

3 Metodologia do estudo

3.1. Descrição da pesquisa

Este estudo está replicando uma pesquisa realizada por Gilliam, Goetz e Hampton (2008), seu objetivo é apresentar uma metodologia de análise visando compreendermos mais sobre o comportamento entre cônjuges em relação à tolerância a risco.

Nesta pesquisa o parâmetro para avaliar a tolerância a risco foi a questão que trata este tema na SCF (“*Survey of Consumer Finances*”), detalhado no item 2.4. Houve uma pequena modificação no texto da questão do SCF retirando a palavra “e seu cônjuge / companheiro”.

A seguir apresentamos a versão modificada:

“Qual das situações abaixo descritas mais se aproxima do tipo de risco que você está disposto a assumir quando investe seus recursos?”

- 1) Assumo risco substancial esperando obter retorno substancial
- 2) Assumo risco acima da média esperando ter retorno acima da média
- 3) Assumo risco mediano esperando obter retorno em torno da média
- 4) Não desejo assumir nenhum risco financeiro

No Anexo 1A apresentamos o questionário original em inglês e sua versão em português com as explicações das adaptações realizadas para o questionário aplicado neste estudo, este último apresentado no Anexo 1B.

O questionário foi enviado por e-mail a vários grupos de pessoas², solicitando que devolvessem o questionário preenchido pelos dois cônjuges separadamente.

A semelhança da pesquisa original, utilizamos uma amostra pequena, de 78 casais (comparado com 110 da pesquisa original), e como tal pode apresentar algum viés, não somente pela quantidade como pela forma como foi coletada, no sentido de não ser representativa da população.

² O critério para distribuição dos questionários foi: colegas de trabalho e pessoas amigas do autor, mailing de alunos cursando pós graduação em Administração na PUC-Rio e na PUC de Curitiba e amigos de amigos do autor que funcionaram como “multiplicadores”.

3.2. Construção do modelo

A variável dependente foi a questão do SCF que avalia a tolerância a risco. No formato original, a resposta que assume risco substancial está codificada como 1 e a que não deseja assumir nenhum risco como 4. Porém para fins de tratamento de dados invertemos a ordem, de maneira que o código 1 representa a alternativa de menor risco e o código 4 a de maior risco, pois, desta forma, quanto maior o valor numérico maior será o risco, o que facilita a análise.

As variáveis independentes foram sexo, idade, nível educacional, participação na renda do casal e participação no patrimônio do casal³. O nível educacional foi dividido em três: 2º Grau completo ou menos, Curso Superior e Curso de Pós Graduação. A participação na renda de cada cônjuge em relação ao total da renda da família define a variável relativa à renda. O mesmo conceito foi utilizado na participação de cada cônjuge na formação de patrimônio, considerando sua participação no total do patrimônio, esta opção contempla a parcela de cada cônjuge isoladamente e o formado pelo casal.

A seguir apresentamos, na Tabela 1, a descrição estatística da amostra:

Tabela 1 – Estatística Descritiva da Amostra

Variáveis	Esposas			Maridos		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Idade	78	44,0	11,4	78	47,9	11,4
2º Grau completo ou menos	22			17		
Curso Superior	32			26		
Curso de Pós Graduação	24			35		
Percentual da Renda Familiar ganho	78	32,8	23,8	78	66,7	24,3
Percentual do patrimônio do casal gerado	78	17,4	19,1	78	44,3	36,0

³ No estudo original inclui, também, o nº de dependentes, porém esta questão não consta do questionário original, daí não sabermos como conseguiu esta informação, e, portanto, não tivemos como tratar esta “variável”.

Na seção 4 iremos comentar cada questão da pesquisa, mas é importante destacarmos algumas características relevantes da amostra. O nível de escolaridade dos Maridos é um pouco superior ao das Esposas, mas os destaques estão na (a) maior participação do marido na renda da família (média=67 e dp=24) comparado com o da esposa (média=33 e dp=24) e (b) maior participação do marido na formação de patrimônio do casal (média=44% e dp=36) comparado com o da esposa (média=17 e dp=19). Na nossa amostra os maridos tem uma participação de quase o dobro da esposa, tanto na renda como na formação do patrimônio. Outro aspecto que corrobora com o item b, é que o máximo (em uma faixa que vai de zero a 100) que as esposas atingiram foi 80%, enquanto que os maridos chegaram a 100% da formação de patrimônio. Lembramos que existe a opção da formação de patrimônio ser pelo casal, porém não se mostrou relevante e será detalhada no item 4.2.5.

