

Agregação de Preferências Eleitorais via Escalagem Psicométrica de Thurstone

Lucas Loureiro Lino da Costa

Julho, 2018

Resumo

Nesse trabalho utilizaremos um método incomum de lidar com dados categóricos ordinais às ciências econômicas, o método de escalagem psicométrica de Thurstone. Usar-lo-emos para determinação de preferências agregadas sobre candidatos em eleições a partir de manifestações individuais sobre pares de alternativas. Geralmente as pesquisas de intenção de voto publicam as preferências coletivas, em termos de frequências relativas, sobre pares de candidatos em determinados momentos: pesquisas pré-eleitorais. O método de escalagem psicométrica de Thurstone é uma das diversas técnicas de escalagem psicométrica unidimensional, que têm como objetivo determinar probabilisticamente as preferências agregadas dos eleitores sobre a lista de candidatos. Como instância de aplicação do método, tomaremos uma lista plausível de candidatos para a eleição presidencial brasileira de 2018, assim como uma amostra representativa da Faculdade de Economia, Administração, Contabilidade e Gestão de Políticas Públicas da Universidade de Brasília, para assim inferirmos as preferências coletivas da população em questão.

Palavras-chave: Agregação de Preferências Coletivas. Escalagem Psicométrica Unidimensional. Método de Thurstone. Eleições Brasileiras 2018.

Keywords: Collectives Aggregated Preferences. Unidimensional Psychometric Scale. Thurstone's method. Brazilian's Election 2018.

1 Introdução

O Método de Escalagem Psicométrica de Thurstone é uma ferramenta utilizada para estimação de preferências entre objetos via frequência observada dos pares comparados, [Thurstone \(1927\)](#), entretanto a sua utilização dentro do campo da Ciência Econômica ainda é escassa, sendo utilizada normalmente ferramentas de inferência econométricas no seu lugar. A fim de preencher essa lacuna epistemológica usar-mos-ei o modelo padrão de escalagem psicométrica unidimensional de Thurstone para elucidar as preferências coletivas ordinais de uma amostra de votantes. Para tanto, vamos utilizar a linguagem de programação *opensource* [R Core Team \(2018\)](#) para compilar e analisar os dados da pesquisa, assim como disponibilizaremos publicamente tanto o código fonte construído para tal (Anexo B), assim como, os dados da pesquisa no repositório online do autor, [Costa \(2018\)](#). Na sessão 2 temos a definição da amostra utilizada durante a análise. Na sessão 3 é apresentado o método de escalagem psicométrica unidimensional, referindo-se a [Borg e Groenen \(2005\)](#) para escalagem multidimensional e [Souza \(1988\)](#) para apresentação de variedades de métodos de escalagem unidimensional e multidimensional. A sessão 4 refere-se à aplicação do modelo de escalagem psicométrica unidimensional de Thurstone na amostra populacional. E finalmente a sessão 5 conclui este trabalho.

2 Definição da amostra

A Faculdade de Economia, Administração, Contabilidade e Gestão de Políticas Públicas da Universidade de Brasília (FACE), segundo último anuário estatístico, publicado pelo Decanato de Planejamento, Orçamento e Avaliação Institucional, apresentava em seu quadro 2.958 alunos de graduação regulares ativos registrados no segundo semestre de 2016, [Brasília \(2017\)](#)¹.

Utilizando-se do método de amostragem aleatória simples sem repetição, com a captura dos dados realizado via formulário preenchido pelos participantes do estudo (Anexo A), assim como, tomando como base um erro amostral padrão de 5% e um nível de confiança também padrão de 95%, temos que o tamanho mínimo da nossa amostra para que seja considerada representativa deve ser de 341 participantes, condição a qual foi saciada devido a amostra utilizada conter 360 participantes.

¹ Por carácter de simplificação, tomaremos que a população se manteve estática até o período da análise, o primeiro semestre de 2018.

