uma decisão fixa, mas sim uma sugestão para ação e estudos adicionais para desenvolver *framework* de validação e estruturas de ação;
- Os métodos de validação não podem mudar consoante o ciclo económico, a menos que estejam documentados de forma clara e completa. (BCBS, 2011)

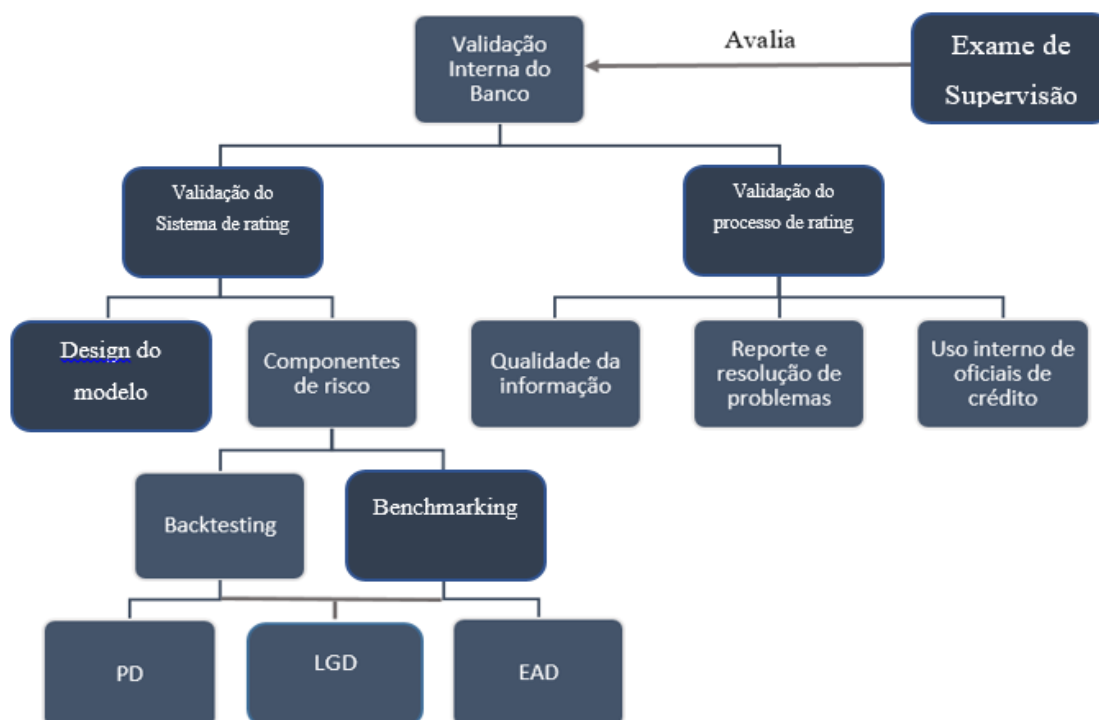
No que diz respeito ao processo de validação de modelos de estimativas do risco de crédito, o BCBS (2017), refere que tanto os modelos para estimação como os de validação de PD se encontram numa fase mais avançada do que para LGD e EAD. Além disso, refere também que quando se trata da validação de modelos de PD, o processo deve abranger dois aspetos principais: capacidade discriminante e a calibração.

Fernandes (2014) refere que a capacidade discriminante de um sistema de classificação de risco indica a habilidade de distinguir, de maneira *ex ante*, contrapartes que entrarão em incumprimento de contrapartes que não entrarão. O poder discriminante do sistema deve ser testado em amostra segregada (*out-of-sample*). Caso contrário, corre-se o risco de o poder discriminante ser superestimado.

Blöchlinger (2012) define a calibração de um sistema interno de classificação de risco como sendo a capacidade de um modelo de corresponder às taxas de incumprimento previstas e observadas em todo o diferencial de dados. Tasche (2006) afirma que a calibração é geralmente entendida como o fato de se encontrar as PD's "verdadeiras" para os graus de classificação correspondentes.

De acordo com BCBS (2017), a análise estatística de sistemas internos de classificação de risco de crédito está predominantemente baseada na separação das contrapartes em dois tipos: contrapartes que entrarão em incumprimento em um horizonte de tempo, denominadas (más) e contrapartes que não entrarão em incumprimento em tal horizonte de tempo denominadas (boas).

Figura 1: Esquema de validação de modelos



Fonte: Adaptado do BCBS (2011)

2.4.2 Principais técnicas estatísticas de validação de modelos de cálculo de probabilidades de incumprimento

Conforme referido anteriormente, a validação de um modelo centra-se no poder discriminante e na calibração. Quanto às técnicas para avaliar o poder discriminante, autores como Blöchlinger, 2012; Tasche, 2006; Fernandes, 2014; Tsukahara et al., 2016; Antão & Lacerda, 2008; Burgt, 2007; Hanley & McNeil, 1982; Stein, 2007; Vieira, Barboza, Sobreiro & Kimura, 2019 & BCBS, 2011, nas suas literaturas descrevem vários métodos, nomeadamente:

- Matriz de confusão;
- Curva *ROC*, índice *AUC*;
- *Pietra Index*
- Curva *CAP Accuracy Ratio*
- *BRIER Score*

2.4.2.1 Matriz de confusão

Os resultados de uma previsão podem ser apresentados usando uma matriz de confusão com as classificações do modelo e as observações efetivas da amostra divididas em 2 grupos, conforme o quadro abaixo:

Tabela 3: Representação de matriz de confusão

		Classificação do modelo		Totais
		$Y_i = 1$	$Y_i = 0$	
Observações	Com $Y_i = 1$	TP (True Positive)	FN (False Negative)	n_i
	Com $Y_i = 0$	FP (False Positive)	TN (True Negative)	n_0
Totais		PP	PN	n

Fonte: Adaptado de Wang, Ma, Lihua e Kaiquan (2011)

De acordo com Wang, et al. (2011), a matriz de confusão (MC), para a validação de modelos, baseia-se em medidas padrão estabelecidas nos domínios da pontuação de crédito, tais medidas incluem precisão média, erro de tipo I e erro de tipo II, conforme se descrevem abaixo:

Precisão – Corresponde à percentagem de amostras positivas classificadas de forma correta sobre a soma de amostras classificadas como positivas:

$$Precisão = \frac{TP}{TP+FP} = \frac{TP}{PP} \quad (10)$$

Sensibilidade – corresponde à proporção das observações positivas classificadas pelo modelo como positivas, é a capacidade do modelo predizer de forma correta a condição para casos que realmente a têm:

$$Sensibilidade = \frac{TP}{TP+FN} = \frac{TP}{n_i} \quad (11)$$

Neste sentido o Erro do tipo I é dado pela seguinte fórmula:

$$Erro\ Tipo\ I = \frac{FN}{TP+FN} = \frac{FN}{n_i} \quad (12)$$

Especificidade – corresponde a percentagem de amostras negativas identificadas corretamente, é a capacidade do modelo predizer de forma correta a ausência da condição para casos que realmente não a têm:

$$Especificidade = \frac{TN}{TN+FP} = \frac{TN}{n_0} \quad (13)$$

Sendo que o Erro do tipo II é dado pela seguinte fórmula:

$$\text{Erro Tipo II} = \frac{FP}{TN+FP} = \frac{FP}{n_0} \quad (14)$$

Capacidade Total de Acerto ou Acurácia (CTA): proporção de acertos de um modelo. Ou seja, é a proporção de verdadeiros-positivos e verdadeiros-negativos em relação a todos os resultados possíveis.

$$CTA = \frac{TP+TN}{TP+FP+TN+FN} \quad (15)$$

2.4.2.2 Receiver Operating Characteristic – ROC e o Índice Pietra

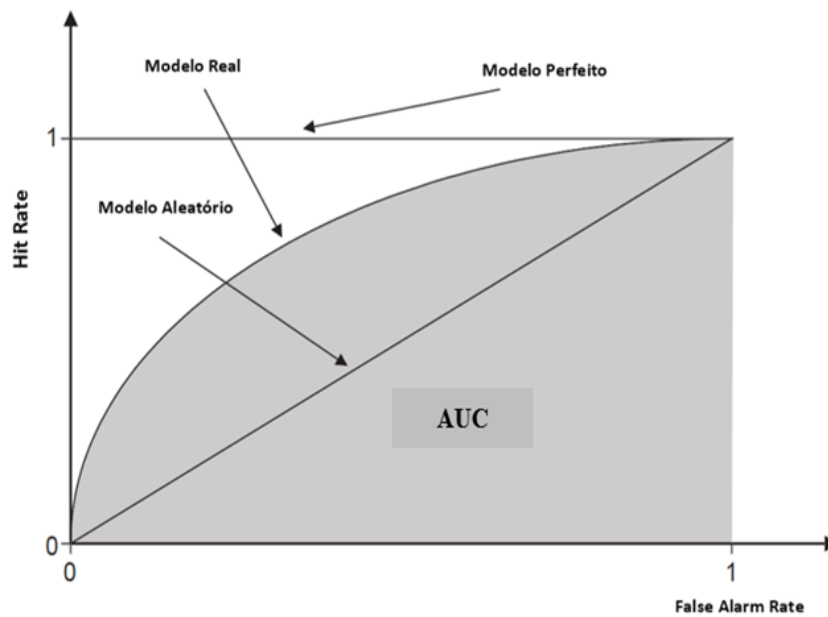
De acordo com Hanley e McNeil (1982) a curva ROC é usada para julgar a capacidade discriminante de vários modelos estatísticos que combinam várias pistas, resultados de testes, com fins preditivos.

Engelmann, Hayden e Tasche (2003), definem a curva ROC como sendo um gráfico que busca mostrar a relação, normalmente antagónica, entre a sensibilidade e a especificidade. A área da curva está relacionada com a distribuição de frequência de eventos de incumprimento e de não incumprimento e permitem quantificar a exatidão de um modelo de classificação, pois, quanto maior a área sob a curva ROC, maior é a sensibilidade.

O método ROC pode ser construído a partir de duas amostras de scores, um para casos normais, e outro para casos anormais (devedores incumpridores). Fernandes (2014) afirma que espera-se que a curva ROC seja côncava, o que significa dizer que há maior concentração de instituições más nos piores scores e de instituições boas nos melhores scores, sendo mais importante o índice derivado da análise de ROC, a área sob a curva ROC (AUC).

Os acordos de Basileia demonstram alguma preferência no ROC para a verificação do poder discriminante de um modelo de classificação de risco de crédito, pela facilidade que o este modelo apresenta no cálculo dos intervalos de confiança, pois a determinação de intervalos de confiança também é fundamental para que se possam comparar os desempenhos de valores distintos de AUC provenientes de diferentes modelos de classificação de risco. (Fernandes, 2014)

Figura 2: Representação da Curva ROC



Fonte: Engelmann et al. (2003)

Engelmann et al. (2003) definem *Hit rate* $HR(C)$ através da seguinte fórmula:

$$HR(C) = \frac{H(C)}{N_D} = \frac{TP}{(TP+FN)} = \text{sensibilidade} \quad (16)$$

Onde:

- $H(C)$ corresponde ao número de incumprimentos corretamente previstos não superior a C ;
- N_D corresponde ao total dos incumprimentos da amostra; e

O *False Alarm Rate* $FAR(C)$ é dado por:

$$FAR(C) = \frac{F(C)}{N_{ND}} = \frac{FP}{(FP+TN)} = 1 - \frac{TN}{(FP+TN)} = 1 - \text{especificidade} \quad (17)$$

Onde:

- $F(C)$ corresponde ao número de falsos alarmes, ou seja, o número de não incumprimentos que foram classificados incorretamente como incumprimentos, utilizando o valor C como ponto de corte entre incumpridores e cumpridores. O valor C mais habitual é 0.5, mas podem ser usados quaisquer outros valores.
- N_{ND} corresponde ao número total de não incumpridores na amostra. A curva ROC é construída da seguinte forma. Para todos os valores de corte C que estão contidos no

intervalo das pontuações das quantidades de $HR(C)$ e os $FAR(C)$ são calculados. A curva ROC é uma parcela de $HR(C)$ versus $FAR(C)$. Conforme referido anteriormente quanto maior a área abaixo da curva ROC, melhor é o modelo, assim, a área AUC, pode ser calculada da seguinte fórmula:

$$AUC = \int_0^1 HR(FAR)d(FAR) \quad (18)$$

Este índice, denotado AUC para simbolizar os seus alicerces gaussianos, varia de 0,5 (sem sensibilidade aparente) a 1.0 (sensibilidade perfeita) Andrade e Oliveira (citados por Fernandes, 2014) consideram que o poder discriminante do modelo é aceitável para valor de AUC acima de 70%, como mostra a Tabela 4:

