

# Seminario Modelli di Scelta Discreta



Università degli Studi di Cagliari  
Facoltà di Ingegneria

Dott. Ing. Eleonora Sottile, PhD

Cagliari, 4 maggio 2018

# Modelli

**Rappresentazione semplificata di una parte del mondo reale** (sistema di interesse) attraverso la definizione di elementi considerati importanti per la sua analisi e secondo un particolare punto di vista:

- Modelli fisici (usati in architettura, nella meccanica dei fluidi, *etc.*)
- Modelli astratti (mentali, matematici, *etc.*)

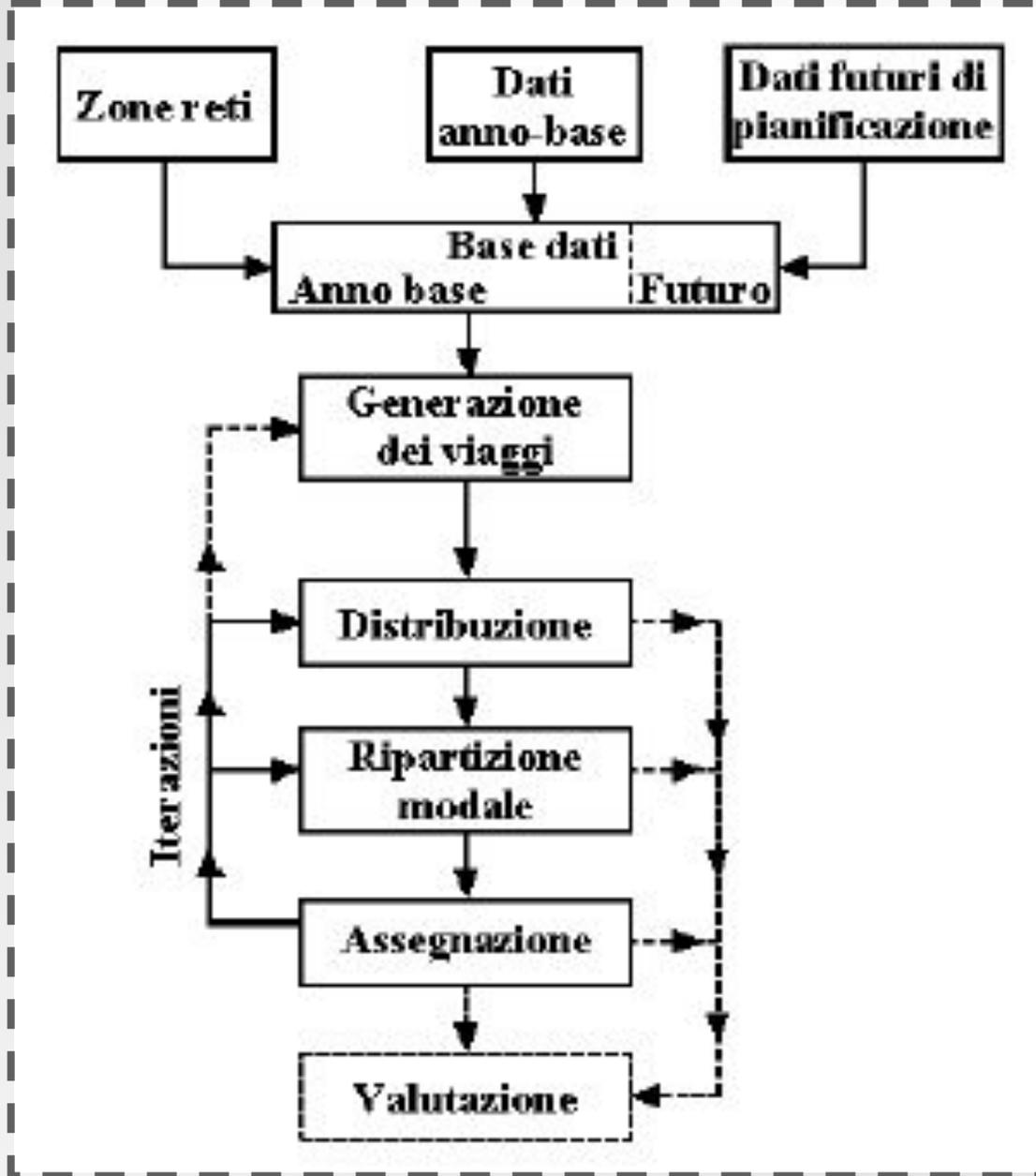
## **Modelli matematici**

- ✓ tentano di replicare il sistema di interesse per mezzo di equazioni matematiche basate su certe ipotesi teoriche
- ✓ hanno validità solo per un certo insieme di problemi sotto specifiche condizioni
- ✓ l'abilità di scegliere e adattare i modelli ai particolari contesti rappresenta uno dei più importanti elementi nel complessivo bagaglio tecnico di un pianificatore

# Approccio modellistico

- contesto decisionale: comporta l'adozione di una particolare *prospettiva* e la scelta di uno *scopo* o di un particolare aspetto del sistema in esame
- il dettaglio richiesto
- la disponibilità di dati appropriati
- lo stato dell'arte della modellizzazione
- le risorse disponibili per lo studio: in particolare il tempo e il livello di comunicazione con il decisore politico e con la collettività
- requisiti dell'elaborazione dei dati
- livelli di preparazione e capacità dell'analista

# Modelli di domanda: il modello a 4 stadi



Sequenza di 4 sub-modelli:

1. Generazione
2. Distribuzione
3. Ripartizione modale
4. Assegnazione

# Modelli di simulazione della domanda di trasporto

**Definizione:** relazione matematica che consente di associare ad un dato sistema di attività e di offerta di trasporto il valore medio del flusso di domanda in un determinato periodo di riferimento con le sue caratteristiche rilevanti

**Obiettivo:** individuare le variabili, interne ed esterne al sistema, che influiscono maggiormente nella determinazione della domanda di trasporto

1. dipendenti o endogene (per esempio il numero di spostamenti)