3 O método de escalagem psicométrica

Segundo o método de escalagem psicométrica unidimensional de Thurstone, quando um indivíduo se depara com um conjunto de escolhas, ele individualmente designa um valor dentro de uma escala para cada alternativa. Entretanto cada designação dentro da escala é por natureza uma variável aleatória. O processo de designação dos valores adentra o que os psicólogos chamam de *psychological continuum*, onde a construção da escala é oriunda de uma sub-divisão finita deste último, composto por sub-intervalos. Quando o valor dessa variável aleatória cai sobre um dos sub-intervalos do *psychological continuum*, é essa posição do sub-intervalo que determina a resposta do indivíduo a pergunta. Resposta esta categórica e ordenada da menor para a maior dentro do espectro. Ou seja, quando economistas falam sobre utilidade e psicometristas falam de escala de valores, eles estão ambos falando sobre a mesma coisa.

Considerando um conjunto finito $\Gamma = \{\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_n\}$ de n diferentes candidatos, além disso supondo que cada eleitor tenha suas relações de preferência estritas \succ em \mathbf{X} , ou seja, completas, transitivas e anti-reflexivas nas suas relações binárias \succ sobre o conjunto de candidatos em \mathbf{X} . O eleitor designa para cada candidato \mathbf{A}_i um número real μ_i chamado de escala de valor, ou utilidade, de \mathbf{A}_i , de maneira que $\mathbf{A}_i \succ \mathbf{A}_j$ se, e somente se, $\mu_i > \mu_j$, para $i \neq j$ e $i, j = 1, \dots, n$. O valor de escala μ_i é representado por uma variável aleatória distribuída normalmente $\mathbf{X}_i \sim N(\mu_i, \sigma_i^2)$, onde $\mu_i \in \mathbb{R}$ e variância $\sigma_i^2 > 0$. Dado que um grande número de eleitores independentemente designa valores de escala para um candidato, podemos assumir, pela Lei dos Números Grandes, que a media de μ_i é uma aproximação consistente para o verdadeiro valor de escala μ_i atribuído coletivamente para o candidato \mathbf{A}_i . A suposição de normalidade é aceitável se um grande número de eleitores independentes é assumido.

Apesar dos eleitores agirem independentemente, assume-se que as variáveis aleatórias são correlatas. Onde, $\sigma_{i,j} = \text{cov}(X_i, X_j)$ é a covariância entre X_i e X_j e sendo $p_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$ o coeficiente de correlação entre X_i e X_j , exatamente pelo fato que $p_{ij} \neq 0$, para $i \neq j$ e $i, j = 1, \dots, n$. Como correlação entre diferentes alternativas estão presentes no modelo, o Axioma de Interdependência das Alternativas Irrelevantes de Arrow, [Arrow \(2012\)](#), não se aplica, assim como demonstrado por [Tanguiane \(2012\)](#), exatamente pelo fato deste ser uma construção meramente algébrica e não estatística. Sendo assim, por consequência o Teorema de impossibilidade de Arrow também não é válido.

Devido a aleatoriedade do processo de escalagem psicométrica, a noção de preferência é substituída pela de probabilidade de preferência.

Onde $\phi_{ij} = Prob[A_i \succ A_j]$ seja a probabilidade que A_i seja preferido a A_j , ou seja, $\phi_{ij} = Prob[X_i > X_j]$. Como variáveis aleatórias distribuídas normalmente são contínuas, definimos $\phi_{ij} = Prob[X_i \geq X_j]$. O processo de normalização do evento $[X_i \geq X_j]$ é demonstrado por [Marques, Peñaloza et al. \(2011\)](#)

