



UNIVERSIDADE FEDERAL DO CEARÁ
PROGRAMA DE ECONOMIA PROFISSIONAL
MESTRADO PROFISSIONAL EM ECONOMIA DE EMPRESAS

JOÃO ATYLLA MARTINS DA SILVA

**PERFIL DOS APLICADORES DE POUPANÇA DO ESTADO DO CEARÁ ATRAVÉS
DE ANÁLISE MULTIVARIADA DE DADOS.**

FORTALEZA

2021

JOÃO ATYLLA MARTINS DA SILVA

PERFIL DOS APLICADORES DE POUPANÇA DO ESTADO DO CEARÁ ATRAVÉS DE
ANÁLISE MULTIVARIADA DE DADOS.

Dissertação apresentada ao Programa de Pós-Graduação em Economia, da Universidade Federal do Ceará - UFC, como requisito parcial para a obtenção do título de Mestre em Economia. Área de concentração: Economia de Empresas

Orientador: Prof. Dr. Paulo Rogério Faustino de Matos

FORTALEZA

2021

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

S58p Silva, João Atylla Martins da.
Perfil dos Aplicadores de Poupança do Estado do Ceará Através de Análise Multivariada de Dados /
João Atylla Martins da Silva. – 2021.
43 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Faculdade de Economia, Administração,
Atuária e Contabilidade, Mestrado Profissional em Economia de Empresas, Fortaleza, 2021.
Orientação: Prof. Dr. Paulo Rogério Faustino de Matos.

1. Análise Discriminante. 2. Perfil de Consumo. 3. Poupança. I. Título.

CDD 330

JOÃO ATYLLA MARTINS DA SILVA

PERFIL DOS APLICADORES DE POUPANÇA DO ESTADO DO CEARÁ ATRAVÉS DE
ANÁLISE MULTIVARIADA DE DADOS.

Dissertação submetida à Coordenação do Programa de Economia Profissional – PEP, da Universidade Federal do Ceará - UFC, como requisito parcial para a obtenção do grau de Mestre em Economia. Área de concentração: Economia de Empresas

Aprovada em: ___/___/_____.

BANCA EXAMINADORA

Prof. Dr. Paulo Rogério Faustino de Matos (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Sergio Aquino de Souza
Universidade Federal do Ceará (UFC)

Prof. Dr. Ricardo Brito Soares
Universidade Federal do Ceará (UFC)

AGRADECIMENTOS

Quando decidi realizar o curso não pensei muito nos benefícios que teria em fazê-lo ou se estava em condições de, naquele momento, dedicar parte do meu tempo, das minhas finanças e da minha sanidade mental nesta formação. Na verdade, só tinha a convicção que deveria cursá-lo. Portanto, meus agradecimentos irão para àqueles que influenciaram de alguma forma para que eu tenha tomado essa decisão de maneira tão automática, tão natural e tão assertiva, que foi a de realizar o Mestrado Profissional em Economia da UFC.

Inicialmente queria agradecer à minha querida companheira de 18 anos e amada esposa, Amanda. Filha, muito obrigado pelo apoio incondicional e por ter cuidado sozinha da nossa filhinha recém-nascida enquanto eu estava em aula ou em casa estudando, sei bem do esforço e sacrifício que foi de segurar “as pontas”, mas quero que saiba que tudo foi pensando em nós.

Ao grande amor da minha vida e filha, Maria Luíza (Malú), você trouxe sentido, inspiração, motivação, sorrisos, futuro e, acima de tudo, trouxe muito amor! Te amo minha menininha...

Aos meus pais Abinadabe e Marileida, pois apesar da vida humilde em Quixadá/CE, sempre zelaram pela educação dos filhos, ao ponto que nunca houve muita cobrança por parte deles em sermos, como diz minha Mãe, “alguém de futuro”, e sem termos muito, minhas irmãs e eu fomos atrás de nosso “lugarzinho ao sol”.

Aos meus grandes amigos, em especial ao Ademar, que me apoiaram e me incentivaram a realizar este curso, não poderia deixar de agradecê-los.

Ao professor e orientador Paulo Matos, quero que saiba que é um professor exemplar, tanto pelo cuidado com os alunos, como pela organização, conhecimento e compromisso que apresentou nas poucas aulas que tivemos.

À Universidade Federal do Ceará – UFC e a FECOMERCIO por terem acreditado neste projeto.

Por último, e mais importante, agradeço ao meu senhor e salvador Jesus Cristo, pois de fato sinto que foi através D’ele que os rumos da minha vida convergiram para este momento.

RESUMO

O presente trabalho objetiva criar um modelo que melhor represente a estimativa do saldo de reservas financeiras das pessoas considerando algumas de suas características individuais, como também, discriminar quais das variáveis representam melhor esse modelo. As variáveis utilizadas foram idade, gênero, renda, cidade residente e a existência ou não de outros investimentos. O estudo se restringe a análise de perfil dos poupadores do estado do Ceará, para tanto, foram levantadas informações cadastrais descaracterizadas de correntistas de caderneta de poupança da Caixa Econômica Federal que residem no Estado. A escolha do ativo a ser analisado foi baseada em estudos recentes apresentados pela Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais – ANBIMA, onde evidencia que o investimento preferido dos brasileiros é a caderneta de poupança. A análise foi realizada utilizando uma técnica de análise multivariada de dados denominada análise discriminante. Aproveitamos e executamos a modelagem em dois períodos, julho de 2019 e julho de 2020, com intuito de verificar se houve alterações nos padrões de poupança da população cearense em consequência da crise sanitária causada pela COVID-19. Os resultados apresentados demonstram que as informações básicas coletadas pelos bancos no processo de abertura de conta-poupança produziram uma função discriminante capaz de classificar corretamente 36,1% das amostras no ano de 2020 e 36,7% no ano de 2019, além de que, não é identificada, até o momento deste estudo, alterações significativas nas características do poupador acarretadas pelo efeito da pandemia.

Palavras-chave: Análise Discriminante. Perfil de Consumo. Poupança.

ABSTRACT

The present work aims to create a model that best represents the estimated balance of financial reserves of people considering some of their individual characteristics, as well as to discriminate which variables best represent this model. The variables used were age, gender, income, city of residence and the existence or not of other investments. The study is restricted to the analysis of the profile of savers in the state of Ceará, for which uncharacterized registration information was collected from Caixa Econômica Federal savings account holders who reside in the state. The choice of the asset to be analyzed was based on recent studies presented by the Brazilian Association of Financial and Capital Market Entities - ANBIMA, which shows that the preferred investment of Brazilians is the savings account. The analysis was performed using a multivariate data analysis technique called discriminant analysis. We took advantage of this and performed the modeling in two periods, July 2019 and July 2020, in order to verify whether there were changes in the savings patterns of the Ceará population as a result of the health crisis caused by COVID-19. The results presented show that the basic information collected by banks in the process of opening a savings account produced a discriminant function capable of correctly classifying 36.1% of the samples in the year 2020 and 36.7% in the year 2019, besides that, it is not identified, up to the moment of this study, significant changes in the characteristics of the saver brought about by the effect of the pandemic.

Keywords: Discriminant Analysis. Consumption Profile. Savings.

LISTA DE QUADROS E FIGURAS

Figura 1	- Principais Aplicações no País (em milhões)	23
Figura 2	- Divisão do Saldo de Poupança no Brasil.....	24
Figura 3	- Distribuição do Saldo por Município.....	25
Figura 4	- Divisão do Saldo de Poupança no Ceará.....	25
Figura 5	- Função Discriminante Canônica para 2019.....	43
Figura 6	- Função Discriminante Canônica para 2020.....	43

LISTA DE TABELAS

Tabela 1	- Faixa de Renda, Saldo e Idade.....	21
Tabela 2	- Distribuição da Amostra por Gênero.....	26
Tabela 3	- Distribuição da Amostra por Faixa Etária.....	26
Tabela 4	- Distribuição da Amostra por Faixa de Renda.....	27
Tabela 5	- Distribuição da Amostra por Faixa de Saldo.....	28
Tabela 6	- Distribuição da Amostra por Investimentos.....	28
Tabela 7	- Testes de igualdade de médias de grupo.....	29
Tabela 8	- Matriz Covariância e Correlação dentro de um pool para o ano de 2019.....	30
Tabela 9	- Matriz Covariância e Correlação dentro de um pool para o ano de 2020.....	30
Tabela 10	- Resultados do teste M- Box.....	31
Tabela 11	- Autovalores.....	32
Tabela 12	- Lambda de Wilks.....	32
Tabela 13	- Resultados da classificação prevista para os grupos.....	35
Tabela 14	- Estatísticas dos Grupos por ano.....	42

LISTA DE ABREVIATURAS E SIGLAS

ANBIMA	Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais
BACEN	Banco Central do Brasil
CEF	Caixa Econômica Federal
COVID-19	<i>Corona Virus 19</i>
TR	Taxa Referencial
FGC	Fundo Garantidor de Crédito

SUMÁRIO

1	INTRODUÇÃO	10
2	REFERENCIAL TEÓRICO	13
2.1	Determinantes de Poupança	14
3	FUNDAMENTAÇÃO TEÓRICA E METODOLOGICA	16
3.1	Estatística Multivariada - Análise Discriminante	16
3.2	O evento a ser modelado	20
3.3	A escolha do ativo analisado	21
3.3.1	<i>Contexto sobre caderneta de poupança</i>	22
3.3.2	<i>A caderneta de poupança no Estado do Ceará</i>	24
4	ANÁLISE DOS RESULTADOS	26
4.1	Análise Descritiva dos Dados	26
4.2	Análise no Modelo Discriminante	28
4.2.1	<i>Seleção das Variáveis Discriminantes</i>	29
4.2.2	<i>Matriz de Correlação e Covariância</i>	29
4.2.3	<i>Teste de Homogeneidade (M de box)</i>	30
4.2.4	<i>Os Autovalores e Lambda de Wilks</i>	34
4.2.5	<i>Coefficiente da Função Discriminante Canônica e da Função de Classificação</i>	33
4.2.6	<i>Resultado da Classificação</i>	34
5	CONCLUSÃO	36
	REFERÊNCIAS	39
	APÊNDICE	42

1 INTRODUÇÃO

A preocupação com a formação de poupança para aposentadoria vem sendo objeto de estudo de grande relevância no meio acadêmico, uma vez que demonstra o quanto de esforço as famílias dedicam a manutenção do padrão de consumo atual com o intuito de manter o poder de compra no momento que o ente provedor do sustento familiar sair da fase produtiva e começar a despoupar (aposentadoria). Essas e outras informações sobre avanços nos estudos que dizem respeito à poupança e aposentadoria, no campo das finanças comportamentais, têm motivado inúmeras pesquisas.