No estudo original foram revistas algumas variáveis, agrupando algumas por conta de baixa representatividade (tais como nível educacional, onde utilizou dois níveis “com e sem Curso superior” pela pequena representatividade em alguns níveis) e outras por conta de sua alta correlação⁴ (por ex. idade).

Neste estudo avaliamos as correlações entre as variáveis independentes e somente a idade apresentou correlação significativa com sexo de forma que, também, iremos tratar esta variável agrupada.

Posteriormente se procurou identificar as correlações entre as variáveis independentes, o método utilizado está descrito no próximo item.

3.3. Modelo GLM (“*Generalized Linear Model*”)

Na estatística o modelo linear generalizado (GLM) é uma generalização flexível das ordinárias de regressão dos mínimos quadrados. O GLM generaliza regressão linear, permitindo que o modelo linear de estar relacionada com a variável resposta, através de uma **função de ligação** e permitindo que a magnitude da variação de cada medida a ser uma função do seu valor previsto.

⁴ A comparação através do coeficiente de correlação é indicado para variável numérica contínua, quando a variável for intervalar ou categórica ou nominal o mais indicado é o Teste Qui Quadrado.

Nelder e Wedderburn (1972) desenvolveram os Modelos Lineares Generalizados, que é uma extensão dos Modelos Normais Lineares em duas direções:

- a) A distribuição de $E(Y/X)$ não tem que ser normal, podendo ser qualquer distribuição da família exponencial (Normal, Poisson, Binomial e Gama) e
- b) a relação entre o valor esperado da variável resposta (ou dependente) Y e as variáveis explicativas (ou independentes) X é dada por uma função diferenciável que tem como argumento uma combinação linear das variáveis explicativas.

A idéia básica consiste em abrir o leque de opções para a distribuição da variável resposta, permitindo que a mesma pertença à família exponencial de distribuições, bem como dar maior flexibilidade para a relação funcional entre a média da variável resposta (μ) e o preditor linear η .

A ligação entre a média e o preditor linear não é necessariamente a identidade, podendo assumir qualquer forma monótona não linear.

Nelder e Wedderburn propuseram também um processo iterativo para a estimação dos parâmetros e introduziram o conceito de desvio que tem sido largamente utilizado na avaliação da qualidade do ajuste dos Modelos Lineares Generalizados, bem como no desenvolvimento de resíduos e medidas de diagnósticos. Este modelo tem como uma de suas vantagens ser uma maneira de unificar vários outros modelos estatísticos de qualquer distribuição da família exponencial.

O GLM consiste de três elementos:

1. Uma distribuição de função f , de uma família exponencial;
2. Um Preditor Linear $\eta = X\beta$, onde β é uma combinação linear de parâmetros desconhecidos;
3. Uma Função de Ligação g como $E(Y) = \mu = g^{-1}(\eta)$, onde $E(Y)$ é o valor esperado de Y .

Nesta estrutura, a variância é uma função V , que se expressa

$\text{Var}(Y) = V(\mu) = V(g^{-1}(X\beta))$, é conveniente se V pertença a uma distribuição da família exponencial, mas podemos simplesmente dizer que a variância é uma função do valor previsto.

Os parâmetros desconhecidos, β , são estimados através da Máxima Verossimilhança, Máxima quase Verossimilhança ou técnicas Bayesianas.

A seguir descrevemos cada um destes elementos.