Tabela 4: Valor de AUC e a capacidade discriminante

Valor da AUC	Capacidade discriminante
$AUC \leq 0.5$	Não existe discriminação
$0.7 \leq AUC \leq 0.8$	Discriminação aceitável
$0.8 \leq AUC < 0.9$	Excelente discriminação
$AUC \geq 0.9$	Discriminação acima do comum

Fonte: Andrade e Oliveira (citados por Fernandes, 2014)

Segundo Fernandes (2014), o índice de *Pietra* é a outra medida bastante importante derivada da análise ROC. Ele é definido como a metade da distância máxima entre a curva ROC e a reta de 45°. A Figura 1 constitui exemplo ilustrativo de construção da curva ROC. Tasche (2006) define matematicamente o índice de *Pietra* como sendo:

$$Pietra\ Index = \frac{\sqrt{2}}{4} \max_c |HR(C) - FAR(C)| \quad (19)$$

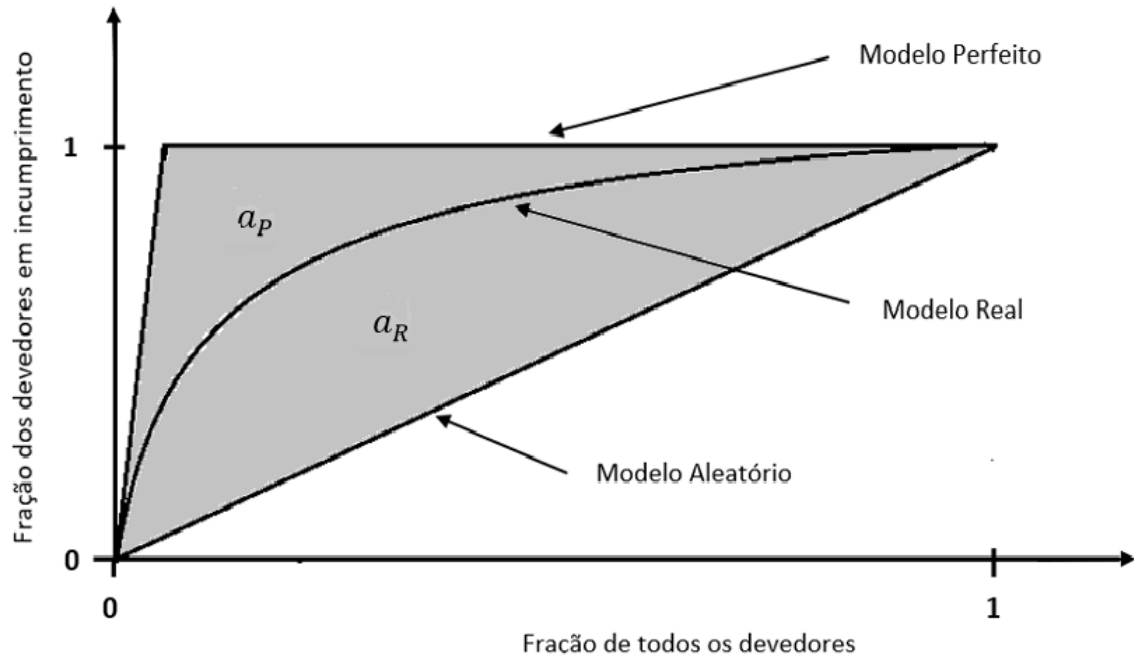
2.4.2.3 Cumulative Accuracy Profile (CAP) – Accuracy Ratio

Segundo Tasche (2006), o Perfil de Precisão Acumulada (curva CAP) é uma ferramenta gráfica útil para a investigação do poder discriminante dos sistemas de classificação.

Para construir uma curva CAP, os devedores devem ser ordenados de acordo com as suas pontuações, isto é, do devedor com a pontuação mais arriscada para o devedor com a pontuação menos arriscada. Depois de ordenados, estabelece-se para dada pontuação de corte a fração de devedores com pontuação menos arriscada do que o ponto de corte e a fração de incumprimentos com pontuação menos arriscada do que o ponto de corte. Calculando-se essas

frações para todos pontos de corte possíveis, a curva CAP é obtida através de um gráfico-fração das contrapartes em incumprimento versus fração de contrapartes. Um modelo de classificação perfeito atribuirá as notas mais baixas aos incumpridores. (Tsukahara et al., 2016; & Engelmann et al., 2003)

Figura 3: Representação da Curva CAP



Fonte: Baseado em Engelmann et al., (2003)

De acordo com Engelmann et al. (2003), a qualidade de um sistema de classificação é medida pela relação de precisão *accuracy ratio* AR . É definido como o rácio da área a_R entre a curva CAP do modelo de classificação que está a ser validado e a curva CAP do modelo aleatório, e a área a_P entre a curva CAP do modelo perfeito e a curva CAP do modelo aleatório, dado por:

$$AR = \frac{a_R}{a_P} \quad (20)$$

Quanto mais próximo de 1 o resultado de AR , maior é a capacidade discriminante do modelo.

2.4.2.4 *BRIER Score*

Brier (1950) desenvolveu um método de avaliação da qualidade da previsão de uma probabilidade: o *brier score*. Apesar do presente modelo ter sido desenvolvido para validar a acurácia das probabilidades meteorológicas, também poder ser usado para mensurar a capacidade discriminante de um modelo de PD.

Se p_0, p_1, \dots, p_k corresponderem as PD's estimadas para contrapartes em K níveis de classificação de risco. O *score* de *Brier* será definido como:

$$BRIER = \frac{1}{N} \sum_{i=1}^N (p_i - O_i)^2 \quad (21)$$

Onde:

- P_i corresponde a probabilidade de ocorrência do evento
- O_i corresponde a uma variável binária (1/0) que significa que o evento observou-se ou não;
- N corresponde ao número de contrapartes da amostra

Um modelo de PD perfeito indicaria um para os eventos de incumprimento observados, e zero caso contrário, assim para um modelo perfeito a relação seria dada por:

$$P_i - O_i = 0 \quad (22)$$

Quanto menor ou próximo de zero estiver o score de BRIER para um conjunto de previsões, melhores são as previsões.

2.5 Estudos anteriores realizados usando o modelo de regressão logística para o cálculo de Probabilidade de *Default* (PD)

A análise do crédito envolve vários fatores de natureza qualitativa e quantitativa. Os fatores qualitativos na análise de crédito são representados pelos cinco (5) C's de crédito (capacidade, caráter, capital, condições e colateral), por outro lado, os itens quantitativos a serem avaliados utilizam modelos estatísticos para auxiliar na decisão de crédito. (Ribeiro, Zani & Zanini, 2009)

Estes modelos estatísticos também podem ser denominados *Credit Scoring*. Brown et al. (2010), definem o *credit scoring* como a pontuação de crédito que se traduz numa medida conclusiva das informações qualitativas e quantitativas recebidas sobre o mutuário. Esta medida sumária de pontuação ou *score* é considerada como um indicador do risco de crédito no futuro (normalmente 1 ano).

De acordo com de Oliveira Rita, Gorla e Hein (2015), os resultados dos modelos estatísticos servem como apoio à análise de crédito, pois seus resultados mostram a probabilidade de ocorrer ou não o incumprimento. A análise global de crédito deve ter bases sólidas para minimizar o incumprimento e, conseqüentemente, a perda do crédito.

A literatura sobre gestão apresenta variados modelos estatísticos para essa previsão, sendo que para a presente dissertação será usado o modelo de regressão logística. Este modelo tem sido usado por vários autores, por motivos que se apresentam a seguir.

Autores como Bertucci, Guimarães e Bressan (2003), Araújo e Carmona (2007), Camargos, Camargos e Araújo (2012), Alves e Camargos (2014), avaliaram os fatores condicionantes de incumprimento nas operações de crédito concedidas por instituições de crédito brasileiras, tendo para tal usado o modelo de regressão logística.

Silva, Lopes, Correia e Faria (2020), justificam o uso do modelo de regressão logística para avaliar o risco de incumprimento dos empréstimos ao consumo em uma instituição financeira portuguesa, pelo fato deste modelo fornecer um tratamento apropriado a existência de correlações nos dados utilizados, pois tal correlação invalida a utilização de modelos estatísticos que requerem uma suposição de observações independentes.

Paiva (2015), no estudo que objetiva propor a utilização de um modelo de regressão logística para estimar esta probabilidade de perda, concluiu que a utilização do modelo de regressão logística pela instituição financeira poderá evitar grupos de clientes com índices de incumprimento altos, a fim de se minimizar prejuízos financeiros e aumentar sua rentabilidade.

Monteiro, Vereda e Pizzinga (2019) justificam a escolha do modelo *logit* na previsão de incumprimento afirmando que “o modelo *logit* apresenta diversas propriedades que o tornam bastante atrativo. Por exemplo, o processo de estimação não envolve maiores complicações computacionais. Outra vantagem deste modelo é a sua maior amplitude de interpretação, já que pode ser de interesse não só analisar a probabilidade p_i , mas também a chance definida por $\frac{p_i}{1-p_i}$. No modelo *logit*, esta chance é facilmente obtida tomando-se antilogaritmos nos dois lados da função de ligação.”

Martins (2018), desenvolveu um modelo de *rating* para medir a capacidade das empresas cumprirem com as suas responsabilidades. Ao selecionar o modelo referiu que, o melhor modelo a aplicar depende da amostra em análise, sendo que para Portugal de acordo com as práticas observadas no sector bancário (*benchmarking*), o modelo mais utilizado para a atribuição de notações de risco a empresas é o modelo de regressão logística.

3 Metodologia

O presente estudo pretende selecionar a técnica estatística mais adequada para validação de modelos de cálculo de probabilidade de incumprimento. Serão desenvolvidos três (3) modelos de regressão logística para prever a probabilidade de incumprimento das empresas não financeiras que atuam no mercado português que fazem parte da amostra. De seguida, serão aplicadas as técnicas estatísticas de Matriz de confusão, Curva ROC, índice AUC, *Pietra Index*, Curva CAP *Accuracy Ratio* e *BRIER Score* de modo a verificar qual destas técnicas valida de forma mais segura os modelos de previsão de PD desenvolvidos, também verificar qual modelo pode ser considerado melhor.

3.1 Modelo de Regressão Logística

Os modelos *logit* e *probit* são os mais adequados para aplicar em análise de variáveis dependentes binárias afirma Witzany (2017), e acrescenta que dentre os dois, o modelo *Logit*, ou *Logistic Regression*, é o mais popular na prática bancária, bem como em pesquisas académicas.

De acordo com Fávero et al (citado por de Oliveira Ritta et al. 2015) “a regressão logística é uma técnica estatística utilizada para descrever o comportamento entre uma variável dependente binária e variáveis independentes métricas e não métricas”.

Como a variável dependente é binária (0 ou 1) possibilita rejeitar (0), e/ou aceitar (1) conceder crédito e, também, determinada interpretação em termos de probabilidade do fenómeno investigado ocorrer ou não ocorrer, por isso, a regressão logística é muito utilizada para análise de risco de crédito (*Credit Scoring*), pois possibilita classificar o cliente de forma prévia como bom ou mau pagador. (de Oliveira Ritta et al., 2015)

Para descrever a regressão logística, Hair (citado por Caeiro 2011), afirma que uma representação dos modelos em que a variável dependente é dicotómica é a que pressupõe que a variável Y , (com $Y=1$ ou $Y=0$), é apenas uma indicação observável de uma variável não observável Y^* (variável latente) tal que:

$$Y_i^* = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + U_i = \beta X_i + U_i \quad (23)$$

De acordo com Witzany (2017), a pontuação é decomposta num valor conhecido (esperado) βX_i (classificação real) e uma alteração futura desconhecida U_i da classificação que se assume

ter média zero e uma distribuição conhecida. O valor da variável Y_i^* desencadeia o evento de incumprimento obedecendo à seguinte regra:

$$Y_i = \begin{cases} 1, & \text{se } Y_i^* \geq 0 \\ 0, & \text{se } Y_i^* < 0 \end{cases} \quad (24)$$

Neste sentido, a probabilidade de incumprimento condicionada irá resultar da seguinte expressão:

$$p_i = Pr[y_i = 1|x_i] = Pr[u_i + \beta * X_i \geq 0] = 1 - F_i(-\beta * X_i) \quad (25)$$

Onde: F_i é a função de distribuição cumulativa da variável aleatória u_i . Isto é, a função de ligação F_i transforma o *score* $z_i = \beta * x_i$, de modo a tomar valores no intervalo de $(-\infty; +\infty)$, para a correspondente probabilidade de incumprimento, valores em $(0, 1)$. Um *score* alto implica uma probabilidade de incumprimento baixa e *score* baixo significa uma maior probabilidade de incumprimento.

