



UNIVERSIDADE CATÓLICA PORTUGUESA

O impacto dos determinantes socioeconómicos na fertilidade em Portugal, e as suas implicações políticas

Henrique Moreira

Universidade Católica Portuguesa, Católica Porto Business School
Abril, 2025



UNIVERSIDADE CATÓLICA PORTUGUESA

O impacto dos determinantes socioeconómicos na fertilidade em Portugal, e as suas implicações políticas

Trabalho Final na modalidade de Dissertação
apresentado à Universidade Católica Portuguesa
para obtenção do grau de mestre em Gestão

por

Henrique Moreira

sob orientação de
Ricardo Ribeiro

Universidade Católica Portuguesa, Católica Porto Business School
Abril, 2025

Agradecimentos

A entrega deste Plano Final de Mestrado, é possível apenas pelo apoio de algumas pessoas e instituições a quem gostaria de deixar o meu agradecimento.

À DGEEC e ao INE pelo acesso aos dados utilizados no desenvolvimento desta tese.

Aos meus pais e tios, quer por me concederem a oportunidade de frequentar este grau de ensino superior, quer pela motivação e incentivo que sempre me transmitiram para as diferentes etapas.

Aos meus amigos e à minha namorada pelo apoio incondicional.

Aos meus colegas de mestrado pela ajuda e simpatia sempre apresentada.

Ao Professor Ricardo Ribeiro, pela disponibilidade e pelo acompanhamento próximo, que permitiu a entrega desta dissertação.

Resumo

A tendência decrescente da fertilidade em Portugal nas últimas décadas apresenta consequências a nível social, demográfico e económico. O envelhecimento da população, as alterações no plano económico e a capacidade de produção e inovação, são potenciais consequências desse declínio.

Esta dissertação tem como objetivo analisar os determinantes socioeconómicos que impactam o número de filhos das famílias em Portugal. Para alcançar os resultados propostos, foi desenvolvido um modelo econométrico composto por variáveis das diversas áreas mencionadas, utilizando como base os dados recolhidos no Inquérito à Fecundidade, desenvolvido pelo INE em 2019.

Os resultados deste estudo revelam que fatores como a idade, o nível de escolaridade e a estabilidade no emprego influenciam significativamente o número de filhos. Também a inclusão de regiões na estimação do modelo permitiu retirar conclusões relativamente à sua distribuição geográfica.

Esta investigação propõe contribuir para o entendimento da Fertilidade em Portugal, associando-se à literatura nacional e internacional dedicada à compreensão deste tema, tendo como objetivo final a sugestão de políticas de atuação para a mitigação das suas consequências.

Palavras-chave: Número de filhos; Determinantes Socioeconómicos; Fertilidade; Famílias; Portugal.

Abstract

The downward trend in fertility in Portugal over recent decades presents social, demographic, and economic consequences. The aging of the population, changes in the economic landscape, and the capacity for production and innovation are potential consequences of this decline.

This dissertation aims to analyse the socioeconomic determinants that impact the number of children in Portuguese families. To achieve the proposed results, an econometric model was developed, comprising variables from the various aforementioned areas, based on data collected in the Fertility Survey conducted by the Portuguese National Statistics Institute (INE) in 2019.

The results of this study reveal that factors such as age, education level, and employment stability significantly influence the number of children. The inclusion of regional effects in the model estimation also allowed for conclusions to be drawn regarding the geographical distribution of fertility across Portugal.

This research purpose is to contribute to the understanding of fertility in Portugal, aligning with both national and international literature dedicated to this topic, with the ultimate goal of suggesting policy measures to mitigate its consequences.

Keywords: Number of Children, Socioeconomic Determinants, Fertility, Families, Portugal

Índice

Agradecimentos	iv
Resumo	v
Abstract	vi
Índice	viii
Índice de Figuras.....	xi
Índice de Tabelas	xiii
Introdução.....	15
 Capítulo 1 (Revisão de Literatura)	20
1.1 Determinantes Socioeconómicos	20
1.1.1 Educação	20
1.1.2 Participação Feminina no Mercado de Trabalho	22
1.1.3 Políticas Públicas de Apoio à Natalidade	22
1.1.4 Rendimento Familiar e Estabilidade no Emprego.....	23
1.1.5 Cultura	24
1.1.6 Urbanização.....	25
1.1.7 Idade.....	25
1.1.8 Desigualdades Regionais	26
1.2 Modelos da Literatura	26
 Capítulo 2 (Método)	29
 Capítulo 3 (Análise Empírica).....	31
3.1 Fonte dos Dados	31
3.2 Descrição das variáveis	31
3.3 Limpeza dos Dados	36
3.4 Estatística Descritiva.....	36
 Capítulo 4 (Resultados).....	38
4.1 Resultados	38
 Capítulo 5 (Conclusões).....	43
 Bibliografia.....	46

Anexos.....	49
-------------	----

Índice de Figuras

Figura 1: Taxa Bruta de Natalidade em Portugal (1950-2022)	15
--	----

Índice de Tabelas

Tabela 1: Código das variáveis no inquérito à natalidade do INE (2019)	35
Tabela 2: Estatística Descritiva.....	38
Tabela 3: Estimação de Modelos.....	38

Introdução

Atualmente, o declínio da natalidade é um fenómeno quase universal (Lewis-Kraus, 2024). A evolução demográfica nas últimas décadas mostra que Portugal não é exceção. A Figura 1, retirada do centro de estatística do Banco de Portugal, ilustra a trajetória da taxa bruta de natalidade em Portugal entre 1950 e 2022, expressa em permilagem (nascimentos por 1.000 habitantes).



Fonte: Banco de Portugal

Figura 1: Taxa Bruta de Natalidade em Portugal (1950-2022)

Após um período de relativa estabilidade durante a década de 1950 observa-se uma queda acentuada a partir da década de 1960. Esta tendência decrescente manteve-se ao longo do tempo, com reduções particularmente pronunciadas em meados dos anos 1970, e novamente na chegada aos anos 2000. Desde 2010, a taxa estabilizou em níveis historicamente baixos, refletindo um padrão demográfico marcado por famílias mais pequenas, maior planeamento familiar e o adiamento da maternidade.

Para além de “quase-universal”, o fenómeno do declínio da natalidade impacta diversas áreas e setores, quer ao nível da economia, como a nível social e da sustentabilidade das gerações futuras. Platt e Sterling (2024) mencionam que para além das preocupações económicas, fatores como o apoio à população mais envelhecida, a inovação e a produção, saem prejudicados com a redução da população mais jovem. Lewis-Kraus (2024) menciona ainda que este paradigma pode provocar um aumento dos encargos sociais, associado a uma redução da força de trabalho. Mason (1988) e Wang, Yip e Scotese (1994) oferecem perspetivas complementares, mas conceptualmente divergentes sobre o impacto da taxa de natalidade no desempenho económico. Mason (1988) argumenta que a redução da fertilidade pode favorecer o crescimento económico ao aumentar a taxa de poupança agregada. O autor menciona que a redução da fertilidade pode originar um “bónus demográfico”, uma vez que esta se traduz numa maior proporção de população ativa, permitindo ganhos de produtividade e maior acumulação de capital. Por outro lado, Wang, Yip e Scotese (1994) demonstram empiricamente que aumentos de fertilidade têm impacto direto sobre a produção, provocando um aumento do crescimento económico.

A tendência decrescente da fertilidade em Portugal pode ser explicada pelo processo de envelhecimento da população portuguesa e pela conjuntura económica e social de instabilidade e precaridade no emprego (Silva, 2021). Mendes et al (2016) reforça o argumento do impacto da idade na fertilidade, mencionando que a idade é uma variável central na explicação do comportamento da fecundidade.

Neste sentido, o presente estudo propõe analisar a forma como os fatores socioeconómicos influenciam o número de filhos por indivíduo em Portugal, recorrendo a uma abordagem microeconómica com base nos dados do Inquérito à Fecundidade realizado pelo INE em 2019. Através da criação e desenvolvimento de um modelo econométrico que contemple variáveis

consideradas importantes pela literatura, pretende-se identificar relações e efeitos que ajudem a compreender as dinâmicas reprodutivas no contexto português, bem como o seu peso nas mesmas.

Este estudo pretende associar-se à literatura nacional e internacional, através do debate em torno da problemática da natalidade, no contexto português. A sua contribuição é relevante na medida em que estuda variáveis das diferentes áreas da sociedade, retiradas de um inquérito com uma amostra significativa e atual, e que estima diferentes modelos permitindo considerar abordagens e conclusões distintas sobre o efeito no número de filhos. Para além do desenvolvimento e formalização da temática em si, também é feita a indicação de políticas e intervenções públicas, que visem mitigar a tendência decrescente, com base nos resultados do estudo.

Com base nos resultados obtidos nesta investigação, podemos retirar que fatores como o nível de escolaridade, o rendimento familiar, a estabilidade no emprego e a habitação, bem como o contexto cultural do inquirido apresentam uma influência significativa sobre o número de filhos. Estas conclusões procuram servir de base empírica para a formulação de políticas públicas direcionadas à promoção da natalidade em Portugal.

Esta investigação está estruturada em 4 principais capítulos. Na revisão de literatura são analisados artigos de diferentes épocas e geografias, de forma a perceber e avaliar as linhas gerais da literatura e os diferentes métodos utilizados, no âmbito da temática da natalidade. No Capítulo 2, é apresentada a equação proposta para estimação do impacto dos determinantes no número de filhos. No capítulo seguinte é descrita a fonte dos dados para este estudo, são descritas as variáveis e a sua formulação, é explicada a limpeza dos dados e ainda é apresentada a estatística descritiva das variáveis. Posteriormente, no Capítulo 4 são analisados os resultados dos modelos estimados e apresentadas limitações associadas ao mesmo. Por fim, no Capítulo 5, são apresentadas as conclusões

gerais do trabalho, juntamente com sugestões de políticas de incentivo à natalidade, pensadas com base nos resultados do modelo.

