



FACULDADE DE CIÊNCIAS
Departamento de Matemática e Informática

Trabalho de Licenciatura em
Estatística

**Análise dos Factores que Influenciam no não Pagamento
das Contribuições dos Trabalhadores por Conta Própria
Inscritos no Sistema de Segurança Social Obrigatória**

Autor: Anibal Alexandre Costa Azevedo

Maputo, Outubro de 2025



FACULDADE DE CIÊNCIAS
Departamento de Matemática e Informática

Trabalho de Licenciatura em
Estatística

**Análise dos Factores que Influenciam no não Pagamento
das Contribuições dos Trabalhadores por Conta Própria
Inscritos no Sistema de Segurança Social Obrigatória**

Autor: Anibal Alexandre Costa Azevedo

Supervisor: Zacarias Mutombene, MSc., UP

Maputo, Outubro de 2025

Dedicatória

Dedico esta conquista, acima de tudo, a DEUS, que foi meu refúgio e minha fortaleza. Nos dias de luta, foi Ele quem me sustentou; nas noites de dúvida, quem me acalmou. Sem Sua presença constante, certamente que eu não teria chegado até aqui.

À minha mãe, mulher incansável, cujo amor não conhece limites. Seus gestos de carinho, suas palavras de incentivo, sua fé inabalável foram os alicerces que me mantiveram firme. Esta vitória é sua também.

À minha família, meu suporte diário, minha rede de apoio e esperança. Cada sorriso, cada abraço e cada "vai dar certo" me impulsionaram a seguir, mesmo quando tudo parecia difícil demais. Obrigado por estarem comigo em cada passo.

Declaração de Honra

Declaro por minha honra, que o presente trabalho de conclusão do curso é resultado da minha investigação e que o processo foi concebido para ser submetido apenas para a obtenção do grau de Licenciado em Estatística, na Faculdade de Ciências da Universidade Eduardo Mondlane.

Maputo, Outubro de 2025

(Anibal Alexandre Costa Azevedo)

Agradecimentos

Ao nosso Senhor Jesus Cristo em primeiro lugar vai o meu agradecimento, fonte de minha fé e que sempre esteve presente na minha jornada e sustentando-me nos momentos bons e difíceis. Aos meus familiares e amigos, que com eles aprendi que a caminhada é longa, porém possível de ser trilhada com muita sabedoria, paz e amor.

Aos meus queridos País... , por me terem educado e num lar com muito amor e respeito, que me ensinaram que o mais relevante é investir na educação e o retorno vem em prosperidade e felicidade.

Ao corpo todo de Docentes do Departamento de Matemática e Informática (DMI), mais particularmente ao supervisor Zacarias Mutombene, pela paciência, transmissão de conhecimento e dedicação que tornou factível a realização de um grande sonho.

Aos meus colegas, pela amizade, apoio, atenção, incentivo e por todo o conhecimento repassado e aos meus amigos por existirem e pelo privilégio que tive de os conhecer.

Por último não por ser menos importante, mas porque tudo nele subsiste agradeço a DEUS pelo dom da vida, pelo seu amor incondicional, pela família onde nasci, cresci e fui educado, pelos amigos que na sua generosidade me concedeu e pela oportunidade que tive de ingressar neste curso.

Muito obrigado

Resumo

O fraco nível de empregabilidade em Moçambique contribui para o aumento da informalidade, levando muitos cidadãos com baixa escolaridade a se tornarem trabalhadores por conta própria. Embora muitos desses trabalhadores já tenham sido sensibilizados a inscrever-se no Sistema de Segurança Social Obrigatória (INSS), a maioria não realiza os pagamentos das contribuições, mesmo com campanhas de sensibilização e facilitação dos meios de pagamento. Este estudo teve como objectivo analisar os factores que influenciam no não pagamento das contribuições dos trabalhadores por conta própria inscritos ao Sistema de Segurança Social Obrigatória. Foram inqueridos 375 trabalhadores por conta própria da província de Maputo. A maioria dos inqueridos é do sexo feminino (57,6%) e tem entre 40 e 49 anos (56,8%), atuando principalmente na pesca (25,8%) e agropecuária (20,27%). Em análise da associação, apenas a fonte, palestra, importância, prioridade, canal, motivos e benefícios tem alguma associação com o pagamento das contribuições do INSS, por conseguinte, usando a análise de regressão logística verificou-se que os factores como a percepção da importância, prioridade, canal de pagamento e motivos têm impacto significativo na decisão do pagamento das contribuições. Das variáveis significativas, verificou-se que o conhecimento da importância do INSS reduz em 77,2% as chances do TCP pagar as contribuições, enquanto considerar o pagamento prioritário reduz em até 92,3%, verificou-se ainda uma redução nas chances dos TCP que fazem transferência bancária pagarem as contribuições quando comparados aos que usam o Internet banking e uma redução de chances (62%) desses trabalhadores que tem falta de tempo comparativamente aos que não possuem condições para pagar. Com estes resultados, espera-se que o INSS possa se inspirar e desenhar estratégias adequadas para o processo de difusão do Sistema de Segurança Social, sensibilização e persuasão dos trabalhadores por conta própria, podendo assim garantir a participação massiva no pagamento das contribuições.

Palavras-chave: Trabalhador por conta própria, Pagamento das contribuições, Sistema de segurança social obrigatório, Análise de associação, Regressão Logística.

Abstract

The low level of employability in Mozambique contributes to the increase in informality, leading many citizens with low levels of education to become self-employed. Although many of these workers have already been encouraged to register with the Mandatory Social Security System (INSS), the majority do not pay their contributions, even with awareness-raising campaigns and facilitation of payment methods. This study aimed to analyze the factors that influence the non-payment of contributions by self-employed workers registered with the Mandatory Social Security System. A total of 375 self-employed workers from Maputo province were surveyed. The majority of respondents are female (57.6%) and are between 40 and 49 years old (56.8%), working mainly in fishing (25.8%) and agriculture (20.27%). In the analysis of the association, only the source, lecture, importance, priority, channel, reasons and benefits have some association with the payment of INSS contributions. Therefore, using logistic regression analysis, it was found that factors such as the perception of importance, priority, payment channel and reasons have a significant impact on the decision to pay contributions. Of the significant variables, it was found that knowledge of the importance of INSS reduces by 77.2% the chances of TCP not paying contributions, while considering priority payment reduces by up to 92.3%. There was also a reduction in the chances of TCPs who make bank transfers not paying contributions when compared to those who use Internet banking and a reduction in chances (62%) of these workers who lack time compared to those who do not have the conditions to pay. With these results, it is expected that INSS can be inspired and design appropriate strategies for the process of disseminating the Social Security System, raising awareness and persuading self-employed workers, thus ensuring massive participation in the payment of contributions.

Keywords: Self-employed, Payment of contributions, Mandatory social security system, Association analysis, Logistic regression.

Abreviaturas

- INE** - Instituto Nacional de Estatística;
INSS - Instituto Nacional de Segurança Social
OIT - Organização Internacional de Trabalho;
ONU - Organização das Nações Unidas;
Promenino - Programa de enfrentamento e combate ao trabalho infantil;
SPSS - Statistical Package for Social Science
SSB – Segurança Social Básica
SSC – Segurança Social Complementar
SSO – Segurança Social Obrigatória
SSSO – Sistema de Segurança Social Obrigatória
SISSMO – Sistema de Informações da Segurança Social de Moçambique
OR - Odds Ratio
TCO – Trabalhador por Conta de Outrem **TCP** – Trabalhador por Conta Própria
UNICEF - Fundo das Nações Unidas para a Infância.

Índice

1	Introdução	1
1.1	Contextualização	1
1.2	Definição do problema	2
1.3	Objectivos	3
1.3.1	Objectivo Geral	3
1.3.2	Objectivos específicos:	3
1.4	Justificação	3
2	Revisão de Literatura	4
2.1	Protecção Social	4
2.1.1	Conceitos básicos	6
2.1.2	Objectivos da Segurança Social Obrigatória	7
2.2	Trabalhadores por Conta Própria (TCP)	7
2.2.1	Regime dos Trabalhadores Por Conta Própria	8
2.2.2	Caracterização dos trabalhadores por conta própria	8
2.2.3	Prestações dos trabalhadores por conta própria	8
2.2.4	Processo de inscrição dos trabalhadores por conta própria	9
2.2.5	Barreiras para a canalização das contribuições	10
2.3	Análise da associação das variáveis	11
2.4	Regressão Logística Binária	13
2.4.1	Interpretação dos coeficientes estimados	14
2.4.2	Método de estimação	15
2.4.3	Verificação da qualidade do ajuste	16
2.4.4	Diagnóstico do modelo ajustado	17
2.5	Estudos relacionados	19
3	Material e métodos	20
3.1	Material	20
3.1.1	Tipo de pesquisa	20
3.1.2	População, amostragem e tamanho da amostra	21
3.1.3	Variáveis em análise	22

3.2	Métodos	23
3.2.1	Teste de independência	23
3.2.2	Testes de significância dos coeficientes	23
3.2.3	Verificação do modelo ajustado	25
3.2.4	Valores dos Pseudo - R^2	27
3.2.5	Diagnóstico do modelo estimado	27
4	Resultados e discussão	28
4.1	Resultados	28
4.2	Descrição do perfil da amostra	28
4.3	Teste de independência das variáveis	30
4.4	Regressão logística binária	34
4.4.1	Análise do modelo nulo	34
4.4.2	Análise do modelo ajustado	35
4.4.3	Avaliação de ajuste do modelo	36
4.4.4	Significância individual dos coeficientes	38
4.4.5	Interpretação e estimação do modelo ajustado	40
4.4.6	Análise residual	41
4.5	Discussão	43
5	Conclusões e Recomendações	44
5.1	Conclusões	44
5.2	Recomendações	46

Lista de Figuras

2.1	Função distribuição de probabilidades	13
4.1	Distribuição dos TCP's segundo o sexo	28
4.2	Distribuição dos TCP's segundo a idade	28
4.3	Distribuição dos TCP's segundo o nível acadêmico	29
4.4	Distribuição dos TCP's segundo o agregado familiar	29
4.5	Distribuição dos TCP's segundo a profissão	29
4.6	Representação da curva de ROC	36
4.7	Representação gráfica dos resíduos estandardizados	41
4.8	Representação gráfica dos resíduos estandardizados	42
5.1	A população economicamente activa, segundo o processo laboral, 2019	53

Lista de Tabelas

2.1	Área abaixo da curva ROC	17
3.1	Variáveis usadas no estudo	22
3.2	Tabela de classificação	26
4.1	Contingência entre Fonte×Pagamento	30
4.2	Contingência entre Palestra*Pagamento	30
4.3	Contingência entre Importância*Pagamento	30
4.4	Contingência entre Vantagens×Pagamento	31
4.5	Contingência entre Prioridades×Pagamento	31
4.6	Contingência entre Meios×Pagamento	31
4.7	Contingência entre Canal×Pagamento	32
4.8	Contingência entre Motivo×Pagamento	32
4.9	Contingência entre Empresa×Pagamento	32
4.10	Contingência entre Motivação×Pagamento	33
4.11	Contingência entre Benefícios×Pagamento	33
4.12	Contingência entre Acessibilidade×Pagamento	33
4.13	Codificação do pagamento	34
4.14	Equação do modelo nulo	34
4.15	Tabela de classificação inicial	34
4.16	Teste de significância dos coeficientes (Omnibus)	35
4.17	Resumo do modelo ajustado	35
4.18	Tabela de contingência para teste de Hosmer e Lemeshow	35
4.19	Teste Hosmer & Lemeshow	36
4.20	Tabela de classificação final	36
4.21	Área sob a curva	37
4.22	Significância individual dos coeficientes	38
4.23	Significância individual dos coeficientes - Continuação	39
4.24	Casos de anomalias listados	41
5.1	Trabalhadores por conta própria inscritos por província em moçambique, no período de 2018 - 2022	52

Capítulo 1

Introdução

Neste capítulo, fez-se uma descrição do tema do estudo, sobre o não pagamento das contribuições dos Trabalhadores por Conta Própria (TCP) inscritos para ao sistema de segurança social obrigatória, onde para além da contextualização foi definido o problema, os objectivos, bem como a justificativa para a escolha do tema do presente estudo.

1.1 Contextualização

Uma das principais conquistas da política social ao nível mundial no último século, foi a redução dos níveis de pobreza, em particular na velhice. Este resultado foi alcançado, pela introdução e desenvolvimento de sistemas de segurança social, na maioria dos casos associados aos mecanismos de seguro social, complementados com a criação de programas de assistência social, não contributivos, direccionados, sobretudo, para a protecção dos jovens e idosos que não preenchiam as condições de elegibilidade, por não terem condições para contribuir no sistema de segurança social ou que se encontravam em situação de especial vulnerabilidade. (Manuel, 2016)

Em Moçambique, o governo tem considerado a necessidade de usar a Segurança Social como instrumento redistributivo capaz de reduzir a pobreza, a desigualdade e promover o crescimento económico inclusivo. Segundo Brito (2020), o compromisso do Governo de Moçambique para responder às necessidades relativas à Protecção Social é demonstrado pelo facto de Moçambique ser signatário de várias convenções e declarações internacionais e de ter aprovado várias políticas e estratégias nacionais com foco na segurança social da população.

A protecção Social é um direito consagrado pela constituição da República de Moçambique. A Lei 4/2007, de define as bases em que assenta a Protecção Social e organiza o respectivo sistema. Segundo Brito (2020), o compromisso do Governo de Moçambique para responder às necessidades relativas à Protecção Social é demonstrado pelo facto de Moçambique ser signatário de várias convenções e declarações internacionais e de ter aprovado várias políticas e estratégias nacionais com foco na segurança social da população.

Assumindo a possibilidade da não diversificação de meios de pagamento, estaria por detrás da falta de pagamento das contribuições, e na tentativa de estimular ao Trabalhador por conta própria (TCP) a canalizar as suas contribuições mensalmente para o Sistema de Segurança Social, o Instituto Nacional de Segurança Social (INSS), adoptou diversas plataformas serviços bancários de transações, nomeadamente a carteira móvel, (M-Pesa, IZI), e Internet Banking, entre outros, que permitem que, mesmo se encontrando no seu posto de trabalho, o empregado doméstico ou outro, possa fazer o pagamento das suas contribuições, sem necessariamente se deslocar aos balcões do INSS ou do Banco.

Embora os esforços levados a cabo pela instituição, os relatórios ainda apontam que dos trabalhadores por conta própria inscritos no Sistema de Segurança Social, até o ano de 2022, apenas 18% dos inscritos, e que canalizam as suas contribuições (INSS, 2022). O que mostra de forma clara que, grande parte dos trabalhadores por conta própria não cumprem com suas obrigações. Nesse contexto, o presente estudo analisa os factores que influenciam no não pagamento das contribuições dos TCP inscritos ao Sistema de Segurança Social Obrigatório.

1.2 Definição do problema

O Governo da República de Moçambique criou a possibilidade de integração dos trabalhadores por conta própria no sistema de segurança social obrigatória gerido pelo Instituto Nacional de Segurança Social. Apesar dos índices de inscrição terem superado as metas previstas no Programa Quinquenal do Governo (PQG). Por exemplo até Dezembro de 2022, encontravam-se inscritos em todo o País 42.638 TCP, e destes apenas 7.544 (que representa 18%) é que efectivamente pagaram as contribuições.

Embora existam diversos mecanismos bancários, como transferência bancária, internet banking e serviços de carteira móvel (M-pesa, e-Mola e M-kesh) existentes no País, o relatório anual do INSS de 2022, revela que apenas 18% do universo dos Trabalhadores por Conta Própria inscritos que efectivamente canalizam as contribuições para o Sistema de Segurança Social embora a sua maioria tenha consciência da importância da Segurança Social nas suas vidas, uma proporção considerada insignificante, para aquilo que são os anseios da instituição.