Sendo assim, o método de Thurstone de escalagem psicométrica unidimensional consistem em achar os valores críticos de z_{ij}^* para os quais $\hat{\phi}_{ij} = Prob[N(0, 1) \geq z_{ij}^*]$, onde $\hat{\phi}_{ij}$ é um estimador consistente para $Prob[A_i \succ A_j]$, ou seja, a frequência relativa do evento $[A_i \succ A_j]$ e por consequência $\hat{\mu}_i = -\frac{1}{n} \sum_{j=1}^n z_{ij}^*$. Assim que os otimizados valores de escala $\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_n$ são obtidos, podemos ordená-los do maior para o menor, $\mu_{(1)} > \mu_{(2)} > \dots > \mu_{(n)}$, onde $\mu_{(1)} = \max\{\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_n\}$, $\mu_{(2)} = \max\{\hat{\mu}_i : \hat{\mu}_i \neq \mu_{(1)}\}$ e assim por diante, constituindo assim a ordem das estatísticas. Finalmente, a relação estatística de preferências coletivas é dada por $A_{(1)} \succ A_{(2)} \succ A_{(n)}$.

4 Aplicação na amostra populacional

Até o período da confecção desse trabalho, os cinco principais candidatos à eleição brasileira de 2018 eram Jair Bolsonaro, Marina Silva, Ciro Gomes, Geraldo Alckmin e Guilherme Boulos ², sendo assim, estes formam o conjunto finito de candidatos os quais os participantes do estudo tinha acesso. Os eleitores eram então inqueridos a manifestar as suas intenções de votos para cada par de candidatos possíveis (Anexo A). Sendo assim, temos a seguinte tabela de frequência relativa por pares de candidatos.

$A_i \ A_j$	Bolsonaro	Marina	Ciro	Alckmin	Boulos
Bolsonaro	-	0.2111111	0.2722222	0.2805556	0.2861111
Marina	0.7888889	-	0.5916667	0.7111111	0.6055556
Ciro	0.7277778	0.4083333	-	0.5805556	0.5111111
Alckmin	0.7194444	0.2888889	0.4194444	-	0.3833333
Boulos	0.7138889	0.3944444	0.4888889	0.6166667	-

Os candidatos foram dados as seguintes nomes $A_1 = \text{Bolsonaro}$, $A_2 = \text{Marina}$, $A_3 = \text{Ciro}$, $A_4 = \text{Alckmin}$, $A_5 = \text{Boulos}$. A entrada $(A_1, A_2) = \text{significa}$ que 21,11% dos eleitores escolheram Bolsonaro contra Marina. Perceba que a estrada simétrica (A_2, A_1) significa exatamente

² Respectivamente candidatos pelo Partido Social Liberal (PSL), partido Rede Sustentabilidade, Partido Democrático Trabalhista (PDT), Partido da Social Democracia Brasileira (PSDB) e Partido Socialismo e Liberdade (PSol)

o contrário, ou seja, que 78,88% dos eleitores escolheram Marina frente a Bolsonaro. As demais entradas são interpretadas analogamente. Portanto temos a seguinte matriz de probabilidade de preferências:

$$\Phi = \begin{pmatrix} 0 & 0.2111111 & 0.2722222 & 0.2805556 & 0.2861111 \\ 0.7888889 & 0 & 0.5916667 & 0.7111111 & 0.6055556 \\ 0.7277778 & 0.4083333 & 0 & 0.5805556 & 0.5111111 \\ 0.7194444 & 0.2888889 & 0.4194444 & 0 & 0.3833333 \\ 0.7138889 & 0.3944444 & 0.4888889 & 0.6166667 & 0 \end{pmatrix}$$

Note que $\phi_{12} + \phi_{21} = 1$, já que, foi assumido que as preferências são completas, além do que, o mesmo é válido para todos os pares de candidatos.

Para calcularmos os valores de $z_{i,j}^*$, basta usarmos uma tabela de valores críticos para distribuição normal padrão. Sendo assim, temos a seguinte matriz Z^* de valores críticos:

$$Z^* = \begin{pmatrix} 0 & -0.8025719 & -0.60610588 & -0.5811919 & -0.56478175 \\ 0.8025719 & 0 & 0.23183436 & 0.5566336 & 0.26775373 \\ 0.6061059 & -0.2318344 & 0 & 0.2033149 & 0.02785503 \\ 0.5811919 & -0.5566336 & -0.20331493 & 0 & -0.29673784 \\ 0.5647817 & -0.2677537 & -0.02785503 & 0.2967378 & 0 \end{pmatrix}$$