Capítulo 1

Revisão de Literatura

A presente secção pretende identificar as opiniões da literatura nacional e internacional, sobre a temática da natalidade e a sua relação com diferentes fatores socioeconómicos. O trabalho desenvolvido no momento da revisão de literatura permitiu apresentar o estudo proposto, tendo por base uma coleção de artigos nacionais, internacionais, com abordagens micro e macro, com argumentos complementares ou opostos, e que visam corroborar os argumentos teóricos com resultados empíricos.

1.1 Determinantes Socioeconómicos

1.1.1 Educação

A análise do fator é complexa e implica uma contemplação do contexto e das variáveis envolventes, uma vez que existem argumentos e estudos que demonstram ambos os efeitos.

Mendes e Rego (2006) destacam que a progressão no percurso educativo conduz, frequentemente, a uma entrada mais tardia no mercado de trabalho, o que, por sua vez, contribui para o adiamento da formação de família. Este adiamento está, geralmente, relacionado com uma maior valorização da estabilidade profissional e financeira. Contudo, a relação entre escolaridade e fecundidade não é uniforme. Testa (2014) argumenta que mulheres com níveis de educação mais elevados tendem a ter maior capacidade de planeamento de

vida reprodutiva e maior acesso a recursos económicos e sociais, o que lhes permite conciliar mais facilmente a vida profissional com o desejo de ter filhos.

Neste sentido, Da Silva (2021), num estudo focado no caso português, demonstra que, em certos contextos, o nível de escolaridade dos progenitores, pode influenciar positivamente a probabilidade de ter três ou mais filhos. Este efeito revela-se estatisticamente significativo tanto entre indivíduos com escolaridade até ao ensino básico completo como entre aqueles com formação superior. No sentido contrário, Mendes e Rego (2006) concluem que a educação influencia o adiamento da fecundidade, visível no aumento da idade média de maternidade, para os países do sul da Europa.

1.1.2 Participação Feminina no Mercado de Trabalho

A crescente participação das mulheres no mercado de trabalho, com particular enfoque para os últimos 60 anos com a mudança do paradigma da sociedade até então, tem sido vista como um dos fatores mais relevantes na explicação da queda da fecundidade nos países desenvolvidos. Mendes e Rego (2006) referem que a crescente inserção das mulheres no mercado de trabalho tem contribuído para a redução da natalidade, sobretudo em contextos onde as estruturas laborais e sociais não facilitam a conciliação entre emprego e vida familiar. Neste sentido, a dificuldade em compatibilizar a progressão profissional com a maternidade tem conduzido muitas mulheres a adiar ou mesmo a reavaliar a intenção de constituir família.

Da Silva (2021) e Salari (2020) sublinham que, em cenários onde o mercado de trabalho continua a penalizar implicitamente a maternidade, bem como a própria dificuldade de conciliação das duas vertentes, a fecundidade tende a ser negativamente afetada. Em particular, da Silva (2021) sublinha que o aumento

do rendimento da mulher tende a elevar o custo de oportunidade da maternidade, na medida em que o tempo dedicado aos filhos implica, frequentemente, uma interrupção na trajetória profissional, perda de rendimento e menor investimento no desenvolvimento de carreira

Do ponto de vista empírico, Mendes e Rego (2006) mostram que, em países do sul da Europa a crescente inserção feminina no emprego está frequentemente associada à redução da fecundidade, sobretudo quando não existem políticas eficazes de conciliação entre vida profissional e familiar.

1.1.3 Políticas Públicas de Apoio à Parentalidade

Em contextos os apoios à parentalidade são escassos ou insuficientes, os casais enfrentam maiores dificuldades em conciliar trabalho e família, o que pode levar ao adiamento ou redução da fecundidade (Mendes & Rego, 2006). Numa análise aos países do sul da Europa, Rydell (2002) refere que a falta ou ineficácia de políticas públicas de apoio existentes acabam por reforçar a dependência das redes familiares tradicionais, desincentivando, ou, pelo menos, não incentivando, a criação de novas famílias.

Adicionalmente, na literatura internacional, Chen (2011) e Hong (2016) defendem que incentivos fiscais bem estruturados podem ajudar a aumentar a fecundidade ao aliviar o peso financeiro associado a ter filhos.

Nos resultados do seu estudo, Rydell (2002) refere que, apesar de Portugal ter um volume de apoios públicos à infância superior ao de países como Itália ou Espanha, o seu impacto continua a ser limitado. As taxas de fecundidade persistentemente baixas refletem não só a vulnerabilidade económica das famílias, mas também a ineficácia das políticas públicas em garantir condições para exercer a parentalidade.

Chen (2011) demonstra que, após a introdução do sistema de quociente familiar em França, os agregados com maiores rendimentos, e, portanto, maior benefício fiscal, registaram um aumento estatisticamente significativo na probabilidade de ter um segundo ou terceiro filho, sugerindo uma potencial sensibilidade da fecundidade a incentivos económicos.

Por fim, Hong et al. (2016) ao analisarem dados de 225 distritos da Coreia do Sul entre 2005 e 2010, constataram que um aumento de 1.000 dólares nos subsídios à natalidade estava associado a um crescimento de 4,4% na taxa bruta de natalidade. O impacto verificado foi mais forte no nascimento do primeiro e segundo filhos, e continuou relevante mesmo entre mulheres com mais de 35 anos.

1.1.4 Rendimento Familiar e Estabilidade no Emprego

Tal como a Educação, o efeito do Rendimento Familiar na decisão de ter filhos não é unívoco, e está sujeito a diferentes contextos e perceções.

Silva (2021) e McDonald (2000) defendem que um aumento no rendimento tende a ter um efeito positivo sobre a fecundidade, uma vez que proporciona aos casais maior capacidade para suportar os custos diretos e indiretos associados à parentalidade.

Contudo, na ótica da teoria do custo de oportunidade desenvolvida por Becker (1991), é possível compreender a existência de uma relação negativa entre rendimento e número de filhos. Esta teoria sugere que, à medida que o rendimento aumenta, os custos indiretos associados à parentalidade, como a perda de rendimento futuro ou interrupções na carreira, tornam-se mais significativos. Assim, casais com rendimentos mais elevados podem ter maior propensão a adiar ou a optar por não ter filhos, precisamente por valorizarem a

estabilidade financeira e o estilo de vida que a parentalidade poderia comprometer.

Da Silva (2021) reforça esta ambiguidade ao demonstrar, com base em dados do contexto português, que rendimentos situados entre 1300€ e 2200€ valores aumentam a probabilidade de ter três ou mais filhos. Rendimentos mais baixos estão geralmente associados a uma maior probabilidade de não ter filhos, sugerindo que a limitação económica constitui um obstáculo relevante à fecundidade. Contudo, o mesmo estudo também revela que rendimentos excessivamente elevados podem, simultaneamente, ter um efeito negativo na decisão de ter filhos, refletindo uma maior aversão ao risco financeiro.

1.1.5 Cultura

Segundo Morales (2021), os pais transmitem aos filhos crenças, normas e preferências que fazem parte do seu legado cultural, influenciando assim o comportamento reprodutivo da geração seguinte. A autora utiliza o “número de filhos dos pais” como *proxy* da cultura para capturar o efeito do contexto e da influência intergeracional nas decisões reprodutivas.

Os resultados da autora mencionada, indicam que, o aumento de uma unidade no número médio de filhos dos pais está associado a um acréscimo de aproximadamente 0.004 filhos nos filhos. O impacto desta variável é mais expressivo entre os indivíduos com baixos níveis de escolaridade, para os quais o aumento no número médio de filhos dos pais está associado a um aumento de cerca de 0.009 filhos, sendo a probabilidade de ter uma família numerosa superior em cerca de dez pontos percentuais.

1.1.6 Urbanização

Segundo Zhang e Li (2017), a urbanização transforma os contextos institucionais e de mercado nos quais as decisões de fertilidade são tomadas. O argumento central dos autores passa pela teoria do custo de oportunidade, mencionando que os custos diretos e indiretos de ter filhos são superiores em contexto urbano.

Após uma análise empírica Zhang, e Li (2017) mostram que zonas não urbanas passaram a apresentar taxas de fertilidade superiores às zonas urbanas, validando a hipótese teórica desenvolvida, em particular, para o contexto do sul da Europa.

1.1.7 Idade

Morales (2021) argumenta que indivíduos com mais idade, têm uma janela temporal mais longa para ter filhos, independentemente de outros fatores contextuais, contribuindo positivamente para a natalidade,

A análise empírica presente no estudo da autora, demonstra que a idade exerce um efeito positivo sobre a fecundidade até aos 44 anos, sugerindo que o número de filhos está positivamente correlacionado com o avanço da idade, pelo menos até ao final do ciclo fértil.

Esta conclusão complementa a anterior, reforçando a ideia de que a fecundidade deve ser compreendida no tempo e no espaço, complementarmente aos determinantes socioeconómicos a si inerentemente associados.

1.1.8 Desigualdades Regionais

A fecundidade não se distribui de forma homogênea, apresentando variações que não podem ser inteiramente explicadas por fatores socioeconómicos. Estas diferenças regionais evidenciam a relevância da dimensão geográfica na análise do comportamento reprodutivo, revelando padrões que sugerem a existência de dinâmicas reprodutivas específicas de cada território (Mendes & Rego, 2006). Mendes e Rego (2006) mencionam ainda que, no caso português, bem como noutros países do sul da Europa, a fecundidade caracteriza-se por uma expressiva heterogeneidade regional revelando que o território funciona como uma unidade com dinâmicas reprodutivas próprias

A nível empírico, Mendes e Rego (2006) demonstram que, mesmo entre países com níveis globais baixos de fecundidade, como os do sul da Europa, existem regiões com padrões significativamente distintos, nomeadamente no que se refere à idade média da maternidade e à distribuição da fecundidade ao longo do ciclo fértil. Castro (2007), por sua vez, mostra que algumas regiões apresentam valores fora da tendência europeia. Estas conclusões confirmam que o território tem um contexto e cultura próprios a ele associado, que também podem pesar no momento do estudo da natalidade dos países.