Neste sentido, no presente estudo pretende-se responder principalmente a seguinte questão de pesquisa: Quais são os factores que influenciam no não pagamento das contribuições dos TCP inscritos para ao Sistema de Segurança Social Obrigatória?

1.3 Objectivos

1.3.1 Objectivo Geral

Analisar os factores que influenciam no não pagamento das contribuições dos trabalhadores por conta própria inscritos no Sistema de Segurança Social Obrigatória.

1.3.2 Objectivos específicos:

- Descrever o perfil sócio-económico e demográfico dos trabalhadores por conta própria inscritos no Sistema de Segurança Social Obrigatória, que fazem parte da amostra;
- Identificar os factores associados ao não pagamento das contribuições pelos trabalhadores por conta própria inscritos no Sistema de Segurança Social Obrigatória;
- Investigar quais são os factores que influenciam no não pagamento das contribuições pelos trabalhadores por conta própria inscritos ao Sistema de Segurança Social Obrigatória;
- Estimar os parâmetros do modelo logística para calcular a probabilidade dos trabalhadores por conta própria inscritos não pagarem as contribuições ao Sistema de Segurança Social Obrigatória.

1.4 Justificação

A escolha dos trabalhadores por conta própria para este estudo, é sustentado pelo facto destes, representarem cerca de 61% dos trabalhadores que fazem parte da população economicamente activa como descrito na figura (5.1) em anexo, e embora a legislação Moçambicana tenha criado um Sistema de Segurança Social e alargado sua cobertura a esta categoria de trabalhadores, os registos apontam para fraca adesão da inscrição no Sistema e ainda um fraco nível de aderência ao pagamento das contribuições. Deste modo, espera também poder ajudar a comunidade académica a perceber a aplicabilidade das técnicas estatísticas em diversos contextos de estudos científicos.

Espera –se igualmente que os resultados alcançados deste estudo, possam auxiliar ao INSS no sentido de desenhar melhores estratégias para captar maior número de trabalhadores por conta própria, e consciencializa-los a pagarem regularmente as contribuições ao Sistema de Segurança Social, que constitui um requisito indispensável para usufruto dos benefícios concedidos pela Segurança Social. Aos trabalhadores por conta própria, o contributo deste estudo está directamente ligado a percepção e consciencialização deste grupo alvo, sobretudo no que diz respeito a necessidade indispensável de ter Proteção Social nas suas vidas, devido aos riscos sociais presentes em todas as actividades que realiza no dia-a-dia, bem como dos benefícios decorrentes do cumprimento da obrigação contributiva ao Sistema de Segurança Social.

Capítulo 2

Revisão de Literatura

Neste capítulo é feita uma abordagem teórica dos conceitos relacionados a segurança social, onde mais sobre os pagamentos feitos pelos contribuintes ao sistema de segurança social obrigatória. Também foram abordados alguns conceitos sobre os métodos ou técnicas da estatística multivariada usadas para o desenvolvimento desse estudo.

2.1 Protecção Social

De acordo com a OIT o termo “protecção social” refere-se ao conjunto de medidas públicas que uma sociedade oferece aos seus membros, para os proteger de dificuldades económicas e sociais que sejam causadas pela ausência ou uma redução substancial do rendimento do trabalho como resultado de várias contingências (doença, maternidade, acidentes de trabalho, desemprego, invalidez, velhice e morte do ganha-pão); fornecimento de cuidados de saúde; e o fornecimento de benefícios para famílias com crianças. Portanto, no geral, segundo esta organização internacional, a protecção social deverá ser associada a instituições públicas, regime regulamentar e iniciativas destinadas a proteger indivíduos, agregados familiares e quaisquer outras unidades de relevância.

Para Quive (2007), “A protecção social é parte integrante dos direitos sociais veiculados pela Política Social de cada país”. A protecção social pode ser melhor compreendida se for classificada em duas dimensões, uma de natureza ampla e a outra restrita. A dimensão ampla diz respeito aos direitos de todos cidadãos, visando uma protecção social condigna e facilitadora de iniciativas e mecanismos numa vasta e efectiva rede de elos interligados, dispositivos e ramificações institucionais; a segunda dimensão é relativa aos direitos restritos, consagrados a grupos sociais particulares, nesta destaca-se os grupos carenciados e vulneráveis, serviços públicos e privados de segurança social a trabalhadores por conta de outrem ou por conta própria, seguros diversos, poupanças, crédito, fundos solidários, mutualidades, prevenções de riscos (Quive, 2007).

A protecção social é definida também como o conjunto de mecanismos, iniciativas e programas com o objectivo de garantir uma segurança humana digna, libertando os cidadãos de dois medos cruciais no ciclo da vida humana: 1) Medo da carência, sobretudo alimentar e profissional, seja accidental, crónica ou estrutural; 2) Medo da agressão e desprotecção física e psicológica (Francisco, 2010)

A Lei 4/2007, de 7 de Fevereiro, define que os cidadãos têm direito à Protecção Social, independentemente da cor, raça, sexo, origem étnica, lugar de nascimento, religião, grau de instrução, posição social, estado civil dos pais ou profissão. Segundo a mesma Lei, a protecção Social está estruturada em três níveis:

1. A Segurança Social Obrigatória (contributiva)

Segurança Social Obrigatória, gerida pelo Instituto Nacional de Segurança Social (INSS), abrange os trabalhadores por conta de outrem e os trabalhadores por conta própria, nacionais e estrangeiros residentes em território nacional e as respectivas entidades empregadoras;

2. A Segurança Social Básica (não contributiva)

Segurança Social Básica, gerida pelo Ministério que superintende a área da Acção Social, abrange os cidadãos nacionais incapacitados para o trabalho, sem meios próprios para satisfazer as suas necessidades básicas. Esses grupos, compostos por pessoas em situação de pobreza extrema, e mais especificamente crianças e idosos vivendo em situação de pobreza absoluta;

3. A Segurança Social Complementar

O Artigo 31 da Lei de PS estabelece que a Segurança Social Complementar (SSC) abrange, com carácter facultativo, as pessoas inscritas no sistema de segurança social obrigatória: E o Artigo 32 esclarece que a SSC visa reforçar as prestações da segurança social obrigatória, através de modalidades sujeitas à homologação pelo órgão de supervisão por proposta da entidade gestora.

2.1.1 Conceitos básicos

Proteção social - é um conjunto de políticas e programas destinados a reduzir vulnerabilidades e garantir o bem-estar da população. Seu objetivo é oferecer suporte a indivíduos e famílias em situações de risco, como pobreza, desemprego, doença ou exclusão social;

Segurança Social é um sistema de proteção pública que tem como objetivo garantir o bem-estar básico dos cidadãos, especialmente em momentos de vulnerabilidade. Ela funciona como uma rede de apoio financeiro e social para trabalhadores, famílias e pessoas em situação de risco;

Declaração de remunerações – documento que contém a informação mensal de cada trabalhador ao serviço de uma entidade empregadora ou de trabalhador por conta própria, o valor das remunerações que constituem a base de incidência contributiva e os períodos de trabalho que lhe correspondem;

Remuneração convencional – valor definido pelo trabalhador por conta própria que serve de base para a contribuição à segurança social obrigatória;

Dívida à Segurança Social – é a que diz respeito a não cumprimento ou contribuição de uma ou mais das várias obrigações contributivas das entidades empregadoras e dos trabalhadores por conta própria perante as instituições que constituem o sistema nacional da Segurança Social;

Indexante dos Apoios Sociais (IAS) -Valor de referência usado para calcular e determinar apoios sociais;

Certificado de Incapacidade Temporária (CIT) – Também conhecido por baixa médica, é o documento que comprova a incapacidade temporária para o trabalho. É enviado à Segurança Social para que possa receber o subsídio de doença;

Índice de profissionalidade – O número mínimo de dias que tem de ter trabalhado nos últimos meses para ter direito ao subsídio de doença;

Prazo de garantia – Período mínimo de trabalho com descontos para a Segurança Social necessário para poder ter acesso a determinado benefício;

Contribuinte activo - É a empresa, estabelecimento ou o trabalhador por conta própria que cumpre com as suas obrigações, ou seja, envia as folhas de remunerações e as devidas contribuições ao sistema de segurança social;

Contribuinte inscrito - É a empresa ou estabelecimento registado e titular de direitos e obrigações no sistema de segurança social;

Pensão de invalidez - É atribuída a um trabalhador que, na sequência de doença ou acidente de origem não profissional, sofreu uma diminuição permanente das suas faculdades físicas ou mentais, devidamente certificada por junta de saúde, que o torne incapaz de ganhar mais do que um terço da remuneração que um trabalhador com a mesma formação pode auferir pelo seu trabalho;

Pensão de sobrevivência - Prestação concedida aos familiares sobreviventes (viúva, viúvo, inválido e menores) pela morte de um beneficiário/pensionista do sistema, mas que reúnam determinadas condições exigidas por lei;

Pensão de velhice - Prestação mensal atribuída a um beneficiário, que tenha atingido a idade de 55 e 60 anos de idade, quer seja mulher ou homem, respectivamente, ou que não tendo atingido as idades indicadas, reúna o requisito de 30 anos de inscrição no sistema e tendo dado 300 meses de entrada de contribuições.

2.1.2 Objectivos da Segurança Social Obrigatória

De acordo com o artigo 2 da Lei 5/89, de 18 de Setembro, o Sistema de Segurança Social em Moçambique, visa essencialmente:

- Garantir a subsistência dos trabalhadores nas situações de falta ou diminuição de capacidade para o trabalho;
- Garantir a subsistência dos familiares, por morte do trabalhador ou pensionista.

A Segurança Social Obrigatória rege-se pelo princípio de solidariedade, o que significa que os trabalhadores no activo contribuem para o sistema, para suportar as despesas com prestações as pessoas em situações de ausência de capacidade para o trabalho.

2.2 Trabalhadores por Conta Própria (TCP)

Fazem parte deste grupo, as pessoas que exploram uma actividade económica, com carácter permanente ou temporário, sem colaboradores e as que prestam serviço de carácter individual a uma ou mais empresas mediante contrato de prestação de serviços. A categoria TCP nasceu como forma de enfrentar a precarização imposta pela concorrência do mercado de trabalho, posto que, como o grande contingente dos trabalhadores por conta própria encontra-se em condição de desemprego, acaba acatando as imposições capitalistas a fim de não ficar totalmente sem meios para adquirir sua subsistência (Lorena, 2013).

2.2.1 Regime dos Trabalhadores Por Conta Própria

O Artigo 58 do RSSO estabelece que: São obrigatoriamente abrangidos pelo regime estabelecido no presente Regulamento os seguintes trabalhadores:

- a) Pessoa física que explora uma actividade económica, com carácter permanente ou temporário, sem colaboradores;
- b) Quem presta serviço de carácter individual a outrem mediante contrato de prestação de serviços.

2.2.2 Caracterização dos trabalhadores por conta própria

Os trabalhadores que pertencem a categoria dos trabalhadores por conta própria, são os que no exercício das suas actividades:

- a) Podem escolher os processos e meios de trabalho, sendo estes da sua propriedade, no todo ou em parte;
- b) Não estão sujeitos a horários de trabalho, salvo se os mesmos resultarem da lei ou regulamento;
- c) Não se integram na estrutura produtiva ou cadeia hierárquica de uma única empresa, nem constituem elemento essencial ao desenvolvimento dos objectivos de qualquer entidade empregadora;
- d) Podem fazer-se substituir livremente (Artigo 59 da Lei de PS).

2.2.3 Prestações dos trabalhadores por conta própria

- a) Na doença o subsídio por doença e o subsídio por internamento hospitalar;
- b) Na maternidade, o subsídio por maternidade;
- c) Na invalidez, a pensão por invalidez;
- d) Na velhice, a pensão por velhice e a pensão reduzida;
- e) Na morte, o subsídio por morte, o subsídio de funeral e a pensão de sobrevivência.

Nota: A semelhança dos trabalhadores assalariados, embora seja complicado estipular um valor, o montante mensal das contribuições dos trabalhadores por conta própria é determinada pela aplicação da taxa de 7% sobre a remuneração convencional escolhida pelo próprio trabalhador. Os TCP devem declarar e pagar as contribuições até ao dia 10 do mês seguinte ao de referência.

2.2.4 Processo de inscrição dos trabalhadores por conta própria

Num prazo de 30 dias, que são contados a partir da data do início da actividade profissional, os trabalhadores devem declarar o respectivo exercício da actividade económica para efeitos de enquadramento e, se for caso disso, de inscrição, comprovado pela declaração do início de actividade ou de outra de natureza análoga.

Para a inscrição dos contribuintes ao Sistema de Segurança Social, os TCP devem preencher o Boletim de Identificação disponível nos serviços do INSS e no website (www.inss.gov.mz) e apresentar os seguintes documentos:

- Bilhete de identidade ou certidão de nascimento ou cédula pessoal;
- Licença de exercício de actividade ou documento emitido pelas entidades competentes equiparável a licença;
- Número Único de Identificação Tributária (NUIT).

Sempre que, à data da declaração de exercício da actividade, os trabalhadores já se encontrem inscritos no Sistema de Segurança Social, devem indicar o seu número de inscrição. Os TCP devem declarar o respectivo exercício para efeitos de enquadramento, no prazo de 60 dias a contar daquela data.

2.2.5 Barreiras para a canalização das contribuições

O desejável é que todo o cidadão que exerça qualquer actividade por conta própria, se inscreva e contribua regularmente para Sistema de Segurança Social, que de certa forma tornaria o Sistema cada vez mais robusto financeiramente, o que implicaria na melhoria das prestações que actualmente são concedidas, bem como o alargamento do leque das prestações. Segundo Organização Internacional de Trabalho (OIT, 2020), passa-se a apresentar algumas barreiras que o INSS enfrenta:

➤ **O fraco acesso à informação fidedigna**

Uma das principais barreiras das contribuições e adesão de novos TCP inscritos à segurança social contributiva é a falta de informação compreensiva, fidedigna e acessível. As campanhas de sensibilização pelo INSS aos TCP sobre a importância destes tornarem-se contribuintes têm sido demasiado esporádica, sendo necessário que se dê continuidade a este trabalho-chave.

Dada a complexidade do subsistema da segurança social contributiva, a falta de familiaridade com a lógica do seguro social, e os elevados níveis de analfabetismo, é necessário ter um acompanhamento mais próximo do INSS usando línguas nacionais ou locais. A falta de clareza sobre a segurança social contributiva, inclusive entre os TCP inscritos no INSS, também tem um impacto negativo sobre a sua continuidade com as contribuições.

➤ **Procedimentos administrativos ineficazes e ineficientes**

Para além da disseminação de informação, a forma mais eficaz de ganhar a confiança dos TCP é proporcionar serviços de qualidade que respondam às suas necessidades. Apesar do INSS ter facilitado o processo de inscrição, o processo de pagamento de contribuições são morosos. O contribuinte tem que levantar uma declaração de pagamento ao INSS, preenchê-la e deslocar-se a um banco para fazer o pagamento. Entretanto para receber o benefício, é necessário apresentar vários documentos e não existe um mecanismo eficaz principalmente deixar ficar as suas reclamações.