Se a distribuição de u_i for normal, então podemos utilizar a função de distribuição normal padrão cumulativa $F_i(x) = \Phi(x)$ que conduz ao modelo *Probit*. Em alternativa, assumindo que os resíduos seguem a distribuição logística, então leva ao modelo *logit*:

$$F_i(x) = \Lambda(x) = \frac{e^x}{1+e^x} = \frac{1}{e^{-x}+1} \quad (26)$$

$$Y_i \sim B(1, \mu_i) \quad \ln\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_n X_{ni} \quad (27)$$

3.2 Dados

Os dados usados foram disponibilizados pelo Laboratório de Investigação de Microdados do BdP (BPLIM), que iniciou a sua atividade em 2016, com a principal missão de apoiar a produção de projetos de investigação e estudos sobre a economia portuguesa, tanto por investigadores internos como externos. A estes últimos, são permitidos um acesso remoto a micro conjuntos de dados bem documentados e anonimizados, adaptados às necessidades particulares, e têm também a possibilidade de utilizar os recursos computacionais disponíveis no laboratório.

Foram disponibilizadas 3 diferentes bases de dados, sendo a primeira, a *CRC – Central Credit Responsibility*/Central de Responsabilidades de Crédito. Esta base de dados reporta todos empréstimos concedidos por todas instituições de crédito Portuguesas. Os dados são recolhidos mensalmente com o objetivo de apoiar os bancos na avaliação do risco do crédito no âmbito

da concessão do crédito. Os microdados disponíveis para investigadores externos consistem num conjunto de ficheiros com dados agregados ao nível da empresa a partir de 1999, sendo que os dados são atualizados anualmente. A informação reporta sobre o tipo de crédito, caracterizando cada responsabilidade em termos da natureza da operação, em outras palavras, esta base de dados nos permitirá conhecer o valor global do empréstimo, a maturidade e a data de conceção, estando as variáveis detalhadamente definidas na Tabela nº 9. Os dados utilizados cobrem o período de Janeiro de 2013 a Dezembro de 2018. Em termos de sociedades não financeiras (objeto de estudo), temos cerca de 3.407.024 registos em 2013 (para cerca de 321.236 empresas).

Na segunda base de dados, *CRC_A_FRM_MMMYY_COVER_V01*, está contida a informação sobre o setor de atividade das empresas que tenham empréstimos, sendo que entre 2006 e 2008, as empresas reportavam a sua principal atividade de acordo com o código da "*Classificação Portuguesa de Economia Actividades - Revisão 2.1*" (CAE 2.1). Desde 2009, as empresas reportam a sua principal atividade de acordo com o código de "*Classificação Portuguesa das Actividades Económicas - Revisão 3*" (CAE 3), além do setor de atividade, este ficheiro contém a localização por município e distrito e por último a data de constituição da empresa.

A terceira e última fonte de dados é a Central de Balanços que contem o registo de todas as sociedades residentes extintas ou em atividade e sobre a sua situação financeira. Esta base de dados contém, entre outras informações, a dimensão (1-Microempresas, 2-Pequenas empresas, 3-Empresas de média dimensão e 4-Grandes empresas), a natureza jurídica da empresa, o volume de negócio, o número de trabalhadores, grau de endividamento e a autonomia financeira.

Devido à confidencialidade da informação usada, o acesso às bases de dados acima indicadas foi possível através do aplicativo de acesso remoto ao ambiente de trabalho do Bplim "*NoMachine*" (**Anexo II**), tendo sido usado o Rstudio para manipular os dados.

3.2.1 Manipulação dos dados

O primeiro passo para o cálculo do incumprimento foi a união das 3 bases de dados que resultou num número total de observações de cerca de 21.054.846 e 51 variáveis (colunas) para um total de 218.687 empresas não financeiras. De seguida, para tornar a base de dados menos pesada e de fácil manipulação, foram selecionadas para a base de dados apenas as variáveis que influenciam diretamente o incumprimento, tendo sido excluídas, variáveis/colunas repetidas assim como variáveis com informação dos anos anteriores a 2013 (ano de início da informação

analisada). E esta ação reduziu o número de variáveis na base de dados de 51 para 29 (ver ponto 2.2.3), tendo mantido constante o número de observações. Ainda com o objetivo de tornar a base de dados mais leve, foram removidos todos N/A's, isto é, foram removidas todas linhas com falta de qualquer um dos dados numa das colunas, tendo reduzido o número de observações para cerca de 9.952.320, para 104.295 empresas não financeiras.

3.2.2 Definição da variável *default* (incumprimento)

Foi usada, como definição de incumprimento, a definição da Central de Responsabilidades de Crédito (CRC): “relativamente ao capital, decorridos no máximo 30 dias após o seu vencimento, sem que se tenha verificado a respetiva regularização, sem prejuízo das instituições poderem classificá-lo em incumprimento logo que considerem esgotadas as possibilidades de uma regularização imediata. Pode dizer-se que uma empresa incumpriu num determinado ano se durante esse mesmo ano ocorreu um evento de incumprimento. É possível que a mesma empresa possa ter mais do que um incumprimento durante o período de análise”.

Para a presente dissertação, teve-se como referência a definição acima, no entanto fez-se uma adaptação para se considerar incumprimento quando uma empresa tiver crédito vencido maior que 0 não pago há 60 dias consecutivos.

Neste sentido, de modo a satisfazer as condições acima indicadas para criar o *default*, foi necessário criar uma variável auxiliar designada “*flag*” que serviu como um indicador de existência de atrasos de pagamento, tendo sido estratificada da seguinte forma:

flag = 0 – quando o crédito vencido não pago no final de 2 meses consecutivos for igual a zero ou N/A

flag = 1 – quando o crédito vencido não pago no final de 2 meses consecutivos for maior que 0

A principal variável incumprimento foi chamada de “*default*” tendo sido assinalada com o valor de 1 quando uma empresa tem crédito vencido há mais de 60 dias consecutivos (ou seja, verifica-se flag=1 em dois meses consecutivos) e 0 no caso contrário. Na Tabela 5 apresentam-se os dados totais incluindo o incumprimento:

Tabela 5: Total de empresas com *default* e as respetivas operações

<i>Default</i>	Empresas		Operações	
	Nr de Empresas	Percentagem	Nr de Operações	Percentagem
1	275	0,3%	2.886	0,03%
0	104.020	99,7%	9.949.434	99,97%
Total	104.295	100%	9.952.320	100%

Fonte: Elaboração própria

Conforme pode notar-se na tabela acima, a percentagem de *default* é quase insignificante face ao número de operações assim como quanto ao número de empresas, tendo sido necessário extrair da amostra empresas sem incumprimento para o desenvolvimento do modelo.

3.2.3 Seleção da amostra

Para selecionar a amostra, optou-se por estratificar o crédito concedido em 4 categorias de acordo com o montante total de responsabilidade dos devedores. Foi criada uma variável designada por “*d*”, conforme ilustra a Tabela 6. Esta estratificação possibilitou ter de forma distribuída o *default* por todas as categorias e facilitar a seleção da amostra.

Tabela 6: Classes de dimensão do crédito

(valores em euros)

Classe	1	2	3	4
Definição	$d < 10.000$	$10.000 \leq d < 50.000$	$50.000 \leq d < 1.000.000$	$d \geq 1.000.000$

Tabela 7: Número de empréstimos e de empresas (com e sem *default*) para amostra

Classe	1	2	3	4	Total
Operações s/ <i>default</i>	1.859.202	2.740.272	4.368.048	981.912	9.949.434
Operações c/ <i>default</i>	498	648	1.296	444	2.886
Empresas s/ <i>default</i>	35.579	41.054	45.803	7.394	129.830
Empresas c/ <i>default</i>	40	78	120	37	275

Fonte: elaboração própria

De modo a não enviesar a amostra devido o número reduzido de *default*, a mesma foi selecionada de modo a manter o número de *default* em cada categoria, isto é, foi usada a função “*sample*” do Rstudio para selecionar as empresas de forma aleatória em cada categoria mas garantindo que todas as empresas com *default* fossem mantidas na amostra. Além de manter a quantidade de *default* em cada categoria, a seleção da amostra foi feita de modo a garantir que cada classe tivesse $default > 3\%$. Neste sentido a amostra foi constituída na totalidade por 3,37

por cento das empresas da classe 1; 5,35 por cento da classe 2; 7,62 por cento da classe 3 e 14,80 por cento da classe 4, conforme indica a Tabela 8 abaixo:

Tabela 8: Composição da amostra

Classe	1	2	3	4	Total
Empresas <i>s/default</i> na amostra	1.160	2.122	3.260	1.063	7.725
Empresas <i>c/default</i> na amostra	40	78	120	37	275
Total da amostra	1.200	2.200	3.500	1.100	8.000
Total dos dados	35.619	41.132	45.923	7.431	130.105
Percentagem da amostra em relação aos dados originais por classe (total da amostra/total dos dados)	3,37%	5,35%	7,62%	14,80%	6,15%
Percentagem de <i>default</i> na amostra (empresas <i>c/default</i> na amostra pelo total da amostra)	3,33%	3,55%	3,43%	3,36%	3,44%

Fonte: Elaboração própria

A característica de dimensão será considerada no desenho econométrico do modelo de incumprimento, de forma a garantir que a amostra não enviesada de forma significativa os resultados.

3.2.4 Apresentação das variáveis

A união das bases de dados, anteriormente descritas, resultou num total de 51 variáveis, tendo sido selecionadas 29 para compor a base de dados usada no trabalho (total); as variáveis encontram-se especificadas em características do empréstimo, e características das empresas.

As variáveis que descrevem os empréstimos, são no total 16, descritas na tabela 9, abaixo:

Tabela 9: Variáveis que caracterizam o empréstimo

Nº	Nome da variável na base de dados	Definição da variável	
1	<i>Tina</i>	<i>Anonymized firm identifier</i>	Nome da empresa (anonimizado)
2	<i>ano</i>		Ano do empréstimo
3	<i>Date</i>	<i>Reference month of the data</i>	Data de referência (mensal)
4	<i>valor_global</i>	<i>Total available credit that a firm can access</i>	Total de crédito
5	<i>valor_efectivo</i>	<i>Credit that a firm used effectively</i>	Crédito usado

Nº	Nome da variável na base de dados	Definição da variável	
6	<i>valor_potencial</i>	<i>Credit that a firm can access because of irrevocable commitments of the participating entities</i>	Limite de crédito que uma empresa pode ter acesso.
7	<i>valor_vencido</i>	<i>Non-performing credit of a firm</i>	Crédito vencido
8	<i>valor_curto_o</i>	<i>Credit with an original maturity of less than or equal to 1 year</i>	Crédito c/ maturidade não superior a 1 ano
9	<i>valor_longo_o</i>	<i>Credit with an original maturity of more than 1 year</i>	Crédito c/ maturidade superior a 1 ano
10	<i>prazomedia_o</i>	<i>The weighted average original debt maturity</i>	Maturidade média ponderada
11	<i>valor_g1</i>	<i>Credit secured by real collateral mortgaged</i>	Crédito com garantias reais hipotecárias
12	<i>valor_g2</i>	<i>Credit secured by real collateral not mortgaged</i>	Crédito garantido por garantias reais não hipotecárias
13	<i>valor_g3</i>	<i>Credit secured by financial collateral</i>	Crédito com garantia financeira
14	<i>valor_g4</i>	<i>Credit secured by personal guarantee provided by firm or individual</i>	Crédito garantido por garantia pessoal prestada por empresa ou pessoa
15	<i>valor_g5</i>	<i>Credit secured by personal guarantee granted by the state or financial institution</i>	Crédito garantido por garantia pessoal concedida pelo Estado ou instituição financeira
16	<i>valor_g6</i>	<i>Credit secured by other guarantees</i>	Crédito garantido por outras garantias

Fonte: Adaptado do Manual CRC BPLIM (2021)