1.2 Modelos da Literatura

Na secção da revisão de literatura, foram analisados artigos que trabalhavam a natalidade e os seus fatores socioeconómicos, recorrendo a diferentes variáveis explicadas, quer a nível micro, quer a nível macro. Fernandez-Crehuet et al. (2020) e Greulich (2020) utilizam a taxa de fecundidade total como variável

dependente, captando o número médio de filhos por mulher ao longo do seu ciclo reprodutivo. Testa (2014) utiliza intenções de fecundidade como variável explicada, procurando captar preferências reprodutivas. A nível micro, Morales (2021), utiliza a variável dependente número de filhos, para procurar o impacto das variáveis no momento da decisão ao nível das famílias.

A literatura empírica tem recorrido a diferentes métodos de estimação, de acordo com a natureza dos dados e os objetivos do contexto da própria investigação.

Morales (2021) utiliza três principais métodos para a estimação do número de filhos no caso espanhol: um modelo OLS, um modelo por efeitos fixos de região e um modelo Tobit. O modelo OLS assume uma distribuição contínua e não censurada do número de filhos. No entanto, dada a elevada proporção de observações com zero filhos na amostra, a autora opta também por estimar um modelo Tobit, de forma a captar simultaneamente a decisão extensiva de ter ou não ter filhos, bem como a quantificação do número de filhos. O modelo de efeitos fixos por região atribui uma visão complementar, ao isolar efeitos inerentes ao contexto do espaço

No estudo de Salari (2020) são utilizados modelos OLS, Probit e modelos de efeitos fixos por região. O modelo OLS é aplicado para estimar o efeito das variáveis culturais sobre o número de filhos. Paralelamente, é estimado um modelo Probit para captar a probabilidade de participação no mercado de trabalho, tratando-se de uma variável binária. A inclusão do modelo de efeitos fixos ao nível das regiões metropolitanas dos Estados Unidos, permite isolar o impacto das variáveis culturais e socioeconómicas dos fatores estruturais locais.

Zhang e Li (2017) no seu estudo sobre normas culturais de fertilidade e o estatuto socioeconómico das mulheres na China, adotam uma estratégia econométrica composta por modelos OLS e de efeitos fixos ao nível do condado.

Por fim, a dissertação de Silva (2021), uma análise da natalidade também no contexto de Portugal, recorre maioritariamente a modelos de regressão linear múltipla (OLS), aplicados a dados agregados por concelho ou região.

Capítulo 2

Método

O presente capítulo propõe explicar o método utilizado para estudar o impacto das variáveis identificadas na natalidade em Portugal. Este adota uma abordagem e quantitativa de base microeconómica com o objetivo de analisar os principais fatores socioeconómicos que influenciam o número de filhos por indivíduo em Portugal.

A especificação base do modelo é a seguinte:

$$NFILHOS_{ir} = \beta_0 + \beta_1 EDU_{ir} + \beta_2 EMP_{ir} + \beta_3 REN_{ir} + \beta_4 CULT_i + \beta_5 URB_{ir} + \beta_6 IDADE_{ir} + \delta_r + \varepsilon_{ir},$$

a variável dependente $NFILHOS_{ir}$ representa o de filhos do indivíduo i na região r (Morales, 2021). EDU_{ir} corresponde ao nível de escolaridade do indivíduo i na região r (Mendes & Rego, 2006). Por sua vez, a variável EMP_{ir} capta a presença, ou não, do indivíduo i no mercado de trabalho, na região r (Mendes & Rego, 2006). REN_{ir} representa o rendimento do agregado do indivíduo i na região r (Da Silva, 2021). $CULT_i$ representa a dimensão cultural, aproximada pelo número de irmãos do indivíduo i (Morales, 2021). A variável URB_{ir} responde à questão da urbanização, indicando se o local de residência do indivíduo é numa zona urbana ou não (Zhang & Li, 2017). AGE_{ir} expressa a idade do indivíduo i em anos completos, na região r (Morales, 2021). Por fim, a variável δ_r representa a região de residência do indivíduo. O termo ε_{ir} representa o erro aleatório, captando os efeitos não observados que podem afetar o número de filhos.

Com base nas variáveis e nos modelos revistos na revisão de literatura, foram especificados quatro modelos econométricos com diferentes níveis de complexidade. Nos seus estudos, Salari (2020) e Zheng (2017), optam por estimações quer através de modelos OLS, quer de modelos fixos por região, com diferenças ao nível das variáveis que os compõe e do seu formato. A equação econométrica desenvolvida seguirá os exemplos mencionados da literatura, estimando através de modelos OLS e por efeitos fixos de região.

A especificação do modelo base inclui variáveis socioeconómicas como idade, nível de escolaridade, rendimento, emprego, número de irmãos e nível de urbanização da residência.

A estimação desenvolvida contempla ainda a inclusão de efeitos fixos regionais. Estes são operacionalizados através da introdução de variáveis. Esta abordagem procura isolar o efeito específico de cada região na determinação da fecundidade.

Capítulo 3

Análise Empírica

3.1 Fonte dos Dados

O presente estudo é desenvolvido tendo utilizado como base de dados um inquérito à fecundidade, realizado em 2019, conduzido pelo Instituto Nacional de Estatística (INE).

A população do inquérito utiliza como referência a idade fértil, contemplando mulheres com idades compreendidas entre os 18 e os 49 anos, e homens com idades compreendidas entre os 18 e os 54. O inquérito inclui informações ao nível individual, representando uma visão micro relativamente ao número de filhos, e contemplando diversos universos do contexto pessoal onde este se insere, desde a situação conjugal, condições económicas, estabilidade no trabalho, questionários de importância e preferência, entre outros.

A amostra inicial contempla cerca de 7000 inquiridos, sendo 2531 do sexo masculino e 5178 do feminino, distribuídos por todo o território nacional, incluindo Portugal Continental e as Regiões Autónomas dos Açores e da Madeira

3.2 Descrição das variáveis

Nesta secção são descritas a variável dependente e as variáveis independentes, presentes na equação em análise. Para melhor interpretação e compreensão das mesmas, é necessário ter em conta que parte significativa das variáveis utilizadas

na modelação econométrica deriva de perguntas formuladas exclusivamente a indivíduos que já tinham filhos, como por exemplo as variáveis que questionam os motivos para decisão do primeiro filho, ou sobre a situação profissional do inquirido no momento do nascimento do primeiro filho. Nos casos em que o inquirido não tem filhos, essas questões não lhe foram colocadas, resultando, naturalmente, em valores omissos nessas variáveis.

A variável dependente do modelo, será representada através do número total de filhos biológicos reportado por cada indivíduo inquirido i .

No inquérito do INE, no contexto da educação, o inquirido pode seleccionar as opções “1- Nenhum e básico”, “2 - Secundário e pós-secundário”, “3 – Superior” e “9 - Não sabe”. Para melhor compreensão deste fenómeno, foram criadas duas variáveis, que representam, respetivamente, a conclusão do ensino secundário ou pós-secundário e do ensino superior (ESC_SEC , ESC_SUP). Ambas são variáveis do tipo *dummy*, assumindo o valor 1 se o indivíduo possui o nível de escolaridade correspondente e 0 caso contrário.

Para analisar o fator emprego foram consideradas as respostas à questão do inquérito do INE sobre a condição perante o trabalho à data do nascimento do 1.º, 2º e 3º filho biológico. O inquirido tinha opção de seleccionar uma das seguintes opções: “1 - Tem um emprego ou trabalho”; “2 - Está desempregado(a)”; “3 - É estudante”; “4 - Está reformado(a) do trabalho”; “5 - É incapacitado(a) permanente (impossibilidade permanente para o trabalho)”; “6 - É doméstico(a)”; “7 - tem outra situação de inatividade”; “8 - Tem outra reforma/pensão”. Foi construída uma variável com a proporção de nascimentos (até ao terceiro filho) em que o inquirido se encontrava empregado(a). Para tal, foi criada uma *dummy* que assume o valor 1 para o caso de estar empregado no momento em questão, e valor 0 na seleção de qualquer outra opção. Posteriormente foram somados os momentos em que o inquirido estava efetivamente empregado no momento do nascimento dos filhos, e de seguida

divididos pelo número de filhos efetivamente tidos. Para contornar a problemática dos valores omissos associados a esta variável, para os casos cujo número de filhos é igual a zero, foi atribuído o valor zero a todas as variáveis que reportam a presença de motivos para a parentalidade.