➤ **A falta de capacidade contributiva**

Para além da falta de informação e morosos processos de administração, a falta de capacidade contributiva constitui uma barreira importante no acesso à segurança social por parte dos TCP. Segundo o Decreto nº 14/2015, o rendimento mensal declarado não pode ser inferior ao salário mínimo do sector em que o trabalhador está enquadrado. Enquanto os TCO contribuem 3% do seu rendimento mensal, os TCP contribuem actualmente 7%, pois, são ambos empregadores e trabalhadores, economia informal é extremamente diferenciada o que leva a considerar que a contribuição de 7% é insustentável para os TCP.

2.3 Análise da associação das variáveis

Segundo Triola (2005), um dos principais objetivos de se construir uma distribuição conjunta de duas variáveis qualitativas é descrever a associação entre elas, isto é, quando se quer conhecer o grau de dependência entre elas, de modo que, se possa prever o resultado de uma delas quando se conhece a realização da outra. Para verificar a existência ou não de associação entre as variáveis usa-se o teste de independência Qui-quadrado, Exemplo da sua aplicação: Pode-se estimar a renda média de uma família moradora na cidade de Maputo, conhecendo a classe social a que ela pertence, pois sabe-se que existe uma dependência entre as variáveis: renda familiar e classe social.

Teste independência Qui-quadrado

O teste do Qui-quadrado para independência compara duas variáveis qualitativas, em que as categorias de uma das variáveis, encontram-se nas linhas e a outra variável nas colunas em uma tabela de contingência, para verificar se elas estão relacionadas (Agresti, 2002).

Os pressupostos para aplicação do teste χ^2 são usadas como regras para ajudar a reduzir o erro da decisão. Se o número total de observações:

- É $n \leq 20$ então as frequências esperadas e_{ij} não devem ser inferiores a 5 unidades;
- É $n > 20$ então não devem existir mais de 20% de células com frequências esperadas e_{ij} inferiores a 5 unidades e nenhuma célula deve ter frequência esperada inferior a 1.

Segundo Beigelman (1996), o teste Qui-quadrado é utilizado para verificar se a frequência com que um determinado acontecimento observado em uma amostra se desvia significativamente ou não da frequência com que ele é esperado. A estatística de teste é dada por:

$$\chi_{obs}^2 = \sum_{i,j=1}^C \sum_{i,j=1}^R \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \sim \chi_{(r-1)(c-1)}^2 \quad (2.1)$$

Em que:

- n_{ij} representa a frequência observada pelo investigador para a célula (i,j);
- e_{ij} representa a frequência esperada assumindo que H_0 é verdade (independência entre factores);
- r é o número de linhas (número de grupos do factor A);
- c é o número de colunas (número de grupos do factor B).

Para realização do teste, se faz necessário calcular o valor esperado (e_{ij}) que é dada pelo produto dos dois respectivos totais marginais ($n_{i\bullet} = \sum_{j=1}^C O_{ij}$ e $n_{\bullet j} = \sum_{i=1}^R O_{ij}$), de cada célula. Supondo-se que as variáveis sejam independentes, os valores esperados, na estatística do qui-quadrado de Pearson, são dados por:

$$e_{ij} = \frac{n_{i\bullet} \times n_{\bullet j}}{n} \quad (2.2)$$

Hipóteses e regras de decisão

H_0 : As variáveis em análise são independentes; vs.

H_1 : As variáveis em análise não são independentes.

Para efeitos da decisão sobre a existência ou não de alguma relação de independência, foram usadas as seguintes regras:

1. Comparando a estatística χ^2_{cal} com o valor tabulado do $\chi^2_{(L-1)(C-1)}$:
 - (a) Se o valor de χ^2 for maior que o valor crítico, rejeita-se a hipótese nula e conclui-se que as duas variáveis estão associadas;
 - (b) Se o valor de χ^2 for menor ou igual ao valor crítico, não se rejeita a hipótese nula, ou seja, as duas variáveis não estão associadas. No caso concreto em que $\chi^2 \neq 0$, a associação entre as duas variáveis existe mas ela não é significativa. E no caso em que $\chi^2 = 0$, há ausência de associação entre as duas variáveis.
2. Através do p-valor associado ao teste: Se o p-valor for menor que $\alpha = 5\%$, que é o nível de significância previamente estabelecido, então rejeita-se a hipótese nula. Se o p-valor for maior que $\alpha = 5\%$, não se rejeita a hipótese nula.

2.4 Regressão Logística Binária

Segundo Batistela *et al.*, (2009), a regressão logística é um método estatístico que descreve as relações entre a ocorrência do evento de interesse da variável resposta e uma ou mais variáveis independentes. A variável resposta (Y_i) é dicotômica, ou seja, ela apresenta duas possibilidades de resposta (sucesso ($Y=1$) e fracasso ($Y=0$)) e as variáveis independentes podem ser qualitativas ou quantitativas.

Assim obtemos as probabilidades, como $\pi_i = P(Y = 1|X = x_i)$ para sucesso $1 - \pi_i = P(Y = 0|X = x_i)$ para fracasso. Para descrever a média condicional de Y dado X com a distribuição logística, é utilizada a notação π_i (Jr; Lemeshow; Sturdivant, 2013). Como $\pi(x)$ varia entre 0 e 1, a representação gráfica da sua função pode ser descritos através de uma forma senoide (S), como ilustra a figura 2.1.

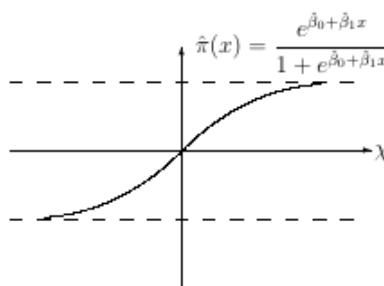


Figura 2.1: Função distribuição de probabilidades

A semelhança da regressão linear, a regressão logística também pode ser classificada como simples, quando existe apenas uma variável independente e múltipla em casos da análise envolver mais de uma variável independente, o foco desse estudo é a análise de regressão logística múltipla, pois, acredita se que, a ocorrência do não pagamento das contribuições pode ser explicado por mais de uma variável independente.

Como se trata de uma sequência de eventos com distribuição de Bernoulli, a soma do número de sucessos ou fracassos neste experimento terá distribuição Binomial de parâmetros n (número de observações) e p (probabilidade de sucesso). A função de distribuição de probabilidades Binomial é dada por:

$$P(y|n, \pi_i) = \binom{n}{y} \pi_i^y (1 - \pi_i)^{1-y} \quad (2.3)$$

onde a probabilidade da ocorrência do sucesso π_i é dada por:

$$\hat{\pi}_i = P(Y = 1|X = x) = \frac{e^{\hat{\beta}_0 + \sum_{i=1}^k \hat{\beta}_i x}}{1 + e^{\hat{\beta}_0 + \sum_{i=1}^K \hat{\beta}_i x}} \quad (2.4)$$

Deste modo a probabilidade da ocorrência do fracasso ($Y=0$), que é o objectivo desse estudo, uma vez que, pretende se estudar a probabilidade do não pagamento das contribuições ao SNSS é dada por:

$$1 - \hat{\pi}_i = P(Y = 0|X = x) = \frac{1}{1 + e^{\hat{\beta}_0 + \sum_{i=1}^K \hat{\beta}_i x}} \quad (2.5)$$

Assim sendo, já conhecidas as chances o sucesso e fracasso, segundo Hosmer e Lemeshow (2000), pode se obter o valor da razão de chances a partir da seguinte equação:

$$\mathbf{RC} = \frac{\frac{\hat{\pi}(1)}{1 - \hat{\pi}(1)}}{\frac{\hat{\pi}(0)}{1 - \hat{\pi}(0)}} = \frac{\frac{e^{(\hat{\beta}_0 + \hat{\beta}_i)}}{1 + e^{(\hat{\beta}_0 + \hat{\beta}_i)}}}{\frac{1}{1 + e^{(\hat{\beta}_0 + \hat{\beta}_i)}}} \div \frac{\frac{1}{1 + e^{\hat{\beta}_0}}}{\frac{e^{\hat{\beta}_0}}{1 + e^{\hat{\beta}_0}}} = \frac{e^{\hat{\beta}_0 + \hat{\beta}_i}}{e^{\hat{\beta}_0}} = e^{\hat{\beta}_i} \quad (2.6)$$

Assim sendo, a transformação logística, também denominada conhecida por *logit* pode ser interpretada como sendo o logaritmo da razão de probabilidades, sucesso versus fracasso ou vice versa, onde a regressão logística nos dará uma ideia do risco do contribuinte do SNSS não pagarem as contribuições dado o efeito de algumas variáveis explicativas que serão introduzidas mais à frente. A função de ligação *logit*, que consiste em linearizar o modelo logístico é dado pela seguinte equação:

$$\eta_i = \log \left(\frac{\hat{\pi}_i}{1 - \hat{\pi}_i} \right) = \log \left(\frac{\frac{e^{\hat{\beta}_0 + \sum_{i=1}^K \hat{\beta}_i x}}{1 + e^{\hat{\beta}_0 + \sum_{i=1}^K \hat{\beta}_i x}}}{\frac{1}{1 + e^{\hat{\beta}_0 + \sum_{i=1}^K \hat{\beta}_i x}}} \right) = \log e^{\hat{\beta}_0 + \sum_{i=1}^K \hat{\beta}_i x} = \hat{\beta}_0 + \sum_{i=1}^K \hat{\beta}_i x \quad (2.7)$$

2.4.1 Interpretação dos coeficientes estimados

Diferente da regressão linear, em que os coeficientes são fáceis de interpretar, as estimativas produzidas no modelo logístico são menos intuitivas (interpretadas directamente). Isso porque a transformação *logit* informa o efeito da variável independente sobre a variação do logaritmo natural da chance da variável dependente.

- Se $\hat{\beta}_{ij}$ é positivo, então a razão de chances aumenta, o que significa que, o aumento unitário em $\mathbf{X}$, aumentam as chances em $\exp(\hat{\beta}_{ij})$ vezes da ocorrência do evento em interesse.
- Se $\hat{\beta}_{ij}$ é negativo, então a razão de chances diminuí, o que significa que o aumento unitário em $\mathbf{X}$, reduzem as chances em $\exp(\hat{\beta}_{ij})$ da ocorrência do evento em interesse.

Por exemplo: Ao se considerar um coeficiente ou $\hat{\beta}_i = 0,6$, espera-se um acréscimo de 0,6 unidades no *logit* de Y - que representa a probabilidade de ocorrência de um evento de interesse Y , sempre que X aumenta uma unidade.

2.4.2 Método de estimação

Um método de estimação de parâmetros bastante utilizado, quando a forma funcional da distribuição de probabilidade das variáveis aleatórias envolvidas na modelagem é conhecida, é o de Máxima Verossimilhança (MV). A ideia básica do método, como diz seu nome, é encontrar um conjunto de estimadores que maximizem os parâmetros populacionais à luz da amostra dos dados observados. Para isso, a função de probabilidade conjunta da amostra, sob o modelo especificado, é avaliada para cada uma das observações da variável resposta, sendo tratada como uma função dos parâmetros do modelo.

O método de Máxima Verossimilhança pode é o utilizado para a estimação em regressão logística, já que se conhece a distribuição de probabilidade associada à variável resposta binária. Se houver apenas uma repetição para cada variável aleatória Y_i , então se está trabalhando sob a distribuição de probabilidade de Bernoulli descrito por (2.3). Com base nesse modelo probabilístico, e assumindo-se que as variáveis aleatórias Y_i são independentes, a função de probabilidade conjunta dos Y_i equivale à:

$$l(\beta) = \prod_{i=1}^n \hat{\pi}(x_i)^{y_i} [1 - \hat{\pi}(x_i)]^{1-y_i} \quad (2.8)$$

Onde $\pi_i = \pi(x_i)$ e aplicando o logaritmo natural ($\ln$) na função de probabilidade conjunta (*log likelihood*) descrita em (2.8), obtêm-se a seguinte equação:

$$L(\beta) = \ln [l(\beta)] = \sum_{i=1}^n \{y_i \ln [\pi(x_i)] + (1 - y_i) \ln [1 - \pi(x_i)]\} \quad (2.9)$$

Para encontrar os valores de β_i estimados que maximizam o poder preditivo do logaritmo da função de verossimilhança, diferencia-se a equação (2.9) em função dos respectivos parâmetros desconhecidos (β_0 e β_i), em que as expressões resultantes são igualados a zero. Essas equações também são conhecidas como equações de verossimilhança e são descritas de seguida:

$$\frac{\partial L(\beta)}{\partial \beta_0} = 0 \implies \sum [1 - \pi(x_i)] = 0 \quad (2.10)$$

E a outra equação é:

$$\frac{\partial L(\beta)}{\partial \beta_i} = 0 \implies \sum x_i [1 - \pi(x_i)] = 0 \quad (2.11)$$

Já que na regressão logística, as equações de verossimilhança (2.10) e (2.11) são não lineares nos seus parâmetros β_0 e β_i , o que requer a utilização de métodos especiais para solucioná-los, os referidos métodos são de natureza iterativa, por isso, torna-se necessário empregar pacotes estatísticos que são computacionais (SPSS, R ou STATA).

2.4.3 Verificação da qualidade do ajuste

Após estimar os coeficientes é necessário assegurar a significância das variáveis no modelo. Isto envolve a formulação de teste de hipóteses para determinar se as variáveis explicativas são significativamente relacionadas com a variável dependente. Veremos a seguir alguns testes:

Teste da razão de verossimilhança

Esse teste faz uma análise conjunta dos parâmetros estimados, isto é, testa simultaneamente se os coeficientes $\hat{\beta}_i$, com exceção do $\hat{\beta}_0$, são todos nulos. A comparação entre valores observados e preditos usando a função de verossimilhança é baseada na seguinte expressão:

$$D = -2 \ln \left[\frac{\text{verossimilhança do modelo ajustado}}{\text{verossimilhança do modelo saturado}} \right] \quad (2.12)$$

O motivo de se utilizar $-2 \ln$ é matemático e é necessário para se obter uma distribuição que é conhecida e, portanto, pode ser utilizada para fins de teste hipótese. Este teste é chamado teste da razão de verossimilhança (Paiva, 2015).

$$D = -2 \ln \sum_{i=1}^n \left[y_i \ln \left(\frac{\hat{\pi}}{y_i} \right) + (1 - y_i) \ln \left(\frac{1 - \hat{\pi}}{1 - y_i} \right) \right] \quad (2.13)$$

Onde o $\hat{\pi}_i$ é a estimativa de verossimilhança de π_i .

A função D, também conhecida como “deviance”, sempre será positiva e quanto menor, melhor será o ajuste do modelo. Um modelo é dito saturado quando se contém todas as variáveis, um modelo ajustado corresponde apenas as variáveis ajustada ao modelo.

Para estimar a significância de uma variável independente, deve-se comparar o valor do modelo com ou sem a variável. A mudança do valor de com a inclusão da variável é obtida por:

$$G = D \left[\frac{\text{modelo sem a variável}}{\text{modelo com a variável}} \right] \quad (2.14)$$

Teste de Wald

O teste Wald também é utilizado na regressão logística para verificar a significância estatística dos coeficientes no modelo. Desta forma, este teste verifica se cada uma das variáveis independentes apresenta uma relação estatisticamente significativa com a variável dependente. O teste de Wald é obtido comparando-se a estimativa de máxima verossimilhança do coeficiente com a estimativa do seu erro padrão:

$$Wald = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \quad (2.15)$$

Diante do comportamento inconsistente, que consiste em rejeitar a H_0 , quando o coeficiente β_i é significativo. Jennings (1986), sugere o uso do teste de verossimilhança.