2. indipendenti o esogene:
  - a. **variabili politiche**, attributi del Livello di Servizio (LdS): tempi e costi
  - b. attributi socio-demografici del contesto e dell'individuo (*popolazione, numero di famiglie, età, genere, possesso d'auto, etc.*)

# Dati

Per costruire un modello si ha quindi bisogno di poter disporre di dati, di alta qualità, su tutte le variabili che si intende valutare (dipendenti ed indipendenti).

La tipologia del dato da reperire varia in funzione del tipo di modello che si intende costruire:

- **Modelli aggregati** o di prima generazione: mirano a rappresentare il comportamento di più di un individuo
- **Modelli disaggregati** o di seconda generazione: cercano di rappresentare il comportamento individuale, quindi è necessario disporre di un'informazione per ogni singolo individuo

# Raccolta dei dati

- ◎ **Indagini qualitative di gruppo**, mirano alla conoscenza del fenomeno e/o all'analisi dettagliata di alcuni aspetti particolari
  - > Focus Group (FG)
  - > Metaplan (MP)
  
- ◎ **Indagini quantitative individuali**
  - > Preferenze Rivelate (PR), consentono la descrizione delle scelte e dei comportamenti di viaggio nella situazione ATTUALE
  - > Preferenze Dichiarate (PD), consentono la descrizione delle scelte e dei comportamenti di viaggio in situazioni IPOTETICHE

# Costruzione del modello

Il modello è una funzione matematica di variabili  $X$  e di parametri  $\theta$

$$Y = f(X, \theta)$$

- 1. Specificazione:** struttura, forma funzionale, specificazione variabili
- 2. Stima:** determinare i valori dei parametri  $\theta$  che riproducono più probabilmente i dati osservati
- 3. Calibrazione:** ricercare i valori dei parametri  $\theta$  in modo che il fenomeno riprodotto con il modello sia il più vicino possibile al fenomeno osservato
- 4. Validazione:** verificare quanto il fenomeno riprodotto con il modello si avvicina al fenomeno osservato (confronto del modello con informazioni non utilizzate nel processo di stima)