No inquérito do INE, no âmbito da análise do rendimento, numa das perguntas apresentadas, os inquiridos deveriam responder qual o intervalo de rendimento médio auferido, entre os seguintes escalões de resposta: “1 - até 500 euros”, “2 - 501 a 700 euros”, “3 - 701 a 900 euros”, “4 - 901 a 1 100 euros”, “5 - 1 101 a 1 300 euros”, “6 - 1 301 a 1 500 euros”, “7 - 1 501 a 1 800 euros”, “8 - 1 801 a 2 200 euros”, “9 - 2 201 a 2 900 euros”, “10 - 2 901 euros ou mais”. Para tratar a variável rendimento, foi desenvolvida uma abordagem contínua e uma abordagem categórica. Na primeira, foram atribuídos valores médios representativos a cada escalão, permitindo uma aproximação contínua desta variável, com exceção dos escalões 1 e 10, dado que se tratam de intervalos abertos. Foram criadas duas variáveis *dummy* para esses escalões, *ESC_REND_1* e *ESC_REND_10*, que assumem o valor 1 caso o inquirido tenha um rendimento pertencente ao respetivo escalão, e 0 caso o rendimento não se encontre no escalão em questão. Desta forma, a variável contínua contempla apenas valores do segundo ao nono escalão, atribuindo o valor 0 aos extremos. A segunda, foi tratada como variável categórica, com a criação de *dummies* para cada escalão. As variáveis foram ainda divididas por mil para uniformização de resultados

Foram desenvolvidas ainda três variáveis *dummy*, que traduzem motivações específicas associadas à decisão de ter o primeiro filho. A variável da condição de habitação capta a importância atribuída às condições de habitação do indivíduo *i*, a variável que estuda a conciliação entre a vida pessoal e profissional procura descrever a importância desta vertente para o indivíduo *i* no momento da decisão de ter filhos, e a variável de emprego estável indica a importância conferida à estabilidade no emprego indivíduo *i* ou do cônjuge. Estas variáveis

foram construídas com base em questões de opinião incluídas no inquérito com as seguintes respostas possíveis, “1 - Muito importante”, “2 – Importante”, “3 - pouco importante”, ou “4 - Nada importante”. As variáveis assumem o valor 1 quando o inquirido considera a razão “muito importante” na decisão de ter o primeiro filho, e 0 nos restantes casos. À semelhança da abordagem levada a cabo no estudo da variável que representa o emprego, também foi implementada uma lógica de proporção de nascimentos até ao 3º filho, onde são somas os valores 1, que correspondem à opinião “Muito importante”, e divididos pelo número efetivo de filhos. Foi também atribuído o valor 0 a estas variáveis, nos casos cujo número de filhos do inquirido é zero.

No inquérito do INE foi também respondida uma pergunta relativa ao número de irmãos do inquirido, com as seguintes opções de resposta “0 - 0 irmãos”; “1 - 1 irmão”; “2 - 2 irmãos”; “3 - 3 irmãos”; “4 - 4 irmãos”; “5 - 5 ou + irmãos”. Neste estudo, este é utilizado como *proxy* cultural, e é também dividida em diferentes abordagens. A variável corresponde ao número de irmãos e é tratada de forma contínua para os valores de 0 a 4, complementarmente, foi desenvolvida uma variável *dummy* que identifica os inquiridos com 5 ou mais irmãos (CULT_5M), dado que se trata de um intervalo aberto. No âmbito desta análise foi ainda criado um conjunto de *dummies* para os diferentes escalões de número de irmãos, sendo 0 irmãos a referência.

Ao nível da urbanização, foi desenvolvida uma variável *dummy* que assume o valor 1 se o local de residência for urbano, e 0 caso o local de residência seja um local não urbano.

A variável da idade é medida em anos completos e capta o efeito do ciclo de vida sobre as decisões reprodutivas.

Por fim, a região é estudada a partir de uma variável categórica e corresponde à região de residência do indivíduo i , de acordo com o critério de distribuição

geográfica NUTS II: Norte, Algarve, Centro, Área Metropolitana de Lisboa, Alentejo, Região Autónoma dos Açores, Região Autónoma da Madeira.

A tabela seguinte, associa as variáveis presentes no modelo econométrico, ao código que identifica a questão presente no inquérito à fecundidade do INE, e utilizada como base de recolha de informação para a análise da respetiva variável.

VARIÁVEL	CÓDIGO DE INQUÉRITO
Número de filhos	Nfilhosnv
Educação	NivelEscola_rec
Emprego	D0420_1_FBO; D0420_3_FBO; D0420_3_FBO
Rendimento	ESC_REND
Condições de habitação	D0444_1_FBO; D0444_2_FBO; D0444_3_FBO
Conciliação vida profissional	D0443_1_FBO; D0443_2_FBO; D0443_3_FBO
Estabilidade no emprego	D0442_1_FBO; D0442_2_FBO; D0442_3_FBO
Cultura	B0100_Cod
Urbanização	tipau1_rec
Idade	Idadeindiv
Região	N2Cod

Tabela 1: Código das variáveis no inquérito à natalidade do INE (2019)

3.3 Limpeza dos Dados

Para melhor análise e interpretação dos resultados, foi necessário proceder a uma limpeza dos dados, que não só excluísse os valores omissos.

Esta abordagem visou garantir que apenas os casos efetivamente incompletos fossem excluídos da análise, minimizando a perda de observações e assegurando a representatividade da amostra.

Após a aplicação dos procedimentos de limpeza dos dados, a amostra foi reduzida de 7709 para 6948 observações, na maioria das variáveis, o que corresponde a uma perda de 761 observações, associadas a valores inconsistentes, inválidos ou ausentes.

3.4 Estatística Descritiva

A presente secção tem como objetivo analisar, do ponto de vista estatístico, as variáveis incluídas no modelo econométrico de estimação da fecundidade. A estatística descritiva foi gerada após, e com base no resultante do processo de limpeza e transformação dos dados.

De forma a não duplicar a análise e a manter a interpretação da estatística descritiva simples e direta, foram apenas incluídas as abordagens contínuas das variáveis associadas ao Rendimento (*REN*, *ESC_REN_1*, *ESC_REN_10*) e Cultura (*CULT_CONT* e *CULT_5M*).

A Tabela 2 apresenta a estatística descritiva das variáveis utilizadas na estimação econométrica, e inclui informações sobre a média, mediana, desvio-padrão, valor mínimo e valor máximo.

Variável	Média	Mediana	Desvio Padrão	Min	Max
NFILHOS	0,997	1	1,01	0	7
EDU_SEC	0,344	0	0,475	0	1
EDU_SUP	0,315	0	0,464	0	1
EMP	0,281	0,333	0,306	0	1
REN_ESC_1	0,026	0	0,160	0	1
REN	1,149	1,200	0,770	0	2,550
REN_ESC_10	0,083	0	0,277	0	1
COND_HAB	0,091	0	0,199	0	1
CONCIL	0,091	0	0,197	0	1
EMP_ESTAVEL	0,105	0	0,209	0	1
CULT_CONT	1,33	1	1,11	0	4
CULT_5M	0,117	0	0,321	0	1
URB	0,269	0	0,443	0	1
IDADE	37,5	40	10,1	18	54

Nota: Tabela construída com base em 6948 observações

Tabela 2: Estatística Descritiva

Através da análise da variável dependente *NFILHOS* podemos concluir que o inquirido mediano *i* tem 1 filho. Das variáveis *EDU_SEC* e *EDU_SUP*, retiramos que este, com base nas respostas do inquérito, tem o ensino básico, não tem ensino, ou que não sabe. O indivíduo mediano *i* estava empregado no momento do nascimento de 33,3% dos filhos, e tinha um rendimento de 1200 euros, não pertencendo a nenhum dos extremos. Ao analisar as variáveis de emprego *COND_HAB*, *CONCIL*, *EMP_ESTAVEL*, podemos concluir que estas variáveis não têm muita importância na decisão de ter mais filhos, para os inquiridos já com filhos. Conclui-se ainda que inquirido mediano *i* tem 1 irmão, não vive em zonas urbanas e tem 40 anos.

Capítulo 4

Resultados

A presente secção tem como propósito dar a conhecer os resultados estimados, enquadrá-los dentro dos modelos desenvolvidos e analisar globalmente os valores estimados, de forma a retirar conclusões sobre o impacto dos determinantes socioeconómicos na variação do número de filhos.

Variável	(1)	(2)	(3)	(4)
EDU_SEC	-0,154*** (0,027)	- 0,149*** (0,027)	-0,094*** (0, 015)	-0,092*** (0,015)
EDU_SUP	-0,258*** (0,029)	- 0,259*** (0,029)	-0,115*** (0,017)	-0,115*** (0,017)
EMP	-	-	0,217*** (0,024)	0,615*** (0,023)
REN_ESC_1	-0,008 (0,070)	0,001 (0,070)	0,037 (0,040)	-
REN	-0,143*** (0,020)	-0,134*** (0,019)	-0,047*** (0,011)	-
REN_ESC_2	-	-	-	-0,003 (0,046)
REN_ESC_3	-	-	-	-0,038 (0,041)
REN_ESC_4	-	-	-	-0,114* (0,046)
REN_ESC_5	-	-	-	-0,137*** (0,040)
REN_ESC_6	-	-	-	-0,099* (0,046)
REN_ESC_7	-	-	-	-0,148*** (0,040)
REN_ESC_8	-	-	-	-0,125** (0,040)

REN_ESC_9	-	-	-	-0,115** (0,042)
REN_ESC_10	0,354*** (0,050)	0,383*** (0,050)	-	-0,107** (0,043)
COND_HAB	-	-	0,157*** (0,047)	-0,139*** (0,036)
CONCIL	-	-	-0,013 (0,049)	-0,013 (0,049)
EMP_ESTAVEL	-	-	0,207*** (0,045)	-0,206*** (0,045)
CULT_CONT	0,060*** (0,011)	0,058*** (0,011)	0,065*** (0,015)	-
CULT_5M	0,404*** (0,038)	0,404*** (0,039)	0,203*** (0,023)	-
CULT_FATOR1	-	-	-	0,002 (0,019)
CULT_FATOR2	-	-	-	0,054* (0,022)
CULT_FATOR3	-	-	-	0,069** (0,025)
CULT_FATOR4	-	-	-	0,054** (0,020)
CULT_FATOR5	-	-	-	0,181*** (0,026)
URB	0,061** (0,023)	0,060** (0,023)	0,012 (0,014)	0,013 (0,014)
IDADE	0,048*** (0,001)	0,006*** (0,001)	0,007*** (0,001)	0,006*** (0,001)
REGIAOAlgarve	-	0,287*** (0,022)	0,075*** (0,021)	0,083*** (0,022)
REGIAOCentro	-	0,029 (0,020)	0,022 (0,019)	0,030 (0,020)
REGIAOAMLisboa	-	0,035 (0,019)	0,015 (0,013)	0,033 (0,019)
REGIAOAlentejo	-	0,053* (0,023)	0,054* (0,013)	0,052* (0,023)
REGIAOAAAÇORES	-	0,157** (0,020)	0,159*** (0,020)	0,168*** (0,022)
R²	0,285	0,288	0,771	0,772
Ajusted R²	0,284	0,287	0,771	0,771
N	6948	6948	6948	6948

NOTA: Os valores indicados em parênteses respeitam aos erros-padrão. *** denota p-values<0,01; ** denota p-values<0,05; * denota p-values<0,1. N corresponde ao número de observações.