2.4.4 Diagnóstico do modelo ajustado

Conforme Souza (2006) é importante que se faça uma análise dos resíduos e diagnósticos do modelo ajustado a fim de detectar possíveis problemas, como por exemplo: Presença de observações discrepantes e presença de observações influentes. As diversas técnicas de análise de diagnóstico têm como objetivo avaliar a qualidade do modelo ajustado, que passam-se a apresentar:

1. Curva do ROC

Num teste de diagnóstico existem dois tipos de erro que podem ocorrer na decisão: a escolha de uma falha (no sentido de declarar um contribuinte inactivo como activo) ou a escolha de um falso alarme (declarar um TCP activa como inactivo). Para contornar este tipo de situações, foram desenvolvidas medidas alternativas de diagnóstico com propriedades mais robustas do que a sensibilidade e a especificidade. A análise **ROC**¹ foi a técnica desenvolvida para tornar este tipo de problema (Braga, 2001).

Segundo Fávero *et al.* (2009), a curva ROC é feita por meio de um gráfico que permite estudar a variação da sensibilidade e especificidade para cada valor de corte (cut-off). A sensibilidade é apresentada no eixo das ordenadas e no eixo das abcissas. Quanto maior a área abaixo da curva ROC, maior é a capacidade de o modelo discriminar os grupos. Quanto mais próximo a curva ROC estiver da recta diagonal, pior é o poder discriminatório do modelo.

Tabela 2.1: Área abaixo da curva ROC

Área	Interpretação
$\text{Área} \leq 0.5$	Sem poder discriminativo
$[0.5 \text{ à } 0.7 [$	Discriminação fraca
$[0.7 \text{ à } 0.8 [$	Discriminação aceitável
$[0.8 \text{ à } 0.9 [$	Discriminação boa
≥ 0.9	Discriminação excepcional

Fonte: Maroco *et al.* (2007)

2. Análise residual

A análise dos resíduos compreende a um conjunto de técnicas baseadas na leitura e interpretação dos resíduos, e são utilizadas para auxiliar na verificação da validade das suposições de um modelo de regressão e, conseqüentemente, analisar a aderência e a adequação da distribuição considerada na formulação do modelo (Pereira, 2019).

¹Receiver Operating Characteristic

Segundo Cabral (2013), os resíduos de Pearson, são definidos como uma comparação entre a diferença de um valor observado e o seu respectivo valor estimado com o desvio padrão estimado para uma observação e é dado pela expressão:

$$r(y_j; \hat{\pi}_j) = \frac{(y_j - \hat{\pi}_j)}{\sqrt{\hat{\pi}_j(1 - \hat{\pi}_j)}} \quad (2.16)$$

Onde:

$\hat{\pi}_j = \hat{\pi}(x_j)\hat{y}_j$, que são os valores estimados das probabilidades para y_j . Por sua vez, o resíduo padronizado de Pearson para um padrão de covariável x_j é definido por:

$$r_{ij} = r(y_j; \hat{\pi}_j) / \sqrt{1 - h_j} \quad (2.17)$$

Onde: h_j representa o j -ésimo elemento da diagonal da matriz $(\hat{H})$, que é dada por: $\hat{H} = \hat{V}^{\frac{1}{2}} X (X' \hat{V} X)^{-1} X' \hat{V}^{\frac{1}{2}}$, e $\hat{V}$ é uma matriz diagonal com n elementos $\hat{\pi}_j(1 - \hat{\pi}_j)$ e X a matriz das variáveis independentes.

3. Distância de Cook

Segundo Agranonik (2005), essa medida é calculada através da diferença entre $\hat{\beta}$ e $\hat{\beta}_{(j)}$, que representam as estimativas de máxima verossimilhança calculadas usando todos os j padrões de covariáveis e excluindo as observações com padrão j e ainda padronizada pela matriz da covariância de $\hat{\beta}$, sendo que, para a regressão logística essa estatística é dada pela seguinte equação:

$$\Delta_j = \frac{r_j^2}{p(1 - h_j)} \times \frac{h_j}{(1 - h_j)} \quad (2.18)$$

4. DFFITS

DFFITS é uma medida alternativa à distância de Cook, e tem como objectivo medir a influência da j -ésima observação nos parâmetros de locação e a escala de modelo. O nome DFFITS vem do inglês *Difference in Fit* - é uma função do resíduo studentizado r_{ij} e da medida h_j dada por:

$$\mathbf{DFFITS}_j = r_{ij} \left\{ \frac{h_j}{(1 - h_j)} \right\}^{1/2} \quad (2.19)$$

2.5 Estudos relacionados

Na pesquisa realizada pelo Brito (2020), foi possível aferir que do total de 52780 TCP inscritos em acumulados na SSO, apenas 11785 estão activos que corresponde apenas à 22.32% dos TCP inscritos sendo que nenhum destes é Pescador artesanal e dos 29 pescadores artesanais entrevistados apenas 3 estão inscritos na SSO mas actualmente não canalizam suas contribuições para a SSO como pescadores artesanais por falta de entendimento sobre os procedimentos e as vantagens da canalização de suas contribuições para a SSO.

Segundo Quive (2009), o fraco conhecimento dos procedimento exigidos para ter acesso a segurança social por parte dos beneficiários faz com que um grande número destes se veja obrigado a prescindir dos seus direitos. O mesmo acontece devido a excessiva burocracia e a deficiente descentralização dos serviços e à localização dos postos de pagamento da segurança social distantes dos beneficiários. Outro problema tem haver com o facto dos beneficiários sentirem que as pensões que recebem não cobrem as suas necessidades básicas. Este facto contraria, em muitos casos, um dos princípios básicos da segurança social que é o da manutenção da qualidade de vida que o beneficiário tinha quando trabalhador activo.

Capítulo 3

Material e métodos

Neste capítulo foram apresentados os passos trilhados para o alcance dos objectivos do estudo, onde foi apresentado o tipo de pesquisa, a amostragem e variáveis a usar no estudo. Além disso, foram arrolados todos os métodos e testes realizados para a descrição da amostra e as técnicas da estatística multivariada, sendo que, o nível significância escolhido para decisão dos testes realizados é de 5% ($\alpha = 0.05$).

3.1 Material

Para a realização do estudo, foram usados dados primários, em que informação dos trabalhadores por conta própria inscritos ao sistema de segurança social obrigatória foram colhidos por intermédio de um questionário, estruturado em 3 partes, em que a 1^a apresenta questões de perfil dos inquiridos, a 2^a parte é caracterizada pelas questões relacionadas com os factores determinantes na contribuições dos TCP ao SNSS, e por conseguinte, a 3^a parte apresenta a questão sobre a contribuição de cada TCP inquirido.

Para o processamento dos dados será usado o pacote estatístico SPSS¹ na sua versão 20.0, e para redigir tanto o relatório do trabalho, como os slides de apresentação será utilizado o software de digitação denominado L^AT_EX.

3.1.1 Tipo de pesquisa

Existem vários aspectos ou abordagens que podem levar a classificação de um determinado estudo. Assim sendo, Prodanov e Freitas (2013), a presente pesquisa pode ser classificada em:

1. **Do ponto de vista da sua natureza**

A presente pesquisa é aplicada, pois, a semelhança desse tipo de pesquisas, esta tem como objectivo gerar conhecimentos para aplicação prática dirigidos à solução de problemas específicos que estejam relacionados com a pesquisa desenvolvida.

¹Statistical Package for Social Science

2. Do ponto de vista de seus objetivos

Este estudo pode ser classificado como uma pesquisa exploratória, pois, ela foi planejada, permitindo o estudo do tema sob diversas vertentes. Ela envolveu o levantamento bibliográfico; as entrevistas com pessoas que tiveram experiências práticas com o problema pesquisado; e foram feitas análise de um caso que estimulou a compreensão.

3. Do ponto de vista dos procedimentos técnicos

Essa é uma pesquisa experimental pelo facto dos dados terem sido fornecidos por pessoas, que também pode ser vista como um levantamento, pois, esse tipo de pesquisa ocorre quando se envolve a interrogação directa das pessoas cujo comportamento desejamos conhecer através de algum tipo de um questionário (Gil, 2008).

4. Do ponto de vista da forma de abordagem do problema

Esta é uma pesquisa quantitativa, pois, tudo foi quantificado, o que significa traduzir em números opiniões e informações para classificá-las e analisá-las, onde foram descritas percentagens, analisada a associação entre as variáveis e feita a análise de regressão.

3.1.2 População, amostragem e tamanho da amostra

A população alvo desse estudo é composta por todos os trabalhadores por conta própria inscritos no sistema nacional de segurança social, no entanto, devido ao elevado número de trabalhadores, foi usada apenas os TCP da província de Maputo onde foi empregue a amostragem aleatória simples. Antes de estimar o número de elementos a ser inquiridos por província é necessário determinar o tamanho mínimo aceitável para realizar o estudo. Segundo Crespo (1998), se a população em análise é finita ou tem número conhecido, pode-se estimar o tamanho mínimo aceitável da amostra, através da seguinte fórmula:

$$n = \frac{N \times (Z_{\frac{\alpha}{2}})^2 \times \hat{p} \times \hat{q}}{\varepsilon^2(N - 1) + (Z_{\frac{\alpha}{2}})^2 \times \hat{p} \times \hat{q}} \quad (3.1)$$

Onde: n - Representa o tamanho mínimo da amostra a seleccionar, N representa o tamanho da População em estudo, ε representa o erro da estimação, $Z_{\frac{\alpha}{2}}$ valor crítico da distribuição normal correspondente a um certo grau de confiança, $\hat{p}$ Representa a proporção dos trabalhadores que já sofreram acidente de trabalho; $\hat{q}$ é dado por $1 - \hat{p}$. Neste estudo, foi considerado o nível de confiança de 95% que corresponde $Z_{\alpha/2} = 1,96$, a população fornecida pelo INSS é finita e composta por $N = 10.218$ trabalhadores por conta própria da cidade de Maputo, o erro da estimação escolhido para esse estudo foi de $\varepsilon = 0,05$ e proporção foi determinado de acordo com os dados das tabelas (5.2).

Por conseguinte obteve-se a seguinte proporção $\hat{p} = \frac{38017}{46720} = 0,814$ e a $\hat{q} = 1 - 0,814 = 0,186$. Assim sendo, foi aplicando a formula (3.1) e obtém-se o seguinte tamanho da amostra:

$$n = \frac{10218 \times (1,96)^2 \times 0,814 \times 0,186}{0,04^2 \times (10218 - 1) + (1,96)^2 \times 0,814 \times 0,186} = \frac{5\,943,13}{16,93} = 351,04$$

Segundo Crespo (1998), se o tamanho da amostra calculado não for um número inteiro, deve-se arredondá-lo para próximo inteiro mais elevado, esse é um dos motivos que levou a definir o tamanho da amostra e nesse estudo é $n = 375$ trabalhadores que façam parte dos TCP. Desta feita, depois de estimado o tamanho da amostra. Já que os dados obtidos no INSS estavam enumerados através dos números de trabalhadores por conta própria, foi usada a amostragem aleatória simples para a seleção de cada unidade amostral.

3.1.3 Variáveis em análise

As variáveis podem ser vistas como características que variam de objecto para objecto ou indivíduo para indivíduo de acordo com o que pretende se estudar. Na presente investigação, o objecto de estudo é cada contribuinte que seja trabalhador por conta própria e as características ou variáveis escolhidas para analisar a sua participação contributiva foram:

Tabela 3.1: Variáveis usadas no estudo

Variável	Descrição da variável	Classificação
Número	Número de contribuinte	Discreta
Sexo	Sexo do contribuinte	Nominal
Idade	Intervalo das idades dos contribuintes	Contínua
Escolaridade	Nível de escolaridade do contribuinte	Categórica
Agregado	Numero de agregado familiar do contribuinte	Discreta
Profissão	Profissão do contribuinte	Categórica
Remuneração	Valor básico declarado por pagar mensalmente	Contínua
Fonte	Fonte de motivação que lhe levou a fazer a inscrição no INSS	Categórica
Palestras	Tem tido palestras sobre o INSS após sua inscrição	Categórica
Importância	Sabe a importância da SS para o cidadão	Categórica
Vantagens	Sabe as vantagens de pagar as contribuições de SS	Categórica
Prioridade	Nível de prioridade do pagamento de contribuições a SS	Categórica
Meios	Conhece os diversos meios de pagamentos de contribuições	Categórica
Canal	Canal que usa para fazer o pagamento das suas contribuições	Categórica
Motivo	Motivo que levam para o não pagamento das contribuições	Categórica
Empresa	Se é descontado no SNSS por uma empresa	Categórica
Benefício	Já teve algum benefício que resultou das contribuições	Categórica
Acessibilidade	Acha que o valor que os TCP descontam são acessíveis	Categórica
Motivação	Qual é a prestação que lhe motiva a descontar para o SNSS?	Categórica
Pagamento	Se efetuou pagamento no INSS nos últimos 3 meses	Binária

3.2 Métodos

Nesta secção apresentaram-se as técnicas da estatística multivariada e os respectivos testes realizados para o alcance de cada objectivo específico, onde usou-se o nível de significância de 5% ($\alpha = 0,05$) para a tomada de decisão de cada teste realizado. O perfil dos trabalhadores por conta própria que fizeram parte da amostra, foi descrito através de gráficos circulares e tabelas com as frequências relativas de cada uma das variáveis sócio-econômicas disponíveis na realização do presente estudo.

3.2.1 Teste de independência

Neste estudo, foi usado o teste de independência através da estatística χ^2 de Pearson, que permitiu investigar em cada duas variáveis qualitativas (das quais, uma independente e outra dependente), com objectivo de verificar se existe alguma relação de dependência ou não entre as em análise.

Sob a hipótese nula de que, as variáveis em análise são independentes, neste estudo usou-se o p-valor associado ao teste χ^2 de Pearson para a tomada da decisão, onde se o p-valor for superior ao nível de significância ($\alpha = 0,05$), não rejeita-se a hipótese nula, levando assim a concluir que, as variáveis em análise são independentes, ou em outras palavras, estas variáveis não tem relação alguma de dependência.

Análise de regressão logística

Para verificar a influência das variáveis independentes sobre a variável dependente (Contribuição), foi usada a regressão logística binária, pois, esta variável apresenta duas possíveis respostas que podem assumir os valores 1 (Para os que continuam contribuindo no sistema do INSS) e 0 (Para os que já não contribuem no sistema do INSS).

3.2.2 Testes de significância dos coeficientes

Depois de estimados através do método de máxima verossimilhança é necessário avaliar a significância das no modelo. Essa avaliação é feita de duas maneiras distintas mas que tem o mesmo objectivo, que é de verificar se os coeficientes do modelo são significativamente diferentes de zero, ou seja existe alguma influencia da variável independente para a ocorrência de algum evento da variável dependente.

1. Teste de razão de máxima verossimilhança

No processo de verificar-se a significância do coeficiente de uma variável em qualquer modelo relaciona-se com a seguinte questão: Será que o modelo incluindo as variáveis

independentes (modelo ajustado) explicam mais sobre a variável dependente, quando comparado com o modelo que não inclui as variáveis explicativas (modelo nulo)?

Segundo Hosmer e Lemeshow (1989), para responder a questão colocada acima, comparam-se os valores observados da variável resposta com aqueles preditos, por cada um dos dois modelos; o primeiro com a variável presente e o segundo sem essa variável. A comparação entre os valores preditos e observados é expressa pela equação (2.13). Essa estatística é chamada de **Deviance** e avalia o valor ajustado na regressão logística.