Distribuição de Função:

As distribuições da família exponencial são as distribuições de probabilidade, parametrizadas por θ e τ , com função de densidade f (ou função de probabilidade de massa, para o caso de uma distribuição discreta) pode ser expressa da seguinte forma:

$f_Y(y, \theta, \tau) = \exp((a(y) b(\theta) - c(\theta)) / h(\tau) + d(y, \tau))$, onde τ , chamado de parâmetro de dispersão, é conhecido e usualmente relacionado a variância da distribuição. As funções a , b , c , d e h são conhecidas. Muitas, mas não todas, as distribuições de função são comuns nesta família.

θ está relacionado com a média da distribuição. Se a é uma função identidade, então a distribuição é dita ter a forma canônica. Se, além disto, b é a identidade e τ é conhecido, então θ é chamado de parâmetro canônico e é relacionado à média através de $\mu = E(Y) = c'(\theta)$.

Neste cenário, a variância da distribuição pode ser apresentada como

$$\text{Var}(Y) = c''(\theta) h(\tau).$$

Preditor Linear:

O preditor linear é a quantidade que incorpora a informação sobre as variáveis independentes no modelo. O símbolo η é geralmente utilizado para denominar o preditor linear. Está relacionado ao valor esperado dos dados através da função de ligação.

η é expressa como combinações lineares dos parâmetros desconhecidos β . Os coeficientes das combinações lineares são representados pela matriz das variáveis independentes X . η pode ser expresso como $\eta = X\beta$.

Os elementos de X tanto podem ser medidos pelos experimentadores ou estipulados por eles no processo de desenho da modelagem.

Função de Ligação:

A Função de Ligação permite o relacionamento entre o preditor linear e a média da função de distribuição. Existem muitas funções de ligação comumente utilizadas, e sua escolha pode ser algo arbitrária. Pode ser conveniente para coincidir com o domínio da função de ligação para o alcance da função de distribuição de média.

Ao usar uma função de distribuição com um parâmetro canônico θ , existe uma função de ligação que permite que $X^T Y$ seja uma estatística suficiente para β . Isto ocorre quando a função de ligação iguala o θ e o preditor linear. Abaixo apresentamos a tabela de Função de Ligação Canônica e seus inversos (também chamados de Função Média) utilizados em várias distribuições da família exponencial.

Distribuição	Nome	Função de Ligação	Média da Função
Normal	Identidade	$X\beta = \mu$	$\mu = X\beta$
Exponencial	Inversa	$X\beta = \mu^{-1}$	$\mu = X\beta^{-1}$
Gamma			
Gauss Inversa	Quadrado inverso	$X\beta = \mu^{-2}$	$\mu = (X\beta)^{-1/2}$
Poisson	Log	$X\beta = \ln(\mu)$	$\mu = \exp(X\beta)$
Binomial	Logit	$X\beta = \ln(\mu / (1 - \mu))$	$\mu = \exp(X\beta) / (1 + \exp(X\beta)) = 1 / (1 + \exp(-X\beta))$
Multinomial			

Nos casos das distribuições Gama e Exponencial, o domínio das Funções de Ligação Canônica não é a mesma que a da faixa da média. Em particular o preditor linear pode ser negativo, o que daria uma média negativa que seria impossível. Quando maximizamos a verossimilhança, devemos tomar as precauções para evitar isto. Uma alternativa é utilizar uma função de ligação não canônica.

Medidas de Ajustamento:

- a) *Máxima Verossimilhança*: Estimativas por Máxima Verossimilhança podem ser encontradas utilizando iterações reponderadas do algoritmo dos mínimos quadrados.

Pode ser realizado tanto pelo método de Newton-Raphson com atualizações da forma abaixo descrita:

$\beta^{(t+1)} = \beta^{(t)} + \tau^{-1}(\beta^{(t)}) \mu(\beta^{(t)})$, onde $\tau(\beta^{(t)})$ é a matriz de informação observada (o negativo da matriz de Hessian) e $\mu(\beta^{(t)})$ é a função “score”.

Como pelo Método de Fisher, descrito a seguir:

$\beta^{(t+1)} = \beta^{(t)} + \tau^{-1}(\beta^{(t)}) \mu(\beta^{(t)})$, onde $\tau(\beta^{(t)})$ é a matriz informação de Fisher. Se a função de ligação canônica for usada, então os dois métodos são iguais.

