Universidade de São Paulo
Escola Politécnica
Departamento de Engenharia de Transportes

Estimativa de Parâmetros de Modelos de Escolha Discreta

PTR 5789
Planejamento de Transporte Urbano

Prof. Orlando Strambi
Prof. Cassiano Isler
Prof. Mateus Humberto Andrade
2023

Agenda

- Estimativa de Parâmetros
- Testes Estatísticos sobre Parâmetros e Modelos
- Exemplos

ESTIMATIVA DE PARÂMETROS

A modelagem de escolha discreta permite calcular as probabilidades de escolha de alternativas sob diferentes hipóteses para a distribuição da probabilidade dos erros aleatórios das respectivas funções utilidade.

Dado que o modelo matemático de cálculo de probabilidade é conhecido (Logit ou Probit), e assumindo que o modelo da parcela determinística da função utilidade da alternativa i para o indivíduo q (V_{iq}) também é previamente definido (geralmente sobre formato aditivo), surge o problema de estimativa dos parâmetros β_{ik} dessa função matemática.

$$V_{iq} = \sum_{k} \beta_{ik} \cdot x_{iqk}$$

Estimativa de Parâmetros

Formalmente, o problema é caracterizado pela necessidade de estimativa dos parâmetros das funções utilidade, representada por uma equação em relação aos atributos das alternativas e dos indivíduos envolvidos no processo de escolha.

Essa estimativa em geral é realizada a partir das escolhas reportadas por indivíduos que participaram de pesquisas específicas.

O método de estimativa de β_{ik} mais recorrente na literatura é o de maximização de verossimilhança, cujas estimativas de parâmetros resultam nos valores de probabilidade de escolha mais próximos daqueles reportados pelos indivíduos na pesquisa.

Em geral, considera-se uma proporção da amostra de respostas para calibração dos parâmetros β_{ik} (por exemplo 75%) e a outra parcela para validação das estimativas (no caso, 25%).

Seja P_{iq} a probabilidade de escolha da alternativa i pelo indivíduo q e g_{iq} um parâmetro binário tal que:

$$g_{iq} = \begin{cases} 1, \text{ se o indivíduo } q \text{ escolhe a alternativa } A_i \\ 0, \text{ caso contrário.} \end{cases}$$

Como as observações são independentes entre indivíduos, uma função de verossimilhança pode ser escrita por:

$$L(\beta) = \prod_{q=1}^{Q} \prod_{A_i \in A(q)} (P_{iq})^{g_{iq}}$$

Estimativa de Parâmetros

Para maximizar essa função de verossimilhança geralmente utiliza-se a sua versão linearizada pelo operador de logaritmo, tal que:

$$l(\beta) = log\,[L(\beta)] = \sum_{q=1}^{Q} \sum_{A_i \in A(q)} g_{iq} \cdot log(P_{iq})$$

O conjunto de valores $\hat{\beta}$ (estimadores de máxima verossimilhança) que resulta no valor máximo da função de verossimilhança $l(\beta)$ é aquele que mais aproxima as probabilidades de escolha daquelas realmente reportadas pelos indivíduos.

O valor máximo de $l(\beta)$ pode ser obtido analiticamente pela resolução de um sistema de equações dado pelas derivadas parciais dessa função em relação aos parâmetros $\hat{\beta}$ a serem estimados.

Estimativa de Parâmetros

Porém, dependendo do modelo utilizado para calcular P_{iq} (por exemplo, Logit ou Probit), a estimativa de parâmetros β que maximizam $l(\beta)$ pode ser difícil analiticamente .

Por isso, na prática, a estimativa de $\hat{\beta}$ envolve um processo iterativo de inferências sobre β e cálculo de $l(\beta)$ até um valor admitido como máximo da função de verossimilhança.

Se P_{iq} é calculado como um modelo Logit (Binomial ou Multinomial), a função $l(\beta)$ é bem definida e esse processo iterativo pode ser facilmente implementado computacionalmente. Existem softwares capazes de estimar parâmetros de funções utilidade pelo modelo Logit como Biogeme (Python), Apollo (Software R) e SPSS.

Por exemplo, considere três cenários de uma pesquisa para escolha de dois modos de transporte (1 e 2), cada um caracterizado por um respectivo atributo apresentado ao indivíduo q, dados por x_{1q} e x_{2q}.