Neste sentido, temos os estudos de Keynes (1936) que afirmam que é a renda quem vai determinar a composição de poupança e consumo das pessoas, ou seja, quanto maior a renda, maior será o valor poupado. Friedman (1957), Ando e Modigliani (1963) assumem que os agentes econômicos tomam decisões financeiras de consumo, poupança, investimento e financiamento de forma racional, buscando maximizar a utilidade ao longo do tempo. Shefrin e Thaler (1988) relatam que a maximização da utilidade está diretamente ligada a propensão que as pessoas possuem em manter determinado nível de poupança para assegurar a capacidade de consumo durante a aposentadoria. Shefrin e Statman (1985) mencionam que os indivíduos são avessos a riscos para ganhos, mas propensos a riscos para perdas. Kahneman e Tversky (1979) concluíram que os indivíduos, por natureza, são avessos à perda e, assim, eles sentem mais a “dor” causada pela perda do que os benefícios gerados pelo ganho de mesma proporção.

A relação consumo e poupança também produziu fundamentações teóricas relacionadas à sua importância no crescimento econômico de um país a longo prazo. O modelo de Solow relata que um aumento na taxa de poupança provoca aumento permanente no nível de renda do indivíduo, sugerindo que uma das formas em tornar o país mais rico seria elevando sua taxa de poupança.

As informações geradas nesse campo de conhecimento podem servir para os agentes públicos e privados dimensionarem esforços com vistas a redução dos riscos relativos à previdência social dos países, já que o desequilíbrio na previdência é atribuído à falta de planejamento dos agentes (públicos e privados), como também das famílias, além de produzir conhecimento a respeito da adequada de formação de reservas financeiras para garantia da manutenção do poder consumo na aposentadoria.

No Brasil, dentro das pesquisas relacionadas ao tema, destaca-se o estudo realizado pela Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais – ANBIMA cujo título é “*Raio X do Investidor Brasileiro*”. Através de dados coletados pelo instituto Data-folha, o trabalho tem como objetivo entender o comportamento dos investidores e da população em geral quando lidam com suas finanças e as motivações e preocupações na hora de investir, ou não, em produtos financeiros. Segundo dados do estudo, dentre os ativos financeiros investidos no ano de 2019, 84,2% dos investidores alocaram seus recursos em caderneta de poupança, frente aos 88% no ano de 2018 e 89% em 2017, portanto, a caderneta de poupança é o investimento de maior popularidade do Brasil. Cabe ressaltar que, o mesmo estudo evidencia que 28% dos entrevistados mencionaram espontaneamente a caderneta de poupança como investimento, demonstrando que há sim, por parte dos investidores, conhecimento sobre outros ativos financeiros.

Considerando a democratização e o alcance da caderneta de poupança na sociedade, parece fazer sentido analisar o perfil do poupador brasileiro através das informações cadastrais de seus correntistas. No entanto, o levantamento de tais informações, ao ponto de identificar características pessoais relevantes que possa ser objeto de estudo, não são divulgadas pelas instituições financeiras, e quando divulgadas são apresentadas de forma genérica, afinal, nenhuma instituição financeira gosta de expor sua carteira de clientes ao mercado sem necessidade.

Conforme dados registrados no Banco Central do Brasil – BACEN, posição 03/2020, a Caixa Econômica Federal é a instituição que detém maior percentual de saldo de poupança no país, equivalente a 37% do volume total aplicado, enquanto no estado do Ceará, para o mesmo período, esse percentual aumenta para 49%.

É imperativo considerar no estudo, buscando evidenciar informações e resultados atuais na análise, a dimensão que a crise econômica acarretada pela pandemia SARS-COV-2, popularmente denominada COVID-19, acarretou no comportamento do poupador, pois, sem reservas financeiras e com a perda ou diminuição de renda, a manutenção ou formação de poupança perde espaço dentro do orçamento das famílias em contrapartida das necessidades primárias de subsistência, tais como, despesas com saúde, alimentação, água, luz, gás, etc.

Por outro lado, a crise também traz ensinamentos relevantes para quem poupa, obrigando as pessoas a terem uma atenção maior com orçamento familiar, ocasionando

controle melhor das receitas e despesas, alterações no estilo de vida, ajuste de gastos, mudanças ou adiamento de projetos e reajuste de dívidas bancárias.

Diante do exposto, e tendo em vista a importância envolvida em torno do tema, esta pesquisa busca evidenciar características de perfil do poupador cearense por meio de técnicas de estatística multivariada de análise discriminante, utilizando como fonte de dados para análise informações cadastrais dos correntistas de conta poupança da Caixa Econômica Federal no Ceará.

O trabalho está dividido em mais 5 seções, além dessa introdução. Na seção 2 é apresentado o referencial teórico, enquanto na seção 3 é evidenciada a fundamentação teórica e metodológica utilizada, detalhando a base de dados. Na seção 4 e 5 são realizadas a análise dos dados e a conclusão obtida, respectivamente. Por fim, na seção 6 realizamos as considerações finais sobre o estudo.

2 REFERENCIAL TEÓRICO

É reconhecido que o nível de poupança afeta não apenas a qualidade de vida de um indivíduo, mas a perspectiva de crescimento econômico do país. No entanto, o ato de poupar é um fenômeno complexo que envolve fatores psicológicos, comportamentais e sociais, os quais influenciam na capacidade de planejamento das pessoas e, assim, impedem que esta decisão siga uma linha estritamente racional e econômica.

Alguns pesquisadores têm estudado as determinantes do comportamento de poupança por meio da análise de variáveis socioeconômicas, tais como: idade, educação, renda, patrimônio e etnia, já outros direcionam seus estudos na influência do fator psicológico na tomada de decisão, por exemplo, tolerância a risco, incertezas, autocontrole e preferências temporais.

Keynes (1936) foi pioneiro ao levantar oito diferentes motivos para poupar, aos quais Browning e Lusardi (1996) adicionaram mais um motivo, sendo estes: (1) motivo de precaução contra imprevistos; (2) motivo de provisão para futuras disparidades entre receitas e despesas; (3) motivo de cálculo intertemporal para ganhos com juros; (4) motivo de melhoria do padrão de vida; (5) motivo de independência; (6) motivo de empreender quando o cenário for favorável; (7) motivo de legado para deixar uma herança; (8) motivo de avareza; e (9) motivo de pagamento de novas aquisições. Esta lista mostra a considerável heterogeneidade nos motivos para economizar, contudo, os motivos de poupança não são necessariamente mutuamente exclusivos, pois as famílias podem economizar por razões de precaução, mas havendo a expectativa confiável de que as reservas não serão utilizadas, os recursos poderão ter outra destinação, por exemplo, uma viagem (Dyner, Skinner e Zeldes; 2004; Smith, 1999). É pouco provável que um motivo seja suficiente para a mesma pessoa por um longo período, além de que muitos motivos podem ser complementares.

Considerando inúmeras propriedades e fatores que colaboram à tomada de decisão dos indivíduos em fomentar o acúmulo de reservas financeiras, questionamos se é possível criar qualquer estimativa onde se possa, em termos quantitativos, gerar uma previsibilidade de poupança das pessoas considerando algumas de suas características.

2. 1. Determinantes de Poupança

Inevitavelmente, quando falamos em formação de poupança já relacionamos como característica determinante a renda, parece até mesmo ser uma variável inerente ao tema. A maior parte dos estudos constataram que, em condições constantes, o aumento da renda nas famílias repercute no aumento do consumo e na formação de poupança. Segundo Keynes (1936), a taxa de poupança aumenta com a renda corrente, já Friedman (1957) argumenta que a taxa de poupança pode depender indiretamente da renda corrente, onde é inserida a influência da renda transitória na decisão de poupança.

A Teoria do Prospecto indica que as pessoas aumentarão a poupança quando a renda for alta e economizarão menos quando a renda for baixa (Kahneman e Tversky, 1979). Por sua vez, Bowman *et. al.* (1999) estenderam a Teoria do Prospecto evidenciando que as pessoas resistem a reduzir o consumo em resposta a uma diminuição na renda atual ou futura quando a incerteza da renda está presente, mas são mais propensas a aumentar imediatamente o consumo em resposta a um aumento na renda atual ou futura.

Outra variável bastante presente nas pesquisas é a idade, quer seja pela perspectiva racional da formação de poupança, onde os indivíduos ajustam seus comportamentos e postergam o consumo para o período de aposentadoria de forma planejada, quer seja pela ótica de que as pessoas suavizam o consumo de acordo com a idade e, assim, geraria um saldo orçamentário maior produzindo uma conseqüente formação de reservas. A idade é abordada por Modigliani e Brumberg (1954), Friedman (1957) e Shefrin e Thaler (1988) como uma das principais variáveis determinantes. Estudos mais recentes também incluem a idade como um dos fatores determinantes na formação de poupança (Silveira e Moreira, 2017; Mendes-Da-Silva *et. al.*, 2015; Fisher e Montalto, 2011; Lee e Hanna 2015).

Pesquisas demonstram que o nível de conhecimento financeiro das pessoas traz uma relação positiva com comportamentos de poupança, investimentos e preparação para aposentadoria. Bernheim, Garrett e Maki (2001) mostraram que indivíduos que participaram de programas de educação financeira melhoraram seu comportamento de poupança e apresentaram um patrimônio líquido mais valioso na fase adulta. Babiarz e Robb (2014) estudaram a relação entre competência financeira e a manutenção de reservas para emergências, e concluíram que o nível de conhecimento financeiro repercute significativamente sobre a criação de reservas de emergência.

Entender o comportamento entre homens e mulheres ajudam muitos planejadores na definição de estratégias para formação de poupança, por conta disso, sexo é uma variável utilizada com relativa frequência. Fisher (2010) argumenta que mulheres são menos propensas a poupar no curto prazo. Além disso, as mulheres tendem a exibir menor tolerância a risco, comparativamente aos homens. Lusardi e Mitchell (2007) descobriram que as mulheres geralmente tinham menos conhecimento financeiro do que os homens, e que a alfabetização financeira afetou tanto a poupança quanto a escolha da carteira. Pesquisas mostram que as preferências de homens e mulheres diferem, afetando as decisões de trabalho e consumo (Croson e Gneezy, 2004), sendo evidenciado que mulheres investem de maneira diferente dos homens.