Para testar a significância conjunta dos coeficientes, em que usando o software estatístico SPSS, a tabela resultante desse teste é denominado como **teste Omnibus**. Esse teste foi desenvolvido com vista a decidir sobre as seguintes hipóteses:

$$H_0 : \hat{\beta}_1 = \hat{\beta}_2 = \hat{\beta}_i = 0;$$

$$H_1 : \exists \hat{\beta}_i \neq 0$$

Regra de decisão

Neste teste a regra para a tomada de decisão, foi usado o p-valor associado ao teste de razão de verossimilhança, em que, o nível de significância escolhido é de $\alpha = 0.05$, onde rejeita-se a hipótese nula de que todos os coeficientes são iguais a zero, quando o p-valor é superior a $\alpha = 0.05$.

2. Teste de Wald

Assim como a regressão linear múltipla, a regressão logística também possui um teste de significância dos coeficientes, usado para testar a hipótese nula de que o coeficiente é estatisticamente diferente de zero, onde a razão de desigualdade não muda e a probabilidade não é afectada. O teste **Wald** fornece a significância estatística para cada coeficiente estimado ($\hat{\beta}_i$), assim como acontece na regressão múltipla pelo teste T (Hair et al., 2005).

A estatística de Wald é dada pela equação (2.15) e realiza-se sob as seguintes hipóteses:

$$H_0 : \hat{\beta}_1 = 0, \hat{\beta}_2 = 0, \dots, \hat{\beta}_i = 0;$$

$$H_1 : \exists \hat{\beta}_i \neq 0$$

Regra de decisão

Para a tomada de decisão, foram usados os p-valores associados aos testes de razão de verossimilhança e de Wald, em que, a um nível de significância de 0.05, rejeita-se a hipótese nula quando o p-valor é superior a $\alpha = 0.05$.

3.2.3 Verificação do modelo ajustado

Sempre que se constrói um modelo de regressão é fundamental, antes de passar a tirar conclusões, verificar se o modelo obtido se ajusta efectivamente aos dados observados para ser usado nas estimativas seguintes. No caso da regressão logística uma ideia bastante intuitiva é calcular a probabilidade de ocorrência do acontecimento, verificar se as classificações do modelo estimado foram feitas de forma correcta, para todos os indivíduos da amostra. Assim sendo, para efeitos de verificação e avaliação do modelo ajustado foram realizados os seguintes testes e avaliações:

1. Teste Hosmer e Lemeshow

Segundo Maroco (2007), o teste de Hosmer e Lemeshow é usado para avaliar o ajuste do modelo, através da qual testa se as classificações previstas para cada grupo, dos 10 grupos criados a partir da amostra são iguais aos valores observados, onde por meio da estratificação das observações em faixa (decis) e da aplicação de um teste Qui-quadrado (χ^2) para avaliar se há diferenças significativas entre as frequências observadas e esperadas em cada faixa, e a estatística do teste é dada pela equação:

$$X_{HL}^2 = \sum_{i=1}^g \frac{(O_i - E_i)^2}{E_i} \sim \chi_{(g-2)}^2 \quad (3.2)$$

Onde: g - é o número de grupos criados; O_i - é o número de “sucessos” observados da variável dependente em cada classe; E_i - é o número de “sucessos” esperados em cada classe. O teste de Hosmer e Lemeshow testa as seguintes hipóteses:

H_0 : As classificações previstas pelo modelo e os valores observados são iguais;

H_1 : As classificações previstas pelo modelo e os valores observados não são iguais.

Regra de decisão: Se o p-valor associado ao teste de Hosmer e Lemeshow for superior ao nível de significância ($\alpha = 5\%$) não se rejeita a hipótese nula de que as classificações previstas e as observadas não se diferem.

2. Tabela de classificação

Para a criação da tabela de classificações, cria-se uma tabela cruzando os valores observados de y_i com os valores encontrados no procedimento anterior, $\hat{y}_i$ e verifica-se quantas classificações foram feitas correctamente, sendo isto um bom preditor para $\hat{y}_i = 0$ ou $\hat{y}_i = 1$. Nesta tabela de classificação, com características idênticas a tabela (3.2), pode-se tirar algumas medidas relevantes para avaliar o ajuste que são a Sensibilidade, Especificidade e a classificação global do modelo estimado:

Tabela 3.2: Tabela de classificação

		Valores previstas ($\hat{Y}$)		Percentagem
Categorias		Não	Sim	Correcta
Valores observadas (Y)	Não	A	B	Especificidade
	Sim	C	D	Sensibilidade
Percentagem global				classificação

Da tabela acima, é importante saber das medidas que quantificam as classificações correctas feitas pelo modelo ajustado, tais medidas são apresentadas de seguida:

➤ **Especificidade:**

É a medida que representa a taxa de acertos para “0”, ou seja, a probabilidade de estimar o fracasso dado que o valor real observado é realmente 0, essa medida pode ser obtida a partir da seguinte forma:

$$\text{Especificidade} = \frac{A}{A+B} \times 100 \quad (3.3)$$

➤ **Sensibilidade:**

Representa a taxa de acertos para “1”, ou seja, a probabilidade de estimar o sucesso dado que o valor real observado é realmente 1, para poder estimar o seu valor usa-se a equação (3.4).

$$\text{Sensibilidade} = \frac{D}{C+D} \times 100 \quad (3.4)$$

➤ **Classificação global:**

Esta medida é a que quantifica todos os casos correctamente classificados pelo modelo ajustado, expressando as classificação correctas tanto para o fracasso=0, assim como para o sucesso=1.

$$\text{Classificação} = \frac{A+D}{A+B+C+D} \times 100 \quad (3.5)$$

3. Curva ROC (Receiver Operating Characteristic)

Na regressão logística um dos procedimentos mais utilizadas para validar o desempenho do modelo estimado é usando a curva ROC, a qual obtê-se gerando um gráfico da especificidade e sensibilidade (taxas de acerto) das previsões do modelo e considerando diferentes pontos de corte no modelo.

Segundo Hosmer e Lemeshow (2000), a regra geral para avaliação do resultado da área sob a curva ROC de modelos estimados são descritas na Tabela (2.1), que qualifica o poder discriminativo do modelo logístico.

3.2.4 Valores dos Pseudo - R^2

Segundo Salotti e Yamoto (2005), os valores “pseudo- R^2 ”, são assim denominados, porque se assemelha ao coeficiente de determinação utilizado no modelo de regressão linear mas estes são valores de R^2 padronizados em relação ao tipo de dados usados na regressão logística.

A tabela do resumo do modelo, apresenta a estimativa Cox-Snell R^2 baseia-se no Valor de Verossimilhança, porém, é situado em uma escala que se inicia em zero, mas não chega ao limite superior 1 e o valor do Nagelkerke R^2 que é um ajuste desenvolvido no indicador anterior para que o mesmo chegue ao limite máximo de 1, Ainda segundo Corrar *et al.* (2007), o teste Nagelkerke R^2 situa-se numa escala que vai de 0 a 1 e tem a mesma finalidade do indicador Cox-Snell R^2 , é por isso que nesse estudo foi usado este valor de pseudo- R^2 para interpretar.

3.2.5 Diagnóstico do modelo estimado

A análise residual gráfica em regressão logística é usada na verificação do ajuste do modelo final aos dados e detecção de casos influentes, a representação gráfica é uma ferramenta útil, pois, já que variável resposta é dicotômica, posto isso, o resíduo do modelo de regressão logística é dado pela diferença entre a probabilidade encontrada pelo modelo e o valor observado do evento, tal como afirma (Guimarães, 2006).

Assim sendo, foram usados resíduos padronizados e a probabilidade estimada ou prevista, com a linha de **Lowess** sub imposta no gráfico, sendo que, se o modelo é bem ajustado, o gráfico dos resíduos padronizados e a probabilidade estimada ou prevista, deve mostrar uma linha de Lowess aproximadamente horizontal com intercepto igual a 0. Qualquer desvio significativo sugere que o modelo é inapropriado. Além disso são usados esses mesmos resíduos para detectar valores extremos. Na análise residual há que considerar que para o nível de confiança de 95%, sendo que os valores dos resíduos estandardizados $|r_{ij}|$ devem ser inferiores a 1.96, e qualquer observação com o $|r_{ij}|$ superior a $1.96 \approx 2$, poder ser classificado como um **outlier** ou valor atípico para $\alpha = 0.05$ ($Z_{0.95} = 1.96$).

As influencias das observações no modelo estimado foram estudadas através da distância do Cook $\Delta\chi_j^2$ descrita pela equação (2.18) em função das probabilidades estimadas, as observações mais influentes são as que tem os valores de $\Delta\chi_j^2$ superiores a 3.84 para o modelo, considerando que, a um nível de significância de $\alpha = 0.05$ o valor de $\chi_{0.95;(1)}^2 = 3.84$.

Capítulo 4

Resultados e discussão

4.1 Resultados

Nesta secção, foram apresentados os resultado obtidos com o processamento das 375 observações, composta pelos Trabalhadores por Conta Própria que contribuem no INSS da província de Maputo. Com esses resultados, espera-se responder os objectivos previamente definidos na introdução do presente trabalho acadêmico.

4.2 Descrição do perfil da amostra

De acordo com a figura (4.1), verificou se que, cerca de 57,60% são do sexo feminino e a percentagem complementar destes inquiridos são do sexo masculino, foi verificado ainda que a maioria dos trabalhadores por conta própria tem 40 à 49 anos, seguido dos 27,47% tem as suas idades compreendidas dos 50 à 59 anos.

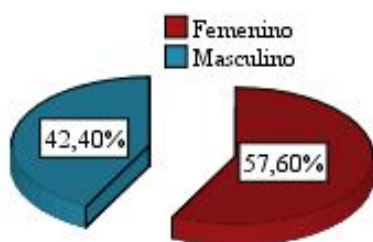


Figura 4.1: Distribuição dos TCP's segundo o sexo

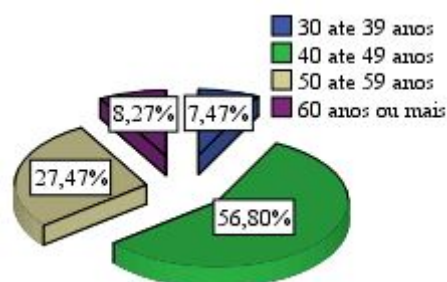


Figura 4.2: Distribuição dos TCP's segundo a idade

Em relação ao nível académico, pode se verificar pela figura (4.3) que o destaque foi para os TCP's com o nível médio de escolaridade, que detêm um percentual de 29,07%, seguido dos que têm o nível básico de escolaridade com cerca de 27,73%, seguem se os TCP's que apenas completaram o ensino primário com cerca de 25,33% dos TCP's, os 15,30% e 2,67% deste TCP's tem o nível superior e pós-graduação respectivamente.



Figura 4.3: Distribuição dos TCP's segundo o nível acadêmico

De acordo com o número de membros da família, verificou-se que, cerca de 35,47% dos TCP's tem um agregado de 5 familiares, seguido dos 29,07% destes TCP's que tem uma família composta por 4 membros e 20,27% destes tem 6 membros nas suas famílias como ilustra a figura (4.4).

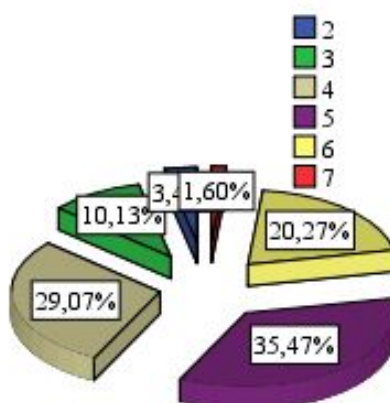


Figura 4.4: Distribuição dos TCP's segundo o agregado familiar

Por conseguinte, na figura (4.5) foi apresentada a distribuição dos TCP's segundo a profissão, onde verificou-se que, cerca de 25,87% são pescadores e 20,27% dos TCP's operam na área de actividades não financeiras e 10,6% na agro-pecuária.

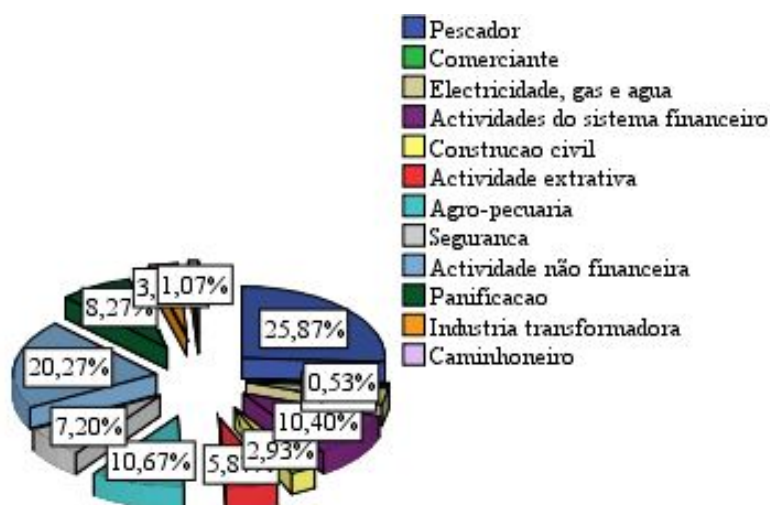


Figura 4.5: Distribuição dos TCP's segundo a profissão

4.3 Teste de independência das variáveis

Foi verificada a existência de associação entre as variáveis que podem influenciar nas contribuições e a própria variável **pagamento**, onde diversas tabelas de contingência e testes de associação foram tabulados. Em relação a fonte de informação e o pagamento, verificou-se na tabela (4.1), que cerca de 79 TCP's que teve acesso a informação do INSS a partir da radio não pagou as prestações nos últimos 3 meses.

Tabela 4.1: Contingência entre Fonte×Pagamento

	Pagou			χ^2 de Pearson	P-valor
Fonte	Não	Sim	Total	19,145	0,000
Voluntariamente	45	56	101		
Radio/publicidade	79	77	156		
Palestra	63	22	85		
Colegas/Amigos	15	18	33		
Totais	202	173	375		

Realizado o teste de independência, verificou se que o p-valor associado ao teste χ^2 de Pearson é inferior ao nível de significância ($\alpha = 5\% = 0.05$), desta forma, conclui-se que existem evidências para rejeitar a hipótese nula, o que leva a confirmar que as variáveis fonte e pagamento tem alguma associação.

Tabela 4.2: Contingência entre Palestra*Pagamento

	Pagou			χ^2 de Pearson	P-valor
Palestra	Não	Sim	Total	5,916	0,045
Sim	96	61	157		
Mais ou menos	81	83	164		
Não	25	29	54		
Totais	202	173	375		

O mesmo teste foi realizado para verificar-se a variável palestra está associada ao pagamento das contribuições, desta feita, de acordo com a tabela (4.2), o p-valor associado ao teste é inferior ao $\alpha = 0.05$, o que levou a rejeição da hipótese nula, afirmando assim que, as variáveis Palestra e pagamento estão associadas.

Tabela 4.3: Contingência entre Importância*Pagamento

	Pagou			χ^2 de Pearson	P-valor
Importância	Não	Sim	Total	10,662	0,005
Sim	135	90	225		
Mais ou menos	52	72	124		
Não	15	11	26		
Totais	202	173	375		

O mesmo fenômeno foi verificado ao nível da variável Importância, onde cerca de 135 dos TCP's que não pagam a 3 meses estão cientes da importância dessas contribuições.

O teste cuja estatística de teste χ^2 de Pearson = 10.662 e o p-valor associado ao teste (0.005) como foi tabulado em (4.3), esses resultados levaram a rejeição da hipótese nula, demonstrando que desse modo que, as variáveis importância e o pagamento tem alguma associação.