Considere ainda a proposição de parcelas determinísticas de funções utilidade com apenas um parâmetro β comum a todos os atributos:

$$V_{1q} = \beta \cdot x_{1q}$$

$$V_{2q} = \beta \cdot x_{2q}$$

Sob essas condições, a probabilidade de escolha das alternativas 1 e 2 por um modelo Logit Binomial é dada por:

$$P_{1q} = \frac{1}{1 + e^{V_{2q} - V_{1q}}} = \frac{1}{1 + e^{\beta(x_{2q} - x_{1q})}} \quad \text{e} \quad P_{2q} = 1 - P_{1q} = \frac{1}{1 + e^{\beta(x_{1q} - x_{2q})}}$$

Após coletar a escolha modal de três indivíduos em uma pesquisa, obteve-se a seguinte tabela paras diferentes níveis de atributos:

Observação (q)	x_{1q}	x_{2q}	**Escolha Modal**
1	5	3	1
2	1	2	1
3	3	4	2

Nesse caso, os parâmetros g_{iq} podem ser caracterizados por:

	Alternativa	
Indivíduo (q)	$i =$**1**	$i =$**2**
1	1	0
2	1	0
3	0	1

A função de verossimilhança linearizada para um modelo Logit Binomial de estimativa das probabilidades de escolha desses modos de transporte, para cálculo do parâmetro $\hat{\beta}$ que mais aproxima essas probabilidades às escolhas dos três cenários da pesquisa, é dado por:

$$\begin{aligned} l(\beta) = {} & g_{11} \cdot log(P_{11}) + g_{21} \cdot log(P_{21}) + g_{12} \cdot log(P_{12}) \\ & + g_{22} \cdot log(P_{22}) + g_{13} \cdot log(P_{13}) + g_{23} \cdot log(P_{23}) \end{aligned}$$

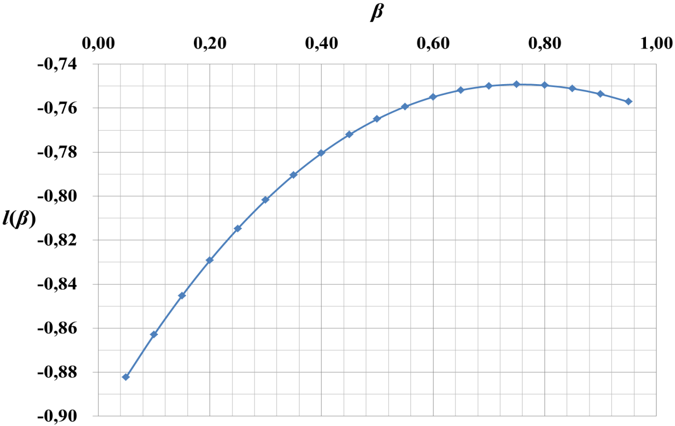
Substituindo os valores de g_{iq} na equação tem-se que:

$$\begin{aligned} l(\beta) &= log(P_{11}) + log(P_{12}) + log(P_{23}) \\ &= log\left[\frac{1}{1+e^{\beta(x_{21}-x_{11})}}\right] + log\left[\frac{1}{1+e^{\beta(x_{22}-x_{12})}}\right] + log\left[\frac{1}{1+e^{\beta(x_{13}-x_{23})}}\right] \\ &= log\left[\frac{1}{1+e^{\beta(3-5)}}\right] + log\left[\frac{1}{1+e^{\beta(2-1)}}\right] + log\left[\frac{1}{1+e^{\beta(3-4)}}\right] \end{aligned}$$

Simplificando a equação de $l(\beta)$ em função de β tem-se:

$$l(\beta) = 10 \cdot \beta - log\left(e^{5\cdot\beta} + e^{3\cdot\beta}\right) - log\left(e^{\beta} + e^{2\cdot\beta}\right) - log\left(e^{3\cdot\beta} + e^{4\cdot\beta}\right)$$

A função de verossimilhança pode ser representada graficamente para identificação do valor de β que maximiza $l(\beta)$.



O estimador de verossimilhança é $\hat{\beta} = 0,756$.

TESTES ESTATÍSTICOS SOBRE PARÂMETROS E MODELOS

- **Teste de Significância dos Atributos**

Quando $l(\beta)$ é maximizada, um conjunto de parâmetros $\hat{\mathbf{B}} = (\hat{\beta}_1, \cdots, \hat{\beta}_k)$ é obtido, representados por funções de distribuição de probabilidade Normal com média $\hat{\beta}_k$ e desvio padrão s_{kk}.