As variáveis que descrevem as empresas, no total 13, são descritas na tabela 10, descrita abaixo:

Tabela 10: Variáveis que caracterizam as empresas

Nº	Nome da variável na base de dados	Definição da variável	
17	<i>nb_relacao</i>	<i>The number of active bank relationships of a firm</i>	Número de relações bancárias ativas de uma empresa
18	<i>cae3</i>	<i>Portuguese classification of Economic activities (version 3)</i>	Nomenclatura Portuguesa das Atividades Económicas (versão 3)
19	<i>district</i>	<i>District</i>	Distrito

Nº	Nome da variável na base de dados	Definição da variável	
20	<i>si_final</i>	<i>Institutional sector (SEC 2010)</i>	Sector institucional (SEC 2010)
21	<i>mindate</i>	<i>Start date of the information</i>	Data de início da informação
22	<i>ancon</i>		Ano de constituição da empresa
23	<i>Pervvn</i>		Volume de negócio da atividade
24	<i>dimcomissao</i> (1,2,3,4)		Categoria de dimensão da empresa (1-micro, 2-pequena, 3-média e 4-grande)
25	<i>VF16320</i>		Autonomia financeira
26	<i>VF16323</i>	<i>Assets to equity ratio</i>	Taxa de endividamento (Rácio de ativos em relação ao capital próprio)
27	<i>VF16360</i>	<i>Degree of other financial income leverage</i>	Grau de alavancagem de outros rendimentos financeiros
28	<i>VF16375</i>	<i>Employee</i>	Pessoal
29	<i>VF17747</i>		Volume de negócios/Ativo

Fonte: Adaptado do Manual CB BPLIM (2021)

Tendo as variáveis nas tabelas acima, foram criados três (3) modelos de regressão diferentes. Estes modelos foram criados de modo a conter as características tanto da empresa como dos empréstimos. A seleção das variáveis obedeceu a um processo de escolha, que numa primeira fase foi aleatória, mas a permanência para compor o modelo dependeu da significância que cada variável tinha para o mesmo (a significância devia ser no máximo de 5%).

4 Resultados

A presente secção do trabalho divide-se em três partes, a primeira apresenta os resultados obtidos pelo desenvolvimento dos modelos de regressão logística, a segunda parte apresenta os resultados da aplicação das técnicas de validação aos modelos e na terceira parte discute-se os resultados obtidos.

4.1 Resultados dos modelos de regressão logística

De modo a ter uma análise mais abrangente dos dados da amostra e possibilitar uma escolha em termos de melhor modelo de regressão, foram desenvolvidos três modelos de regressão logística diferentes.

Começamos por estimar um modelo de base (*modeloorg*-Anexo I) composto por uma variável dependente *default*, e todas as variáveis independentes da amostra, tendo sido excluídas as variáveis que não influenciam diretamente o *default* ou que foram usadas para definir o incumprimento, nomeadamente: o nome da empresa (*tina*), ano do empréstimo (*ano*), a data de referência (*date*), valor vencido (*valor_vencido*), atividade económica (*cae3*), localização da empresa (*district*), setor institucional (*si_final*), data de início de informação da empresa (*mindate*) e o ano de constituição da empresa (*ancon*). O *modeloorgn* deu-nos a conhecer o nível de influência de cada variável para que haja ou não *default*. As variáveis com *p_value* menores foram as primeiras a compor os modelos, obedecendo um nível de significância de 1% e 5%.

Do *modeloorg*, foram desenvolvidos cerca de 15 modelos de regressão logística diferentes que correspondem a diferentes combinações entre as variáveis independentes, mantendo a variável dependente acima definida. Dos modelos desenvolvidos foram selecionados 3 modelos que foram os que tiveram todas as variáveis significativas ao nível de significância desejado e um *Akaike Information Criterion* (AIC)⁶ baixo, estes modelos têm como variáveis independentes comuns o valor global e o prazo do empréstimo.

⁶ Para comparar modelos não encaixados usa-se o *Akaike Information Criterion* (AIC):

$$AIC = -2 \ln(L) + 2k$$

Pretende-se um modelo que minimize o AIC

4.1.1 Variável dependente

O modelos de regressão logística foram desenvolvidos de acordo com a equação (25), e usados para calcular a probabilidade de incumprimento das empresas que compõem a amostra, tendo como variável dependente o Incumprimento (variável *default*) definida no ponto 2.2.2, esta variável foi calculada usando o crédito vencido, estando algumas empresas assinaladas com o número 1 *default* (mau pagador), e as outras assinaladas com 0 (bom pagador).

4.1.2 Modelo 1

As variáveis independentes usadas para estimar o modelo 1 são:

a) Variáveis independentes que caracterizam o empréstimo:

valor_global – esta variável corresponde ao crédito total da operação, a mesma foi dividida em 4 classes (d1, d2, d3 e d4) de acordo com o montante de crédito dos clientes (ver ponto 2.2.2.2), por isso a sua leitura nos resultados será feita considerando eventuais alterações da classe d2, d3 e d4 em relação a d1.

Prazomedia_o – esta variável corresponde ao prazo do empréstimo em anos tendo um mínimo de 1 ano e o máximo de 5 anos. A análise ao prazo será em unidades, isto é, aumentando ou diminuindo um ano no prazo do empréstimo até que ponto influenciará o incumprimento.

Nb_relacao – esta variável diz respeito ao número de IC com as quais a empresa tem alguma relação tendo um mínimo de 1 e máximo 33. Importa referir que os dados nessa variável concentram-se até o 3º quartil igual a 3, sendo que o máximo de 33 acima descrito é *outlier*. A análise a esta variável será de acordo com o seu aumento ou diminuição em termos de unidades para influenciar o incumprimento.

Garantias reais hipotecárias, garantias reais não hipotecárias, garantias financeiras, garantias pessoais prestadas por empresa ou pessoa singular, garantias pessoais concedidas pelo Estado ou instituição financeira e outras garantias – as variáveis que correspondem às garantias foram transformadas em variáveis binárias para *garthipo*, *garnthipo*, *garfinanc*, *garpessoa*, *garestado* e *garoutras*, respetivamente, com ponderação (1 ou 0). A sua análise corresponderá à sua presença, 1, ou não, 0, no modelo de modo a verificar a influência positiva e negativa no incumprimento.

b) Variáveis independentes que caracterizam a empresa:

Autonomia Financeira (VF16320) – corresponde à relação entre os capitais próprios e o ativo da empresa. O seu valor permite-nos compreender qual a percentagem dos ativos que está a ser financiada pelo capital próprio (Silva, E., 2011). Esta variável tem um mínimo de 0,02% e um máximo de 988% (*outlier*), neste sentido o máximo pode ser considerado até o 3º quartil que é de 244%. Será analisada quanto ao aumento ou diminuição de uma unidade de modo a verificar até que ponto influencia o incumprimento e é dada pela fórmula:

$$\text{autonomia financeira} = \frac{\text{Capital Próprio}}{\text{Ativo total}} * 100 \quad (24)$$

Volume de negócios (prevvn) – Esta variável diz respeito ao valor do volume de negócios em euros. A sua análise será consoante o aumento ou diminuição de um euro de modo a verificar até que ponto influencia o incumprimento.

Pessoal (VF16375) – esta variável corresponde ao número de trabalhadores na empresa com um mínimo na base de dados de 5, e máximo de 76. A análise a esta variável será consoante o aumento ou diminuição de um trabalhador de modo a verificar até que ponto influencia o incumprimento.

antiguidade – esta variável corresponde à antiguidade com a qual a empresa atua no mercado, foi criada com base no ano de constituição da empresa “*ancon*” pela subtração entre o ano de constituição e o ano atual 2018, variando de um mínimo de 5 anos e máximo de 136 anos. Será analisada com base no aumento de um ano de modo a verificar o grau de influência no incumprimento.

Tabela 11: Estimação dos parâmetros do modelo1

Coefficientes	Valor Estimado	Erro	P-Valor	exp(β_i) ⁷
Constante	-5.912	0.3316	<0.0001	0.002
<i>factor(d)2</i>	-1.030	0.0787	<0.0001	0.426
<i>factor(d)3</i>	-1.319	0.0765	<0.0001	0.221
<i>factor(d)4</i>	-2.365	0.1206	<0.0001	0.083
<i>prazomedia_o</i>	0.2895	0.0220	<0.0001	1.357
<i>nb_relacao</i>	0.2075	0.0109	<0.0001	1.218
<i>Autonomia Financeira</i>	-1.120	0.0867	<0.0001	0.310
<i>Volume de Negócios</i>	0.0072	0.0032	0.0220	1.012
<i>pessoal</i>	-1.576	0.1316	<0.0001	0.221

⁷ Interpretação dos coeficientes: o aumento de uma unidade numa variável preditora X_i , mantendo constantes as restantes variáveis predictoras, traduz-se numa multiplicação do odds ratio pelo fator exp(β_i)

Coefficientes	Valor Estimado	Erro	P-Valor	exp(β_i)⁸
<i>factor(garhipo)1</i>	0.2991	0.0484	<0.0001	1.459
<i>factor(garnthipo)1</i>	-0.3283	0.0610	<0.0001	0.643
<i>factor(garfinanc)1</i>	-0.7860	0.0637	<0.0001	0.495
<i>factor(garpessoa)1</i>	0.2661	0.0510	<0.0001	1.221
<i>factor(garestado)1</i>	-1.228	0.0648	<0.0001	0.261
<i>factor(garoutras)1</i>	-0.6101	0.0767	<0.0001	0.546
<i>Antiguidade</i>	0.0049	0.0014	0.0008	1.003

Fonte: Adaptado dos resultados do Rstudio

As variáveis acima foram usadas para compor o modelo1 (**ANEXO I⁹**), tendo produzido os resultados constantes na tabela 11:

De acordo Silva et al. (2020), nos modelos de regressão logística, em vez de olhar para os coeficientes per se β_i (*P_valor*), é mais importante centrar-se nos valores $exp(\beta_i)$, porque representam a influência que o aumento de uma variável independente X_i tem sobre a probabilidade da variável dependente Y tornar-se 1.

Neste sentido os resultados do *modelo1* descritos na tabela 11 mostram que o volume de negócios é uma variável significativa no modelo para 5% de significância e todas as restantes variáveis são significativas ao nível de significância de 1%.

Os resultados deste modelo indicam que aumentado o crédito na classe 4, assim como na classe 3 e/ou na classe 2 em relação à classe 1, mantendo as restantes variáveis constantes, reduz-se em 92%, 78% e 57%, respetivamente, a probabilidade de haver incumprimento relativamente à probabilidade de não haver.

Da mesma forma que, aumentando um ano no prazo do empréstimo, mantendo as restantes variáveis constantes, aumentam-se as chances de haver incumprimento em 36%, relativamente às chances de não haver.

Aumentando uma unidade no n° de relações da empresa com os outros bancos, mantendo as restantes variáveis constantes, aumentam-se as chances de haver incumprimento em 22%, relativamente às chances de não haver.

⁸ Interpretação dos coeficientes: o aumento de uma unidade numa variável preditora X_i , mantendo constantes as restantes variáveis predictoras, traduz-se numa multiplicação do odds ratio pelo fator $exp(\beta_i)$

⁹ Anexo I – composto pelos scripts dos modelos desenvolvidos

Aumentando um ano na antiguidade da empresa mantendo as restantes variáveis constantes, aumentam as chances de haver incumprimento em 0,2%, relativamente às chances de não haver.

Aumentando um euro no volume de negócios na empresa, mantendo as restantes variáveis constantes, aumentam-se as chances de haver incumprimento em 0,1% relativamente às chances de não haver.

Aumentando uma unidade no rácio da autonomia financeira da empresa, mantendo as restantes variáveis constantes, diminuem-se as chances de haver incumprimento em 0.69%.

Aumentando um funcionário no pessoal da empresa, mantendo as restantes variáveis constantes, reduzem-se as chances de haver incumprimento em 78%.

Quanto às garantias, o *modelo1* mostra que as garantias reais hipotecárias e as garantias pessoais prestadas por empresas ou pessoa particular aumentam as chances de haver incumprimento, enquanto as garantias pessoais prestadas pelo estado, as garantias financeiras e outras garantias, reduzem as chances de haver incumprimento.

Ao *modelo1* foi também feito o teste de ajustamento de *Pearson (goodness-of-fit)*. Este teve como resultado 1, sendo este valor muito superior a p-valor de 5%, com isso, não foi rejeitada a hipótese nula H_0 que expressa que o modelo se ajusta adequadamente aos dados.