Tabela 3: Resultados de Estimação

O Modelo 1 representa a especificação mais simples da análise desenvolvida. Este modelo não incorpora efeitos fixos de região, nem variáveis construídas com base no número de filhos.

À semelhança da informação retirada na revisão de literatura, no artigo de Mendes e Rego (2006), a educação, tanto ao nível do ensino secundário do superior, tem um efeito negativo na variável dependente, contribuindo para um menor número de filhos.

REN, rendimento mensal médio, está positivamente correlacionado com a fecundidade, sinalizando que maiores rendimentos poderão influenciar positivamente a decisão de ter filhos. O mesmo acontece com a variável ESC_REN_10, indicando que acontece o mesmo para escalões de rendimento acima de 3000. Por sua vez, ESC_REN_1, tem um efeito negativo, porém não estatisticamente significativo.

As variáveis *CULT_CONT* e *CULT_5M* surgem neste modelo com um coeficiente positivo e estatisticamente significativo, refletindo a hipótese de que experiências familiares prévias contribuem para a fertilidade, tal como identificado na Revisão de Literatura, no estudo desenvolvido por Morales (2021).

A variável URB revela um impacto positivo na variável dependente, contrastando com as conclusões da Revisão de Literatura, segundo as quais viver em zonas urbanas está associado a um maior custo de oportunidade, o que tende a reduzir a natalidade.

Os resultados indicam que a idade do inquirido (IDADE) apresenta um efeito positivo e estatisticamente significativo sobre a fecundidade.

O Modelo 2 amplia a especificação anterior através da introdução de efeitos fixos de região.

A introdução dos efeitos fixos permite observar que as diferenças entre os valores das variáveis de região e a sua significância estatística, demonstram a existência de diferenças significativas da fertilidade por região.

As restantes variáveis que transitaram do Modelo 1 mantém, de forma geral, os seus efeitos na variável dependente, com apenas ligeiras alterações ao nível dos valores.

Ao modelo 2, o Modelo 3 adiciona a análise das variáveis que estudam a motivação para ter filhos no modelo.

A variável *EMP*, traduz a situação de trabalho do inquirido no nascimento dos filhos, revela um efeito positivo no número de filhos. Mais uma vez, para as variáveis cuja resposta é apenas possível para quem tem filhos, não é possível tirar nenhuma conclusão significativa exclusivamente a partir da sua análise.

Nesta especificação, as variáveis *COND_HAB* e *EMP_ESTAVEL*, mostram efeitos positivos e estatisticamente significativos, sugerindo que condições habitacionais adequadas e estabilidade laboral do casal são fatores que podem influenciar positivamente a decisão de ter filhos, para casais que já possuem filhos. Tal como a variável associada ao emprego, estas apenas são respondidas por inquiridos com filhos, pelo que não se deve tirar nenhuma conclusão imediata.

A variável *CONCIL*, que mede a importância atribuída à conciliação entre vida profissional e familiar, não se revela significativa.

As variáveis regionais mantêm-se idênticas em termos de sinal e significância.

O Modelo 4 representa a especificação mais completa da análise, adicionando uma abordagem por escalão para as variáveis do rendimento (*ESC_REND*) e do número de irmãos (*CULT_FATOR*).

A análise dos escalões de rendimento revela que, comparativamente com o primeiro escalão, diversos grupos com rendimento superior apresentam um impacto negativo, sendo estes estatisticamente significativos, sobretudo nos escalões intermédios. Este padrão sugere que, para alguns segmentos da população, o aumento de rendimento pode associar-se a um menor número de filhos.

Relativamente à variável cultural, observa-se que o número de irmãos tem um efeito positivo sobre a fecundidade do inquirido, com especial impacto a partir de dois irmãos.

As restantes variáveis mantêm valores coerentes com os modelos anteriores, com ligeiras alterações a nível quantitativo da significância.

A presente investigação utiliza um modelo de regressão linear (OLS) para estimar o efeito de diversas variáveis explicativas sobre o número de filhos dos inquiridos. Nesta análise, é possível observar uma proporção significativa de inquiridos com zero filhos, o que pode enviesar resultados e levar a falsas conclusões. Nestes casos, modelos utilizados em estudos presentes na revisão de literatura, como o modelo *Tobit* ou modelos de zeros inflacionados poderiam constituir alternativas mais robustas. A opção por utilizar o modelo OLS decorre, nesta fase, de uma estratégia de simplicidade, num âmbito de primeira análise exploratória.

Capítulo 5

Conclusões

Com base nos modelos estimados e respetivos resultados, é possível retirar conclusões, ainda que gerais, sobre o impacto de determinantes socioeconómicos na fertilidade, em Portugal

O propósito final deste estudo prende-se, precisamente, pela atuação perante esses resultados. Ao identificar o efeito e a significância dos diversos fatores no número de filhos, a um nível micro, é possível iniciar a discussão sobre medidas de atuação ao nível das políticas públicas.

Conseguimos concluir que maiores níveis de escolaridade estão associados a um menor número de filhos, possivelmente decorrente do adiamento da parentalidade. O emprego, a estabilidade no mesmo e as condições de habitação, têm um impacto positivo e significativo na fertilidade.

Para responder e mitigar o efeito negativo da escolaridade, apesar de se tratar de uma decisão do próprio(a), pode ser sugerida uma melhoria e desenvolvimentos dos serviços públicos de apoio à parentalidade, como creches ou programas de conciliação.

Em relação ao emprego e à sua estabilidade, naturalmente a diminuição do desemprego poderia ter um efeito imediatamente positivo, sobretudo quando associada a uma melhoria dos contratos de trabalho e de benefícios fiscais destinados à parentalidade.

Relativamente às condições de habitação, e, tendo em conta o contexto português neste momento, o investimento em medidas de habitação, como a diminuição do juro associado ao crédito da habitação, medidas de habitação

acessível com incentivos fiscais para jovens e melhoria das redes de transportes, poderiam revelar um efeito positivo na fertilidade ao longo do tempo.

Declaração de IA generativa e tecnologias assistidas por IA no processo de redação

Durante a elaboração do meu trabalho escrito/dissertação, “O impacto dos determinantes socioeconómicos na fertilidade em Portugal, e as suas implicações políticas”, foram utilizadas as ferramentas de Inteligência Artificial ChatGPT e Elicit para as tarefas de Revisão de Literatura, Programação e Análise de Dados tendo sido utilizadas as prompts listadas no final do documento na secção Lista de Prompts. Após a utilização destas ferramentas/serviços, revi e editei o conteúdo conforme necessário e assumo total responsabilidade pelo conteúdo do trabalho apresentado.

Declaro ainda conhecer e respeitar o Código de Conduta de Inteligência Artificial da Católica Porto Business School.

Bibliografia

Becker, G. S. (1960). An economic analysis of fertility. In Demographic and economic change in developed countries (pp. 209-240). Columbia University Press.

Becker, G. S. (1991). A treatise on the family: Enlarged edition. Harvard university press.

Chen, D. L. (2011). Can countries reverse fertility decline? Evidence from France's marriage and baby bonuses, 1929–1981. *International Tax and Public Finance*, 18, 253-272.

da Silva, T. E. F. M. (2021). Incentivos económicos como fatores de aumento na natalidade em Portugal (Master's thesis, ISCTE-Instituto Universitario de Lisboa (Portugal)).

Fernandez-Crehuet, J. M., Gil-Alana, L. A., & Barco, C. M. (2020). Unemployment and fertility: A long run relationship. *Social Indicators Research*, 152, 1177-1196.

Hong, S. C., Kim, Y. I., Lim, J. Y., & Yeo, M. Y. (2016). Pro-natalist cash grants and fertility: A panel analysis. *The Korean Economic Review*, 32(2), 331-354.

Jensen, F. P. (2020). *Does Length Matter for Fertility? A Multilevel Analysis of the Association between Parental Leave Length and Women's Completed Fertility in 10 OECD Countries* (Master's thesis).

Kohler, H. P., Billari, F. C., & Ortega, J. A. (2002). The emergence of lowest - low fertility in Europe during the 1990s. *Population and development review*, 28(4), 641-680.

Lewis-Kraus, G. (2025, March 3). The population implosion. *The New Yorker*. <https://www.newyorker.com/magazine/2025/03/03/the-population-implosion>

Mason, A. (1988). Saving, economic growth, and demographic change. *Population and Development Review*, 113-144.

McDonald, P. (2000). The “toolbox” of public policies to impact on fertility—a global view

Mendes, M. F., & Rego, C. (2006). Baixa fecundidade nos países do Sul da Europa: a importância das desigualdades na educação e na participação no mercado de trabalho, ao nível regional. Estudo em elaboração no âmbito do projecto POCTI/DEM/59445/2004. Acedido em, 19(05), 2010.

Miranda, A., & Trivedi, P. K. (2021). *Econometric models of fertility* (pp. 113-154). Edward Elgar Publishing.

Morales, M. (2021). Intergenerational transmission of fertility outcomes in Spain. *The Manchester School*, 89(4), 315-329

Muller, S. A. (2023). Comparação das Diferenças Nos Resultados das Principais Taxas de Natalidade e Fecundidade Entre as Regiões com Municípios Saudáveis e Não Saudáveis Durante o Primeiro Ano da Pandemia da Covid-19 em Portugal (Master's thesis, Universidade NOVA de Lisboa (Portugal)).