- b) *Método Bayesiano*: Em geral esta distribuição não pode se encontrada em forma fechada, portanto deve ser aproximada, usando as aproximações de Laplace ou algum tipo de cadeia de Markov, Método de Monte Carlo como a Amostragem de Gibbs.

3.3.1. Modelo GLM de medidas repetidas

Na pesquisa original foi aplicado o Modelo Linear Generalizado (GLM – “General Linear Model”) de Medidas Repetidas para utilizar o “within-subject factors” como Tolerância a risco por sexo e o “between-subject factors” como nível educacional da esposa, idade do casal, participação da esposa na formação de patrimônio, participação do marido na formação de patrimônio, participação do marido na renda familiar e participação da esposa na renda familiar.

Este método permite avaliar o efeito de uma ou mais variáveis independentes (ou explicativas) sobre uma variável de interesse (explicada, endógena ou resposta). No “between-subject” o modelo realiza comparações, considerando os efeitos entre os conjuntos de variáveis, as trata separadamente, enquanto que o “within-subject” (ou medidas repetidas) realiza comparações no próprio grupo.

3.4. Resultado

O Teste t de Student, que considera a diferença entre as médias, e é adequada ao tamanho da amostra, foi aplicado na tolerância a risco por sexo. Considerando a variável dependente a questão da tolerância a risco do SCF, identificamos que existe diferença significativa de risco entre o sexo masculino e feminino [t(78)= -3,167, p <0,002], ou seja, podemos afirmar, com 99,8% de segurança de que existe diferença significativa entre a atitude de homens e mulheres em relação a tolerância a risco, onde os homens assumem mais risco que as mulheres.

A seguir apresentamos a Tabela 2, com o cálculo.

Tabela 2 – Teste t de Student

Test T = (t = -3,167; p 0,002)						
Sexo = Masc						
Descriptive Statistics(a)		M			SD	
	N	Minimum	Maximum	Mean	Std. Deviation	Variance
O que se aproxima do tipo de risco que você está disposto a assumir quando investe seus recursos	78	1	4	2,04	0,80	0,63
Valid N (listwise)	78					
Sexo = Fem						
Descriptive Statistics(a)		M			SD	
	N	Minimum	Maximum	Mean	Std. Deviation	Variance
O que se aproxima do tipo de risco que você está disposto a assumir quando investe seus recursos	78	1	4	1,51	0,75	0,56
Valid N (listwise)	78					

Fonte: Própria

Utilizando o GLM de Repetição de Medida, apresentamos a comparação do “within-subject”, na Tabela 3, onde comparamos a variável Tolerância a risco por sexo (agrupada pelo casal) com as demais variáveis independentes, como idade do casal, nível educacional do marido e assim por diante, onde P12 e P13 se referem as questões 12 e 13, respectivamente, participação na renda da família e participação na formação do patrimônio. Nesta aplicação o sistema avalia se existe alguma relevância entre as variáveis analisadas, sem necessariamente estar condicionada a apresentar significância, ou seja, consegue testar. Estas mesmas variáveis (que se consegue testar no “Within-subject”) são, também, testadas no “Between-subject”.

O resultado desta aplicação está apresentado na Tabela 3, descrita a seguir, e que nos mostra que somente duas relações são significativas ($p < 0,05$), são elas:

sexo * MARIDO_P12 = indica que os maridos que possuem maior percentual de renda tem maior risco do que suas esposas.

sexo * MARIDO_P13 = indica que os maridos que possuem maior percentual do patrimônio tem maior risco do que suas esposas.