O estudo feito aos resíduos¹⁰ mostrou que os mesmos não têm influência no modelo1, pois a sua variância é aproximadamente constante de 0,043205, a maior parte dos resíduos encontram-se distribuídos no intervalo de 2 e -2, sendo a percentagem dos resíduos fora do intervalo muito reduzida (0,383%), e a média dos resíduos é de -0,0678, o que significa que pela análise feita aos resíduos do modelo1 ajusta-se aos dados.

4.1.3 Modelo 2

O *modelo2* (**ANEXO I**) tem como variável dependente o incumprimento (*default*), anteriormente definida.

¹⁰ Condições dos resíduos para que o modelo seja considerado adequado

- Num bom ajustamento do modelo GLM, os resíduos devem estar distribuídos em torno de zero, sem ordem constante, com uma amplitude constante para diferentes valores de M

- Contudo, a maioria dos resíduos devem estar entre -2 e 2, a média deve ser aproximadamente zero e a sua variância deve ser unitária.

Para as variáveis independentes que caracterizam o empréstimo manteve-se do *modelo1*, *valor_global* e o *prazo do empréstimo*, tendo-se acrescentado:

valor_potencial – esta variável corresponde ao limite de crédito que uma empresa pode ter acesso, com um mínimo de 0 e máximo de 92.020.304 euros. A sua análise será feita com base no aumento de uma unidade de modo a verificar o quão tal aumento influenciará o incumprimento. (Manual CRC, 2019)

Quanto às variáveis que caracterizam a empresa manteve-se do *modelo1* o *nº de relações da empresa com outros bancos*, tendo-se adicionada ao *modelo2*:

Taxa de endividamento – de acordo com o Manual CRC (2019), esta variável corresponde ao rácio do total de ativos em relação ao passivo.

A análise a esta variável será feita no sentido de aumentar uma unidade para ver até que ponto tal aumento influenciará o incumprimento.

Grau de alavancagem – de acordo com Silva, E. (2011c) este índice indica o grau de financiamento por capitais alheios, se o resultado deste índice estiver próximo de 1 verifica-se um equilíbrio entre capital próprio e alheio e se maior representa um endividamento elevado e menor solidez financeira na empresa, também será analisado quanto ao aumento de uma unidade de modo a verificar qual a percentagem a influenciar o incumprimento e é dado pela fórmula:

$$\text{grau de alavancagem} = \frac{\text{Passivo total}}{\text{Capital próprio}} * 100 \quad (25)$$

Assim o *modelo2* tem os resultados apresentados na tabela 12:

Tabela 12: Estimação dos parâmetros do modelo2

Coefficientes	Valor Estimado	Erro	P-Valor	exp(βi)
Constante	-5.912	0.3316	<0.0001	0.003
factor(d)2	-0.087	0.0620	<0.0001	0.356
factor(d)3	-1.548	0.0645	<0.0001	0.310
factor(d)4	-2.521	0.0914	<0.0001	0.221
prazomedia_o	0.304	0.0218	<0.0001	1.480
valor_potencial	-0.0001	0.0071	<0.0001	0.999
nb_relacao	0.147	0.0099	<0.0001	1.108
Taxa de endividamento	1.104	0.0017	<0.0001	1.011
Grau de alavancagem	0.231	0.0022	<0.0001	0.023

Fonte: Adaptado dos resultados do Rstudio

Os resultados do modelo2, descritos na tabela 12 mostram que todas variáveis são significativas ao nível de significância de 1%.

Neste sentido os resultados do modelo2 indicam que aumentando o crédito na classe 4, assim como na classe 3 e/ou na classe 2 em relação à classe 1, mantendo as restantes variáveis constantes, reduz-se em 78%, 69% e 64%, respetivamente, a probabilidade de haver incumprimento relativamente à probabilidade de não haver.

A tabela 12 mostra também que se aumentarmos um ano no prazo de empréstimo, mantendo as restantes variáveis constantes, aumentam-se as chances de haver incumprimento em 48%, relativamente às chances de não haver.

Aumentando um euro no valor potencial, mantendo as restantes variáveis constantes, reduzem-se as chances de haver incumprimento em <0,0001%.

Se aumentarmos em uma unidade o nº de relações da empresa com outros bancos, mantendo as restantes variáveis constantes, aumentam-se as chances de haver incumprimento em 10%, relativamente as chances de não haver.

Se aumentarmos uma unidade no grau de alavancagem da empresa, mantendo as restantes variáveis constantes, aumentam-se as chances de haver incumprimento em 1%, relativamente as chances de não haver.

Da mesma forma que se aumentarmos uma unidade na taxa de endividamento, mantendo as restantes variáveis constantes, aumentam-se as chances de haver incumprimento em 2%, relativamente as chances de não haver.

As garantias, não foram incluídas no modelo2.

De modo a testar a qualidade do modelo2, foi feito o teste de ajustamento *Pearson (goodness-of-fit)*, este teve como resultado 1, sendo este valor muito superior a p-valor de 5%, não sendo rejeitada a H0: o modelo2 ajusta-se adequadamente aos dados.

O estudo feito aos resíduos mostrou que os mesmos não têm influência no modelo1, pois a sua variância é aproximadamente constante de 0,04311478, a maior parte dos resíduos encontram-se distribuídos no intervalo de 2 e -2, sendo a percentagem dos resíduos fora do intervalo de 0,00384, e a média dos resíduos é de -0,0674, o que significa que pela análise feita aos resíduos do modelo1 ajusta-se aos dados.

4.1.4 Modelo 3

Para desenvolver o modelo3 (ANEXO D) manteve-se a variável dependente incumprimento (definida anteriormente).

A parte explicativa do modelo3 é composta por algumas variáveis que caracterizam os empréstimos definidas no modelo1, como *valor_global do empréstimo*, o *prazo do empréstimo*, sendo que para as garantias, foram selecionadas garantias financeiras (*garfinanc*), pessoais prestadas pelo Estado e IF (*garestado*) e outras garantias (*garoutras*).

Quanto às variáveis explicativas que caracterizam a empresa, manteve-se do *modelo1* a autonomia financeira (*VF16320*) e do *modelo2* Grau de alavancagem (*VF16360*), tendo sido adicionadas:

dimcomissao – esta variável corresponde à dimensão da empresa, subdivide-se em 4 categorias (1-Microempresas, 2-Pequenas empresas, 3-Empresas de média dimensão e 4-Grandes empresas). A variável dimensão será analisada face à primeira categoria.

Volume de negócios/Ativo (VF17747) – esta variável financeira representa a porção do volume de negócios no ativo total da empresa, tendo um mínimo de 0 e um máximo de 86,31.

Neste sentido, os resultados do *modelo3* são apresentados na tabela 13:

Tabela 13:Estimação dos parâmetros

Coeficientes	Valor Estimado	Erro	P-Valor	exp(β_i)
<i>Constante</i>	-5.3339	0.06590	<0.0001	0.0048251
<i>factor(d)2</i>	-0.7888	0.06123	<0.0001	0.4543445
<i>factor(d)3</i>	-1.1936	0.06135	<0.0001	0.3031217
<i>factor(d)4</i>	-1.6766	0.08939	<0.0001	0.1870009
<i>prazomedia_o</i>	0.3508	0.02025	<0.0001	1.420295
<i>factor(garfinanc)1</i>	-0.6718	0.06273	<0.0001	0.5107661
<i>factor(garestado)1</i>	-1.3549	0.06605	<0.0001	0.2579651
<i>Factor(garoutras)1</i>	-0,2158	0.07311	0.0031	0.805858
<i>factor(dimcomissao)2</i>	-0.8102	-13.785	<0.0001	0.4447693
<i>factor(dimcomissao)3</i>	-1.4063	-12.481	<0.0001	0.2450297
<i>factor(dimcomissao)4</i>	-14.3625	-0.180	0.8335	insignificante
<i>autonomia financeira</i>	-1.1768	0.08301	<0.0001	0.3082393
<i>grau de alavancagem</i>	0.0191	0.00218	<0.0001	1.019293
<i>volume de negócios/ativo</i>	0.0121	0.00575	0.0341	1.012045

Fonte: Adaptado dos resultados do Rstudio

A tabela 13 mostra os resultados do *modelo3* e que indica que a dimensão da empresa na categoria 4 é uma variável insignificante no modelo. O rácio do volume de negócios/ativo é uma variável significativa no modelo para 5% de significância e todas as restantes variáveis são significativas ao nível de significância de 1%.

Neste sentido, o *modelo3* indicam que aumentado o crédito na classe 4, assim como na classe 3 e/ou na classe 2 em relação à classe 1, mantendo as restantes variáveis constantes, reduz-se em 81%, 70% e 55%, respetivamente, a probabilidade de haver incumprimento relativamente à probabilidade de não haver.

Da mesma forma que se aumentarmos um ano no prazo do empréstimo, mantendo as restantes variáveis constantes, aumenta-se as chances de haver incumprimento em 42%, relativamente às chances de não haver.

Quanto as garantias, os resultados mostram que os empréstimos garantidos por recursos financeiros ou por pessoas representadas pelo Estado ou Instituições Financeira e outras garantias, têm poucas chances de entrar em incumprimento.

A tabela 13 mostra também que quanto maior a empresa em termos de dimensão, reduzem-se as chances de haver incumprimento.

Aumentando uma unidade no grau de autonomia da empresa, mantendo as restantes variáveis constantes, reduzem-se as chances de haver incumprimento em 70%.

Se aumentarmos uma unidade no grau de alavancagem da empresa, mantendo as restantes variáveis constantes, aumentam-se as chances de haver incumprimento em 1,9%, relativamente às chances de não haver.

Aumentando uma unidade no rácio de volume de negócio/ativo da empresa, mantendo constantes as restantes variáveis, aumentam-se as chances de haver incumprimento em 1,2%, relativamente às chances de não haver.

O teste de ajustamento de *Pearson (goodness-of-fit)* para o *modelo3* foi igual a 1, sendo maior do que p-valor de 5%, pelo que não se rejeitou a hipótese de que o modelo se ajusta adequadamente aos dados.

O estudo feito aos resíduos mostra que os mesmos não têm influência no *modelo3* desenvolvido, pois tem a variância aproximadamente constante de 0,043283, a maior parte dos resíduos encontram-se distribuídos no intervalo de -2 a 2, sendo a percentagem dos resíduos

fora do intervalo de 0,00381, e a média dos resíduos é de -0,0675, o que significa que pela análise feita aos resíduos os modelos desenvolvidos ajustam-se aos dados.

Após o desenvolvimento dos modelos, sua aplicação à amostra e a realização dos testes, calculou-se a probabilidade de incumprimento¹¹. O fato da percentagem de *defaults* na amostra ser muito reduzida, cerca de 3% das empresas e 0,38% das observações (ver tabela 8), contribuiu para que as probabilidades de incumprimento calculadas através dos modelos fossem muito baixas conforme ilustra a tabela 14.

Tabela 14: Medidas das probabilidades de incumprimento previstas pelos três modelos

Modelos	Mín	Média	Máx
Modelo1	pd1		
	0,00008447	0,003832	0,06824
Modelo2	pd2		
	0,000128	0,00382	0,05644
Modelo3	pd3		
	0	0,003832	0,06271

Fonte: Adaptado dos resultados do Rstudio

A previsão sobre se um cliente é bom ou mau pagador corresponde a dois números absolutos 0 e 1, respetivamente. Estes números são possíveis de ter para cada cliente definindo um ponto de corte nas probabilidades de incumprimento calculadas. Pela definição do ponto de corte, as probabilidades de incumprimento que estiverem acima do ponto de corte serão rotuladas como 1 e as que estiverem abaixo 0. Deve ser considerado um ponto de corte e comparar cada probabilidade estimada com o ponto de corte criado, o valor mais utilizado nas literaturas é 0,5 (Hosmer & Lemeshow, 2000).

O presente trabalho, teve probabilidades de incumprimento muito baixas, (abaixo de 0,5), então não será este o valor a usar como ponto de corte. Sempre que se aumenta o ponto de corte, a sensibilidade diminui, a especificidade aumenta e vice-versa. Neste sentido, o ponto de corte a usar nos modelos deve ser o que maximiza a sensibilidade e a especificidade e em simultâneo reduz os erros, foi escolhido como ponto de corte para os modelos a probabilidade de incumprimento média 0,0038.