Platt, M., & Sterling, P. (2024, May 3). A world with fewer children? Addressing the despair behind declining fertility. *Penn Today*. <https://penntoday.upenn.edu/news/world-fewer-children-addressing-despair-behind-declining-fertility>

Rydell, I. (2002). Demographic patterns from the 1960s in France, Italy, Spain and Portugal. *Institute for Future Studies*, (2003), 24.

Salari, M. (2020). Culture and heritage language: a study of female labor force participation. *Review of Economics of the Household*, 18(2), 285-306.

Testa, M. R. (2014). On the positive correlation between education and fertility intentions in Europe: Individual-and country-level evidence. *Advances in life course research*, 21, 28-42.

Tomé, L. P. (2015). Why Portugal is not replacing generations?: A period and cohort perspective in a comparative analysis with selected European countries (Doctoral dissertation, Universidade de Evora (Portugal)).

Viera, J. M. F. (2019). Análise de modelos de regressão binária com eventos raros (Master's thesis, Universidade de Lisboa (Portugal)).

Wang, P., Yip, C. K., & Scotese, C. A. (1994). Fertility choice and economic growth: Theory and evidence. *The Review of Economics and Statistics*, 255-266.

Webbink, E., Smits, J., & de Jong, E. (2013). Household and context determinants of child labor in 221 districts of 18 developing countries. *Social indicators research*, 110, 819-836.

Zhang, C., & Li, T. (2017). Culture, fertility and the socioeconomic status of women. *China Economic Review*, 45, 279-288.

Anexos

Script

```
library(haven)
library(dplyr)
library(tibble)
library(tidyr)
library(car)

getwd()

# Iniciar log
sink("output_script.txt", split = TRUE)
cat("Início da execução:", format(Sys.time(), "%Y-%m-%d %H:%M:%S"),
"\n\n")

# Ler base de dados
dados <- read_sav("IFEC2019_BDINV.sav")

# Número de filhos (variável dependente)
dados <- dados %>%
  mutate(N_FILHOS = NFilhosnv)

# Idade
dados <- dados %>% mutate(IDADE = IdadeIndiv)

# Dummies para escolaridade (substitui EDU_NUM)
dados <- dados %>%
  mutate(
    EDU_SEC = case_when(
      NivelEscola_rec == 2 ~ 1,
      NivelEscola_rec %in% c(1, 3) ~ 0,
      TRUE ~ NA_real_
    ),
    EDU_SUP = case_when(
      NivelEscola_rec == 3 ~ 1,
      NivelEscola_rec %in% c(1, 2) ~ 0,
```

```

    TRUE ~ NA_real_
  )
)

#Participação do inquirido no mercado de trabalho

dados <- dados %>%
  mutate(
    emp_1 = case_when(
      D0420_1_FBO == 1 ~ 1,      # Empregada
      D0420_1_FBO == 9 ~ NA_real_, # Não sabe/não responde
      TRUE ~ 0                  # Todos os restantes (incluindo desempregada,
                                # estudante, etc.)
    ),
    emp_2 = case_when(
      D0420_2_FBO == 1 ~ 1,
      D0420_2_FBO == 9 ~ NA_real_,
      TRUE ~ 0
    ),
    emp_3 = case_when(
      D0420_3_FBO == 1 ~ 1,
      D0420_3_FBO == 9 ~ NA_real_,
      TRUE ~ 0
    ),

    emp_validos = rowSums(!is.na(cbind(emp_1, emp_2, emp_3))),
    emp_soma = rowSums(cbind(emp_1, emp_2, emp_3), na.rm = TRUE),

    EMP = case_when(
      N_FILHOS == 0 ~ 0,          # Inquiridos sem filhos
      emp_validos > 0 ~ emp_soma / emp_validos, # Média proporcional
      TRUE ~ NA_real_
    )
  )
)

#Região
dados <- dados %>%
  mutate(
    REGIAO = factor(N2Cod,
      levels = c(11, 15, 16, 17, 18, 20, 30),
      labels = c("Norte",
        "Algarve",
        "Centro",

```

```

        "Área Metropolitana de Lisboa",
        "Alentejo",
        "Região Autónoma dos Açores",
        "Região Autónoma da Madeira"))
)

```

Versão contínua (usada nos modelos 1 a 3)

```

dados <- dados %>%
  mutate(

    CULT_CONT = case_when(
      B0100_COD %in% 0:4 ~ B0100_COD,
      B0100_COD == 5 ~ 0,
      TRUE ~ NA_real_
    ),

    CULT_5M = case_when(
      B0100_COD == 5 ~ 1,
      B0100_COD %in% 0:4 ~ 0,
      TRUE ~ NA_real_
    )
  )
)

```

Versão categórica completa

```

dados <- dados %>%
  mutate(
    CULT_CONT = case_when(
      B0100_COD %in% 0:4 ~ B0100_COD,
      B0100_COD == 5 ~ 0,
      TRUE ~ NA_real_
    ),
    CULT_5M = case_when(
      B0100_COD == 5 ~ 1,
      B0100_COD %in% 0:4 ~ 0,
      TRUE ~ NA_real_
    ),
    CULT_FATOR = case_when(
      B0100_COD %in% 0:5 ~ as.character(B0100_COD),
      TRUE ~ NA_character_
    )
  )
)

```

```

    ),
    CULT_FATOR = factor(CULT_FATOR, levels = as.character(0:5))
  )

# Definir "0 irmãos" como categoria de referência
dados$CULT_FATOR <- relevel(dados$CULT_FATOR, ref = "0")

#Residência em zona urbana

dados <- dados %>%
  mutate(
    URB = case_when(
      tipau1_rec == 1 ~ 1,
      tipau1_rec == 2 ~ 0,
      TRUE ~ NA_real_
    )
  )

# Definir os valores médios dos escalões
rend_medias <- c(400, 600, 800, 1000, 1200, 1400, 1650, 2000, 2550, 3000)

# --- Rendimento Familiar - Médias ---
rend_medias <- c(400, 600, 800, 1000, 1200, 1400, 1650, 2000, 2550, 3000)

dados <- dados %>%
  mutate(
    REN = ifelse(ESC_REND %in% 2:9, rend_medias[ESC_REND] / 1000, 0) #
    divide por 1000 e associa 1 e 10 a zero
  )

# --- Rendimento Familiar - Dummies ---
dados <- dados %>%
  mutate(
    ESC_REND = factor(ESC_REND, levels = 1:10)
  )

# Definir escalão 1 como referência (opcional, caso uses ESC_REND como fator
em modelos)
dados$ESC_REND <- relevel(dados$ESC_REND, ref = "1")

```

```
# Dummies para extremos do rendimento, com estrutura de limpeza igual à  
CULT
```

```
dados <- dados %>%  
  mutate(  
    REN_ESC_1 = case_when(  
      ESC_REND == 1 ~ 1,  
      ESC_REND %in% 2:10 ~ 0,  
      TRUE ~ NA_real_  
    ),  
    REN_ESC_10 = case_when(  
      ESC_REND == 10 ~ 1,  
      ESC_REND %in% 1:9 ~ 0,  
      TRUE ~ NA_real_  
    )  
  )
```

```
# --- Variáveis adicionais relacionadas ao rendimento (dummies) ---
```

```
# Condições de habitação
```

```
dados <- dados %>%  
  mutate(  
    cond_1 = case_when(D0444_1_FBO == 1 ~ 1, D0444_1_FBO == 9 ~ NA_real_,  
      TRUE ~ 0),  
    cond_2 = case_when(D0444_2_FBO == 1 ~ 1, D0444_2_FBO == 9 ~ NA_real_,  
      TRUE ~ 0),  
    cond_3 = case_when(D0444_3_FBO == 1 ~ 1, D0444_3_FBO == 9 ~ NA_real_,  
      TRUE ~ 0),
```

```
    cond_validos = rowSums(!is.na(cbind(cond_1, cond_2, cond_3))),  
    cond_soma = rowSums(cbind(cond_1, cond_2, cond_3), na.rm = TRUE),
```

```
    COND_HAB = case_when(  
      N_FILHOS == 0 ~ 0,  
      cond_validos > 0 ~ cond_soma / cond_validos,  
      TRUE ~ NA_real_  
    )  
  )
```

```
# Conciliação vida pessoal/profissional
```

```
dados <- dados %>%  
  mutate(  
    
```

```

    concil_1 = case_when(D0443_1_FBO == 1 ~ 1, D0443_1_FBO == 9 ~ NA_real_,
TRUE ~ 0),
    concil_2 = case_when(D0443_2_FBO == 1 ~ 1, D0443_2_FBO == 9 ~ NA_real_,
TRUE ~ 0),
    concil_3 = case_when(D0443_3_FBO == 1 ~ 1, D0443_3_FBO == 9 ~ NA_real_,
TRUE ~ 0),

    concil_validos = rowSums(!is.na(cbind(concil_1, concil_2, concil_3))),
    concil_soma = rowSums(cbind(concil_1, concil_2, concil_3), na.rm = TRUE),

    CONCIL = case_when(
      N_FILHOS == 0 ~ 0,
      concil_validos > 0 ~ concil_soma / concil_validos,
      TRUE ~ NA_real_
    )
  )
)

# Estabilidade do EMP
dados <- dados %>%
  mutate(
    estavel_1 = case_when(D0442_1_FBO == 1 ~ 1, D0442_1_FBO == 9 ~
NA_real_, TRUE ~ 0),
    estavel_2 = case_when(D0442_2_FBO == 1 ~ 1, D0442_2_FBO == 9 ~
NA_real_, TRUE ~ 0),
    estavel_3 = case_when(D0442_3_FBO == 1 ~ 1, D0442_3_FBO == 9 ~
NA_real_, TRUE ~ 0),

    estavel_validos = rowSums(!is.na(cbind(estavel_1, estavel_2, estavel_3))),
    estavel_soma = rowSums(cbind(estavel_1, estavel_2, estavel_3), na.rm =
TRUE),

    EMP_ESTAVEL = case_when(
      N_FILHOS == 0 ~ 0,
      estavel_validos > 0 ~ estavel_soma / estavel_validos,
      TRUE ~ NA_real_
    )
  )
)

# --- Estatísticas descritivas (Antes da Limpeza dos Dados) ---
vars_modelo <- dados %>%
  select(N_FILHOS, IDADE, EDU_SEC, EDU_SUP, REN, EMP, CULT_CONT,
CULT_5M, URB, COND_HAB, CONCIL, EMP_ESTAVEL)