3 FUNDAMENTAÇÃO TEÓRICA E METODOLÓGICA

3.1 Estatística Multivariada - Análise Discriminante

Para Hair Jr. *et. al.* (2005), a análise discriminante múltipla é a técnica multivariada adequada para estudar problemas em que a variável estatística é dicotômica e, portanto, não métrica. O valor previsto da função discriminante é o escore discriminante, o qual é calculado para cada objeto na análise.

$$Z = \alpha + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \dots + \alpha_n X_n \quad (3.1.1)$$

Sendo:

Z = escore discriminante

α = intercepto

α_i = peso ou coeficiente discriminante para a variável i , com $i = 1, 2, 3, \dots, n$

X_i = variável independente i

As variáveis independentes relevantes, formadoras da função discriminante, são obtidas por recurso à estatística Lambda de Wilks. Aplica-se também o cálculo da F univariada para cada uma das variáveis independentes e observar a significância a 5%, onde a rejeição da hipótese nula para variável é um indicativo favorável sobre ela. Para uma determinada variável independente X_l , com N objetos e de combinações lineares pertencentes M grupos, temos:

a) $\text{Lambda de Wilks} = \frac{SQB}{SQT}$

b) $\text{Razão } F = \frac{SQB}{SQW}$

c) SQT (Soma dos Quadrados Totais) = SQB (Soma dos quadrados intragrupos) + SQW (Soma dos quadrados intragrupos)

d) $SQB (M - 1) = \sum_{j=1}^M (\underline{X_{jl}} - \underline{X_l})^2$

- e) $SWQ(N - M) = \sum_{j=1}^M \sum_{i=1}^{n_j} (X_{ijl} - \underline{X}_{jl})^2$
- f) $SQT(N - 1) = \sum_{j=1}^M \sum_{i=1}^{n_j} (X_{ijl} - \underline{X}_j)^2$
- g) $\underline{X}_{jl} = \text{Média da variável independente } X_l \text{ no grupo } j$
- h) $\underline{X}_l = \text{Média de todos } n \text{ do grupo } M$

Os coeficientes $\alpha_1, \alpha_2, \dots, \alpha_n$ são estimados de modo que a variabilidade dos escores (D_i) da função discriminante seja máxima entre os grupos e mínima dentro dos grupos. Eles representam uma medida relativa da importância das variáveis originais na função estimada. Quanto maior for o coeficiente das variáveis independentes, maior será a sua contribuição na discriminação entre os grupos.

Dessa forma, dados p -variáveis e M -grupos, é possível estabelecer $m = (M - 1; p)$ funções discriminantes que são combinações lineares das p -variáveis, tal que:

$$D_i = \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \dots + \alpha_n X_n, \text{ com } i = 1, 2, \dots, n \quad (3.1.2)$$

Algumas condições são necessárias para o uso da análise discriminante, as suposições mais importantes são a de normalidade das variáveis independentes e igualdade nas matrizes de dispersão e covariância dos grupos. A falta de normalidade pode causar problemas na estimação da função discriminante, e matrizes de covariâncias desiguais podem afetar negativamente o processo de classificação (Hair Jr. *et al*, 2005). Outra característica dos dados que deverá ser evitada é a multicolinearidade entre variáveis independentes.

O pressuposto de linearidade das relações entre as variáveis está implícito na função discriminante, pois relações não lineares não são refletidas na função, a menos que transformações específicas de variáveis sejam executadas para representar efeitos não lineares.

Teste comumente utilizado para avaliar a suposição de normalidade univariada é o de Kolmogorov-Smirnov, ou ainda, Jarque-Bera (1980 e 1981), e para a gaussianidade conjunta é utilizado os testes de Mardia e Henze-Zikler, enquanto o mais utilizado para evidenciar a hipótese de homogeneidade das matrizes de covariância é o M de Box.

Antes da aplicação de testes estatísticos, devemos estabelecer o nível de significância α . Trata-se de uma medida a partir do qual assumimos como real o resultado da estatística obtido no estudo, ou seja, é a probabilidade de que uma prova estatística apresente um valor (p-valor) que rejeite a hipótese nula, H_0 , quando esta é verdadeira. A regra geral é rejeitar H_0 , se o p-valor $\leq \alpha$. Quanto menor o p-valor, mais forte será a evidência contra H_0 . Para este estudo, foi definido o nível de significância de 5%.

Dois métodos computacionais podem ser utilizados para determinar uma função discriminante: o método simultâneo e o método sequencial.

Na estimação simultânea, a função discriminante é computada de modo que todas as variáveis independentes sejam analisadas em conjunto, sem considerar o poder discriminatório de cada uma delas. Assim, a função discriminante é computada com base no conjunto inteiro de variáveis independentes, sem consideração do poder discriminatório de cada uma dessas variáveis.

No método sequencial, as variáveis são escolhidas conforme seu poder discriminatório. Primeiramente, é escolhida aquela variável que melhor discrimina os grupos. Em seguida, ela é combinada com cada uma das outras variáveis, para que seja selecionado o par de variáveis que melhora o poder discriminatório. Esse procedimento é repetido até que seja encontrado o melhor conjunto de variáveis para compor a função discriminante. A partir do momento em que novas variáveis são inseridas, é possível que algumas das incluídas anteriormente sejam removidas, caso a informação contida nela esteja disponível em uma combinação linear das demais escolhidas.

Cada uma das técnicas está associada para um momento diferente. Maroco (2003) sugere a utilização das estatísticas Distância de Mahalanobis, V de Rao e Razão F entre grupos para corroborar ou não a escolha efetuada com o uso do critério lambda de Wilks, e para diagnosticar possíveis problemas de multicolinearidade, o que implicaria a existência de viés no modelo, o autor indica a medida tolerância.

A Distância de Mahalanobis, DM_{ij} , entre variáveis i e j , é dada por:

$$DM_{ij} = \sqrt{\left[(X_i - X_j)' S^{-1} (X_i - X_j) \right]} \quad (3.1.3)$$

Onde:

X_i e X_j = são variáveis dos grupos i e j

S^{-1} = Matriz de covariância

O quadrado dessa distância é utilizado para demonstrar quanto um grupo se diferencia de outro. Assim, uma variável é adicionada a função discriminante se a sua inclusão aumentar significativamente a DM^2 . Essa estatística é sugerida quando as variáveis sob estudo apresentam correlações significativas que podem causar desvio no resultado da função discriminante.

O critério V de Rao objetiva maximizar a distância entre os centroides dos grupos. Para tanto, é computado o valor do centroide geral pela fórmula:

$$V = -(n - g) \sum_{i=1}^p \sum_{j=1}^p w_{ij}^{-1} (t_{ij} - w_{ij}) \quad (3.1.4)$$

Onde:

n = dimensão da amostra

g = número de grupos

w_{ij} e t_{ij} = elementos genéricos das matrizes W e T, respectivamente.

A significância da variação do V de Rao, derivada da adição ou remoção de uma variável, pode ser avaliada por aproximação a distribuição qui-quadrado com $g-1$ graus de liberdade. A razão F entre grupos é uma transformação da distância de Mahalanobis considerando as diferentes dimensões entre os grupos, de maneira que os grupos de maior dimensão tenham maior peso na análise. Dados dois grupos a e b, a razão F, F_{ab} , é computada como:

$$F_{ab} = \frac{(n - p - g)n_a n_b}{p(n - p)(n_a + n_b)} DM_{ab}^2 \quad (3.1.5)$$

Onde:

n = dimensão da amostra

n_a = número de elementos do grupo a

n_b = número de elementos do grupo b

g = número de grupos

p = número de variáveis

DM_{ab}^2 = Distância de Mahalanobis entre os grupos a e b.

A medida de tolerância de uma variável X_i é definida como $T = 1 - R_i$, onde R^2 é o coeficiente de determinação do modelo composto da variável dependente, X_i , e as variáveis independentes, X_j . Quanto mais próximo de 0 (zero) é o valor da medida de tolerância, maior é o problema de correlação entre variáveis e mais instável será o modelo. Por outro lado, quanto mais próximo de 1 (um), mais estável será o modelo.

3.2 O evento a ser modelado

O evento a ser modelado neste trabalho é o saldo de aplicação em poupança da população cearense através de características individuais dos correntistas.

Para tanto, fora levantada uma base de dados com informações cadastrais descaracterizadas de 6953 clientes pessoas físicas do Estado detentores de conta poupança na Caixa Econômica Federal, posição de julho/2020, e de mais 6243 clientes com posição de julho/2019. Os dados levantados contemplam o saldo, cidade de residência, gênero, faixa etária e faixa de renda, acrescentamos também a informação se estes clientes possuem ou não outros investimentos além da poupança, tais como, fundos de investimento, aplicações em renda fixa, tesouro SELIC, ações e outros.

Os dados referentes à renda, saldo e idade foram extraídos e segmentados por faixa, tais como é descrito na tabela 1.

Tabela 1: Faixas de Renda, Saldo e Idade

Renda	Saldo	Idade
Até R\$ 1.000,00	Até R\$ 1.000,00	Até 18 anos
R\$ 1.000,00 até R\$ 2.000,00	R\$ 1.000,00 até R\$ 5.000,00	18 a 30
R\$ 2.000,00 até R\$ 3.000,00	R\$ 5.000,00 até R\$ 20.000,00	30 a 40
R\$ 3.000,00 até R\$ 5.000,00	R\$ 20.000,00 até R\$ 35.000,00	40 a 50
R\$ 5.000,00 até R\$ 10.000,00	R\$ 35.000,00 até R\$ 50.000,00	Acima de 50
R\$ 10.000,00 até R\$ 20.000,00 Acima de R\$ 20.000,00	Acima de R\$ 50.000,00	

Fonte: Elaborado pelo autor, com dados disponibilizados pela CEF

A informação quanto a existência de outros investimentos não é representada por saldo aplicado, uma vez que, o substancial da informação é identificar o conhecimento dos indivíduos em outros ativos financeiros por meio da posse de, no mínimo, outro produto de investimento.

3.3 A escolha do ativo analisado

Segundo dados apresentados pela Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais – ANBIMA, através do trabalho “*Raio X do Investidor Brasileiro*”, 38% dos brasileiros que conseguiram economizar em 2019, 48% deste universo destinaram seus recursos para aplicação em produtos financeiros, por exemplo, renda fixa, títulos públicos, poupança ou mercado de ações, os demais destinaram suas economias para outros fins, como, educação, imóveis, viagens, etc.

O mesmo estudo aponta que 66% das pessoas que alocaram suas economias em ativos financeiros aplicaram no mesmo ativo já existente no portfólio, denotando uma baixa diversificação da carteira ou mudança de perfil do investidor. Outro ponto de destaque é que 50% da população não conhece ou não utilizam nenhum tipo de investimento, e já para os que

investem, a poupança é a principal destinação dos recursos, recebendo 84,2% do montante aplicado.

Na pesquisa, dentre os diferentes produtos, a poupança é a que apresenta maior média de tempo de aplicação, que é de onze anos, à frente dos planos de previdência privada e fundos de investimentos, que foram de sete anos. Essa informação induz que, quando se fala em planejamento de médio e longo prazo, o brasileiro prefere deixar os seus recursos em caderneta de poupança. Bem como, em situações de exigibilidade imediata de capital ou em contexto de elevação de risco, é preferível manter os recursos em poupança em detrimento de outros ativos.