Esse desvio padrão é proveniente de uma matriz de variância $\mathbf{S}$ cujos elementos são iguais ao valor negativo da matriz inversa de derivadas parciais de segunda ordem da função $l(\beta)$.

$$\mathbf{S}^2 = -\left[E\left(\frac{\partial^2 l(\mathbf{B})}{\partial \mathbf{B}^2}\right)\right]^{-1} = -\begin{bmatrix} \frac{\partial^2 l(\mathbf{B})}{\partial \beta_1^2} & \frac{\partial^2 l(\mathbf{B})}{\partial \beta_1 \, \partial \beta_2} & \cdots & \frac{\partial^2 l(\mathbf{B})}{\partial \beta_1 \, \partial \beta_n} \\ \frac{\partial^2 l(\mathbf{B})}{\partial \beta_2 \, \partial \beta_1} & \frac{\partial^2 l(\mathbf{B})}{\partial \beta_2^2} & \cdots & \frac{\partial^2 l(\mathbf{B})}{\partial \beta_2 \, \partial \beta_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 l(\mathbf{B})}{\partial \beta_n \, \partial \beta_1} & \frac{\partial^2 l(\mathbf{B})}{\partial \beta_n \, \partial \beta_2} & \cdots & \frac{\partial^2 l(\mathbf{B})}{\partial \beta_n^2} \end{bmatrix}^{-1}$$

- **Teste de Significância dos Atributos**

Essa variância como valor esperado das derivadas de segunda ordem de $l(\beta)$ em relação a cada β é um estimador não-viesado para grandes amostras (entre 500 e 1.000 observações).

Um teste de hipótese utilizado é a verificação da hipótese nula de que o parâmetro estimado $\hat{\beta}_k$ é estatisticamente igual a zero, tal que:

$$\begin{cases} H_0 : \hat{\beta}_k = 0 \\ H_1 : \hat{\beta}_k \neq 0 \end{cases}$$

- **Teste de Significância dos Atributos**

A estatística calculada para verificação da hipótese nula é:

$$t_k = \frac{\hat{\beta}_k}{s_{kk}}$$

Assim, valores suficientemente grandes de $|t_k|$ (em geral maiores que 1,96 para níveis de confiança de 95%) incorrem na rejeição hipótese nula $H_0 : \hat{\beta}_k = 0$ e, portanto, na não-rejeição da hipótese de que k-ésimo estimador da função utilidade afeta significativamente as escolhas dos modos de transporte.

Uma análise complementar pode ser realizada em termos de nível de significância do parâmetro estimado. Por exemplo, para nível de confiança de 95%, a hipótese nula $H_0 : \hat{\beta}_k = 0$ é rejeitada se o $p - value$ do $k-$ésimo atributo é menor que 5% ($p - value_k \leq 5\%$).

- **Teste de Significância sobre os Modelos**

Além do teste estatístico de significância sobre os coeficientes dos atributos das funções utilidade, existem outros testes que podem ser realizados para análise da qualidade de um modelo quanto à parcela determinística da função utilidade, e para comparação de diferentes modelos possíveis.

  - Teste de Generalidade dos Parâmetros
  - Teste de Ajuste Geral
  - Índice ρ^2

- **Teste de Generalidade dos Parâmetros**

Considere um modelo com duas alternativas i e j representadas respectivamente por parcelas determinísticas V_{iq} e V_{jq}, cada uma com apenas um atributo e respectivo parâmetro β_i e β_j.

$$V_{iq} = \beta_i \cdot x_{iq}$$

$$V_{jq} = \beta_j \cdot x_{jq}$$

O teste verifica se ambos os parâmetros são iguais estatisticamente, ou seja, se o modelo pode ser genérico em termos dos parâmetros associados aos seus atributos.

$$\begin{cases} H_0 : \beta_i = \beta_j \\ H_1 : \beta_i \neq \beta_j \end{cases}$$

- **Teste de Generalidade dos Parâmetros**

O teste de hipótese é realizado mediante o cálculo da estatística:

$$t = \frac{\hat{\beta}_i - \hat{\beta}_j}{\sqrt{se_i^2 + se_j^2 + 2 \cdot \rho_{ij} \cdot se_i \cdot se_j}}$$

onde $\hat{\beta}_i$ e $\hat{\beta}_j$ são os respectivos valores estimados para os parâmetros, se_i e se_j são os respectivos erros padrão e ρ é o coeficiente de correlação entre eles.

$$\rho_{ij} = \frac{\text{Covariância}_{ij}}{\sqrt{\text{Variância}_i} \cdot \sqrt{\text{Variância}_j}} = \frac{\sum_k (x_{ik} - \bar{x}_i) \cdot (x_{jk} - \bar{x}_j)}{\sqrt{\sum_k (x_{ik} - \bar{x}_i)^2} \cdot \sqrt{\sum_k (x_{jk} - \bar{x}_j)^2}}$$

$$se(\hat{\beta}_i) = \frac{s_{ii}}{\sqrt{n}} \qquad se(\hat{\beta}_j) = \frac{s_{jj}}{\sqrt{n}}$$

- **Teste de Generalidade dos Parâmetros**

Se o valor de t é maior que um valor crítico - por exemplo, 1,96 ao nível de confiança de 95% - a hipótese nula $H_0 : \beta_i = \beta_j$ é rejeitada e os coeficientes são admitidos estatisticamente diferentes entre si.