```

```
summary_stats <- tibble(

  Variável = names(vars_modelo),
  N = sapply(vars_modelo, function(x) sum(!is.na(x))),
  Média = sapply(vars_modelo, function(x) if(is.numeric(x)) mean(x, na.rm =
TRUE) else NA),
  Mediana = sapply(vars_modelo, function(x) if(is.numeric(x)) median(x,
na.rm = TRUE) else NA),
  Desvio_Padrão = sapply(vars_modelo, function(x) if(is.numeric(x)) sd(x,
na.rm = TRUE) else NA),
  Mínimo = sapply(vars_modelo, function(x) if(is.numeric(x)) min(x, na.rm =
TRUE) else NA),
  Máximo = sapply(vars_modelo, function(x) if(is.numeric(x)) max(x, na.rm =
TRUE) else NA)
)

print(summary_stats)
```

Limpeza: remover observações com NA em qualquer uma das variáveis relevantes

```
dados <- dados %>%
  drop_na(N_FILHOS, IDADE, EDU_SEC, EDU_SUP, REN, EMP,
CULT_CONT, URB,
  COND_HAB, CONCIL, EMP_ESTAVEL, REGIAO)
```

```
# --- Estatísticas descritivas (Após Limpeza dos Dados) ---
vars_modelo <- dados %>%
  select(N_FILHOS, IDADE, EDU_SEC, EDU_SUP, REN, REN_ESC_1,
REN_ESC_10, EMP, CULT_CONT, CULT_5M, URB, COND_HAB, CONCIL,
EMP_ESTAVEL)
```

```
summary_stats <- tibble(

  Variável = names(vars_modelo),
  N = sapply(vars_modelo, function(x) sum(!is.na(x))),
  Média = sapply(vars_modelo, function(x) if(is.numeric(x)) mean(x, na.rm =
TRUE) else NA),
  Mediana = sapply(vars_modelo, function(x) if(is.numeric(x)) median(x,
na.rm = TRUE) else NA),
  Desvio_Padrão = sapply(vars_modelo, function(x) if(is.numeric(x)) sd(x,
na.rm = TRUE) else NA),
```

```

Mínimo = sapply(vars_modelo, function(x) if(is.numeric(x)) min(x, na.rm =
TRUE) else NA),
Máximo = sapply(vars_modelo, function(x) if(is.numeric(x)) max(x, na.rm =
TRUE) else NA)
)

print(summary_stats)

cat("\nFrequência - REGIAO\n")
print(table(dados$REGIAO))

#Modelo OLS (sem variáveis por filhos)

modelo_1 <- lm(N_FILHOS ~ IDADE + EDU_SEC + EDU_SUP +
URB + CULT_CONT + CULT_5M + REN_ESC_1 + REN +
REN_ESC_10,
data = dados)

summary(modelo_1)

#Efeitos Fixos Região (sem variáveis por filhos)

modelo_2 <- lm(N_FILHOS ~ IDADE + EDU_SEC + EDU_SUP +
URB + CULT_CONT + CULT_5M + REN_ESC_1 + REN +
REN_ESC_10 + REGIAO,
data = dados)

summary(modelo_2)

#Efeitos Fixos Região (com variáveis por filhos)

modelo_3 <- lm(N_FILHOS ~ IDADE + EDU_SEC + EDU_SUP +
EMP + URB + CULT_CONT + CULT_5M + REN_ESC_1 + REN +
REN_ESC_10 +
COND_HAB + CONCIL + EMP_ESTAVEL +
REGIAO,
data = dados)

summary(modelo_3)

```

```

#Efeitos Fixos Região (com variáveis por filhos + CULT E ESC_REND)

modelo_4 <- lm(N_FILHOS ~ IDADE + EDU_SEC + EDU_SUP +
  EMP + URB +
  COND_HAB + CONCIL + EMP_ESTAVEL +
  ESC_REND + CULT_FATOR + REGIAO,
  data = dados)

summary(modelo_4)

cat("\n\nFim da execução:", format(Sys.time(), "%Y-%m-%d %H:%M:%S"),
"\n")
sink()
print(summary_stats)

# Limpeza: remover observações com NA em qualquer uma das variáveis
relevantes
dados <- dados %>%
  drop_na(N_FILHOS, AGE, EDU_SEC, EDU_SUP, REN, EMP, CULT_CONT,
  URB,
  COND_HAB, CONCIL, EMP_ESTAVEL, REGIAO)

# --- Estatísticas descritivas (Após Limpeza dos Dados) ---
vars_modelo <- dados %>%
  select(N_FILHOS, AGE, EDU_SEC, EDU_SUP, REN, ESC_REND_1,
  ESC_REND_10, EMP, CULT_CONT, CULT_5M, URB, COND_HAB, CONCIL,
  EMP_ESTAVEL)

summary_stats <- tibble(

  Variável = names(vars_modelo),
  N = sapply(vars_modelo, function(x) sum(!is.na(x))),
  Média = sapply(vars_modelo, function(x) if(is.numeric(x)) mean(x, na.rm =
  TRUE) else NA),
  Mediana = sapply(vars_modelo, function(x) if(is.numeric(x)) median(x,
  na.rm = TRUE) else NA),
  Desvio_Padrão = sapply(vars_modelo, function(x) if(is.numeric(x)) sd(x,
  na.rm = TRUE) else NA),
  Mínimo = sapply(vars_modelo, function(x) if(is.numeric(x)) min(x, na.rm =
  TRUE) else NA),

```

```
Máximo = sapply(vars_modelo, function(x) if(is.numeric(x)) max(x, na.rm =
TRUE) else NA)
)
```

```
print(summary_stats)
```

```
cat("\nFrequência - REGIAO\n")
print(table(dados$REGIAO))
```

```
#Modelo OLS (sem variáveis por filhos)
```

```
modelo_1 <- lm(N_FILHOS ~ AGE + EDU_SEC + EDU_SUP +
EMP + URB + CULT_CONT + CULT_5M + REN,
data = dados)
```

```
summary(modelo_1)
```

```
#Efeitos Fixos Região (sem variáveis por filhos)
```

```
modelo_2 <- lm(N_FILHOS ~ AGE + EDU_SEC + EDU_SUP +
EMP + URB + CULT_CONT + CULT_5M + REN +
REGIAO,
data = dados)
```

```
summary(modelo_2)
```

```
#Efeitos Fixos Região (com variáveis por filhos)
```

```
modelo_3 <- lm(N_FILHOS ~ AGE + EDU_SEC + EDU_SUP +
EMP + URB + CULT_CONT + CULT_5M + REN +
COND_HAB + CONCIL + EMP_ESTAVEL +
REGIAO,
data = dados)
```

```
summary(modelo_3)
```

```
#Efeitos Fixos Região (com variáveis por filhos + CULT E ESC_REND)
```

```
modelo_4 <- lm(N_FILHOS ~ AGE + EDU_SEC + EDU_SUP +
EMP + URB +
```

```
COND_HAB + CONCIL + EMP_ESTAVEL +  
ESC_REND + CULT_FATOR + REGIAO,  
data = dados)
```

```
summary(modelo_4)
```

```
cat("\n\nFim da execução:", format(Sys.time(), "%Y-%m-%d %H:%M:%S"),  
"\n")  
sink()
```

LOG FINAL

Início da execução: 2025-04-10 23:33:00

A tibble: 12 x 7

Variável	N	Média	Mediana	Desvio_Padrão	Mínimo	Máximo
<chr>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1 N_FILHOS	7709	0.997	1	1.01	0	7
2 IDADE	7709	37.5	40	10.1	18	54
3 EDU_SEC	7702	0.344	0	0.475	0	1
4 EDU_SUP	7702	0.315	0	0.464	0	1
5 REN	7709	1.15	1.2	0.770	0	2.55
6 EMP	7709	0.281	0.333	0.306	0	1
7 CULT_CONT	7709	1.33	1	1.11	0	4
8 CULT_5M	7709	0.117	0	0.321	0	1
9 URB	7709	0.269	0	0.443	0	1
10 COND_HAB	7709	0.0918	0	0.199	0	1
11 CONCIL	7709	0.0913	0	0.197	0	1
12 EMP_ESTAVEL	7709	0.105	0	0.209	0	1

A tibble: 14 x 7

Variável	N	Média	Mediana	Desvio_Padrão	Mínimo	Máximo
<chr>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1 N_FILHOS	7702	0.997	1	1.01	0	7
2 IDADE	7702	37.5	40	10.1	18	54
3 EDU_SEC	7702	0.344	0	0.475	0	1
4 EDU_SUP	7702	0.315	0	0.464	0	1
5 REN	7702	1.15	1.2	0.770	0	2.55
6 REN_ESC_1	6948	0.0262	0	0.160	0	1
7 REN_ESC_10	6948	0.0835	0	0.277	0	1
8 EMP	7702	0.281	0.333	0.306	0	1
9 CULT_CONT	7702	1.33	1	1.11	0	4
10 CULT_5M	7702	0.117	0	0.321	0	1
11 URB	7702	0.269	0	0.443	0	1
12 COND_HAB	7702	0.0919	0	0.199	0	1
13 CONCIL	7702	0.0914	0	0.197	0	1
14 EMP_ESTAVEL	7702	0.105	0	0.209	0	1