Tabela 4.4: Contingência entre Vantagens×Pagamento

	Pagou			χ^2 de Pearson	P-valor
Vantagens	Não	Sim	Total	4,355	0,113
Sim	135	90	225		
Mais ou menos	80	51	131		
Não	16	14	30		
Totais	202	173	375		

De acordo com a tabela (4.4), pode-se verificar que cerca de 135 dos TCP's que conhecem as vantagens de contribuir para o INSS, não pagaram as prestações a mais de 3 meses. Realizado o teste χ^2 de Pearson verificou se que este teve a sua estatística igual a 4.355 e o p-valor associado ao teste (0.113) que é superior a $\alpha = 0.05$, condicionando assim, as evidências para a não rejeição da hipótese nula, deste modo conclui-se que, as vantagens e o pagamento são independentes.

Tabela 4.5: Contingência entre Prioridades×Pagamento

	Pagou			χ^2 de Pearson	P-valor
Prioridade	Não	Sim	Total	45,378	0,000
Prioritário	54	103	157		
± Prioritário	127	67	194		
Não prioritário	21	3	24		
Totais	202	173	375		

Em análise das variáveis prioridades dos TCP's e o pagamento, pode se verificar na tabela (4.5) que, o p-valor associado ao teste χ^2 de Pearson (45,378) é inferior a $\alpha = 0.05$, assim sendo, não rejeita-se a hipótese nula e conclui-se que as variáveis são dependentes ou associadas. De forma contrária ao fenômeno anterior, verificou se que entre as variáveis meios×pagamento, cerca de 137 dos TCP's conhecem os vários meios usados para o pagamento das contribuições mas mesmo assim, estão a mais de 3 meses sem contribuir para o INSS.

Tabela 4.6: Contingência entre Meios×Pagamento

	Pagou			χ^2 de Pearson	P-valor
Meios	Não	Sim	Total	0,025	0,874
Sim	137	116	253		
Mais ou menos	65	57	122		
Totais	202	173	375		

Em análise da existência de associação entre os Meios e a variável Pagamento, foi revelado na tabela (4.6) que, a estatística de teste χ^2 de Pearson é igual à 0,025 e o p-valor associado ao teste (0,874) supera ao nível de significância ($\alpha = 0,05$), tornando-se a evidência necessária para a não rejeição da hipótese nula, o que leva a concluir que as duas variáveis em análise são independentes.

Tabela 4.7: Contingência entre Canal×Pagamento

	Pagou			χ^2 de Pearson	P-valor
Canal	Não	Sim	Total	12,887	0,002
Transferência bancária	47	70	117		
Balcão do banco	147	97	244		
Internet Banking	8	6	14		
Totais	202	173	375		

Verificou se ainda se o canal usado para o pagamento das contribuições está associado ao pagamento, onde os resultados foram tabelado em (4.7). Verificou-se que, cerca de 147 dos TCP's que vinham pagando no balcão do banco, não chegaram de pagar no últimos 3 meses. Pode se verificar ainda que, o χ^2 de Pearson = 12,887 e o p-valor associado ao teste (0,002), levam a concluir que o conhecimentos dos meios usado pelos TCP's e a variável pagamento são dependentes ou associados.

Tabela 4.8: Contingência entre Motivo×Pagamento

	Pagou			χ^2 de Pearson	P-valor
Motivo	Não	Sim	Total	13,882	0,001
Ocupação/falta de tempo	43	38	81		
Não prioritário	74	92	166		
Falta de condições	85	43	128		
Totais	202	173	375		

Por conseguinte analisou se a relação existente entre os motivos do não pagamento e o próprio pagamento. Desta análise, verificou-se que, os que tem falta de condições para efectuar o pagamento não conseguiram materializa-lo, do contrario os 92 contribuintes que não acham esse acto tão prioritário fizeram as suas contribuições nos últimos 3 meses. Ainda na tabela (4.8), verificou se que as variáveis estão associadas, pois, o p-valor associado ao teste (0,001) é inferior ao $\alpha = 0,05$.

Tabela 4.9: Contingência entre Empresa×Pagamento

	Pagou			χ^2 de Pearson	P-valor
Empresa	Não	Sim	Total	3,415	0,065
Sim	21	9	30		
Não	181	164	345		
Totais	202	173	375		

A tabela (4.9), revela a análise feita as variáveis Empresa e Pagamento, onde cerca de 181 dos que não é descontado o INSS de algum empresa não conseguiu fazer os últimos pagamentos, enquanto que 164 destes, acharam melhor pagar. Deste modo, para certificar se da independência, o p-valor associado ao teste é superior ao nível de significância ($p\text{-valor} = 0,065 > \alpha = 0.05$), com essas evidências, não rejeita-se a H_0 e pode se concluir que, as variáveis em análise são independentes.

Tabela 4.10: Contingência entre Motivação×Pagamento

	Pagou			χ^2 de Pearson	P-valor
Motivação	Não	Sim	Total	2,254	0,133
Pensão por velhice	177	142	319		
Subsídio por maternidade	25	31	56		
Totais	202	173	375		

Foi por conseguinte realizado o teste de independência entre as variáveis Motivação e a variável pagamento, onde verificou-se que, o χ^2 de Pearson = 2,254 e associado ao teste o p-valor = 0.133 é superior ao nível de significância $\alpha = 0.05$, o que leva concluir que as variáveis motivação e pagamento são independentes.

Tabela 4.11: Contingência entre Benefícios×Pagamento

	Pagou			χ^2 de Pearson	P-valor
Benefícios	Não	Sim	Total	4,569	0,033
Sim	55	31	86		
Não	147	142	289		
Totais	202	173	375		

Segundo a tabela (4.11), existem evidencias para concluir que as variáveis recepção dos benefícios e o pagamento estão associados, pois, o p-valor associado ao teste ($p\text{-valor}=0,033$) é inferior ao nível de significância ($\alpha = 0,05$). Já a acessibilidade das prestações revelou não estar associada aos pagamentos, visto que, o p-valor associado ao teste é superior ao nível de significância ($p\text{-valor} = 0,767 > \alpha = 0.05$) como ilustra a tabela (4.12).

Tabela 4.12: Contingência entre Acessibilidade×Pagamento

	Pagou			χ^2 de Pearson	P-valor
Acessibilidade	Não	Sim	Total	0,088	0,767
Sim	161	140	301		
Não	41	33	74		
Totais	202	173	375		

Depois destas análises feitas, pode se verificar que de todas as 12 variáveis que foram considerados factores importantes para a contribuição dos TCP no INSS, apenas as variáveis Fonte, Palestras, Importância, Prioridades, Canal, Motivo e Benefícios tem alguma associação com o pagamento das prestações do INSS.

4.4 Regressão logística binária

Para a verificação das variáveis que mais influenciam no pagamento das prestações passa-se a apresentar os resultados da análise de regressão logística, que tem a vantagem de ser menos afetada pelas suposições básicas, particularmente a normalidade das variáveis, não são satisfeitas, pois, ela pode acomodar variáveis não-métricas por meio da codificação. Assim sendo, a verificação dos pressupostos será verificada pela análise residual, para já prossegue-se com a codificar da variável dependente pagamento denotado por "Pagou".

Tabela 4.13: Codificação do pagamento

Valor original	Valor interno
Não pagou	0
pagou	1

4.4.1 Análise do modelo nulo

Em análise do modelo nulo, que contém apenas a constante (β_0), pode se verificar na tabela (4.14), que a constante β_0 revelou-se significativa, pois, o p-valor = 0,0135 e é inferior ao $\alpha = 0,05$, o que sugere a rejeição da hipótese de que $\beta_0 = 0$.

Tabela 4.14: Equação do modelo nulo

	β_0	S.E.	Wald	df	Sig.	Exp(B)
Constante	-0,155	0,104	2,238	1	0,0135	0,856

A tabela de classificação inicial (4.15), onde foi levado em consideração as 375 observações com maior frequência sobre a variável pagou, a categoria de resposta "Não" esteve em destaque e explicam cerca de $\frac{202}{375} = 53.9\%$ dos casos correctamente classificados.

Tabela 4.15: Tabela de classificação inicial

		Pagamentos previstas		Percentagem
Categorias		Não	Sim	Correcto
Pagamentos observados	Não	202	0	100,0
	Sim	173	0	0,0
Percentagem global				53,9

Analisado o modelo nulo, passa-se a apresentar os resultados do modelo ajustado, com vista a certificar se da existência de alguma influência das variáveis independentes sobre o pagamento das contribuições do INSS, com objectivo de compará-lo com o modelo nulo.

4.4.2 Análise do modelo ajustado

Para fazer análise do modelo ajustado foi inicialmente feito teste de razão de verossimilhança, também denominado teste de significância conjunta dos coeficientes. De acordo com a tabela (4.16), pode se verificar que a estatística Qui-quadrado igual a 160,709 e o p-valor associado ao teste (0,000) é inferior ao nível de significância $\alpha = 0,05$, o que leva a rejeitar a hipótese nula, concluindo deste modo que, no modelo ajustado existe pelo menos um coeficiente estimado que é estatisticamente diferente de zero.

Tabela 4.16: Teste de significância dos coeficientes (Omnibus)

	Chi-square	df	Sig.
Step	160,709	45	0,000
Block	160,709	45	0,000
Model	160,709	45	0,000

Por conseguinte, a tabela (4.17), apresenta o resumo do modelo ajustado, que contém informações também usadas para avaliar o modelo, onde pode se ver que, cerca de 36,6% da probabilidade de um contribuinte terem pago as 3 últimas contribuições é explicada pelas variáveis independente significativas.

Tabela 4.17: Resumo do modelo ajustado

Step	-2 Log Likelihood	R ² de Cox & Snell	R ² de Nagelkerke
1	356,907	0,349	0,466

O ajuste do modelo, são comparados os valores observados dos previstos pelo modelo ajustado, onde pode ver na tabela (4.18) que os valores em comparação apresentam diferenças pequenas. Desta forma, fez se o teste formal de para poder inferir sobre o ajuste do modelo.

Tabela 4.18: Tabela de contingência para teste de Hosmer e Lemeshow

Grupos	Pagamento (0) = Não		Pagamento (1) = Sim		Total
	Observado	Esperado	Observado	Esperado	
1	35	36,382	3	1,618	38
2	31	33,763	7	4,237	38
3	31	31,412	7	6,588	38
4	28	27,631	10	10,369	38
5	29	22,971	9	15,029	38
6	23	18,754	15	19,246	38
7	11	13,839	27	24,161	38
8	8	9,444	30	28,556	38
9	6	5,815	32	32,185	38
10	0	1,989	33	31,011	33

Segundo os resultados descritos na tabela (4.19), onde a estatística χ^2 -Quadrado = 12,543 e o p-valor (0.386) associado ao teste de Hosmer & Lemeshow, não rejeita se a hipótese nula, levando a concluir que o modelo estimado se ajusta aos dados.

Tabela 4.19: Teste Hosmer & Lemeshow

Step	χ^2 - Quadrado	df	Sig.
1	12,543	8	0,386

4.4.3 Avaliação de ajuste do modelo

Depois dos resultados do teste de significância conjunta dos coeficientes relevarem a existência de pelo menos um parâmetro significativo para a previsão do pagamento das prestações nos últimos 3 meses, e o modelo estar devidamente ajustado aos dados observados, passa-se para a verificação da capacidade do modelo logístico classificar correctamente os casos.

Tabela 4.20: Tabela de classificação final

		Pagamentos previstos		Percentagem
		Não	Sim	Correcto
Pagamentos observados	Não	166	36	82,2
	Sim	41	132	76,3
Percentagem global				79,5

De acordo com a tabela (4.20), fez-se a classificação dos sujeitos observados e os previstos pelo modelo, onde verificou-se que a sensibilidade do modelo é de 82.2%, e a especificidade como indicador de classificação correcta dos que foram previsto como não tendo pago nos últimos 3 meses, quando de facto não chegaram de pagar com um percentual de 76.3%, por isso, pode-se dizer que cerca de 79.5% dos casos, foram correctamente classificados pelo modelo ajustado.

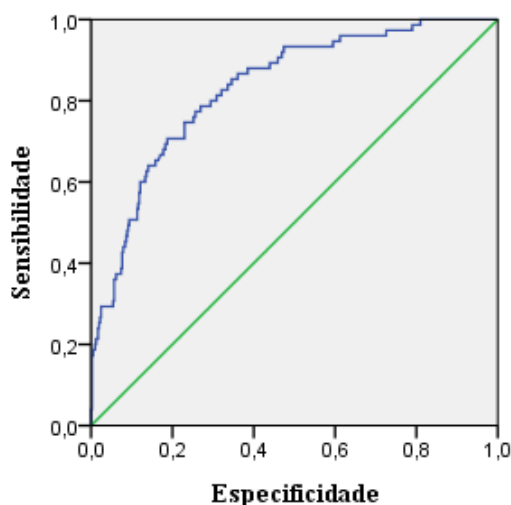


Figura 4.6: Representação da curva de ROC

Para complementar as classificações correctas dos casos, fez-se a análise da curva de ROC, em que de acordo com os resultados obtidos, pode-se verificar que a área abaixo da curva ROC, é maior do que 50%, pois, está acima da recta que faz o corte da área pela metade (0.5), como ilustra a figura (4.6) acima. Como forma de confirmar a representação gráfica acima, verificou-se que, sub a suposição da variável dependente não ser paramétrica, a Tabela (4.21) apresenta o teste formal da área abaixo da curva do ROC. De seguida, passa-se para a verificação da significância individual dos coeficientes estimados.

Tabela 4.21: Área sob a curva

			Intervalo de confiança	
			Assintótico a 95%	
Área	Erro padrão^a	Sig. assintótico^b	inferior	superior
0,831	0,025	0,000	0,783	0,880

a. Sub a suposição da variável não paramétrica

b. Hipótese nula: Área da curva do ROC = 0.5

Como pode se verificar na tabela (4.21), o valor da área abaixo da curva de ROC que é igual a 0.831, já que o valor da área supera o valor do corte $c=0.5$, e o $p\text{-valor} = 0.000 < \alpha = 0.05$, então pode se concluir que o modelo tem algum poder discriminativo. Desta feita, para classificar esse poder basta comparar a área $= 0.831 > 0.8$ (80%), o que leva a concluir que o modelo ajustado têm um bom poder discriminativo.

4.4.4 Significância individual dos coeficientes

A significância de cada coeficiente foi feita através do teste de Wald que foi apresentado nas tabelas (4.22) e (4.23), onde as variáveis significativas foram destacadas e sinalizadas com o sinal (**) como um simbolo de variável ou categoria estatisticamente significativa.