¹¹ A probabilidade de incumprimento foi calculada usando o comando do Rstudio “predict”, de modo a ter os valores previstos de incumprimento *scoring*.

4.2 Resultados da aplicação das técnicas de validação aos modelos de incumprimento

A presente secção apresenta os resultados da aplicação das técnicas estatísticas de validação aos modelos de regressão desenvolvidos. Os resultados da aplicação destas técnicas nos possibilitarão verificar a eficiência dos modelos para estimar a PD.

4.2.1 Matriz de Confusão-MC

Conforme referido anteriormente, a MC é apresentada em forma de tabela, tendo sido criadas 3 matrizes de confusão para cada modelo.

Tabela 15: Modelo 1

		Previstos		
		1	0	Total
Observados	1	2 010 (69,6%)	876 (30,4%)	2 886 (100%)
	0	268 788 (35,8%)	481 470 (64,2%)	750 258 (100%)
	Total	270 798	482 346	753 144

Fonte: Adaptado dos resultados do Rstudio

Tabela 16: Modelo2

		Previstos		
		1	0	Total
Observados	1	1 998 (69,2%)	888 (30,8%)	2 886 (100%)
	0	271 638 (36,2%)	478 620 (63,8%)	750 258 (100%)
	Total	273 636	479 508	753 144

Fonte: Adaptado dos resultados do Rstudio

Tabela 17: Modelo 3

		Previstos		
		1	0	Total
Observados	1	1 950 (67,6%)	936 (32,4%)	2 886 (100%)
	0	288 012 (38,4%)	462 246 (61,6%)	750 258 (100%)
	Total	289 962	463 182	753 144

Fonte: Adaptado dos resultados do Rstudio

A percentagem de observações em que se verificou efetivamente incumprimento é de $2886/753144 = 0,0038$, ou seja apenas se verificou *default* em 0,38% das observações na amostra, pelo que esta proporção levou a ter de se definir este valor como ponto de corte para reduzir os erros na classificação e maximizar a sensibilidade e a especificidade. Neste sentido as percentagens de classificações corretas (CTA) para os três modelos variam entre 61,6% e 64,2% e as taxas aparentes de erro (1-CTA), para cada modelo variam entre 35,8% e 38,4%.

A sensibilidade nos três modelos varia entre 67,6% e 69,6% a capacidade dos modelos prever um incumprimento que realmente é um incumprimento é razoável. Quanto a especificidade, nos três modelos varia entre 61,61% e 64,17%, também pode-se considerar razoável a capacidade dos modelos classificarem um cliente como bom pagador, quando realmente é um bom pagador.

4.2.2 Curva ROC

Figura 4: Curva ROC do modelo1

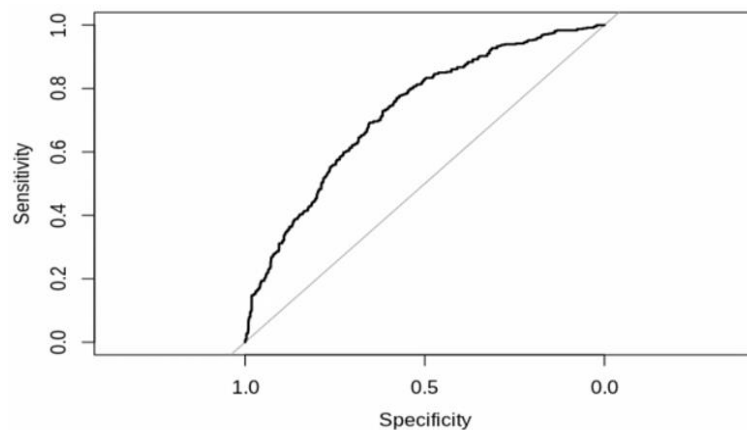


Figura 5: Curva ROC do modelo2

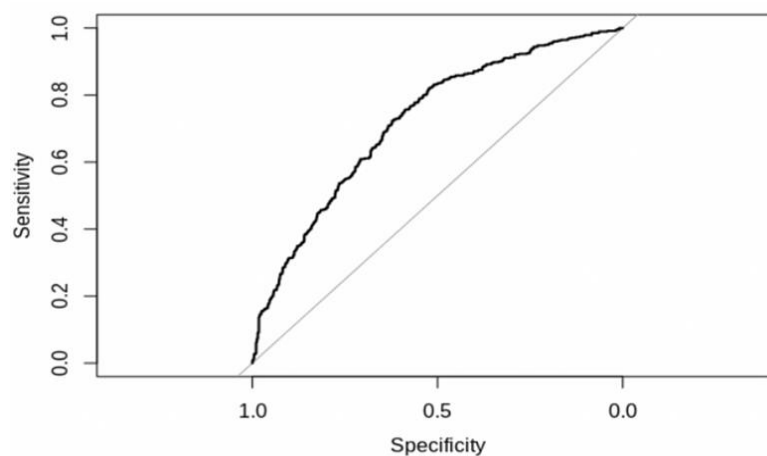
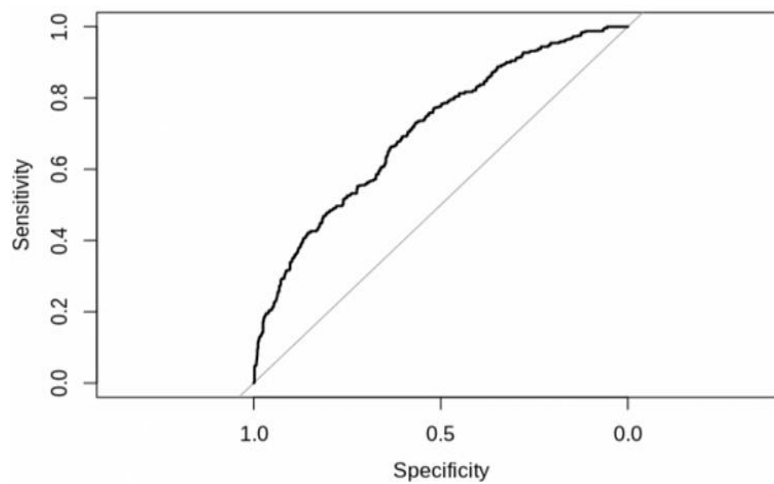


Figura 6: Curva ROC do modelo3



As figuras 4, 5 e 6 apresentam os resultados das curvas ROC dos modelos 1,2 e 3, respectivamente. Nos três gráficos nota-se que a curva ROC encontra-se acima da reta que representa o modelo aleatório, o que significa que estamos perante modelos com algum grau de sensibilidade quanto a discriminação de bons e maus pagadores. Os valores da Área abaixo da curva AUC nos três modelos situa-se entre 64,56% e 66,96%, o que mostra um grau de sensibilidade bom dos modelos.

4.2.3 Curva CAP Accuracy ratio

Figura 7: Curva CAP do modelo1

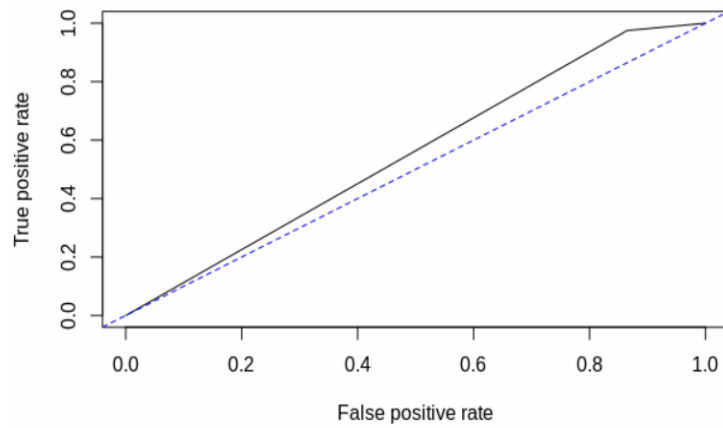


Figura 8: Curva CAP do modelo2

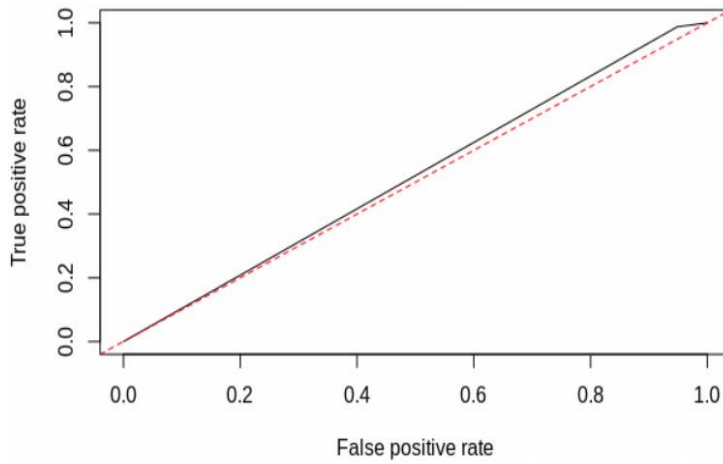
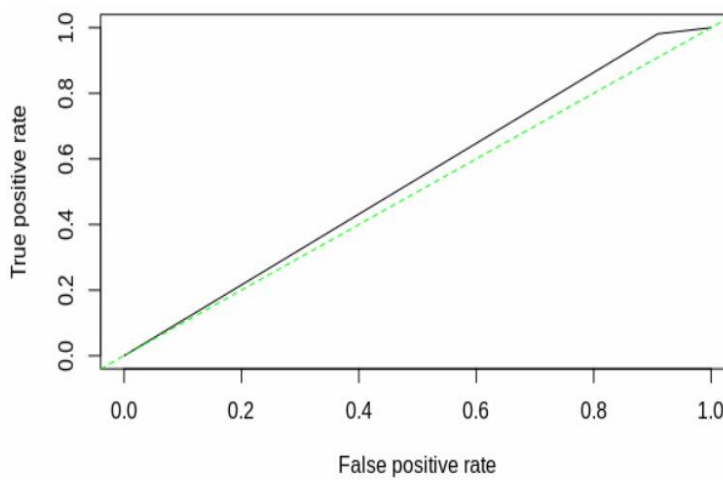


Figura 9: Curva CAP do modelo3



Os gráficos da curva CAP para os três modelos (Figuras 7, 8 e 9), são descritos por curvas que se situam acima do modelo aleatório, sendo que nos três modelos a curva CAP do modelo em

análise é descrita pela cor preta e o modelo aleatório com uma cor diferente e tracejado. Isto significa que os modelos desenvolvidos são considerados válidos de acordo com a análise das curvas CAP apresentadas.

Nota-se a proximidade das curvas CAP nos três modelos à reta de 45° (modelo aleatório), o que pode significar um poder discriminante baixo, sendo esta aproximação mais notável na curva CAP do *modelo2* (que pode ser justificado pelo fato deste modelo ser composto por um número de variáveis reduzidas do que os restantes modelos), isto significa que este modelo poderá ser classificado com uma baixa qualidade na discriminação em relação aos restantes modelos, pois para esta técnica de validação quanto mais distante do modelo aleatório a curva CAP do modelo estiver, maior qualidade tem o modelo.

Tabela 18: Resultado das medidas de validação para os modelos de cálculo da PD

Modelo	Matriz de Confusão		ROC-AUC	Pietra Index	CAP Accuracy ratio	BRIER Score
	Indicadores	Corte 0,0038				
Modelo1	Precisão	0,74%	66,91%	0,120	67,93%	5,557*10 ⁽⁻¹⁹⁾
	Sensibilidade	69,65%				
	Especificidade	64,17%				
Modelo2	Precisão	0,73%	65,61%	0,117	56,98%	3,2796*10 ⁽⁻¹⁹⁾
	Sensibilidade	69,23%				
	Especificidade	63,79%				
Modelo3	Precisão	0,67%	64,56%	0,103	63,18%	1,3407*10 ⁽⁻¹²⁾
	Sensibilidade	67,57%				
	Especificidade	61,61%				

Fonte: Elaboração própria

A Tabela 18 apresenta o resumo dos resultados da aplicação das técnicas estatísticas a cada modelo de regressão logística desenvolvido para calcular as probabilidades de incumprimento. Na secção a seguir apresenta-se a discussão dos resultados obtidos.