La costruzione del modello è un processo di

**“trial and error”**

# Modelli di scelta discreta – scelta modale

- sono modelli di domanda di seconda generazione/disaggregati
- modellizzano scelte discrete, l'individuo sceglie tra un numero finito di alternative
- in generale affermano che:  
*“la probabilità che gli individui scelgano una determinata alternativa è funzione delle loro caratteristiche socioeconomiche e della relativa attrattività (utilità) dell'alternativa”*

# Teoria comportamentale

L'individuo  $q \in Q$  è un *homo economicus*, pertanto:

- conosce tutte le alternative disponibili nel suo insieme di scelta  $A(q) \in A$
- valuta ogni alternativa  $j \in A(q)$  sulla base delle caratteristiche  $X(j)$
- associa ad ogni alternativa un livello di soddisfazione, misurato attraverso l'indice dell'utilità
- confronta le alternative sulla base del livello di soddisfazione ricevuto e sceglie sempre l'alternativa più attrattiva (cioè quella che dà la maggiore soddisfazione) soggetta ai vincoli ambientali

# Teoria comportamentale

## ➤ Punto di vista dell'individuo:

- L'individuo ha una conoscenza perfetta e sceglie sempre l'alternativa con la massima utilità (teoria perfettamente comportamentale)
- Le preferenze degli individui non sono sempre consistenti e transitive (non sempre si osservano comportamenti razionali)

## ➤ Punto di vista del modellizzatore:

- Il modellizzatore non possiede informazioni perfette, pertanto assume che l'utilità  $U_{jq}$  sia definita da 1) una componente sistematica (funzione degli attributi misurati) e da 2) una parte casuale che contiene gli errori commessi dallo stesso modellizzatore e gli aspetti latenti che sottendono la scelta (inerzia, abitudine, avversione, *etc.*)

# Derivazione econometrica

L'individuo  $q$  sceglie l'alternativa  $j$ , se e solo se:

$$U_{jq} \geq U_{iq}, \quad \forall i \neq j, (i, j) \in A(q)$$

Poiché il modellizzatore tratta l'utilità come una variabile casuale, non può conoscere con certezza quale alternativa ha la massima utilità, può solo calcolare la probabilità che l'utilità dell'alternativa  $j$  sia maggiore o uguale dell'utilità delle altre alternative disponibili:

$$P_{jq} = \text{Prob} \left\{ U_{jq} \geq U_{iq}, \quad \forall i \neq j, (i, j) \in A(q) \right\}$$

# Derivazione econometrica

Il modellizzatore pur trattando l'utilità  $U_{jq}$  come una variabile casuale, possiede informazioni sull'utilità associata alle alternative, pertanto scompone l'utilità nel seguente modo:

$$U_{jq} = V_{jq} + \varepsilon_{jq}$$

$$P_{jq} = \text{Prob} \left\{ \varepsilon_{iq} \leq \varepsilon_{jq} + V_{jq} - V_{iq}, \quad \forall i \neq j, (i, j) \in A(q) \right\}$$

$$P_{jq} = \int_{R_N} f(\varepsilon) d\varepsilon$$

$$R_N = \begin{cases} \varepsilon_{iq} \leq \varepsilon_{jq} + (V_{jq} - V_{iq}), & \forall A_i \in \mathbf{A}(q) \\ V_{jq} + \varepsilon_{jq} \geq 0 \end{cases}$$

# Il modello logit multinomiale (MNL)

É il modello più popolare e più semplice e può essere generato assumendo che i residui casuali siano **Identicamente Indipendentemente Distribuiti** (IID) (1- le variabili hanno tutte la stessa distribuzione di probabilità, 2- le variabili sono tutte statisticamente indipendenti)

**Gumbel** (EV1-*Extreme Value type 1*):

$$P_{jq} = \frac{\exp(\lambda V_{jq})}{\sum_{i \in A(q)} \exp(\lambda V_{iq})}$$

$\lambda$  è il parametro di scala  $> 0$

Come tutti i modelli di scelta discreta, il MNL richiede una normalizzazione perché il parametro di scala non può essere identificato (non può essere stimato separatamente dai parametri  $\theta$ ), quindi viene assunto pari ad 1.