A escala para o candidato A_1 no caso seria a seguinte:

$$\hat{\mu}_1 = -\frac{1}{5} \sum_{j=1}^3 z_{1j}^* \\ \hat{\mu}_1 = -0.51093029$$

Analogamente:

$$\hat{\mu}_2 = 0.37175872 \\ \hat{\mu}_3 = 0.12108830 \\ \hat{\mu}_4 = -0.09509889 \\ \hat{\mu}_5 = 0.11318217$$

Relembrando que por construção, a soma das escalas de valores tende a zero. Como $\mu_2 > \mu_3 > \mu_5 > \mu_4 > \mu_1$, o que nos dá

$A_2 \succ A_3 \succ A_5 \succ A_4 \succ A_1$, ou seja, *Marina* \succ *Ciro* \succ *Boulos* \succ *Alckmin* \succ *Bolsonaro*. Portanto essa seria a ordenação de preferência de nossa amostra, com Marina Silva como candidato maximal e Jair Bolsonaro como candidato com a maior taxa de rejeição.

5 Conclusão

O método de escalagem psicométrica unidimensional de Thurstone é um método bastante simples de ser aplicado, além de sua poderosa habilidade para elucidar as ordenações de preferências coletivas de votantes ou consumidores, sendo assim, advocamos que a sua utilização deveria ser mais presente no campo das Ciências Econômicas, assim como no da Teoria da Escolha Social.

Referências

ARROW, K. J. *Social choice and individual values*. [S.l.]: Yale University Press, 2012. v. 12. Citado na página 3.

BORG, I.; GROENEN, P. J. *Modern multidimensional scaling: Theory and applications*. [S.l.]: Springer Science & Business Media, 2005. Citado na página 2.

BRASÍLIA, F. U. de. *Anuário Estatístico da UnB 2017 - Período: 2012 a 2016*. [S.l.], 2017. (Anuário Estatístico da UnB). Citado na página 2.

COSTA, L. L. L. d. *Agregação de Preferências Eleitorais via Escalagem Psicométrica de Thurstone*. [S.l.]: GitHub, 2018. <<https://github.com/loureirolino/Agregacao-de-Preferencias-Eleitorais-via-Escalagem-Psicometrica-de-Thurstone>>. Citado na página 2.

MARQUES, I.; PEÑALOZA, R. et al. *Brazilian 2010 presidential election: psychometric scaling of voting intention*. [S.l.], 2011. Citado na página 4.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2018. Disponível em: <<https://www.R-project.org/>>. Citado na página 2.

SOUZA, J. d. Métodos estatísticos nas ciências psicossociais. vol. v-métodos de escalagem psicossocial: Uni e multidimensional. *Métodos Estatísticos nas Ciências Psicossociais. Vol. V: Métodos de escalagem psicossocial: Uni e Multidimensional*, Thesaurus Brasília, 1988. Citado na página 2.

TANGUIANE, A. S. *Aggregation and Representation of Preferences: Introduction to Mathematical Theory of Democracy*. [S.l.]: Springer Science & Business Media, 2012. Citado na página 3.

THURSTONE, L. A law of comparative judgement. *Psychological Review*, v. 34, p. 273–286, 1927. Citado na página 2.

ANEXO A – Questionário

Agregação de Preferências Eleitorais via Escalagem Psicométrica de Thurstone

Formulário de Pesquisa

Você está prestes a participar de um pesquisa sobre preferências eleitorais, onde a sua participação é voluntária e a sua identidade preservada. As questões a seguir apresentam os cinco principais candidatos à presidência do Brasil em 2018, apresentados em pares, totalizando assim dez questões. Sendo assim, por favor, entre as duas opções apresentadas em cada questão escolha a sua preferível, ou seja, entre o par proposto qual seria o seu candidato maximal entre os propostos (escolha única).

Os candidatos escolhidos para essa pesquisa serão: **Jair Bolsonaro** (PSL), **Marina Silva** (Rede Sustentabilidade), **Ciro Gomes** (PDT), **Geraldo Alckmin** (PSDB) e **Guilherme Boulos** (Psol).