Diante do exposto e considerando a grande representatividade que o investimento em caderneta de poupança possui para os brasileiros, parece ser bastante coerente desenvolver a análise sobre as características do poupador, em sentido amplo, através das informações destes correntistas.

3.3.1 Contexto sobre caderneta de poupança

As cadernetas de poupança foram criadas pelo Imperador Dom Pedro II por meio do decreto de nº 2.723 de 12 de janeiro de 1861 que instituiu e regulou a criação da Caixa Econômica. O artigo 1º do referido decreto menciona que:

“A Caixa Econômica estabelecida na cidade do Rio de Janeiro, em virtude do art. 2º, §§ 1º e 14. a 16. da lei nº 1.083 de 22 de agosto de 1.860, tem por fim receber a juro de 6%, as pequenas economias das classes menos abastadas, e de assegurar, sob garantia do Governo Imperial a fiel restituição do que pertencer a cada contribuinte, quando este o reclamar na fórmula do art. 7º deste Regulamento.”¹

As regras para captação de depósitos em poupança estão atualmente definidas pela Resolução nº 3.459, de 27/03/08, e alterações posteriores.

¹ fonte: www2.camara.leg.br

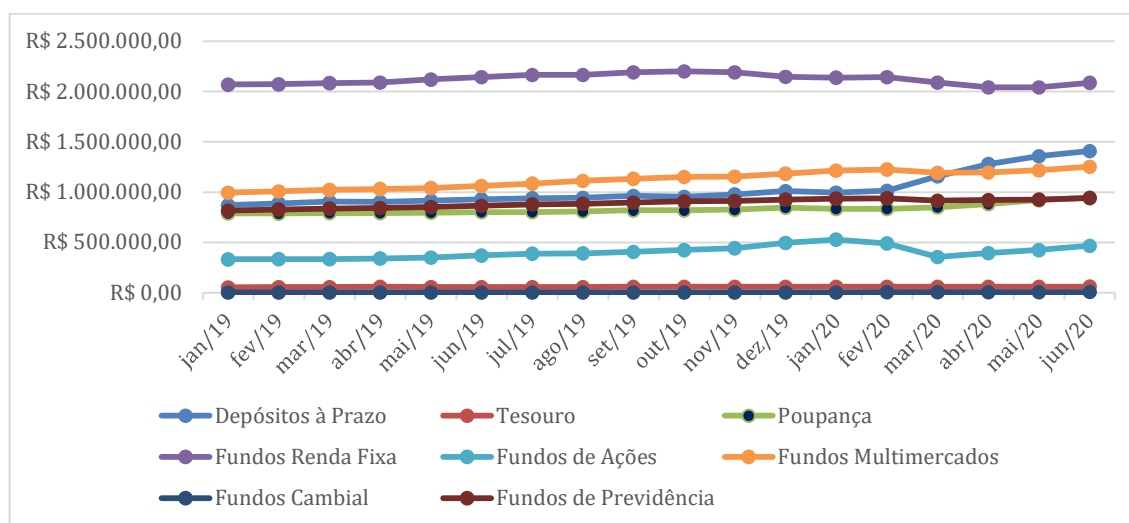
A caderneta de poupança, para os recursos aplicados até 03/05/12, é remunerada mensalmente ou trimestralmente. Quando for o caso, por uma taxa de juros fixa de 0,50% ao mês, atualizados pela taxa referencial (TR), para os valores aplicados a partir da data de 04/05/12, de acordo com a MP nº 567, de 03/05/12, posteriormente convertida pela Lei nº 12.703, de 07/08/12, a remuneração da taxa de juros da poupança será da seguinte forma:

- Taxa SELIC igual ou inferior a 8,5% ao ano, a remuneração da poupança terá uma taxa equivalente a 70% da meta da taxa SELIC ao ano;
- Taxa SELIC superior a 8,5% ao ano, a remuneração dos recursos aplicado em poupança permanece igual a que existia anteriormente à edição da MP.

Atualmente a rentabilidade das cadernetas de poupança pessoa física e pessoa jurídica sem fins lucrativos são isentas de imposto de renda, a incidência de imposto de renda ocorre somente nas cadernetas de poupança para pessoa jurídica. Os recursos aplicados em poupança estão amparados pela proteção do Fundo Garantidor de Crédito – FGC.

A figura 1 apresenta a evolução dos principais investimentos negociados em instituições financeiras do país e, apesar da poupança não possuir relevante alocação de recursos dos investidores pessoa jurídica devido à baixa rentabilidade, como também pela incidência de imposto de renda, evidenciamos expressiva participação do ativo quando comparado aos demais.

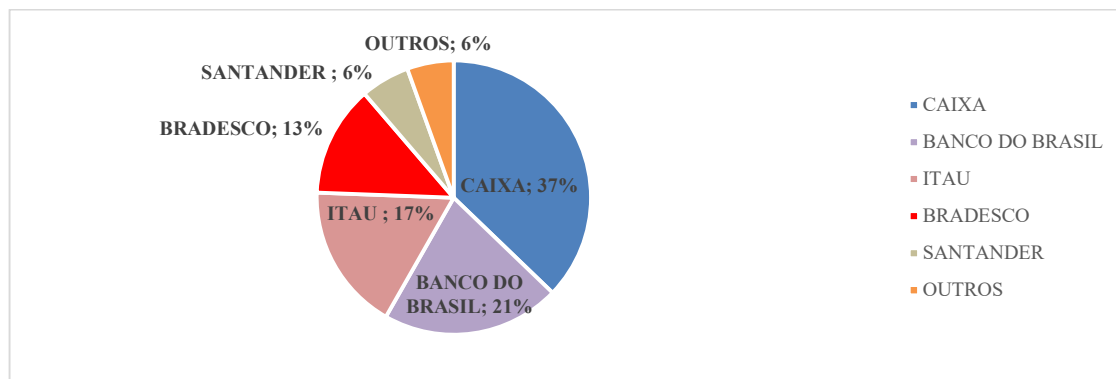
Figura 1: Principais Aplicações no País (em milhões)



Fonte: BACEN, ANBIMA e TESOURO.

Segundo dados levantados junto ao Banco Central, posição março de 2020, com 37% do investimento total, a Caixa Econômica Federal é o Banco que possui maior saldo de poupança no Brasil, seguido pelo Banco do Brasil, Itaú, Bradesco e Santander. (figura 2)

Figura 2: Divisão do Saldo de Poupança no Brasil

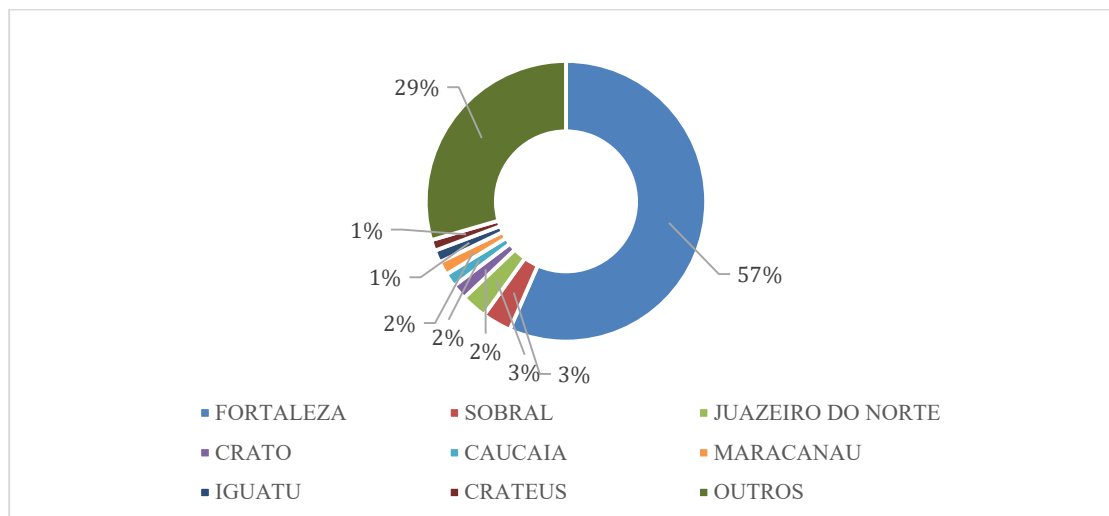


Fonte: BACEN

3.3.2 A caderneta de poupança no Estado do Ceará

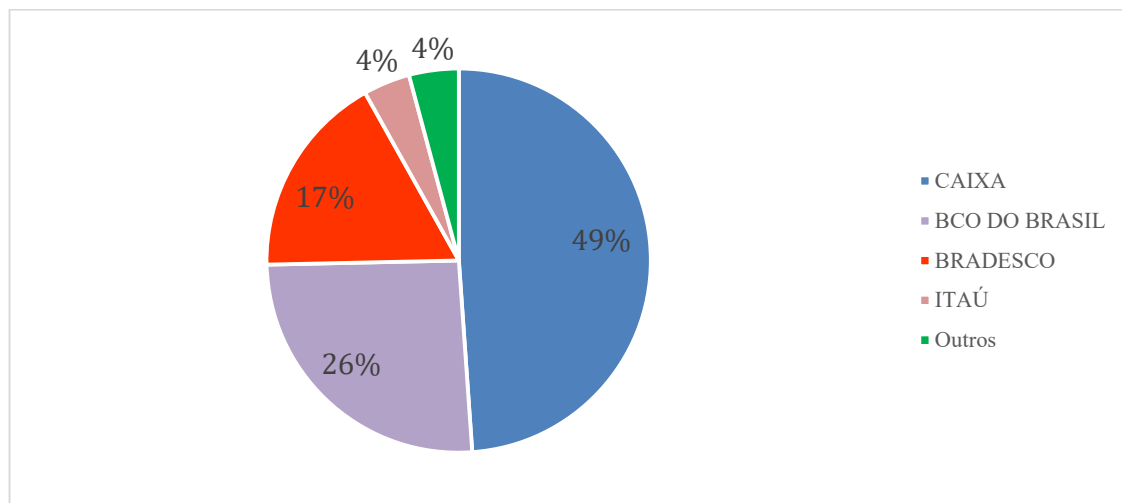
Em março de 2020, o somatório das contas de saldo de poupança das instituições financeiras sediadas no Estado representava o montante de R\$ 19,2 bilhões (fonte BACEN, posição mar/2020).

A figura 3 evidencia como ocorre essa distribuição de saldo por município, onde é identificada uma relevante concentração das aplicações em instituições financeiras da capital, sendo 57% da totalidade dos recursos, seguido por Sobral, Juazeiro do Norte, Crato, Caucaia, Maracanaú, Iguatú e Crateús. Os outros 29% dos saldos aplicados estão distribuídos em instituições de 94 municípios.