4.3 Discussão global dos resultados da aplicação das técnicas de validação aos modelos de incumprimento

Os resultados obtidos pela aplicação da matriz de confusão nos três modelos descritos no ponto 3.2.1, mostram que o modelo 1 teria maior possibilidade de ser escolhido pois tem a sensibilidade e especificidades maiores do que nos outros modelos. Os resultados mostraram também que os modelos desenvolvidos tem a precisão baixa variando entre 0,67% e 0,75%, para os três modelos, o que significa que a percentagem de amostras positivas classificadas de forma correta sobre a soma de amostras classificadas como positivas é muito baixa.

A AUC obtida através da curva ROC mediu a relação entre a sensibilidade e especificidade dos modelos, sendo que nos modelos desenvolvidos pode ser classificada como boa nos três modelos, pois encontra-se acima de 64%. A baixa precisão dos modelos deve-se ao facto da distribuição da amostra ter uma percentagem muito reduzida de observações em que se verificou *default*. Poderíamos tentar resolver este problema aumentando o ponto de corte, mas se aumentar o ponto de corte os modelos ganham a precisão e especificidade, mas perdem sensibilidade.

O cálculo da curva ROC, permitiu calcular a classificação dos modelos para o índice de *pietra*. Este índice é definido como a metade da distância máxima entre a curva ROC e a reta de 45° (modelo aleatório) (ver gráfico 1), por isso quanto maior for este índice maior discriminação tem o modelo. O índice de *pietra* variou entre 0,10 e 0,12, sendo novamente o modelo1 o mais discriminante comparado aos restantes modelos

A acurácia dos modelos foi possível ter através do cálculo do rácio entre $\frac{a_R}{a_P}$.

a_R Representa a área entre a curva CAP do modelo que está a ser validado e a curva CAP do modelo aleatório. Os resultados de a_R nos três modelos variam entre 0,256 e 0,312.

a_P Representa a área entre a curva CAP do modelo perfeito e a curva CAP do modelo aleatória, para os três modelos esta área varia entre 0,42 e 0,46. Estes valores permitiram o cálculo das Acurácia dos modelos pela divisão entre as áreas acima indicadas tendo obtido valores que variam entre 0,57 a 0,67, neste sentido pode-se afirmar que a relação de precisão dos modelos tem uma boa qualidade, pois para a acurácia quanto mais próxima de 1/100% melhor é o modelo.

A quinta e última técnica estatística usada para validar os modelos criados, foi o *Brier score*. Os resultados para este modelo situaram-se entre 1,34077E-12 e 5,55705E-19, isto significa que os modelos de regressão logística estimados têm um poder discriminante muito alto, pois quanto mais próximo de zero, o *Brier score* estiver, maior é o poder discriminante do modelo, de acordo com este modelo o melhor modelo continuaria sendo o modelo1 pois tem o *Brier score* mais baixo.

De um modo geral, embora a estimação da probabilidade de incumprimento ter obtido valores muitos baixos acima da média, pode-se dizer que os resultados dos modelos de regressão logística não são descartáveis sendo considerados satisfatórios, embora não esperados. Quanto aos resultados das técnicas de validação aplicadas aos modelos de regressão podem ser

considerados razoáveis, pois permitiram efetuar as análises que o trabalho tencionava, bem como permitiram concluirmos sobre o tema investigado.

5 Conclusões

A presente pesquisa permitiu ver o quão importante é o processo de validação dos modelos internos de classificação de risco de crédito para as Instituições Financeiras portuguesas, pois para que estas utilizem modelos internos para calcular os requisitos de capital, o Basileia II através do banco central exige que os modelos sejam validados por um departamento interno e independente no banco.

As técnicas estatísticas de validação apresentadas neste trabalho têm como principal objetivo medir o poder discriminante de tais modelos em relação a bons e maus pagadores. Neste sentido, as conclusões desta pesquisa versam sobre a amostra selecionada composta por empresas não financeiras portuguesas, sobre os resultados dos modelos de regressão logística desenvolvidos para prever a probabilidade de incumprimento e sobre a validação dos referidos modelos de modo a decidir entre: a Curva ROC, o índice AUC, *Pietra Index*, Curva CAP *Accuracy Ratio*, Matriz de confusão e BRIER Score, qual destas técnicas estatísticas valida de forma adequada os referidos modelos.

As bases de dados disponibilizadas pelo Bplim contêm vários *outliers* e contêm várias empresas que não contêm informação, o que requer um tratamento de dados muito profundo. Os dados revelaram que no período em observação as instituições não financeiras portuguesas têm probabilidades de incumprimento muito baixas.

Neste trabalho foram apresentados alguns conceitos de regressão logística bem como a sua implementação ao nível computacional além dos conceitos necessários para a avaliação de performance e validação (Curva ROC, o índice AUC, *Pietra Index*, Curva CAP *Accuracy Ratio*, Matriz de confusão e BRIER Score) como ferramentas usadas para a escolha de modelos. Estas técnicas foram aplicadas utilizando o *software* R em todas as fases práticas do trabalho e foram obtidos três modelos de predição satisfatórios de acordo com os critérios, objetivos e comparáveis entre si. Podemos afirmar que os resultados desse trabalho comprovam a capacidade da regressão logística de prever estimativas suficientes para classificação de risco de crédito.

Quanto aos resultados da aplicação das técnicas estatísticas de validação, concluiu-se que as mesmas analisam aspetos importantes e diferentes em cada modelo para qual foram aplicados, sendo que o ROC e CAP analisam os modelos de forma profunda possibilitando uma análise

gráfica sobre a concentração dos bons pagadores em probabilidades de incumprimentos baixas e os maus pagadores em probabilidades de incumprimento altas.

Concluiu-se também que as técnicas de validação aplicadas são complementares, dado que umas privilegiam a identificação correta dos maus pagadores, outras a identificação dos bons pagadores, não sendo possível verificar a importância de uma técnica em detrimento da outra. Esta conclusão coincide com a de Tsukahara (2013).

O presente trabalho foi desenvolvido usando o Rstudio, seja para a manipulação dos dados, assim como o desenvolvimento e validação dos modelos de regressão, pois o acesso às bases de dados do Bplim é remoto com atribuição de um “user” e uma “password” para aceder ao “no machine”. O uso desta ferramenta constituiu uma limitação ao trabalho devido às restrições de acesso a alguns comandos no Rstudio, instalação de packages, copiar e colar informação de e/ou para dentro do ambiente de trabalho criado.

As probabilidades de incumprimento calculadas são muito pequenas (com uma média de 0,0038), tendo limitado a separação da amostra em duas partes, uma para o teste (70%) e a outra para a validação (30%).

Como sugestão para trabalhos futuros, poder-se-ia analisar mais técnicas estatísticas de validação mais modernas como (Medida M, árvore de decisão) de modo a verificar as alterações na decisão sobre a escolha dos modelos mais adequados.

REFERÊNCIAS BIBLIOGRÁFICAS

Referências bibliográficas

Acordos de Basileia

BCBS – Basel Committee on Banking Supervision International (1998)

BCBS – Basel Committee on Banking Supervision International (2005)

BCBS – Basel Committee on Banking Supervision International (2006)

BCBS – Basel Committee on Banking Supervision International (2011)

BCBS – Basel Committee on Banking Supervision International (2017)

Alcarva, P. (2011). *O Guia Completo sobre A Banca e as PME*. Porto-Portugal: Vida Economica Editorial. Retrieved from <http://livraria.vidaeconomica.pt>

Alegre, T. M. (2014). *Probabilidade de default em crédito à habitação Aplicação de técnicas de estimação alternativas*. Porto: Trabalho Final na modalidade de Relatório de Estágio para obtenção do grau de mestre em Banca e Seguros - Universidade Católica Portuguesa. Retrieved from <http://hdl.handle.net/10400.14/19301>

Alves, C. M., & Camargos, M. A. (2014, Março). Fatores Condicionantes da Inadimplência em Operações de Microcrédito. *Revista de Administração e Contabilidade da Unisinos*, 59-74. doi:doi: 10.4013/base

Antão, P., & Lacerda, A. (2008). Avaliação dos Requisitos de Capital Sob Basileia II: O caso português. (B. d. Económicos, Ed.) *Relatório de Estabilidade Financeira*(Parte II), 207-228.

Araújo, E. A., & Carmona, C. U. (2007). Desenvolvimento de Modelos Credit Scoring com Abordagem de Regressão Logística para a Gestão da Inadimplência de uma Instituição de Microcrédito. *Contabilidade Vista & Revista*, 3, 107-131. Retrieved from <https://www.redalyc.org/pdf/1970/197014735006.pdf>

Arriaga, C., & Miranda, L. (2009). Risk and Efficiency in Credit Concession: A Case Study in Portugal. *Managing Global Transitions: International Research Journal*, 7, 307-326.

Retrieved from
<http://widgets.ebscohost.com/prod/customerspecific/ns000290/authentication/index.php?url=https%3a%2f%2fsearch.ebscohost.com%2flogin.aspx%3fdirect%3dtrue%26AuthType%3dip%2cshib%2cuid%26db%3da9h%26AN%3d44774978%26lang%3dpt-pt%26site%3dedd-live%26scope%3dsit>

Banco de Portugal. (2020). *Boletim Económico*. Lisboa: Departamento de Estudos Económicos. Retrieved from <http://www.bportugal.pt>

Banco de Portugal, M. R. (2019). *Central Credit Responsibility Database - Firm Level Data* (Vol. I). Lisboa: Banco de Portugal-Bplim. Retrieved from <https://doi.org/10.17900/CRC.FRM.Jun2019.V1>

Bertucci, L. A., Guimarães, J. B., & Bressan, V. G. (2003). Condicionantes de Adimplência em Processos de Concessão de Crédito a Micro e Pequenas Empresas. *Jour*. Retrieved from https://www.researchgate.net/publication/36407652_Condicionantes_de_Adimplencia_em_Processos_de_Concessao_de_Credito_a_Micro_e_Pequenas_Empresas

Bessis, J. (2011). *Risk Management in Banking*. Reino Unido: John Wiley & Sons, Lda. Retrieved from [https://books.google.com.br/books?hl=pt-PT&lr=&id=oq-MAjw2ezQC&oi=fnd&pg=PT19&dq=Risk+Management+in+Banking+Alcavara,+J.++\(2011\).+&ots=CqdEBtp1w9&sig=h5RXG7tbY1nqPhOdUJveLC-7Rfl#v=onepage&q&f=false](https://books.google.com.br/books?hl=pt-PT&lr=&id=oq-MAjw2ezQC&oi=fnd&pg=PT19&dq=Risk+Management+in+Banking+Alcavara,+J.++(2011).+&ots=CqdEBtp1w9&sig=h5RXG7tbY1nqPhOdUJveLC-7Rfl#v=onepage&q&f=false)

Blöchlinger, A. (2012). Validation of Default Probabilities. *Journal of Financial and Quantitative Analysis*, 1089-1123. doi:10.1017/S0022109012000324

Brier, G. W. (1950). Verification of Forecasts Expressed in Terms of Probability. *Monthly Weather Review*, 78, 1-3.