Tabela 4.22: Significância individual dos coeficientes

Variáveis	β_i	S.E.	Wald	Df	Sig.	Exp(β)
Sexo (Feminino)	-0,48	0,309	2,439	1	0,118	0,617
Masculino (ref.)	—	—	—	—	—	—
Idade			0,824	3	0,844	
30 à 39	-1,271	1,697	0,561	1	0,454	0,281
40 à 49	-2,439	2,436	1,003	1	0,317	0,087
50 à 59	-1,503	1,862	0,652	1	0,419	0,222
60+ (ref.)	—	—	—	—	—	—
Profissão			8,525	11	0,666	
Pescador	-1,271	1,697	0,561	1	0,454	0,281
Comerciante	-2,439	2,436	1,003	1	0,317	0,087
Eletricidade/gás/água	-1,503	1,862	0,652	1	0,419	0,222
Sistema financeiro	-2,153	1,821	1,397	1	0,237	0,116
Construção civil	-1,725	1,835	0,885	1	0,347	0,178
Actividade extrativa	-2,479	1,776	1,947	1	0,163	0,084
Agro-pecuária	-1,365	1,738	0,617	1	0,432	0,255
Segurança	-1,231	1,788	0,474	1	0,491	0,292
Sistema não financeiro	-1,932	1,712	1,274	1	0,259	0,145
Panificação	-1,798	1,759	1,045	1	0,307	0,166
Indústria transformadora	-2,102	1,882	1,247	1	0,264	0,122
Condutor de caminhão (ref)	—	—	—	—	—	—
Escolaridade			2,454	4	0,653	
Primário	-0,075	1,189	0,004	1	0,950	0,928
Nível básico	0,196	1,154	0,029	1	0,865	1,216
Nível médio	-0,350	1,131	0,095	1	0,757	0,705
Licenciatura	0,130	1,069	0,015	1	0,903	1,139
Pós-graduação (ref.)	—	—	—	—	—	—
Remuneração			2,515	4	0,642	
Até 15.000	-0,166	0,539	0,095	1	0,758	0,847
15.000 à 30.000	0,471	0,619	0,578	1	0,447	1,601
30.000 à 45.000	0,453	0,965	0,221	1	0,639	1,573
45.000 à 60.000	-0,169	0,816	0,043	1	0,836	0,844
60.000 + (ref.)	—	—	—	—	—	—
Fonte			30,048	3	0,000**	
Voluntariamente	-1,079	0,432	6,244	1	0,012**	0,340
Radio/Publicidade	-2,619	0,484	29,323	1	0,000**	0,073
Colegas/Amigos	-1,655	0,570	8,429	1	0,004**	0,191
Palestras (ref.)	—	—	—	—	—	—

A tabela (4.23) apresenta a continuação da tabela acima, que contém os resultados do teste de significância individual dos coeficientes estimados (Wald).

Tabela 4.23: Significância individual dos coeficientes - Continuação

Variáveis	β_i	S.E.	Wald	Df	Sig.	Exp(β)
Palestra			3,073	2	0,215	
Sim	0,556	0,328	2,879	1	0,090	1,744
Mais ou menos	0,605	0,518	1,365	1	0,243	1,831
Não (ref.)	—	—	—	—	—	—
Importância			7,697	2	0,021**	
Sim	0,801	0,364	4,855	1	0,028**	0,228
Mais ou menos	-0,625	0,563	1,234	1	0,267	0,535
Não (ref.)	—	—	—	—	—	—
Vantagens Sim	-0,653	0,376	3,007	1	0,083	0,521
Mais ou menos (ref.)	—	—	—	—	—	—
Prioridade			32,199	2	0,000**	
Prioritário	-1,790	0,342	27,369	1	0,000**	0,167
Mais ou menos	-2,567	0,760	11,415	1	0,001**	0,077
Não prioritário (ref.)	—	—	—	—	—	—
Meios Sim	-0,109	0,338	0,104	1	0,747	0,897
Mais ou menos (ref.)	—	—	—	—	—	—
Canal			6,815	2	0,033**	
Transferência bancária	-0,927	0,355	6,805	1	0,009**	0,396
Balcão do banco	-0,571	0,795	0,516	1	0,473	0,565
Internet Banking (ref.)	—	—	—	—	—	—
Motivo			19,913	2	0,000**	
Não prioritário	0,584	0,402	2,118	1	0,146	1,794
Falta de tempo	-0,967	0,430	5,053	1	0,025**	0,380
Falta de condições (ref.)	—	—	—	—	—	—
Empresa Sim	-22,479	40193,840	0,000	1	1,000	0,000
Não (ref.)	—	—	—	—	—	—
Benefícios (Sim)	0,482	0,309	2,439	1	0,118	1,620
Não (ref.)	—	—	—	—	—	—
Acessibilidades (Sim)	0,792	0,431	3,376	1	0,066	2,209
Não (ref.)	—	—	—	—	—	—
Motivação			0,491	2	0,782	0,365
Subsidio de doença	20,415	28195,807	0,000	1	0,999	734780640,911
Subsidio de velhice	-0,321	0,458	0,491	1	0,484	0,726
Sub p/ Maternidade (ref.)	—	—	—	—	—	—
Constante	20,265	40193,840	0,000	1	1,000	632109652,478

Os resultados tabulados nas duas tabelas acima revelam que, as variáveis que influenciam de forma significativa na probabilidade dos TCP's terem pago as 3 últimas prestações são: A Fonte, Importância, Prioridades, Canal e os Motivos, onde os sinais dos coeficientes $\hat{\beta}_i$ significativos, influenciam no aumento ou redução das chances dos TCP's pagarem as prestações do INSS de acordo com o seu sinal.

4.4.5 Interpretação e estimação do modelo ajustado

Assim sendo, passa-se a interpretar os valores de $\text{Exp}(\beta)$ a um nível de confiança de 95% da seguintes maneira:

- **Fonte:** De acordo com a tabela (4.22), Nota-se uma redução na chance do contribuinte pagar as contribuições em cerca de 66% dos que inscreveram-se voluntariamente, 92.7% dos que ouviram a radio/publicidade, 81% dos que ouviram de colegas/amigos quando comparados aos que ouviram palestras, mantendo o resto das variáveis constantes;
- Na variável **Importância**, constatou-se que os TCP's que conhecem a importância do INSS, tem uma redução de até 77,2% das chance de pagarem as prestações quando comparado com os que não conhecem a importância do INSS;
- **Prioridade:** Os TCP's que acham prioritário pagar o INSS, tem uma redução de até 83,3% das chances de pagar e 92,3% dizem ser mais ou menos prioritário comparativamente aos que não vê esse acto como prioritário;
- **Canal:** Em relação ao canal de pagamento, verificou se uma redução de 60,4% das chance de um contribuinte que faz transferência bancária pagar a contribuição quando comparados com os preferem usar o internet Banking;
- **Motivo:** Em relação aos Motivos que dificultam o pagamento, verificou se uma redução de 62% das chance do contribuinte que tem falta de tempo pagar a contribuição quando comparados com aquele que tem falta de condições. Considerando as de mais variáveis constantes.

Usando as variáveis significativas, o modelo de regressão logística pode ser descrita substituindo as variáveis significativas na equação que se segue:

$$P(Y = 0|\mathbf{X} = 0) = 1 - \hat{\pi}(x) = \frac{1}{1 + e^{(\hat{\beta}_0 + \sum_{i=1}^p \hat{\beta}_i x_i)}} \quad (4.1)$$

Onde para determinar a probabilidade do contribuinte não pagar as contribuições, dever se substituir a parte da combinação linear das variáveis influentes como demonstrado de seguida:

$$\begin{aligned} \hat{\beta}_0 + \sum_{i=1}^p \hat{\beta}_i x_i = & \hat{\beta}_0 + \hat{\beta}_{Fon} \times \mathbf{Fonte} + \hat{\beta}_{Imp} \times \mathbf{Importância} + \hat{\beta}_{Pri} \times \mathbf{Prioridades} \\ & + \hat{\beta}_{Can} \times \mathbf{Canal} + \hat{\beta}_{Mot} \times \mathbf{Motivo} \end{aligned} \quad (4.2)$$

Onde:

β_0 é a constante; e os coeficientes estimado das variáveis significativas: $\hat{\beta}_{Fon}$ - estimativa da Fonte, $\hat{\beta}_{Imp}$ - coeficiente estimado da Importância, $\hat{\beta}_{Pri}$ - coeficiente estimado da Prioridade, $\hat{\beta}_{Can}$ - coeficiente estimado do Canal e $\hat{\beta}_{Mot}$ - coeficiente estimado do Motivo. Cada coeficiente é substituído de acordo com a categoria escolhida na variável significativa.

4.4.6 Análise residual

O gráfico (4.7), sugere um bom ajuste do modelo aos dados, visto que, a linha central denominada de "Lowess", apresentam alguns pontos sobre a linha central, e alguns pontos atípicos são os que superam o resíduo " $\varepsilon_i = 2$ ".

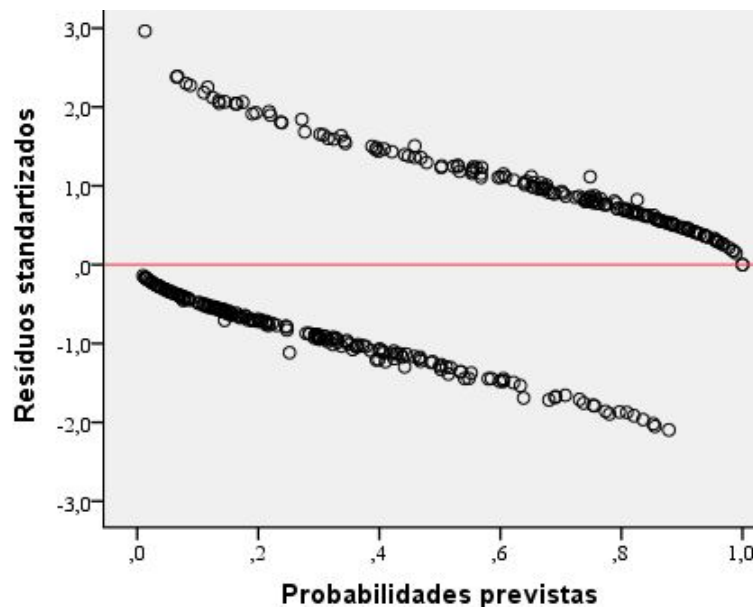


Figura 4.7: Representação gráfica dos resíduos estandardizados

Os tais pontos atípicos representados graficamente em (4.7), foram tabulados em (4.24). De acordo com essa tabela acima, cerca de 15 observações revelam ser atípicas ou *outliers*. De seguida, foram representadas graficamente as observações influentes.

Tabela 4.24: Casos de anomalias listados

Caso	Selecionado	Observado	Previsto	Grupo previsto	Variável temporária	
					Resid	ZResid
117	S	S**	0,089	N	0,911	3,205
120	S	S**	0,059	N	0,941	4,002
122	S	S**	0,121	N	0,879	2,693
138	S	N**	0,869	S	-0,869	-2,576
152	S	S**	0,063	N	0,937	3,845
164	S	N**	0,871	S	-0,871	-2,604
210	S	S**	0,011	N	0,989	9,496
225	S	S**	0,089	N	0,911	3,204
233	S	S**	0,122	N	0,878	2,681
263	S	S**	0,131	N	0,869	2,571
265	S	S**	0,106	N	0,894	2,898
312	S	S**	0,149	N	0,851	2,387
315	S	S**	0,141	N	0,859	2,468
346	S	N**	0,840	S	-0,840	-2,288
362	S	S**	0,078	N	0,922	3,430

De acordo com a representação do quadrado da variação dos resíduos ($\Delta\chi^2$) e a probabilidade prevista, onde verificou-se que, existem observações que superam 4 e são consideradas influentes no modelo de regressão estimado ou ajustado, mas para que essas observações sejam assim consideradas e removidas têm que se verificar a sua quantidade.

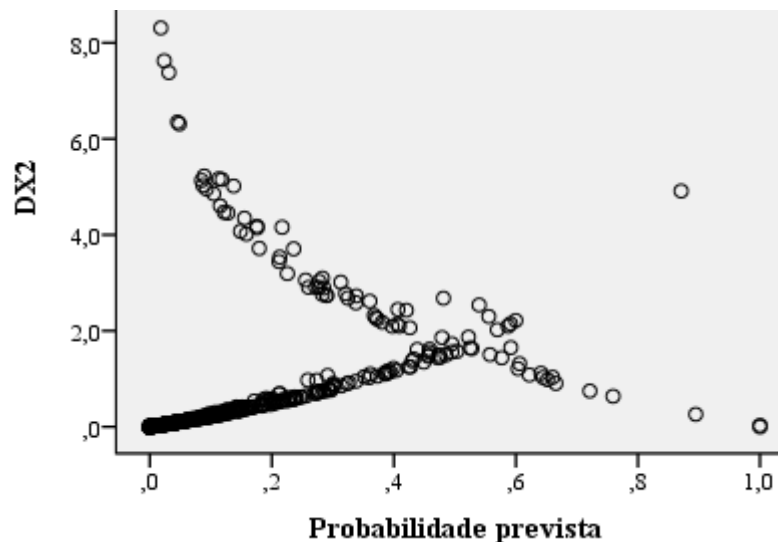


Figura 4.8: Representação gráfica dos resíduos estandardizados

Deste modo, através da figura (4.8), foi possível concluir que, embora as observações influentes, a percentagem destas não superou os 10% da amostra, então, a influência dessas observações não é significativa ou não afectam de forma significativa no modelo ajustado.

4.5 Discussão

O tema em análise ainda não foi muito explorado pelos investigadores, esse facto foi confirmado pela dificuldades obtida para encontrar estudos relacionados com os factores que influenciam no pagamento das contribuições no Sistema Nacional de Segurança Social. No entanto, dos dois estudos foram trazidos para a presente discussão:

De acordo com Brito (2020), que realizou a pesquisa para os TPC no sector da pesca artesanal, verificou que a falta de entendimento sobre os procedimentos e as vantagens da canalização das suas contribuições para o sistema de segurança obrigatório faz com que os pescadores artesanais não canalizem as suas contribuições, em concordância, para o Quive (2009) o conhecimento dos procedimentos para o acesso a segurança social, associado ao excesso de burocracia, bem como a descentralização dos serviços e a localização dos postos de pagamentos podem ser barreiras para o pagamento das contribuições.

Embora descritas em outras palavras no estudo realizado pelo Brito (2020), o conhecimento dos procedimentos pode ser enquadrado em duas variáveis que se revelaram significativas para o pagamento das contribuições que são a fonte de onde recebeu a informação sobre INSS e o canal usado para fazer as contribuições, já o conhecimento das vantagens não foi considerada significativa mas a importância de fazer as contribuições foi considerada importante para o pagamento da segurança social obrigatória. Já em relação ao estudo realizado pelo Quive (2009), não houve algum ponto de concordância mas a localização dos postos de contribuição pode ser comparado a variável Canal usado de pagamento.

As variáveis prioridade e os motivos que levam os contribuinte a não canalizarem as contribuições foram as duas que não foram levadas em consideração, nos dois estudos envolvidos na presente discussão, mas no presente estudo foram tidas com algumas das que influenciam significativamente no pagamento do INSS, por isso, encoraja-se aos investigadores que explorem mais variáveis que possam influenciar no pagamento das contribuições no INSS.

Capítulo 5

Conclusões e Recomendações

5.1 Conclusões

Este estudo tinha como objectivo principal analisar os factores que influenciam no não pagamento das contribuições dos trabalhadores por conta própria inscritos ao sistema de segurança social obrigatória, onde foram inquiridos 375 Trabalhadores por conta própria, através dos quais foram processados e analisados, e dessas análises foram alcançadas as seguintes conclusões:

- Dos 375 TCP's inquiridos, verificou-se que a maioria são do sexo feminino com um percentual de 57,6% e 42,4% do sexo masculino. Em relação a distribuição das idades destes TCP's o destaque vai ao grupo com idades compreendidas dos 40 à 49 (56,8%), seguido dos TCP's nas idades de 50 à 59 anos idade com um percentual de 27,47% como mostra a figura (4.2). Em relação ao nível académico verificou se um equilíbrio na distribuição a maioria do nível médio (29,07%) e quanto mais aumentando nível académico as percentagens vão diminuindo até a menor percentagem (2,67%) dos pós-graduados. A maioria dos TCP's tem 5 membros na sua família. Cerca de 25,8% destes são pescadores, seguido dos TCP's que operam na área da agro-pecuária (20,27%).
- Por conseguinte, foi feita a análise da associação entre as variáveis consideradas determinantes para o pagamento das prestações dos TCP no INSS, desta feita, cerca de 12 variáveis foram analisadas e constatou-se que, apenas as variáveis Fonte, Palestras, Importância, Prioridades, Canal, Motivo e Benefícios tem alguma associação com o pagamento das prestações do INSS. Além disso a análise de influência dessas variáveis foi feita usando a regressão logística binária que revelou que as variáveis: Fonte, Importância, Prioridades, Canal e os Motivos foram as variáveis que de forma significativa influenciaram no não pagamentos das prestações dos TCP no INSS.