Figura 3: Distribuição do Saldo por Município

Fonte: BACEN

Segundo dados apresentados pelo BACEN, posição março de 2020, a Caixa Econômica Federal possui 49% dos saldos aplicados em poupança no Estado, apresentando uma participação maior quando comparada a participação total do país, este saldo é distribuído em 2.989.121 contas².

Figura 4: Divisão do Saldo de Poupança no Ceará

Fonte: BACEN

² Fonte: Caixa Econômica Federal

4. ANÁLISE DOS RESULTADOS

4.1 Análise Descritiva dos Dados

A variável gênero está representada por 3569 amostras do sexo masculino para o ano de 2020 e 3226 para o ano de 2019, já em relação as mulheres, tivemos 3384 amostras no ano de 2020 e 3017 em 2019. Os dados da tabela 2, demonstram que não é identificada distinção relevante entre a quantidade de contas de homens e mulheres nos dois anos.

Tabela 2: Distribuição da Amostra por Gênero

2019	6243	100%
Homem	3226	51%
Mulher	3017	49%
2020	6953	100%
Homem	3569	51%
Mulher	3384	49%

Fonte: Elaborado pelo autor

Quando se observa a distribuição da amostra pela faixa etária, 50% dos correntistas estão entre 30 e 50 anos de idade. Até os 30 anos é representado por 17%, e acima de 50 anos em 33%. Não houve diferenciação percentual relevante entre os dois períodos analisados. Os dados da tabela 3 indicam que a busca pela formação de poupança ou o processo de bancarização é mais forte a partir dos 30 anos de idade, e a informação condiz com o ciclo de aferição de renda das pessoas e a capacidade de geração de reservas.

Tabela 3: Distribuição da Amostra por Faixa Etária

2019	c	100%
00-18	213	3%
18-30	849	14%
30-40	1483	24%
40-50	1560	25%
50+	2138	34%
2020	6953	100%
00-18	211	3%
18-30	963	14%
30-40	1709	25%
40-50	1766	25%
50+	2304	33%

Fonte: Elaborado pelo autor

Como exposto na tabela 4, na faixa de renda, para o ano de 2020, 80% dos clientes estão enquadrados na faixa de renda até R\$ 3.000,00, sendo que, destes, 52% ganham até R\$ 1.000,00 por mês e 9% dos clientes ganham entre R\$ 3.000,00 a R\$ 5.000,00 por mês. Além disso, 7% dos clientes ganham entre R\$ 5.000,00 a R\$ 10.000,00 e 4% dos clientes ganham acima de R\$ 10.000,00.

Tabela 4: Distribuição da Amostra por Faixa de Renda

2019	6243	100%
Renda 1.000-2.000	1023	16%
Renda 10.000-20.000	195	3%
Renda 2.000-3.000	634	10%
Renda 3.000-5.000	602	10%
Renda 5.000-10.000	438	7%
Renda acima 20.000	98	2%
Renda até 1.000	3253	52%
2020	6953	100%
Renda 1.000-2.000	1170	17%
Renda 10.000-20.000	192	3%
Renda 2.000-3.000	731	11%
Renda 3.000-5.000	648	9%
Renda 5.000-10.000	464	7%
Renda acima 20.000	101	1%
Renda até 1.000	3647	52%

Fonte: Elaborado pelo autor

Já para a faixa de saldo, exposta na tabela 5, em julho de 2020, 47% dos correntistas apresentavam saldo de poupança de até R\$ 5.000,00, 28% entre R\$ 5.000,00 e R\$ 20.000,00, 18% apresentaram saldo entre R\$ 20.000,00 e R\$ 50.000,00, e 7% apresentaram saldo acima de R\$ 50.000,00. Para o período de 2019, houve variação negativa de 2% nas faixas de R\$ 5.000,00 a R\$ 20.000,00, ao passo que houve variação positiva em 3% na faixa de até R\$ 1.000,00.

Tabela 5: Distribuição da Amostra por Faixa de Saldo

2019	6243	100%
Poupança 1.000-5.000	1202	19%
Poupança 20.000-35.000	702	11%
Poupança 35.000-50.000	345	5%
Poupança 5.000-20.000	1623	25%
Poupança acima de 50.000	437	9%
Poupança até 1.000	1934	31%
2020	6953	100%
Poupança 1.000-5.000	1307	18%
Poupança 20.000-35.000	813	11%
Poupança 35.000-50.000	406	6%
Poupança 5.000-20.000	1978	28%
Poupança acima de 50.000	495	9%
Poupança até 1.000	1954	28%

Fonte: Elaborado pelo autor

Como descrito na tabela 6, verifica-se que 545 clientes apresentaram outro investimento além da poupança, representando 8% da base. Há um aumento na média de aplicação em outros produtos para investidores com saldo de poupança acima de R\$ 50.000,00. Não houve diferenças percentuais significativas entre os períodos analisados.

Tabela 6: Distribuição da Amostra por Investimentos

2019	6243	100%
Não	5739	91%
Sim	504	9%
2020	6953	100%
Não	6408	92%
Sim	545	8%

Fonte: Elaborado pelo autor

4.2 Análise no Modelo Discriminante

Para o processamento da análise discriminante no processo de modelagem da função, consideramos para os dois períodos, todas as amostras na análise de 2019 e 2020. Toda a amostra foi utilizada para o modelo discriminante, isto é, 6243 indivíduos para o ano de 2019 e 6953 para o ano de 2020. Essa diferença entre os anos se deu pela criação e o fechamento de novas contas na CEF.

O modelo utilizado considera faixa de saldo como escore discriminante, e o

município, gênero, faixa etária, faixa de renda e a existência de investimento como variáveis independentes. O resumo estatístico está disponível na tabela 14, no Apêndice.

O processamento do modelo ocorreu com o apoio do *software* SPSS, onde, tem-se como padrão o ponto de corte em 0,5. Este modelo apresentou um nível de acurácia adequado em relação aos grupos de clientes por faixa de saldo.

4.2.1 Seleção das Variáveis Discriminantes

O teste abaixo, exposto na tabela 7, apresenta a igualdade de médias dos grupos. Este quadro demonstra através da estatística de Lambda de Wilks, as variáveis preditoras com o melhor poder de discriminação, sendo que, quanto menor a estatística da variável, melhor a discriminação. Este fato pode ser comprovado através do nível de significância de cada variável, sendo que, níveis inferiores a 0,05, indicam uma diferença significativa entre as médias do grupo, e representam boas variáveis discriminadoras. No nível de significância de 5%, podemos inferir que somente a variável “gênero” não é discriminante para a função dos dados coletados de 2020 e 2019.

Tabela 7: Testes de igualdade de médias de grupo

Lambda de Wilks	2019			2020	
	Z		Sig.	Z	Sig.
Município	0,984	19,669	0,000	28,078	0,000
Gênero	1,000	0,365	0,873	1,484	0,191
Faixa Etária	0,971	37,377	0,000	29,402	0,000
Faixa Renda	0,960	51,986	0,000	63,070	0,000
investimentos	0,982	22,525	0,000	24,299	0,000

Fonte: Elaborado pelo autor

4.2.2 Matriz de Correlação e Covariância

A matriz de covariância e correlação, apresentada nas tabelas 8 e 9, entre as variáveis, contribui para a avaliação da relação entre as variáveis e a partir dela podemos perceber a presença de multicolinearidade entre as variáveis.

Nesse exemplo de 2020, a maior correlação foi de faixa de renda e investimento, ou seja, quanto maior sua faixa de renda, maior sua probabilidade de ter algum tipo de investimento (renda fixa, fundos de investimento, tesouro direto, ações ou outros investimentos).

As matrizes de covariância para cada um dos grupos, auxilia quanto a percepção de homogeneidade de covariância. Importante salientar que a presença de homogeneidade das matrizes de covariância é um dos pré-requisitos da análise discriminante. Entretanto, será por meio da estatística M de box que teremos condições de verificar se as diferenças observadas são ou não estatisticamente significantes.

Tabela 8: Matriz Covariância e Correlação dentro de um pool para o ano de 2019

		Município	Gênero	Faixa Etária	Faixa Renda	investimentos
Covariância	Município	13951231,016	2,110	232,641	247,093	-52,661
	Gênero	2,110	0,250	0,001	0,032	0,005
	Faixa Etária	232,641	0,001	1,320	0,303	0,031
	Faixa Renda	247,093	0,032	0,303	2,399	0,113
	investimentos	-52,661	0,005	0,031	0,113	0,073
Correlação	Município	1,000	0,001	0,054	0,043	-0,052
	Gênero	0,001	1,000	0,002	0,042	0,040
	Faixa Etária	0,054	0,002	1,000	0,170	0,100
	Faixa Renda	0,043	0,042	0,170	1,000	0,270
	investimentos	-0,052	0,040	0,100	0,270	1,000

A matriz de covariâncias possui 6237 graus de liberdade.

Fonte: Elaborado pelo autor

Tabela 9: Matriz Covariância e Correlação dentro de um pool para o ano de 2020

		Município	Gênero	Faixa Etária	Faixa Renda	investimentos
Covariância	Município	13850441,818	4,202	276,274	297,503	-53,877
	Gênero	4,202	0,250	-0,009	0,024	0,005
	Faixa Etária	276,274	-0,009	1,298	0,278	0,031
	Faixa Renda	297,503	0,024	0,278	2,284	0,107
	investimentos	-53,877	0,005	0,031	0,107	0,071
Correlação	Município	1,000	0,002	0,065	0,053	-0,054
	Gênero	0,002	1,000	-0,016	0,031	0,035
	Faixa Etária	0,065	-0,016	1,000	0,161	0,102
	Faixa Renda	0,053	0,031	0,161	1,000	0,267
	investimentos	-0,054	0,035	0,102	0,267	1,000

A matriz de covariâncias possui 6947 graus de liberdade

Fonte: Elaborado pelo autor

4.2.3 Teste de Homogeneidade (M de box)

O teste de homogeneidade tem como hipótese nula que não há diferença significativa entre os grupos, ou seja, que há homogeneidade das matrizes de covariância para

os grupos em análise. O resultado do teste terá que ser maior que 0.05 para concluirmos que a variância do grupo é igual, se este valor for menor que 0.05, então significa que temos uma variância de grupo desigual e rejeitamos a hipótese nula.

O resultado do teste M de Box, apresentado na tabela 10, determinou a rejeição da hipótese de serem iguais às matrizes de variâncias-covariâncias entre os grupos (heterocedasticidade). Assim, a função derivada pelo método direto foi invalidada, posto que houve violação de um dos principais pressupostos da análise discriminante.

Tabela 10: Resultados do teste M- Box

	2019	2020
M de Box	754,851	938,002
Z Aprox.	10,034	12,474
Sig.	0,000	0,000
Testa hipótese nula de matrizes de covariâncias de população igual.		