Portanto, se $t > t_{crit,95\%} = 1,96$, ou se $p - value < 5\%$, então $\beta_i \neq \beta_j$ e os parâmetros β_i e β_j não devem ser generalizados para um único parâmetro β.

Em geral os softwares apresentam resultados da estatística t para todos os pares de parâmetros considerados no modelo, comparando-os em termos possibilidade de generalização.

- **Teste de Ajuste Geral**

Considere um modelo com duas alternativas de escolha modal, cada uma representada por parcelas determinísticas V_{iq} e V_{jq}. Além disso, cada alternativa tem uma utilidade intrínseca associada a ela que independe de qualquer atributo ou característica do indivíduo que está sujeito à sua escolha.

Essa utilidade intrínseca é caracterizada por uma "Constante Específica da Alternativa" (*ASC - Alternative Specific Constant*), tal que as utilidades podem ser caracterizadas por:

$$V_{iq} = ASC_i$$

$$V_{jq} = ASC_j$$

- **Teste de Ajuste Geral**

Devido às características do modelo Logit, só é possível estimar a diferença entre contantes $(ASC_j - ASC_i)$, ou admitir que a constante de uma $j-$ésima função utilidade seja nula $(ASC_j = 0)$.

Nesse caso, os parâmetros constantes são sempre comparados com o valor nulo daquela alternativa que não possui ASC associado.

Para o exemplo anterior, as funções utilidade passam a ser $V_{iq} = ASC$ e $V_{jq} = 0$, e a função de verossimilhança pode ser escrita por:

$$l(ASC) = \frac{1}{1 + e^{-ASC}} + \frac{1}{1 + e^{-ASC}} + \frac{1}{1 + e^{ASC}}$$

- **Teste de Ajuste Geral**

O **teste de ajuste geral** de significância dos atributos é dado por:

$$\begin{cases} H_0 : \mathbf{B} = 0 \;\forall\; A_j \\ H_1 : \beta_k \neq 0 \text{ para pelo menos um } k \end{cases}$$

A hipótese nula é testada pelo cálculo de uma razão de verossimilhança (*Likelihood Ratio - LR*):

$$LR = -2 \cdot [l(ASC) - l(\mathbf{B})]$$

onde $l(ASC)$ é o valor da função de verossimilhança linearizada considerando somente as constantes específicas das alternativas e $l(\mathbf{B})$ é o valor da função de verossimilhança linearizada com parâmetros $\mathbf{B} = (\beta_1, \cdots, \beta_n)$ não nulos.

- **Teste de Ajuste Geral**

Para modelos em que as mesmas alternativas estão disponíveis para todos os participantes da pesquisa tem-se que:

$$l(ASC) = \sum_i Q_i \cdot log\left(\frac{Q_i}{Q}\right)$$

onde Q_i e o número de indivíduos que escolheu a alternativa i e Q é a amostra de participantes consultados na pesquisa.

A estatística LR é comparada com uma distribuição $\chi^2_{k-c;1-\alpha}$ com $k-c$ graus de liberdade e $(1-\alpha)$ nível de confiança, onde k é o número de parâmetros estimados (incluindo constantes ASC), c é o número de constantes ASC e p geralmente é igual a 95%.

Se o valor de LR é maior que o valor crítico $\chi^2_{k-c;1-\alpha}$ então a hipótese nula é rejeitada, sendo possível admitir que os parâmetros associados aos atributos das funções utilidade são não nulos.

- **Índice** ρ^2

O índice ρ^2 permite comparar a qualidade do modelo (aderência) em relação às respostas observadas na pesquisa, tal que $\rho^2 = 0$ caracteriza um modelo sem aderência e $\rho^2 = 1,0$ indica representação total das respostas pelo modelo.

$$\rho^2 = 1 - \frac{l(\mathbf{B})}{l(0)}$$

onde $l(0)$ é o valor da função de verossimilhança linearizada quando todos os parâmetros são nulos.

- **Índice** ρ^2_{ajust}

Ben-Akiva e Lerman (1985) propuseram um ajuste à equação anterior para considerar a quantidade de parâmetros do modelo, resultando em um parâmetro ρ^2_{ajust}.

$$\rho^2_{ajust} = 1 - \frac{l(\mathbf{B}) - k}{l(0)}$$

onde k é o número de parâmetros estimados (incluindo constantes específicas das alternativas ASC).

Ben-Akiva, M.E. e Lerman, S.R. (1985) Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press, Cambridge, Mass.

Referências

Aula preparada com apoio de:

Ortúzar, J. D., & Willumsen, L. G. (2011). Modelling transport. John wiley & sons.

Ben-Akiva, M. E., Lerman, S. R., & Lerman, S. R. (1985). Discrete choice analysis: theory and application to travel demand (Vol. 9). MIT press.