- | | | |
|--|----------|---|
| 1. <input type="radio"/> Jair Bolsonaro | x | <input type="radio"/> Marina Silva |
| 2. <input type="radio"/> Jair Bolsonaro | x | <input type="radio"/> Ciro Gomes |
| 3. <input type="radio"/> Jair Bolsonaro | x | <input type="radio"/> Geraldo Alckmin |
| 4. <input type="radio"/> Jair Bolsonaro | x | <input type="radio"/> Guilherme Boulos |
| 5. <input type="radio"/> Marina Silva | x | <input type="radio"/> Ciro Gomes |
| 6. <input type="radio"/> Marina Silva | x | <input type="radio"/> Geraldo Alckmin |
| 7. <input type="radio"/> Marina Silva | x | <input type="radio"/> Guilherme Boulos |
| 8. <input type="radio"/> Ciro Gomes | x | <input type="radio"/> Geraldo Alckmin |
| 9. <input type="radio"/> Ciro Gomes | x | <input type="radio"/> Guilherme Boulos |
| 10. <input type="radio"/> Geraldo Alckmin | x | <input type="radio"/> Guilherme Boulos |

ANEXO B – R Script

Agregação de Preferências Eleitorais via Escalagem de Thurstone

Script em R para análise dos dados

Autor: Lucas Loureiro Lino da Costa

Data: Julho de 2018

Instalação e carregamento dos pacotes necessários

```
packages = c("ggplot2", "readr", "tidyverse") if (length(setdiff(packages, rownames(installed.packages()))) > 0) { install.packages(setdiff(packages, rownames(installed.packages()))) } library(readr) library(tidyverse)
```

Carregando os dados da pesquisa no sistema

```
dados_pibic = read_csv(file.choose()) candidatos = c("Jair_Bolsonaro", "Marina_Silva", "Ciro_Gomes", "Geraldo_Alckmin", "Guilherme_Boulos")
```

Criação de um dataframe vazio para receber os dados por pares

```
quadro_parwise = data.frame(Jair_Bolsonaro = numeric(), Marina_Silva = numeric(), Ciro_Gomes = numeric(), Geraldo_Alckmin = numeric(), Guilherme_Boulos = numeric())
```

Alimentando o quadro de escolhas por pares

```
for (i in 1:5){ for (j in 1: 5){ quadro_parwise[i,j] = length(which(dados_pibic[, i] > dados_pibic[,j])) } } row.names(quadro_parwise) = c("Jair_Bolsonaro", "Marina_Silva", "Ciro_Gomes", "Geraldo_Alckmin", "Guilherme_Boulos") quadro_parwise[quadro_parwise == 0] = NA
```

Quadro de escolhas por pares, dados em frequência relativa

```
quadro_parwise_freq = quadro_parwise/nrow(dados_pibic)
```

Quadro para os valores críticos para uma distribuição normal

```
quadro_parwise_critical = quadro_parwise_freq %>% mutate_all(funs(qnorm))
```

Valores de escala de cada candidato

```
escala = rowSums(quadro_parwise_critical, na.rm = TRUE)*(1/5) escala_candidatos = data.frame(candidatos, escala)
```

Agregação de Preferências Eleitorais via Escalagem Psicométrica de Thurstone

Lucas Loureiro Lino da Costa

Julho, 2018

Abstract

In this work we will use an unusual method of dealing with ordinal categorical data in economic sciences, Thurstone's method of psychometric scaling. We will use it to determine aggregate preferences about candidates in elections from individual manifestations about pairs of alternatives. As an instance of application of the method, we will take a plausible list of candidates for the 2018 Brazilian presidential election, as well as a representative sample of the Faculty of Administration, Accounting and Economics of the University of Brasília, in order to infer the collective preferences of the population in question.

Key-words: Collectives Aggregated Preferences. Unidimensional Psychometric Scale. Thurstone's method. Brazilian's Election 2018.