Fonte: Elaborado pelo autor

Apesar de rejeitar-se a hipótese nula, pode-se dar continuidade a análise. Entretanto, é necessário ter muito cuidado com a forma como se interpreta e se percebe os resultados, o que torná-os limitados.

4.2.4 Os Autovalores e Lambda de Wilks

Os autovalores e lambda apresentam os sumários das funções discriminantes canônicas. Os autovalores apresentam o percentual de diferenças entre os grupos, sendo uma medida na qual diferente são os grupos na função discriminante, ou seja, quanto mais afastado de 1 forem os autovalores, maiores serão as variações entre os grupos identificados pela função discriminante. A primeira função discriminante representa um percentual de 74,9% para o ano de 2020 e 69,6% para o ano de 2019. A função 2 também mostra-se com boa variância para os dois anos. Já as funções 3 e 4 não mostram poder discriminante. Para cada autovalor é calculado um Lambda de Wilks, onde identificamos que a primeira e segunda função discriminante são significantes. Os resultados são apresentados na tabela 11.

Tabela 11: Autovalores

Função	2019			2020		
	Autovalor	% de variância	Correlação canônica	Autovalor	% de variância	Correlação canônica
1	,073 ^a	69,6	0,260	,077 ^a	74,9	0,268
2	,028 ^a	26,9	0,165	,023 ^a	22,1	0,149
3	,003 ^a	3,3	0,059	,003 ^a	2,9	0,054
4	,000 ^a	0,2	0,015	,000 ^a	0,1	0,009
5	,000 ^a	0,0	0,001	,000 ^a	0,0	0,001

As primeiras 5 funções discriminantes canônicas foram usadas na análise.

Fonte: Elaborado pelo autor

Em relação à estatística de Lambda de Wilks, referente ao modelo como um todo, conforme a tabela 12, pode-se concluir que o modelo em questão possui capacidade discriminante e, portanto, está adequado frente ao nível de significância de 0,05. Sendo que, em relação ao Lambda de Wilks, quanto mais próximo de zero e do nível de significância, melhor a capacidade de discriminar do modelo.

Tabela 12: Lambda de Wilks

Teste de funções	2019		2020	
	Lambda de Wilks	Sig.	Lambda de Wilks	Sig.
1 até 5	0,904	0,000	0,905	0,000
2 até 5	0,969	0,000	0,975	0,000
3 até 5	0,996	0,007	0,997	0,013
4 até 5	1,000	0,857	1,000	0,966
5	1,000	0,929	1,000	0,951

Fonte: Elaborado pelo autor

4.2.5 Coeficiente da Função Discriminante Canônica e da Função de Classificação

Os coeficientes de função discriminante e de classificação apresentam os coeficientes não padronizados para as funções discriminantes de cada uma das variáveis explicativas e, assim, sendo possível escrever cada função referente. As figuras 5 e 6, disponíveis no Apêndice, mostram a função discriminante canônica para os anos de 2019 e 2020, respectivamente.

Para o ano de 2019, temos as seguintes equações como coeficientes da função discriminante canônica

$$\begin{aligned}
 Z_1 &= -289,293 + 0,00 \text{ Municipio} + (-0,0697) \text{ Gênero} + (-0,5247) \text{ Faixa}_{etária} \\
 &\quad + (0,4730) \text{ Faixa}_{renda} + (0,3691) \text{ Investimento} \\
 Z_2 &= 33,630 + 0,000 \text{ Municipio} + (-0,035) \text{ Gênero} + (0,553) \text{ Faixa}_{etária} \\
 &\quad + (0,184) \text{ Faixa}_{renda} + (2,028) \text{ Investimento} \\
 Z_3 &= -18,274 + 0,000 \text{ Municipio} + (0,590) \text{ Gênero} + (0,357) \text{ Faixa}_{etária} \\
 &\quad + (0,333) \text{ Faixa}_{renda} + (-3,125) \text{ Investimento} \\
 Z_4 &= -536,877 + 0,000 \text{ Municipio} + (-0,353) \text{ Gênero} + (0,268) \text{ Faixa}_{etária} \\
 &\quad + (-0,263) \text{ Faixa}_{renda} + (0,158) \text{ Investimento} \\
 Z_5 &= -105,166 + 0,000 \text{ Municipio} + (1,880) \text{ Gênero} + (-0,064) \text{ Faixa}_{etária} \\
 &\quad + (-0,156) \text{ Faixa}_{renda} + (0,936) \text{ Investimento}
 \end{aligned} \tag{4.2.5.1}$$

Ainda para o ano de 2019, as funções de classificação é apresentado abaixo:

$$\begin{aligned}
 FS_{até1000} &= -192334,897 + 0,167 \text{ Municipio} + (-0,156) \text{ Gênero} + (-26,595) \text{ Faixa}_{etária} \\
 &\quad + (-20,490) \text{ Faixa}_{renda} + (164,834) \text{ Investimento} \\
 FS_{1000-5000} &= -192230,458 + 0,167 \text{ Municipio} + (-0,134) \text{ Gênero} + (-26,547) \text{ Faixa}_{etária} \\
 &\quad + (-20,690) \text{ Faixa}_{renda} + (164,421) \text{ Investimento} \\
 FS_{5000-20000} &= -192175,007 + 0,167 \text{ Municipio} + (-0,150) \text{ Gênero} + (-26,406) \text{ Faixa}_{etária} \\
 &\quad + (-20,813) \text{ Faixa}_{renda} + (164,491) \text{ Investimento} \\
 FS_{20000-30000} &= -192146,785 + 0,167 \text{ Municipio} + (-0,087) \text{ Gênero} + (-26,183) \text{ Faixa}_{etária} \\
 &\quad + (-20,787) \text{ Faixa}_{renda} + (164,600) \text{ Investimento} \\
 FS_{30000-50000} &= -192103,049 + 0,167 \text{ Municipio} + (-0,018) \text{ Gênero} + (-26,116) \text{ Faixa}_{etária} \\
 &\quad + (-20,770) \text{ Faixa}_{renda} + (164,304) \text{ Investimento} \\
 FS_{acima 50000} &= -192171,328 + 0,167 \text{ Municipio} + (-0,212) \text{ Gênero} + (-26,153) \text{ Faixa}_{etária} \\
 &\quad + (-20,674) \text{ Faixa}_{renda} + (165,992) \text{ Investimento}
 \end{aligned} \tag{4.2.5.2}$$

Já para o ano de 2020, os coeficientes da função discriminante canônica são dados por:

$$\begin{aligned}
Z_1 &= -309,994 + 0,00 \text{ Municipio} + (0,006)\text{Gênero} + (-0,461)\text{Faixa}_{\text{etária}} \\
&\quad + (0,484)\text{Faixa}_{\text{renda}} + (0,568)\text{Investimento} \\
Z_2 &= 136,849 + 0,000 \text{ Municipio} + (0,221)\text{Gênero} + (0,542)\text{Faixa}_{\text{etária}} \\
&\quad + (0,169)\text{Faixa}_{\text{renda}} + (2,065) \text{ Investimento} \\
Z_3 &= 217,772 + 0,000 \text{ Municipio} + (0,899)\text{Gênero} + (0,150)\text{Faixa}_{\text{etária}} \\
&\quad + (0,394)\text{Faixa}_{\text{renda}} + (-2,907) \text{ Investimento} \\
Z_4 &= -373,817 + 0,000 \text{ Municipio} + (-0,694)\text{Gênero} + (0,511)\text{Faixa}_{\text{etária}} \\
&\quad + (0,021)\text{Faixa}_{\text{renda}} + (-1,386) \text{ Investimento} \\
Z_5 &= -293,814 + 0,000 \text{ Municipio} + (1,635)\text{Gênero} + (0,088)\text{Faixa}_{\text{etária}} \\
&\quad + (-0,254)\text{Faixa}_{\text{renda}} + (0,589) \text{ Investimento}
\end{aligned}
\tag{4.2.5.3}$$

E as funções de classificação, são definidas como:

$$\begin{aligned}
FS_{\text{até}1000} &= -194272,523 + 0,168 \text{ Municipio} + (-2,914)\text{Gênero} \\
&\quad + (-32,790)\text{Faixa}_{\text{etária}} + (-25,347)\text{Faixa}_{\text{renda}} \\
&\quad + (182,005)\text{Investimento} \\
FS_{1000-5000} &= -194198,678 + 0,168 \text{ Municipio} + (-2,963)\text{Gênero} \\
&\quad + (-32,744)\text{Faixa}_{\text{etária}} + (-25,528)\text{Faixa}_{\text{renda}} \\
&\quad + (181,651)\text{Investimento} \\
FS_{5000-20000} &= -194115,277 + 0,168 \text{ Municipio} + (-2,998)\text{Gênero} \\
&\quad + (-32,597)\text{Faixa}_{\text{etária}} + (-25,687)\text{Faixa}_{\text{renda}} \\
&\quad + (181,539)\text{Investimento} \\
FS_{20000-30000} &= -194060,277 + 0,168 \text{ Municipio} + (-2,907)\text{Gênero} \\
&\quad + (-32,467)\text{Faixa}_{\text{etária}} + (-25,667)\text{Faixa}_{\text{renda}} \\
&\quad + (181,544)\text{Investimento} \\
FS_{30000-50000} &= -193972,734 + 0,168 \text{ Municipio} + (-2,748)\text{Gênero} \\
&\quad + (-32,331)\text{Faixa}_{\text{etária}} + (-25,647)\text{Faixa}_{\text{renda}} \\
&\quad + (181,537)\text{Investimento} \\
FS_{\text{acima } 50000} &= -194094,098 + 0,168 \text{ Municipio} + (-2,934)\text{Gênero} \\
&\quad + (-32,369)\text{Faixa}_{\text{etária}} + (-25,556)\text{Faixa}_{\text{renda}} \\
&\quad + (182,984)\text{Investimento}
\end{aligned}
\tag{4.2.5.4}$$

4.2.6 Resultado da Classificação

Para finalizar a análise discriminante, é apresentado o resultado de classificação prevista para os grupos na tabela 13. Os resultados são expostos em termos de percentagem, tanto para os valores originais, como com validação cruzada

Aplicando as informações das amostras nas funções, para o ano de 2019, obtivemos 36,7% de classificação correta, já para o ano de 2020, 36,1% foram classificados corretamente.