➤ Através dessas variáveis significativas foi possível estimar o modelo que possa prever a probabilidade dos TCP pagarem as suas contribuições junto ao INSS, onde as suas chances variam de acordo com os valores de $\text{Exp}(\beta_i)$ das variáveis que se revelaram ser significativas, desta forma constatou se a redução das chances dos TCP pagarem as contribuições, como passa-se a descrever:

- Redução das chances dos TCP's que inscreveram-se voluntariamente ao SNSS, que ouviram pela radio/publicidade e dos que ouviram pelos colegas/amigos quando comparados com os que ouviram palestras;
- Verificou-se a redução das chances do contribuinte pagar as contribuições na ordem de 77,2%, para os que sabem da importância de estar inscrito e contribuir para o INSS quando comparados com os que não sabem;
- O mesmo acontece para todos os TCP's que consideram que as contribuições ao INSS são a prioridade das suas despesas, estes tem uma redução das chances de pagar a segurança social na ordem de 83,3% e 92,3% que acha que esse pagamento é mais ou menos prioritário quando comparados aos que acham que essa despesa não é prioritária;
- Em relação ao canal usado para canalizar o valor das contribuições, verificou se uma redução das chances dos TCP's que fazem transferência bancária pagarem as contribuições quando comparados aos que usam o Internet banking, mantendo as demais variáveis constantes;
- Quanto aos motivos dos TCP's não conseguirem pagar, verificou-se que, aqueles que tem falta de tempo tem uma redução de chances (62%), destes pagarem o valor das contribuições ao INSS, comparativamente aos que não tem condições de pagar.

Diante desses resultados, fica evidente que revidar todos os esforços para informar detalhadamente o processo de inscrição, importância, difundir mais sobre os diversos canais para efectuar o pagamento das prestações, pois, procedendo dessa maneira mais inscritos por conta própria farão parte do INSS e farão as devidas contribuições.

5.2 Recomendações

➤ Governo

- Realizar campanhas de sensibilização regulares em línguas nacionais com objectivos de centrar-se nas seguintes questões: a lógica e os benefícios da segurança social contributiva, os requisitos para acedê-los, e o processo de requisição de benefícios, incluindo os prazos.
- Introduzir uma linha telefónica gratuita de consultas e de reclamações;
- Desenhar e introduzir uma aplicação com base na tecnologia de USSD/ SMS e API, que permite que os TCP possam consultar o seu histórico contributivo, fazer simulações das várias pensões e dar seguimento aos procedimentos administrativos;
- Expandir a presença do INSS ao nível local através de uma colaboração com o Governo ao nível do distrito, posto administrativo, localidade e bairro;
- Introduzir e divulgar sobre o sistema misto de contribuições e pagamentos de benefícios, que englobe os balcões de atendimento do INSS, as ATMs dos bancos principais, e os serviços financeiros móveis como Mpesa, Mkesh e Mmola;

➤ Aos Trabalhadores por Conta Própria

- Encorajados a inscrever-se ao INSS, pois, a maioria dos TCP, em busca da sua subsistência colocam suas vidas em riscos todos os dias, dado as características e condições do seu ambiente de trabalho.
- Fazer se presente nas palestras e campanhas da INSS, de modo a obter mais conhecimentos sobre estes serviços que poderão garantir a segurança diante de situações de calamidades ou desastres naturais, existe um fundo dentro do INSS criado para reagir a situações do gênero que afectam os TCP, além das prestações já conhecidas que são disponibilizadas em casos de doenças, gravidez, velhice e outros.

➤ A comunidade acadêmica

- A todos os estudantes e investigadores, que desenvolvam mais estudos relacionados com os factores que influenciam para que os Trabalhadores por conta própria pagam ou não as prestações por si declaradas no INSS, o que pode ajudar nas estratégias para a sensibilização desse grupo alvo.
- Encoraja-se ainda, a utilização das diversas técnicas estatísticas para analisar as variáveis que podem ser consideradas determinantes tanto nos pagamentos, nos factores importantes para os TCP manterem se activos e analisar os modelos económicos de previsão das receitas que resultam desse pagamentos.

Referências

1. Agranonik, M. (2005) . Técnicas de diagnostico aplicadas ao Modelo de Regressão Logística. Monografia para obtenção de titulo de Bacharel em Estatística, Porto Alegre, Brasil;
2. Agresti, A. (2002). Categorical Data Analysis, Second Edition. New York: Editora Wiley;
3. Batistela, G. C.; Rodrigues, S. A. e Bononi, J. T. C. M. (2009). Estudo sobre a evasão escolar usando regressão logística: análise dos alunos do curso de administração da fundação educacional de ituverava, Tékhne e Lógos, Botucatu, SP, 1(1);
4. Beiguelman, B. (1996).Curso de Bioestatística Básica. 4ed. Ribeirão Preto: Sociedade Brasileira de Genética;
5. Brito, C. J. (2020). Análise Do Nível De Cobertura Da Segurança Social Obrigatória Em Moçambique: Caso De Trabalhadores Por Conta Própria Do Sector Da Pesca Artesanal Nos Bairros De Guachene E Praia Nova (2015-2019). Monografia apresentada para obter o grau de licenciatura em Administração Pública na Universidade Eduardo Mondlane;
6. Cabral, C. I. S. (2013) . Aplicação do Modelo de Regressão Logística num estudo de Mercado, Dissertação para obtenção de nível de mestre - Universidade de Lisboa, Lisboa;
7. Crespo, T. (1998). Técnicas de Amostragem. Lisboa;
8. Corrar, L. J.; Paulo, E. e Filho, J. M. (2007). Análise multivariada: para cursos de administração, ciências contábeis e economia. FIPECAFI – Fundação Instituto de Pesquisas Contábeis, Atuariais e Financeiras. São Paulo: Atlas;
9. Fávero et al. (2009). Análise de Dados: modelagem multivariada para tomada de decisões. Rio de Janeiro. Editora Elsevier;
10. Gil, A. C. (2008). Métodos e técnicas de pesquisa social. 6^a ed, São Paulo: Atlas;
11. Guimarães, R. V. (2006). Uso de Regressão Logística para Previsão de Fechamento de Operações Financeiras: Temo de moedas, USP, Brasil;
12. Hair et al. (2005) . Estatística multivariada de dados. 5^a edição, Porto Alegre, Bookman;
13. Hosmer, D. W. e Lemeshow, S. (1989). Applied Logistic Regression;
14. Hosmer, D. W. e Lemeshow, S. (2000). Applied Logistic Regression. 2nd edition;
15. Hosmer Jr, D.; Lemeshow, S.; Sturdivant, R. X. (2013). Applied logistic regression. [S.l.]: John Wiley & Sons, 2013. v. 398. Citado 3 vezes nas páginas 13, 14 e 15.

16. <https://inss.gov.mz/77-template-details.html&hl=pt-PT&gl=mz;>
17. INE (2019). IV Recenseamento Geral da População e Habitação 2017: Resultados Definitivos Moçambique. Maputo: INE;
18. Instituto Nacional de Segurança Social (INSS). (2022). Relatório e Contas do Instituto Nacional de Segurança Social: Exercício económico de 2022. Maputo, Moçambique: INSS;
19. Lorena. H. (2013). O Trabalhador por conta própria no Brasil;
20. Manuel, J. E. (2016). Sustentabilidade Económica e Financeira da Segurança Social em Angola. Dissertação para obtenção do Grau de Mestre em Economia Monetária e Financeira, Universidade de Évora, Angola;
21. Maroco, J. (2007). Análise Estatística – Com utilização do SPSS. 3ª edição;
22. Organização Internacional do Trabalho (OIT, 2020). Ultrapassando as barreiras no acesso à segurança social contributiva dos trabalhadores por conta própria em Moçambique;
23. Prodanov, C. C. e Freitas, E. C. (2013). Metodologia do trabalho científico: Métodos e Técnicas da Pesquisa e do Trabalho Acadêmico. 2ª ed, Novo Hamburgo: Feevale;
24. Quive, S. A. (2009). Sistemas formais e informais de protecção social desenvolvimento em Moçambique. Maputo: IESE;
25. República de Moçambique (2022). 4º Boletim Estatístico sobre Protecção Social, Ministério do Trabalho e Segurança Social;
26. RdM (2007). Lei da Protecção Social n.o 2/2007 de 7 de Fevereiro. República de Moçambique;
27. RdM (2015). Decreto 14/2015 de 16 de Julho, concernente a Taxa de Contribuição dos Trabalhadores por Conta Própria;
28. Salotti, B. M. e Yamoto, M. (2005). Informação contábil: estudos sobre sua divulgação no mercado de capitais. São Paulo: Atlas;
29. Souza, R. F. (2006). Impacto do Preço do Petróleo na Política Energética mundial. Dissertação apresentada ao corpo docente da Coordenação dos Programas de Pós-graduação de Engenharia da Universidade Federal do Rio de Janeiro para obtenção do grau de mestre. Disponível em <http://www.ppe.ufrj.br/ppes/production/tesis/souzafr.pdf/>. Acessado em 25 de Outubro de 2021;
30. Triola, M.F. (2005). Introdução à Estatística (9ª ed.) Rio de Janeiro: LTC Editora.

Apêndices 1



Questionário

Universidade Eduardo Mondlane
Faculdade de ciências - DMI
Curso de Licenciatura em Estatística

A presente entrevista, integra-se no âmbito do trabalho final do curso de Licenciatura em Estatística, na Universidade Eduardo Mondlane. As questões a serem colocados, tem como objectivo verificar os factores que influenciam o não pagamento de contribuições dos TCP inscritos no Sistema de Segurança Social em Moçambique. Esta entrevista tem fins académicos, é de natureza sigilosa e confidencial o seu anonimato será respeitado.

Agradecemos a sua colaboração

Número do contribuinte [_____]

Atento as questões relacionadas com o perfil e factores que podem influenciar no pagamento do INSS, assinale com [X] apenas na alternativa correcta, de modo a garantir transparência na análise dos dados.

I. Perfil do contribuinte

1. Sexo	2. Idade	3. Escolaridade	3. Agregado
Masculino []	18 à 28 anos []	[] Ensino primário	[] Menbros
Feminino []	28 à 38 anos []	[] Nível básico	
	38 à 48 anos []	[] Nível médio	
	48+ anos []	[] Licenciatura	
		[] Pós-graduado	
5. Data de inscrição	6. Profissão do contribuinte		
____/____/____	Pescador []	[] Comerciante	
	Eletricidade, gás e água []	[] Act. do sistema financeiro	
	Construção civil []	[] Actividade extrativa	
	Agro-pecuária []	[] Segurança	
	Actividade não financeira []	[] Panificação	
	Industria transformadora []	[] Condutor de caminhão	

II. Desenvolvimento sobre a Segurança Social

1. Conhece a importância da Segurança Social?

☐ Sim

☐ Não

2. Em que intervalo se enquadra o seu rendimento mensal?

☐ Até 15.000

☐ 15.000 à 30.000

☐ 30.000 à 45.000

☐ 45.000 à 60.000

☐ 60.000 ou mais

3. Qual foi a fonte que lhe inspirou a fazer parte do Sistema Nacional de Segurança Social?

☐ Voluntariamente

☐ Radio/Publicidade

☐ Palestra

☐ Colegas/Amigos

4. Tem tido palestras sobre o INSS regularmente após sua inscrição?

☐ Sim

☐ Mais ou menos

☐ Não

5. Conhece a importância da Segurança Social para o cidadão?

☐ Sim

☐ Mais ou menos

☐ Não

6. Conhece as vantagens de pagar todas as contribuições a Segurança Social?

☐ Sim

☐ Mais ou menos

☐ Não

7. Qual é o nível de prioridade do pagamento de contribuições a Segurança Social?

☐ Despesas prioritárias do mês

☐ Despesas por pagar quando sobrar dinheiro

☐ Despesas menos prioritárias do mês

8. Conhece os diversos meios/formas de pagamentos de contribuições para o SNSS?
- [] Sim
- [] Mais ou menos
- [] Não
9. Qual é o canal que usa para fazer o pagamento das suas contribuições no INSS?
- [] Transferência bancária [] Na boca da caixa
- [] Internet Banking [] Carteira móvel (M-pesa,e-mola)
10. Aponte o motivo que faz com que as vezes não consiga fazer as contribuições no INSS
- [] Ocupação/falta de tempo [] Não e sua prioridade
- [] Falta de condições financeiras [] Esquecimento
11. É descontado no SNSS por alguma empresa?
- [] Sim
- [] Não
12. Já teve algum benefício na Segurança Social que decorreu de alguma eventualidade?
- [] Sim
- [] Não
13. Acha que o valor que os TCP descontam são acessíveis?
- [] Sim
- [] Não
14. Qual é a principal prestação que lhe motiva a contribuir para o SNSS?
- [] Subsidio de doença [] Pensão de invalidez
- [] Pensão de velhice [] Subsidio por morte
- [] Pensão de sobrevivência [] Subsidio de maternidade

III. Questão conclusiva

Efetuo o pagamento no INSS nos últimos 3 meses?

- [] Não
- [] Sim

Apêndices 2

Tabela 5.1: Trabalhadores por conta própria inscritos por província em moçambique, no período de 2018 - 2022

Ano Províncias	Inscritos em 2018	Inscritos em 2019	Inscritos em 2020	Inscritos em 2021	Inscritos em 2022	Inscritos de 2018 à 2022
Niassa	41	665	292	403	139	1540
Cabo delgado	240	510	380	366	321	1817
Nampula	129	215	443	356	291	1434
Zambézia	270	518	551	609	470	2418
Tete	256	260	321	619	262	1718
Manica	185	400	353	294	430	1662
Sofala	860	773	1252	572	639	4096
Inhambane	801	638	355	508	593	2895
Gaza	664	619	353	625	627	2888
Maputo	626	499	554	730	807	3216
Cidade de Maputo	1606	3252	1370	636	567	7431
Totais por ano	5678	8349	6224	5718	5146	31115

Anexos

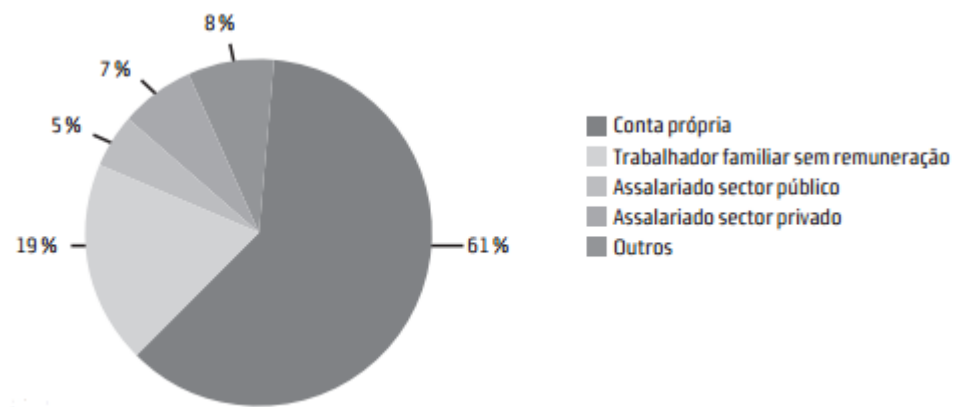


Figura 5.1: A população economicamente activa, segundo o processo laboral, 2019

Fonte: INE (2019)