**NON** possono essere confrontati direttamente i parametri di due modelli diversi (si usa il valore del tempo):

$$\lambda_{MNL} = \frac{\pi}{\sigma \sqrt{6}} \quad \lambda_{MNP} = \frac{1}{\sigma \sqrt{2}}$$

# Proprietà e limiti del MNL

Il MNL **non** consente:

- ⊙ correlazione tra le alternative (indipendenza dalle alternative irrilevanti (IIA))
- ⊙ variazione casuale nelle preferenze (non tiene conto di osservazioni ripetute per uno stesso individuo)
- ⊙ eteroschedasticità
  - > Tra osservazioni:
    - dati diversi (RP, SP)
    - differenza nella percezione degli attributi
  - > Tra le alternative:
    - gli attributi dell'alternativa scelta sono percepiti in modo più preciso di quelli delle alternative non scelte

# Specificazione dell'utilità

Definizione della forma matematica:

- > sufficientemente esplicitiva
- > trattabile matematicamente

La più usata è Lineare nei parametri e negli attributi (LPLA):

$$V_{jq} = \sum_k \theta_{jk} X_{jqk} \quad \theta \text{ costante tra gli individui}$$

Le variabili esplicative possono essere:

- ⊙ generiche: il parametro associato è uguale in tutte le alternative
- ⊙ specifiche: la variabile è presente solo in un'alternativa oppure la stessa variabile ha parametri diversi tra le diverse alternative
- ⊙ costanti specifiche dell'alternativa: possono essere inserite in n-1 alternative

# Stima dei parametri

Metodo della massima verosimiglianza:

$$L = \prod_{q \in Q} \prod_{j \in A(q)} P_{jq}^{d_{jq}}, \quad d_{jq} = \begin{cases} 1 & \text{se } j \text{ è scelta} \\ 0 & \text{altrimenti} \end{cases}$$

La funzione di verosimiglianza  $L$  è il prodotto tra tutti gli individui e tutte le alternative della probabilità che l'individuo scelga l'alternativa che sceglie nella realtà:

- è una funzione concava
- nel caso di utilità lineare nei parametri ha solo un massimo
- $L \in ]0; -\infty[$

$$l = \log L = \sum_{q \in Q} \sum_{j \in A(q)} d_{jq} \log P_{jq}$$

# Test di verifica

- **Il test del t asintotico:** consente di testare singolarmente i parametri del modello
- **Test del rapporto di verosimiglianza (LR):** consente di effettuare un test congiunto su tutti i parametri del modello, fornendo quindi un confronto della funzione di verosimiglianza nel punto di massimo ed il corrispondente valore nell'ipotesi di alcune restrizioni lineari imposte sui parametri del modello
- **L'indice  $\rho^2$ :** consente di confrontare modelli diversi, è un indice compreso tra 0 e 1

# Test del t asintotico

Consente di verificare se ogni singolo parametro  $\theta_k^*$  stimato differisce significativamente da un dato valore di riferimento  $\theta_k$  (generalmente 0):

$$t = \frac{\theta_k^* - \theta_k}{\sqrt{\text{var } \theta^*}} \sim N(0,1)$$

Ricorda:

- i parametri stimati con il metodo della massima verosimiglianza sono asintoticamente normali
- se  $|t| \geq 1.96$ , la probabilità che il parametro differisca dal valore di riferimento è maggiore o uguale al 95% altrimenti la devo calcolare

# Test del rapporto di verosimiglianza: Likelihood ratio test

$$LR = -2 \left\{ l^* (\theta_{res}) - l^* (\theta) \right\} \sim \chi_r^2 \quad r = \text{gradi di libertà (numero restrizioni lineari)}$$

È il valore della funzione di log-verosimiglianza del modello (in studio) calcolato nel punto di ottimo.