Tabela 3 - "Within- Subject"

Tests of Within-Subjects Contrasts				
Measure: MEASURE_1				
Source	df	Mean Square	F	Sig.
sexo	1	4,727	8,497	0,005
sexo * Fx_IDADE CASAL	4	0,238	0,429	0,788
sexo * ESPOSA_ESCOLA	3	0,745	1,339	0,268
sexo * MARIDO_ESCOLA	3	0,159	0,341	0,796
sexo * ESPOSA_P12	11	0,061	0,109	1,000
sexo * MARIDO_P12	11	1,207	2,584	0,008
sexo * ESPOSA_P13	8	0,991	1,782	0,093
sexo * MARIDO_P13	14	0,863	1,846	0,047
sexo * Fx_IDADE * ESPOSA_ESCOLA	1	0,953	1,714	0,194
sexo * Fx_IDADE * MARIDO_P13	1	0,357	0,764	0,385
sexo * ESPOSA_P12 * ESPOSA_P13	2	0,190	0,342	0,712
Error(sexo)	77	0,556		

Fonte: Própria

A Tabela 4, que apresenta o resultado do "Between-subject", onde mostra que não foi encontrada diferença significativa entre os grupos de variáveis em relação a tolerância a risco ($\text{sig.} = p > 0,05$).

Tabela 4 - "Between- Subject"

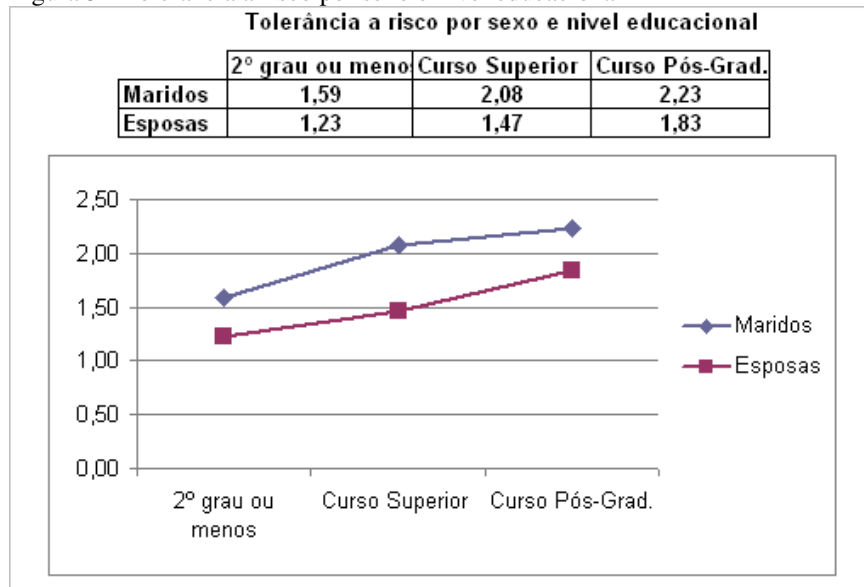
Tests of Between-Subjects Effects				
Measure: MEASURE_1				
Transformed Variable: Average				
Source	df	Mean Square	F	Sig.
Intercept	1	32,277	210,651	0,000
Fx_IDADE CASAL	4	0,115	0,753	0,559
ESPOSA_ESCOLA	3	0,434	1,833	0,064
MARIDO_ESCOLA	3	0,023	0,163	0,921
ESPOSA_P12	11	0,070	0,454	0,926
MARIDO_P12	11	0,261	1,877	0,065
ESPOSA_P13	8	0,248	1,618	0,134
MARIDO_P13	14	0,213	1,531	0,121
Fx_IDADE * ESPOSA_ESCOLA	1	0,463	3,021	0,086
Fx_IDADE * MARIDO_P13	1	0,071	0,514	0,475
ESPOSA_P12 * ESPOSA_P13	2	0,083	0,544	0,583
Error	77	0,153		

Fonte: Própria

Levando em consideração o estudo original que detectou uma relação entre nível educacional e sexo referente à tolerância a risco, o presente estudo também mostrou uma interação significativa entre nível educacional e sexo em relação a tolerância a risco. Mostrando que (a) quanto maior o nível educacional, maior a tolerância a risco e (b) que os maridos possuem maior tolerância a risco que suas esposas, já comentado acima, adicionalmente em todos os níveis educacionais.

Este resultado está apresentado na Figura 3, encontra respaldo de outros estudos como citado no Referencial Teórico. Precisamos atentar que o perfil do nível educacional da amostra é muito alto, bastante diferente da população.

Figura 3 - Tolerância a risco por sexo e nível educacional



Fonte: Própria