Frequência - REGIAO

Norte

Algarve

Centro

	1542	945	1239
Área Metropolitana de Lisboa			
Açores			Alentejo Região Autónoma dos
	1521	713	892
Região Autónoma da Madeira			
	850		

Call:

```
lm(formula = N_FILHOS ~ IDADE + EDU_SEC + EDU_SUP + URB +
  CULT_CONT +
  CULT_5M + REN_ESC_1 + REN + REN_ESC_10, data = dados)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.1843	-0.5108	-0.1269	0.5633	5.6943

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.989941	0.055716	-17.768	< 2e-16 ***
IDADE	0.047501	0.001103	43.050	< 2e-16 ***
EDU_SEC	-0.153698	0.026973	-5.698	1.26e-08 ***
EDU_SUP	-0.257864	0.029281	-8.807	< 2e-16 ***
URB	0.060665	0.023445	2.588	0.00969 **
CULT_CONT	0.059824	0.010688	5.597	2.26e-08 ***
CULT_5M	0.404430	0.038639	10.467	< 2e-16 ***
REN_ESC_1	-0.008249	0.070059	-0.118	0.90627
REN	0.143274	0.019982	7.170	8.27e-13 ***
REN_ESC_10	0.354028	0.049929	7.091	1.47e-12 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8584 on 6938 degrees of freedom
(754 observations deleted due to missingness)

Multiple R-squared: 0.2848, Adjusted R-squared: 0.2838

F-statistic: 306.9 on 9 and 6938 DF, p-value: < 2.2e-16

Call:

```
lm(formula = N_FILHOS ~ IDADE + EDU_SEC + EDU_SUP + URB +
  CULT_CONT +
  CULT_5M + REN_ESC_1 + REN + REN_ESC_10 + REGIAO, data = dados)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.2448	-0.5200	-0.1169	0.5628	5.6402

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.083872	0.059341	-18.265	< 2e-16 ***
IDADE	0.047566	0.001113	42.753	< 2e-16 ***
EDU_SEC	-0.149467	0.027035	-5.529	3.34e-08 ***
EDU_SUP	-0.258504	0.029321	-8.816	< 2e-16 ***
URB	0.039464	0.025209	1.565	0.117524
CULT_CONT	0.057613	0.010813	5.328	1.02e-07 ***
CULT_5M	0.401415	0.039566	10.145	< 2e-16 ***
REN_ESC_1	-0.001450	0.069959	-0.021	0.983470
REN	0.152319	0.020042	7.600	3.35e-14 ***
REN_ESC_10	0.383330	0.050219	7.633	2.60e-14 ***
REGIAOAlgarve	0.144168	0.037957	3.798	0.000147 ***
REGIAOCentro	0.098418	0.034655	2.840	0.004525 **
REGIAOÁrea Metropolitana de Lisboa	0.041399	0.034015	1.217	0.223615
REGIAOAlentejo	0.108591	0.041082	2.643	0.008229 **
REGIAORegião Autónoma dos Açores	0.197338	0.038726	5.096	3.56e-07 ***
REGIAORegião Autónoma da Madeira	0.089139	0.038231	2.332	0.019751 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8567 on 6932 degrees of freedom

(754 observations deleted due to missingness)

Multiple R-squared: 0.2883, Adjusted R-squared: 0.2867

F-statistic: 187.2 on 15 and 6932 DF, p-value: < 2.2e-16

Call:

```
lm(formula = N_FILHOS ~ IDADE + EDU_SEC + EDU_SUP + EMP + URB +  
    CULT_CONT + CULT_5M + REN_ESC_1 + REN + REN_ESC_10 +  
COND_HAB +  
    CONCIL + EMP_ESTAVEL + REGIAO, data = dados)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.7705	-0.2188	-0.1019	-0.0045	4.6250

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.0068737	0.0348629	0.197	0.843706
IDADE	0.0064530	0.0007171	8.999	< 2e-16 ***
EDU_SEC	-0.0942121	0.0153522	-6.137	8.89e-10 ***
EDU_SUP	-0.1148669	0.0166886	-6.883	6.38e-12 ***
EMP	2.6106666	0.0242703	107.566	< 2e-16 ***
URB	0.0124356	0.0143098	0.869	0.384862
CULT_CONT	0.0281591	0.0061414	4.585	4.62e-06 ***
CULT_5M	0.2029547	0.0225081	9.017	< 2e-16 ***
REN_ESC_1	0.0372149	0.0396904	0.938	0.348467
REN	-0.0472389	0.0114934	-4.110	4.00e-05 ***
REN_ESC_10	-0.0748807	0.0287470	-2.605	0.009212 **
COND_HAB	0.1567488	0.0470128	3.334	0.000860 ***
CONCIL	-0.0129242	0.0492068	-0.263	0.792827
EMP_ESTAVEL	0.2067358	0.0454234	4.551	5.42e-06 ***
REGIAOAlgarve	0.0826882	0.0215422	3.838	0.000125 ***
REGIAOCentro	0.0294922	0.0196782	1.499	0.133990
REGIAOÁrea Metropolitana de Lisboa	0.0347637	0.0193012	1.801	0.071728 .
REGIAOAlentejo	0.0520871	0.0233170	2.234	0.025523 *
REGIAORegião Autónoma dos Açores	0.1750361	0.0219718	7.966	1.90e-15 ***
REGIAORegião Autónoma da Madeira	0.0530585	0.0217314	2.442	0.014649 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.486 on 6928 degrees of freedom

(754 observations deleted due to missingness)

Multiple R-squared: 0.7711, Adjusted R-squared: 0.7705

F-statistic: 1228 on 19 and 6928 DF, p-value: < 2.2e-16

Call:

```
lm(formula = N_FILHOS ~ IDADE + EDU_SEC + EDU_SUP + EMP + URB +
    COND_HAB + CONCIL + EMP_ESTAVEL + ESC_REND + CULT_FATOR
+
    REGIAO, data = dados)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.7601	-0.2156	-0.1004	-0.0032	4.6182

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.0611842	0.0483618	1.265	0.205866
IDADE	0.0064709	0.0007161	9.037	< 2e-16 ***
EDU_SEC	-0.0921466	0.0153533	-6.002	2.05e-09 ***
EDU_SUP	-0.1149536	0.0166934	-6.886	6.24e-12 ***
EMP	2.6146924	0.0242525	107.811	< 2e-16 ***
URB	0.0126767	0.0142863	0.887	0.374929
COND_HAB	0.1587320	0.0469535	3.381	0.000727 ***
CONCIL	-0.0127933	0.0491231	-0.260	0.794537
EMP_ESTAVEL	0.2057966	0.0453605	4.537	5.80e-06 ***
ESC_REND2	-0.0025086	0.0406145	-0.062	0.950750
ESC_REND3	-0.0379922	0.0413904	-0.918	0.358703
ESC_REND4	-0.1140264	0.0405815	-2.810	0.004971 **
ESC_REND5	-0.1366014	0.0396640	-3.444	0.000577 ***
ESC_REND6	-0.1135506	0.0401071	-2.831	0.004651 **
ESC_REND7	-0.1478419	0.0404764	-3.653	0.000262 ***
ESC_REND8	-0.1251936	0.0402385	-3.111	0.001870 **
ESC_REND9	-0.1145230	0.0418047	-2.739	0.006169 **
ESC_REND10	-0.1106634	0.0428493	-2.583	0.009826 **
CULT_FATOR1	0.0019420	0.0192831	0.101	0.919782
CULT_FATOR2	0.0358838	0.0215405	1.666	0.095784 .
CULT_FATOR3	0.0686488	0.0254060	2.702	0.006908 **
CULT_FATOR4	0.1054229	0.0304302	3.464	0.000535 ***
CULT_FATOR5	0.1813443	0.0255750	7.091	1.47e-12 ***
REGIAOAlgarve	0.0828041	0.0215140	3.849	0.000120 ***
REGIAOCentro	0.0303436	0.0196634	1.543	0.122840
REGIAOÁrea Metropolitana de Lisboa	0.0328165	0.0192775	1.702	0.088740 .
REGIAOAlentejo	0.0523920	0.0232913	2.249	0.024517 *
REGIAORegião Autónoma dos Açores	0.1681772	0.0219879	7.649	2.31e-14 ***
REGIAORegião Autónoma da Madeira	0.0511621	0.0217053	2.357	0.018445 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.485 on 6919 degrees of freedom
(754 observations deleted due to missingness)

Multiple R-squared: 0.7723, Adjusted R-squared: 0.7714

F-statistic: 838.2 on 28 and 6919 DF, p-value: < 2.2e-16

Lista de Prompts

Elaboração de um script em R para criação de variáveis relevantes para a análise e gerar estatísticas descritivas completas (média, mediana, desvio-padrão, mínimo, máximo e número de observações);

Codificação de variáveis categóricas em formato de *dummies* binárias;

Reestruturação de variáveis com múltiplas observações por filho (até três) para calcular proporções médias de respostas;

Aplicação de procedimentos de limpeza de dados;

Reestruturação e criação de variáveis culturais associadas ao número de irmãos do inquirido, através de diferentes abordagens: contínua, *dummies* e categórica ordinal;

Criação de variáveis regionais e de residência urbana, com codificação apropriada para regressões com ou sem efeitos fixos.;

Apoio no desenvolvimento do código para estimação de diferentes especificações de modelos de regressão linear (OLS), com e sem efeitos fixos de região, com e sem uso de variáveis ajustadas ao número de filhos;

Ajuste de gráfico em formato excel

Apoio na tradução de secções da tese e no alinhamento linguístico e terminológico em português e inglês;

Pesquisa de artigos para integração na revisão de literatura.