Tabela13: Resultados da classificação prevista para os grupos ^{a,c}

Faixa Saldo		Associação ao grupo prevista													
		até 1.000	1.000-5.000	5.000-20.000	20.000-35.000	35.000-50.000	Acima de 50.000	Total	até 1.000	1.000-5.000	5.000-20.000	20.000-35.000	35.000-50.000	Acima de 50.000	Total
		2019							2020						
Original %	até 1.000	68,0	0,0	30,7	0,0	0,0	1,3	100,0	56,0	0,0	42,9	0,0	0,0	1,1	100,0
	1.000-5.000	52,8	0,0	46,2	0,0	0,0	1,0	100,0	43,5	0,0	55,7	0,0	0,0	0,8	100,0
	5.000-20.000	40,0	0,0	59,0	0,0	0,0	0,9	100,0	28,3	0,0	71,1	0,0	0,0	0,6	100,0
	20.000-35.000	36,8	0,0	62,3	0,0	0,0	1,0	100,0	26,9	0,0	72,7	0,0	0,0	0,4	100,0
	35.000-50.000	37,4	0,0	61,7	0,0	0,0	0,9	100,0	24,9	0,0	75,1	0,0	0,0	0,0	100,0
	acima de 50.000	45,8	0,0	50,3	0,0	0,0	3,9	100,0	35,6	0,0	62,0	0,0	0,0	2,4	100,0
Com validação cruzada ^b %	até 1.000	67,7	0,0	31,0	0,0	0,0	1,3	100,0	55,8	0,0	43,1	0,0	0,0	1,1	100,0
	1.000-5.000	52,8	0,0	46,2	0,0	0,0	1,0	100,0	43,5	0,0	55,7	0,0	0,0	0,8	100,0
	5.000-20.000	40,1	0,0	59,0	0,0	0,0	0,9	100,0	28,4	0,0	70,9	0,0	0,0	0,7	100,0
	20.000-35.000	36,8	0,0	62,3	0,0	0,0	1,0	100,0	26,9	0,0	72,6	0,0	0,0	0,5	100,0
	35.000-50.000	37,4	0,0	61,7	0,0	0,0	0,9	100,0	24,9	0,0	75,1	0,0	0,0	0,0	100,0
	acima de 50.000	46,0	0,0	50,3	0,0	0,0	3,7	100,0	36,0	0,0	62,6	0,0	0,0	1,4	100,0
a 36,7% de casos agrupados originais classificados corretamente.								a. 36,1% de casos agrupados originais classificados corretamente.							
b. A validação cruzada é feita apenas para os casos da análise. Na validação cruzada, cada caso é classificado pelas funções derivadas de todos os casos diferentes desse caso.								b. A validação cruzada é feita apenas para os casos da análise. Na validação cruzada, cada caso é classificado pelas funções derivadas de todos os casos diferentes desse caso.							
c. 36,6% de casos agrupados com validação cruzada classificados corretamente.								c. 36,0% de casos agrupados com validação cruzada classificados corretamente.							

Fonte: Elaborado pelo Autor

5 CONCLUSÃO

O presente trabalho apresentou uma função preditora que melhor distribuiu grupos de poupadores entre seis faixas de saldo de aplicação. As variáveis explicativas escolhidas foram: renda, faixa etária, gênero, cidade onde reside e se os clientes possuem ou não outros ativos de investimento (renda fixa, tesouro, ações e fundos de investimentos).

Estudos demonstram que a caderneta de poupança é o investimento mais popular do Brasil, portanto, os resultados apresentados sobre os aplicadores em caderneta de poupança trariam robustez analítica. Nesse contexto, restringimos à amostra aos clientes residentes no estado do Ceará.

Foram utilizados dois períodos na análise, julho de 2019 e julho de 2020, onde apresentamos o saldo de poupança de 6953, para o ano de 2020, e 6243, para o ano de 2019, clientes correntistas da CEF. A divisão em dois períodos fazia-se necessária pois deveríamos considerar o provável impacto da COVID-19 no comportamento de formação de poupança das pessoas. Ressalta-se que a base utilizada é descaracterizada e privada. As variáveis independentes escolhidas são informações básicas coletadas no processo de abertura de conta.

Criar um modelo econômico que gere uma previsibilidade aceitável de capacidade de poupança dos indivíduos, auxiliaria as instituições financeiras a adotarem estratégias mais assertivas no processo de prospecção e na segmentação desses clientes. A segmentação correta dos clientes gera uma atuação das instituições melhor direcionada, promovendo um ganho em eficiência operacional e um consequente incremento de margem.

Atualmente, a segmentação dos clientes nos bancos é fundamentada, basicamente, na renda que o correntista apresenta e/ou volume de negócios já existente, ou seja, são ações essencialmente reativas, e não havendo, por parte do funcionário, habilidade suficiente em identificar no primeiro atendimento um possível “bom” cliente, a maior parte desses correntistas são esquecidos ou não abordados. Antecipar esse movimento é de suma importância para construir estratégias que possibilitem ser um diferencial competitivo dentro do mercado bancário, ainda mais nesse contexto de expansão da rede bancária pelos meios digitais.

Utilizamos para modelagem do estudo uma técnica de análise multivariada de dados denominada de análise discriminante. A técnica objetiva classificar grupos de

indivíduos considerando características (variáveis) levantadas, buscando apresentar uma função que melhor discrimine o grupo analisado, como também, às variáveis mais representativas dentro da função.

A análise nos apresentou o seguinte:

- A variável “gênero” não possui significância estatística na análise. Portanto, o fato do correntista ser homem ou mulher não interfere na estimativa de seu saldo de poupança.
- Comparando o período de 2019 e 2020, não identificamos alteração no perfil de investimento do poupador cearense, portanto, o efeito COVID-19 não trouxe impactos no resultado.
- As variáveis utilizadas no estudo foram somente àqueles essenciais para a abertura de conta, exceto a variável “investimento”, não sendo abordadas variáveis que representam fatores psicológicos/motivacionais dos poupadores.
- As funções discriminaram corretamente 36,1% das amostras no ano de 2020 e 36,7% para o ano de 2019, apesar de usar na análise o mínimo de informações coletadas pelos bancos.
- Vale questionar a qualidade cadastral mantida pelas instituições financeiras, prejudicando o levantamento das informações e na geração do modelo, principalmente no que se refere à presença de *outliers*.
- As variáveis “investimentos” e “renda” discriminam melhor a função. A ação do cliente de possuir outras aplicações, independente de valor, pode ser consequência de um maior nível de educação financeira na busca de melhor retorno do patrimônio.
- Apesar de não ter considerado variáveis psicológicas, o aprofundamento das variáveis estudadas talvez pudesse discriminar melhor a função. Por exemplo, variáveis como “renda per capita familiar”, “se possui ou não emprego formal”, “patrimônio”, “possuir ou não operação de crédito”, etc.
- Variáveis educacionais e motivacionais são comumente utilizadas em estudos que determinam o comportamento da formação de poupança, portanto, características como “nível de escolaridade”, “formação”, “área de atuação”, “aversão a riscos”, etc, poderiam trazer maior poder estatístico ao estudo.

- Em cidades com indicadores de desenvolvimento melhores, sugere-se que sua população tenda a ter um poder aquisitivo maior e, conseqüentemente, reflita na sua capacidade de formação de poupança. Por outro lado, existem diferenças no custo de vida dos municípios, ou seja, duas pessoas que possuem a mesma faixa salarial, faixa etária e gênero, mas uma reside em Fortaleza/Ce e outra em Quixadá/Ce, onde possui custo de vida menor, é possível indicar que o quixadaense terá mais saldo orçamentário passível de formação de poupança do que o fortalezense. Sendo assim, é necessário um aprofundamento maior nestas questões em futuras análises, já que essas informações podem influenciar na interpretação da variável.
- Isolando a variável renda e comparando com o saldo aplicado, as amostras de menores rendas alocaram maiores esforços na formação de poupança. Ou seja, identificar variáveis que gerem essa motivação maior em poupar, poderia trazer variáveis melhor representadas.
- Apesar do presente trabalho utilizar como metodologia análise discriminante, é possível ampliar a discussão futura incorporando outros aspectos metodológicos, tais como, regressões lineares, bem como o uso de modelo *logit* ordenado e *probit* multivariado, até mesmo métodos de *machine learning*, que até o momento não foram encontrados durante a pesquisa do autor. A incorporação de tais metodologias poderá agregar maior robustez aos resultados obtidos no estudo.

Diante do exposto, e considerando somente as características apresentadas, foi possível gerar uma função discriminante capaz de agrupar corretamente 36% da população nos dois períodos analisados. Ou seja, as informações cadastrais básicas e coletadas pelos bancos no processo de abertura de conta-poupança foram suficientes para alocar um pouco mais de 1/3 dos clientes. Portanto, havendo um maior investimento na coleta de variáveis mínimas e confiáveis à serem utilizadas neste processo cadastral, poderíamos evidenciar uma representação maior em níveis estatísticos em relação ao estudo. Outro ponto apresentado é que a crise sanitária acarretada pela COVID-19 não acarretou, até o momento da análise, alterações nos padrões de formação de poupança do poupador cearense.

REFERÊNCIAS

- ANBIMA, **Raio X do Investidor Brasileiro** – 3ª Edição, 2020.
- ANDO, A., e MODIGLIANI, F. (1963). The ‘life-cycle’ hypothesis of saving: aggregate implications and tests. **American Economic Review**.
- BABIARZ, P., E ROBB, C.A. (2014). Financial literacy and emergency saving. **Journal of Family Economic Issues**, 35. doi:10.1007/s10834 -013- 9369-9
- BANCO CENTRAL DO BRASIL. **Resolução nº 3549**, de 27 de março de 2008.
- BERA, A., & JARQUE, C. (1981). Efficient tests for normality, heteroskedasticity, and serial independence of regression residuals: Monte Carlo evidence. **Economics Letters**, 7, 313–318
- BERNHEIM, B. D., GARRETT, D. M., & MAKI, D. M. (2001). Education and saving: The long-term effects of high school financial curriculum mandates. **Journal of Public Economics**, 80(3), 435-465.
- BOWMAN, D., MINEHART, D., E RABIN, M. (1999). Loss aversion in a consumption-savings model. **Journal of Economic Behavior & Organization**, 38. DOI:10.1016/S0167-2681(99)00004-9.
- BRASIL. **Decreto nº 2.723**, de 12 de Janeiro de 1861. Coleção de Leis do Império do Brasil - 1861, Página 11 Vol. 1 (Publicação Original).
- BRASIL. **Lei nº 1.083**, de 22 de Agosto de 1860. Contendo providencias sobre os Bancos de emissão, meio circulante e diversas Companhias e Sociedades.
- BRASIL. **Lei nº 12.703**, de 7 de Agosto de 2012. Altera o art. 12 da lei nº 8.177, de 10 de março de 1991, que estabelece regras para a desindexação da economia e dá outras providências, o art. 25 da lei nº 9.514, de 20 de novembro de 1997, que dispõe sobre o sistema de financiamento imobiliário, institui a alienação fiduciária de coisa imóvel e dá outras providências, e o inciso II do art. 167 da lei nº 6.015, de 31 de dezembro de 1973, que dispõe sobre os registros públicos e dá outras providências. Diário Oficial da União, Brasília, DF, 8 de ago. 2012, p 46, TEXTO ORIGINAL.
- BRASIL. **Medida Provisória nº 567**, de 3 de Maio de 2012. Diário Oficial da União, Poder Executivo, Brasília, DF, 3 mai. 2012. Seção 1 - Edição Extra - Página 1.
- BROWNING, M., & LUSARDI, A. (1996). Household saving: Micro theories and micro facts. **Journal of Economic literature**, 34(4), 1797-1855.
- CROSON, R., & GNEEZY, U. (2004). Gender differences in preferences. **Journal of Economic Literature**, 47, 448-474
- DO AMARAL GURGEL CARNEIRO DE OLIVEIRA, Otávio; FELIPE, Israel José dos Santos; MENDES-DA-SILVA, Wesley. Impacto de Crise sobre o Comportamento de Poupança de Famílias Americanas. **Innovar: Revista de ciencias administrativas y sociales**,

v. 27, n. 66, p. 137-151, 2017.

