离散选择模型

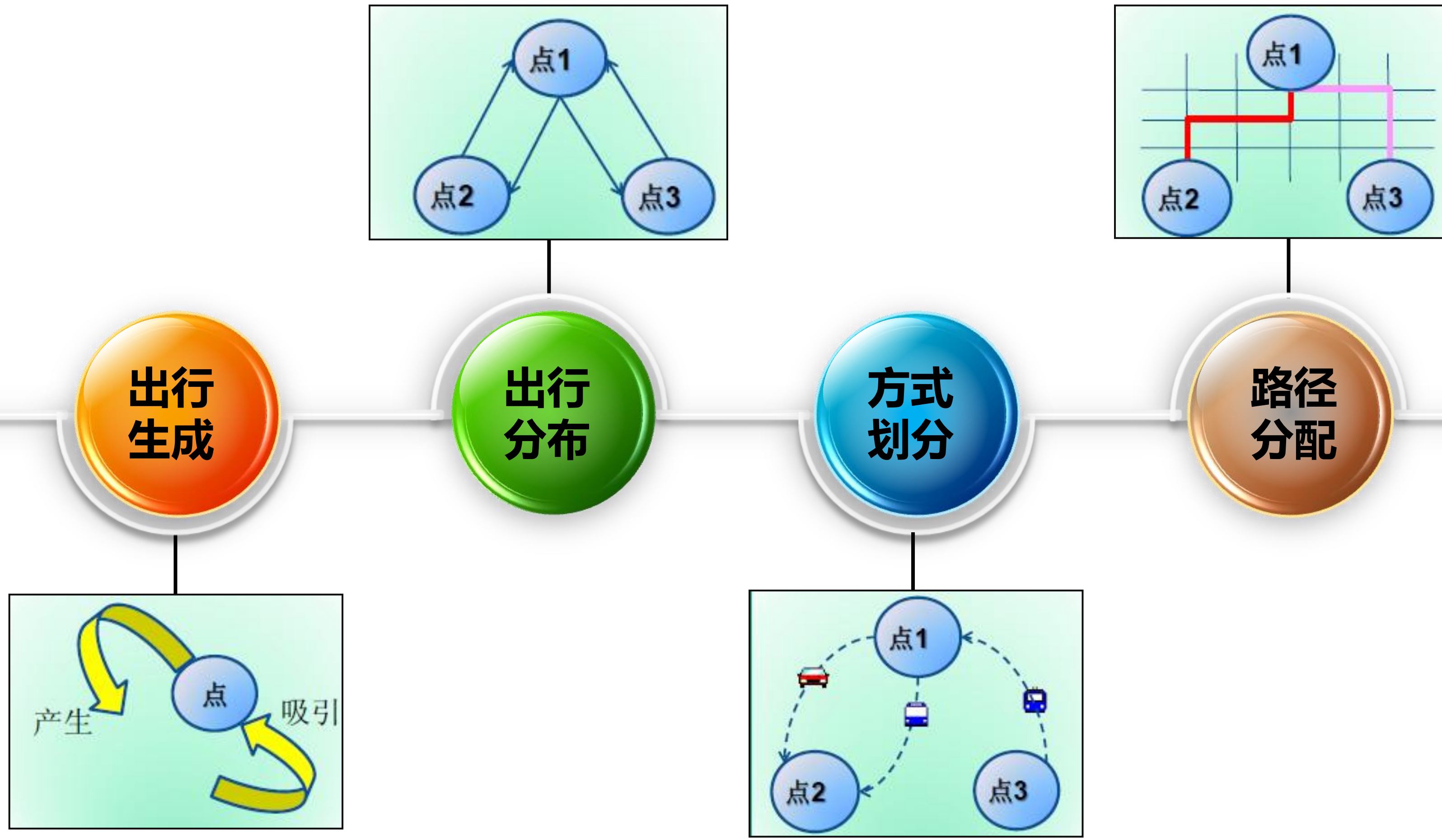
城市分析方法系列课程

苏州大学 王灿

大纲

- 离散选择模型简介
- Logit模型的核心原理
- 应用案例：交通方式选择
- 个体行为模拟初步