- Burgt, M. V. (2007, Março). Calibrating Low-Default Portfolios, using the Cumulative Accuracy Profile. *ABN AMRO/ Group Risk Management/Tools & Modelling*, 1-14.
- Caiado, A. C., & Caiado, J. (2006). *Gestão de Instituições Financeiras* (1ª ed.). Edições Sílabo, Lda.
- Caiado, A. C., & Caiado, J. (2008). *Gestão de Instituições Financeiras*. 2ª Edição. Lisboa: Edições Sílabo
- Caeiro, V. C. (2011). *Avaliação do risco de crédito de clientes empresariais levantamento de requisitos e estimação de modelos*. Lisboa: Dissertação de Mestrado - UNIVERSIDADE TÉCNICA DE LISBOA-INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO. Retrieved from <http://hdl.handle.net/10400.5/4392>
- Colquitti, J. (2007). *Credit Risk Management How to Avoid Lending Disasters and Maximize Earnings* (3 ed.). New YorkChicago, Estados Unidos da America: JoEtta Colquitt. doi:10.1036/0071446605
- Crouhy, M., Galai, D., & Mark, R. (2004). *Gerenciamento de Risco Abordagem conceptual e prática: Uma visão integrada dos risco de crédito, operacional e mercado*. Rio de Janeiro: BY QUALITYMARK.
- de Camargos, M. A., Araújo, E. A., & Camargos, M. C. (2012, Stembro). A Inqdiplência em um Programa de Crédito de uma Instituição Financeira Pública de Minas Gerais: Uma análise utilizando Regressão Logística. *creative commons*, 3, 473-492. doi: 10.5700/rege 475
- de Souza, S. M., & Bruni, A. L. (2008). Risco de Crédito, Capital de Giro e Solvência Empresarial: Um estudo na indústria brasileira de transformação de cobre. (U. R. BLUMEAU, Ed.) *Revista Universo Contábil*, 59-74. Retrieved from <http://www.redalyc.org/articulo.oa?id=117016816005>

Dias, R. M. (2015). *Evolução do crédito concedido e vencido em Portugal, por setores de atividade, no período de 2008 a 2012*. Lisboa: INSTITUTO POLITÉCNICO DE LISBOA - INSTITUTO SUPERIOR DE CONTABILIDADE E ADMINISTRAÇÃO DE LISBOA. Retrieved from <http://hdl.handle.net/10400.21/5323>

Diretivas do Banco de Portugal disponíveis em: <https://www.bportugal.pt/>

2006/48/CE - Relativa ao acesso à atividade das instituições de crédito e ao seu exercício

2006/49/CE - Relativa à adequação dos fundos próprios das empresas de investimento e das instituições de crédito

2007/64/CE - Relativa aos serviços de pagamento no mercado interno

648/2012 - Relativo aos derivados do mercado de balcão, às contrapartes centrais e aos repositórios de transações

575/2013 - Relativo aos requisitos prudenciais para as instituições de crédito e para as empresas de investimento

Encarnação, C. (2005). Conceito e Causas do Risco de Crédito. *Revista TOC* 58: 54-55

Encarnação, C. F. (2010). *Indicadores económico-financeiros: os impactos da alteração normativa em Portugal*. Lisboa: Tese de Mestrado em Contabilidade-ISCTE. Retrieved from <http://hdl.handle.net/10071/1851>

Engelmann, B., Hayden, E., & Tasche, D. (2003). Testing rating accuracy. *Credit risk*. 82-86. doi:10.1.1.70.6163

Falkenstein, e. (2000). *Riskcalctm for private companies: moody's default model*. New York: Crystal Carrafiello.

Fernandes, G. (2014). Validação de sistemas internos de classificação de risco de crédito sob o arcabouço prudencial de Basileia. *Revista do BNDES*, 145-180. Retrieved from <http://web.bndes.gov.br/bib/jspui/handle/1408/3111>

- Filipe, M. M. (2012). *Executou o sistema bancário português as normas orientadores dos Acordos de Basileia I e II?* Lisboa: Trabalho final de mestrado-Universidade Técnica de Lisboa-Instituto Superior de Economia e Gestão.
- Gottschalk, r., & sodré, c. M. (2005). O novo acordo da basileia no brasil e na índia uma análise comparada. *Economia Política Internacional: Análise Estratégica*, 32-41.
- Hosmer, D. W., and Lemeshow, S. 2000. *Applied Logistic Regression*. 2nd ed. New York: Wiley
- Lopes, A. L. (2013). *Risco de crédito num contexto de crise*. Porto-Portugal: Dissertação de mestrado em Contabilidade e Fianças - INSTITUTO POLITÉCNICO DO PORTO - INSTITUTO SUPERIOR DE CONTABILIDADE E ADMINISTRAÇÃO DO PORTO.
- Louhichi, A., & Bougelbene, Y. (2019, Março). Credit risk pricing and the rationality of lending decision-making within dual banking systems: A parametric approach. *Journals & Books Economic Systems*, 44. Retrieved from <https://doi.org/10.1016/j.ecosys.2019.100740>
- Machado, c. E., hórbe, n. A., marchi, j., & moura, l. G. (2017). A utilização conjunta dos modelos de gestão canvas e bsc. *Revista Científica Hermes*, 18, 186-212.
- Marinho de Souza, S., & Leal Bruni, A. (2008). Risco de Crédito, Capital de Giro e Solvência Empresarial: Um estudo na indústria brasileira de transformação de cobre. *Revista Universo Contábil*, 4, 59-74.
- Martins, d. S. (2012). *Impacto do Acordo Basileia II no financiamento bancário das PME*. Viseu: Tese de Mestrado em Finanças Empresariais-Instituto Politécnico de Viseu-Escola Superior de Tecnologia e Gestão de Viseu.
- Martins, R. R. (2018). *Modelos de Notação de Risco de Crédito - Rating de Empresas*. Lisboa: Tese de Mestrado em Matemática Financeira - FACULDADE DE CIÊNCIAS DA

- UNIVERSIDADE DE LISBOA-INSTITUTO SUPERIOR DE CIÊNCIAS DO TRABALHO E DA EMPRESA. Retrieved from <http://hdl.handle.net/10451/36833>
- Moraes, o. M., junior, g. I., neto, c. O., & gonçaves, f. R. (2017). Análise de um projeto de inovação tecnológica. *Iberoamerican Journal of Project Management*, 14-26.
- Moteiro, T. P., Vereda, L., & Pizzinga, A. (2019). Modelos de regressão logística para risco de inadimplência de doadores filantrópicos: variáveis determinantes e predição. *Revista Brasileira de Economia de Empresas / Brazilian Journal of Business Economics*, 19, 27-45. doi:139043163
- Pereira, M. F. (2012). *Abordagem ao risco de crédito no âmbito do acordo de Basileia III em Portugal*. Porto: Dissertação de mestrado-**INSTITUTO POLITÉCNICO DO PORTO-
INSTITUTO SUPERIOR DE CONTABILIDADE E ADMINISTRAÇÃO DO PORTO**.
- Pestana, F. M. (2016). *Crédito bancário concedido a empresas : o caso português*. Lisboa: Dissertação de Mestrado em Finanças - **INSTITUTO UNIVERSITÁRIO DE LISBOA-
ISCTE BUSINESS SCHOOL**. doi:201554291
- Pestana, F. M. (2016). *Crédito Bancário Concedido a Empresas: O caso Português*. Lisboa: Dissertação de mestrado-**INSTITUTO UNIVERSITÁRIO DE LISBOA-ISTEC
BUSINESS SCHOOL**.
- Pires, C. I. (2012). *Modelo de Avaliação de Risco no Sector Bancário Português*. Porto: Dissertação de Mestrado em Contabilidade e Finanças-**INSTITUTO POLITÉCNICO DO PORTO-
INSTITUTO SUPERIOR DE CONTABILIDADE E ADMINISTRAÇÃO DO PORTO**. Obtido de <http://hdl.handle.net/10400.22/7599>
- Portugal, B. d. (2020). *Boletim Económico*. Lisboa: Departamento de Estudos Económicos. Retrieved from https://www.bportugal.pt/sites/default/files/anexos/pdf-boletim/be_dez2020_p.pdf

- Ribeiro, C. F., Zani, J., & Zanini, F. A. (2009, Setembro). Estimação da probabilidade de inadimplência: Uma verificação empírica na Universidade Católica de Pelotas. *XXXIII Encontro da ENPAD*. Retrieved from http://www.anpad.org.br/diversos/down_zips/45/FIN645.pdf
- Ritta, C. d., Gorla, M. C., & Nelso, H. (2015). Modelo De Regressão Logística Para Análise De Risco De Crédito Em Uma Instituição De Microcrédito Produtivo. *Iberoamerica Journal of Industrial Enginnering*, 7, 103-122. doi:10.13084/2175-8018/ijie.v7n13p103-122
- Rodrigues, d. B. (2018). *Impacto do Basileia III no financiamento bancário das empresas portuguesas*. Lisboa: Dissertação para obtenção do grau de mestrado em Controlo de Gestão - Instituto Politécnico de Lisboa - Instituto Superior de Contabilidade e Administração de Lisboa. Retrieved from http://purl.org/coar/access_right/c_abf2
- Satin, R. I. (2010). *Evolução do Regulamento Bancário: do acordo de Basileia I ao acordo de Basileia III*. Porto Alegre: Bacharelato em Ciências Económicas-UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL-FACULDADE DE CIÊNCIAS ECONÓMICAS. Retrieved from <http://hdl.handle.net/10183/28145>
- Schlickenmaier, m. (2012). Basel III and Credit Risk Measurement: Variations Among G20 Countries. *San Diego International Law Journal*, 193-224. doi:85744327
- Silva, E. C., Lopes, I. C., Correia, A., & Faria, S. (2020). A logistic regression model for consumer default risk. *Journal of Applied Statistics*, 47, 13-15, 2879–2894. doi:10.1080/02664763.2020.1759030
- Skoglund, J., & Chen, W. (2015). *Financial Risk Management: Applications in Market, Credit, Asset and Liability Management and Firmwide Risk*. USA: John Wiley & Sons.

- Sousa, Q. H., dos Anjos, E. A., dos Santos, E. A., & Petri, S. M. (2019). Análise dos Fatores Preditivos de Risco para Inadimplência dos Cooperados em uma Cooperativa de Crédito. *revistafsa*, 62-76. Retrieved from <http://dx.doi.org/10.12819/2019.16.5.4>
- Sousa, Q. H., Petri, S. M., & dos Anjos, E. A. (2018). Análise dos Fatores Preditivos de Risco para Inadimplência dos Cooperados em uma Cooperativa de Crédito. *III Congresso de Contabilidade da UFRGS*, 1-13. Retrieved from <https://repositorio.ufsc.br/handle/123456789/192599>
- Stein, R. M. (2007). Benchmarking default prediction models: pitfalls and remedies in model validation. *Journal of Risk Model Validation*, 1, 77-113. Retrieved from http://www.rogermstein.com/wp-content/uploads/BenchmarkingDefaultPredictionModels_TR030124.pdf
- Tasche, D. (2006). Validation of internal rating systems and PD estimates. *arXiv:physics/0606071*, 1-27. Retrieved from <http://arxiv.org/abs/physics/0606071>
- Tsukahara, F. Y. (2013). *Adequação das técnicas de validação dos modelos de probabilidade de default em carteiras simuladas*. Brasil: Dissertação de Mestrado em Administração de Empresas-UNIVERSIDADE PRESBITERIANA MACKENZIE. Retrieved from <http://tede.mackenzie.br/jspui/handle/tede/592>
- Tsukahara, F. Y., Kimura, H., Sobreiro, V. A., & Zambrano, J. C. (2016). Validation of default probability models: A stress testing approach. *International Review of Financial Analysis*, 47, 70-85. Retrieved from <https://doi.org/10.1016/j.irfa.2016.06.007>
- Vieira, J. (2016). Predição do bom e do mau pagador no programa minha casa, minha vida. *Dissertação de Mestrado, Programa de Pós-Graduação em Administração - PPGA, Universidade de Brasília, Brasília*, 88 p.

Wang, G., Ma, J., Lihua, H., & Kaiquan, X. (2011). Two credit scoring models based on dual strategy ensemble trees. *Knowledge-Based Systems*, 61-68. Retrieved from <http://www.elsevier.com/locate/knosys>

Witzany, J. (2010). *Credit Risk Management and Modeling*. European Social Fund Prague & EU: Supporting Your Future. Retrieved from <https://doi.org/10.1007/978-3-319-49800-3>

ANEXOS

ANEXO – I – Scripts dos modelos de regressão logística tiradas do Rstudio

```
modeloorg<-glm(default~factor(d)+pervvn+VF16320+VF16323+VF16360+VF16375+
  VF17747+valor_efectivo_p+valor_potencial_p+valor_curto_o_p+
  valor_longo_o_p+nb_relacao+prazomedia_o+valor_efectivo_p
  +nb_relacao+factor(garhipo)+factor(garnthipo)+factor(garfinanc)
  +factor(garpessoa)+factor(garestado)+factor(garoutras)+
  +(antiguidade),
  family=binomial(link = logit),
  data=dados)
```

```
modelo1<-glm(default~factor(d)+prazomedia_o
  +nb_relacao+
  pervvn+VF16320+VF16360+VF16375+
  factor(garhipo)+factor(garnthipo)+factor(garfinanc)
  +factor(garpessoa)+factor(garestado)+factor(garoutras)+(antiguidade),
  family=binomial(link = logit),
  data=dados)
```

```
modelo2<-glm(default~factor(d)+prazomedia_o+
  nb_relacao+VF16323+valor_potencial_p+
  VF16360,
  family=binomial(link = logit),
  data=dados)
```

```
modelo3<-glm(default~factor(d)+prazomedia_o+
  factor(garfinanc)+factor(garestado)+factor(garoutras)+
  factor(dimcomissao)+
  VF16320+VF16360+VF17747,
  family=binomial(link = logit),
  data=dados)
```

Anexo II – NOMACHINE