È il valore della funzione di log-verosimiglianza di un modello di confronto, costruito imponendo ai parametri del modello da studiare un numero  $r$  di restrizioni lineari

Se  $LR \leq \chi_{r,\alpha}^2$  allora il modello ristretto è preferibile a quello non ristretto

# Test del rapporto di verosimiglianza: Likelihood ratio test

Test di adattamento globale:

1. restrizioni: tutti i parametri uguali a zero  $\rightarrow$  modello equiprobabile
2. restrizioni: tutte le variabili tranne le costanti uguali a zero  $\rightarrow$  Market Share (riproduce la quota di mercato)

# L'indice $\rho^2$

$$\rho^2 = 1 - \frac{l^*(\theta)}{l^*(0)} \begin{cases} 0, & \text{se } l^*(\theta) = l^*(0) \\ 1, & \text{se } l^*(\theta) = 0 \end{cases}$$

Questo indice è costruito come l'indice di determinazione delle regressioni lineari multiple e può essere usato nello stesso modo perché dipende dalla % di individui che sceglie ciascuna alternativa

Pertanto è preferibile usare:

$$\rho_c^2 = 1 - \frac{l^*(\theta)}{l^*(C)}$$

# Analisi microeconomica: segno delle variabili

		Variabili	
		Policy	Other
Segno Corretto	Significativa	<i>Includere</i>	<i>Includere</i>
	Non significativa	<i>Includere</i>	<i>Può essere eliminata</i>
Segno errato	Significativa	<i>Problemi Grossi</i>	<i>Eliminare</i>
	Non significativa	<i>Problemi</i>	<i>Eliminare</i>

# Analisi microeconomica: utilità marginale

$$V_{qj} = \theta_1 T_{qj} + \theta_2 C_{qj} + \theta_3 et\grave{a} + ASC$$

$$MUT_j = \frac{\partial V_j}{\partial T_j} = \theta_1 < 0 \quad \forall q \in Q$$

$$MUC_j = \frac{\partial V_j}{\partial C_j} = \theta_2 < 0 \quad \forall q \in Q$$

Nelle specificazioni LALP, l'UM:

- coincide con il parametro associato alla variabile rispetto alla quale si calcola
- non varia in funzione delle caratteristiche del LDS (indipendentemente dalla durata complessiva del viaggio, 1 minuto addizionale è percepito allo stesso modo da tutti)

# Analisi microeconomica: valore soggettivo del tempo VST

Disponibilità a pagare per migliorare le caratteristiche del viaggio (WTP)

$$\text{VST}_j = \frac{\partial C_j}{\partial T_j} \bigg|_V = \frac{\frac{\partial V_j}{\partial T_j}}{\frac{\partial V_j}{\partial C_j}} > 0$$

# Software: Biogeme

← → ↻ ↗ bioeme.epfl.ch/home.html

App M scholar - Cerca con Google Scopus - Author search InCites™ Gmail sia per - Traduzione in Homepage Home Elsevier Editorial System U-GOV L'Unione Sarda Voli Low Cost, Voli Economici Rejseplanen traffiCA.it

Home Start Help Install Videos Documents Courses Utilities Archives Examples



# BIOGEME

## BIOGEME

Biogeme is an open source freeware designed for the maximum likelihood estimation of parametric models in general, with a special emphasis on discrete choice models. Two versions of the software are available.

### Pythonbiogeme

is designed for general purpose parametric models. The specification of the model and of the likelihood function is based on an extension of the python programming language. A series of discrete choice models are precoded for an easy use.

### Bisonbiogeme

is designed to estimate the parameters of a list of predetermined discrete choice models such as logit, binary probit, nested logit, cross-nested logit, multivariate extreme value models, discrete and continuous mixtures of multivariate extreme value models, models with nonlinear utility functions, models designed for panel data, and heteroscedastic models. It is based on a formal and simple language for model specification.

# Libri di testo

- Cherchi, E. (2003). Il valore del tempo nella valutazione dei sistemi di trasporti: teoria e pratica.
- De Dios Ortuzar, J., & Willumsen, L. G. (2006). Pianificazione dei Sistemi di Trasporto.
- Hensher, D. A., Rose, J. M., & Greene, W. H. (2005). *Applied choice analysis: a primer*. Cambridge University Press.
- Marcucci, E. (2011). *Scelte di trasporto e modelli a scelta discreta*. Franco Angeli.
- Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.



**Fine**