DYNAN, K. E., SKINNER, J., & ZELDES, S. P. (2004). Do the rich save more? **Journal of Political Economy**, 112(2), 397–444.

FISHER, P.J., E MONTALTO, C.P. (2010). Effect of saving motives and horizon on saving behaviors. **Journal of Economic Psychology**, 31. DOI:10.1016/j.joep.2009.11.002

FISHER, P.J., E MONTALTO, C.P. (2011). Loss aversion and saving behavior: Evidence from the 2007 U.S. Survey of Consumer Finance. **Journal of Family and Economic Issues**, 32. doi:10.1007/s10834-010-9196-1

FORTUNA, E. (2017). Mercado Financeiro: Produtos e Serviços. 21ª Edição Revista Atualizada. **Qualitymark** Editora Ltda.

FRIEDMAN, M. (1957). A theory of the consumption function. **Princeton University Press**. Princeton.

FURNHAM, A., & ARGYLE, M. (1998). The psychology of money. New York: **Routledge**.
HAIR JR.; JOSEPH, F.; ANDERSON, R. E.; TATHAM, R. L.; BLACK, W. Análise multivariada de dados. Tradução Adonai Schlup Santánn. Supervisão técnica Maria Aparecida Gouvêa. 6. ed. Porto Alegre: **Bookman**, 2009.

JARQUE, C., & BERA, A. (1980). Efficient tests for normality, heteroskedasticity, and serial independence of regression residuals. **Economics Letters**, 6, 255–259

KAHNEMAN, D., TVERSKY, A. (1979). Prospect Theory: An Analysis of Decision Under Risk. **Econometrica**, 47(2). doi:10.2307/1914185

KEYNES, J.M. (1936). The general theory of employment, interest and money.

LEE, J.M., E HANNA, S.D. (2015). Savings Goals and Saving Behavior from a Perspective of Maslow's Hierarchy of Needs. **Journal of Financial Counseling and Planning**, 26(2), 129-147. Londres: Macmillan

LUSARDI, A., & MITCHELL, L. (2007). Financial literacy and retirement preparedness: Evidence and implications for financial education. **Business Economics**, 42(1), 35-44.

MANLY, BRYAN F. J; ALBERTO, JORGE A. NAVARRO. Métodos estatísticos multivariados: uma introdução. Tradução Carlos Tadeu dos Santos Dias. – 4 ed. – Porto Alegre: **Bookman** 2019

MAROCO, J. Análise estatística com utilização do SPSS. 2. ed. Lisboa: Silabo, 2003.

MATOS, P.; LIMA FILHO, JOSÉ M. OLIVEIRA. O que determina o padrão de contribuição previdenciária em PGBL das famílias nos estados brasileiros?

MODIGLIANI, Franco; BRUMBERG, Richard. Utility analysis and the consumption function: An interpretation of cross-section data. **Franco Modigliani**, v. 1, n. 1, p. 388-436, 1954.

SALVATO, M. A.; GOMES, F. A. R.; DE SANT'ANNA, Pedro H. C. G., Os ricos poupam mais que os pobres no Brasil?

SHEFRIN, H.M., E STATMAN, M. (1985). The disposition to sell winners too early and ride losers too long: theory and evidence. **Journal of Finance**, 40(3), 777-90.

SHEFRIN, H.M., E THALER, R.H. (1988). The behavioral life-cycle hypothesis. **Economic Inquiry**, 26(4). doi:10.1111/j.1465-7295.1988.tb01520.x

SILVEIRA, M. A. C.; MOREIRA, A. R. B. Taxa de poupança e consumo no ciclo de vida das famílias brasileiras: evidência microeconômica. **Texto para discussão**, n. 1997. Brasília: Ipea, 2014.

SMITH, J. P. (1999). Inheritances and bequests. In J. P. Smith & R. J. Willis (Eds.), *Wealth, work, and health: Innovations in measurement in the social sciences: Essays in honor of F. Thomas Juster*. Ann Arbor: **University of Michigan**.

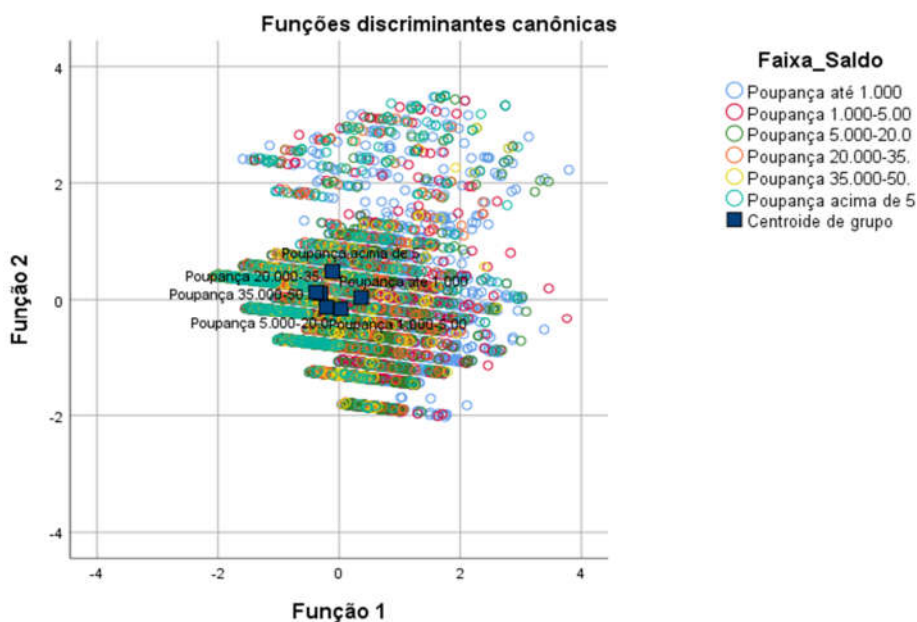
APÊNDICE

Tabela 14: Estatísticas dos Grupos por ano

Faixa Saldo		2019		2020	
		Média	Desvio Padrão	Média	Desvio Padrão
Poupança até 1.000	Gênero	0,52	0,500	0,53	0,499
	Faixa Etária	2,59	1,183	2,59	1,163
	Faixa Renda	2,61	1,683	2,60	1,671
	investimentos	0,10	0,306	0,11	0,310
Poupança 1.000-5.000	Gênero	0,52	0,500	0,51	0,500
	Faixa Etária	2,57	1,194	2,58	1,191
	Faixa Renda	2,09	1,478	2,15	1,491
	investimentos	0,06	0,230	0,07	0,248
Poupança 5.000-20.000	Gênero	0,51	0,500	0,49	0,500
	Faixa Etária	2,71	1,161	2,71	1,147
	Faixa Renda	1,83	1,381	1,81	1,311
	investimentos	0,05	0,223	0,05	0,212
Poupança 20.000-35.000	Gênero	0,53	0,500	0,52	0,500
	Faixa Etária	3,02	1,057	2,88	1,084
	Faixa Renda	1,97	1,475	1,88	1,433
	investimentos	0,07	0,257	0,06	0,229
Poupança 35.000-50.000	Gênero	0,54	0,499	0,55	0,498
	Faixa Etária	3,10	1,086	3,05	1,067
	Faixa Renda	2,00	1,498	1,96	1,495
	investimentos	0,06	0,228	0,06	0,245
Poupança acima de 50.000	Gênero	0,51	0,501	0,52	0,500
	Faixa Etária	3,14	1,006	3,08	1,016
	Faixa Renda	2,41	1,839	2,32	1,765
	investimentos	0,19	0,389	0,17	0,378
Total	Gênero	0,52	0,500	0,51	0,500
	Faixa Etária	2,73	1,166	2,72	1,151
	Faixa Renda	2,19	1,580	2,15	1,545
	investimentos	0,08	0,272	0,08	0,269

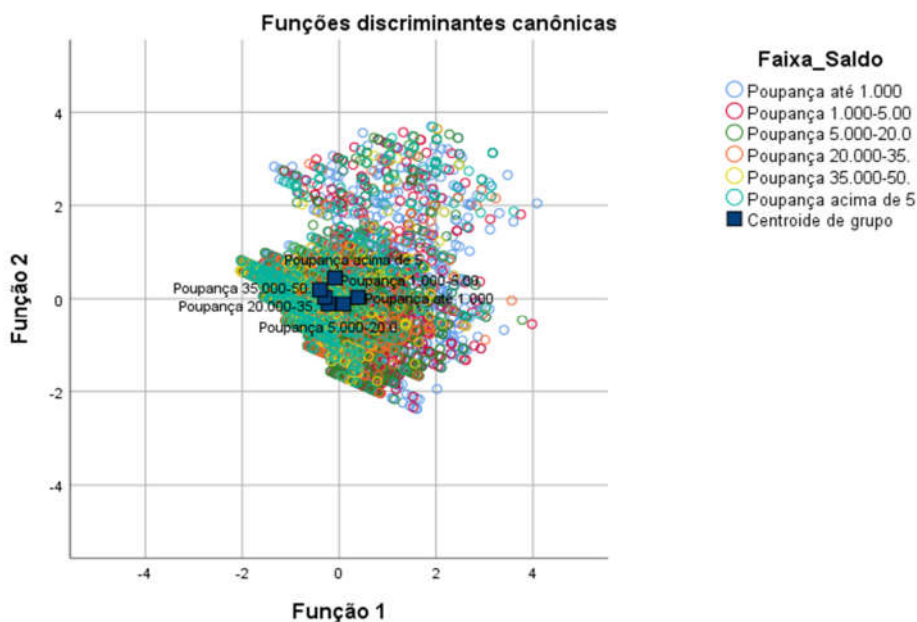
Fonte: Elaborado pelo autor

Figura 5: Função Discriminante Canônica para 2019



Fonte: Elaborado pelo autor

Figura 6: Função Discriminante Canônica para 2020



Fonte: Elaborado pelo autor