离散选择模型简介



离散选择模型简介

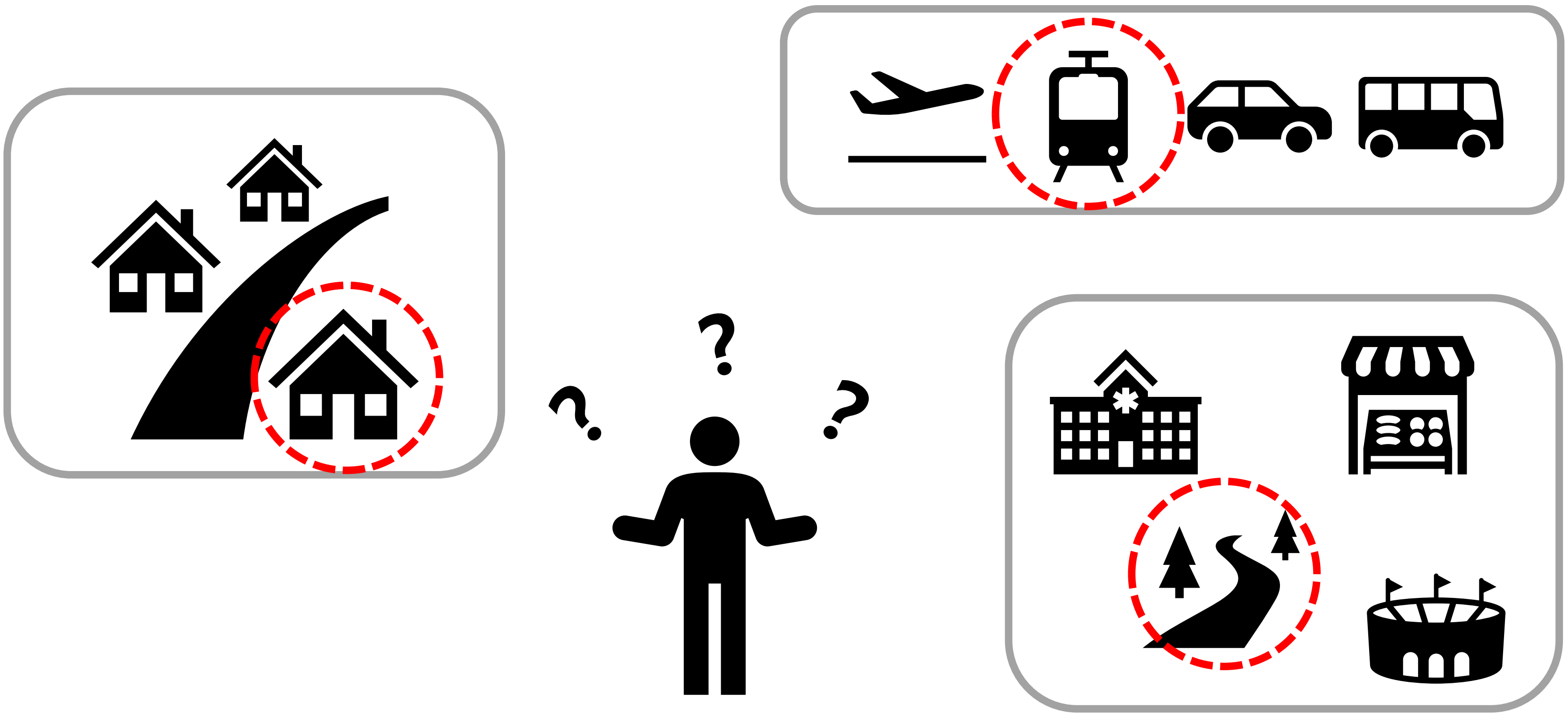
出行生成

路径分配

非集计模型：

对每位个体单独建模，而不是将他们汇总为群体，能够进行更精细的分析，捕捉个体差异。

离散选择模型简介



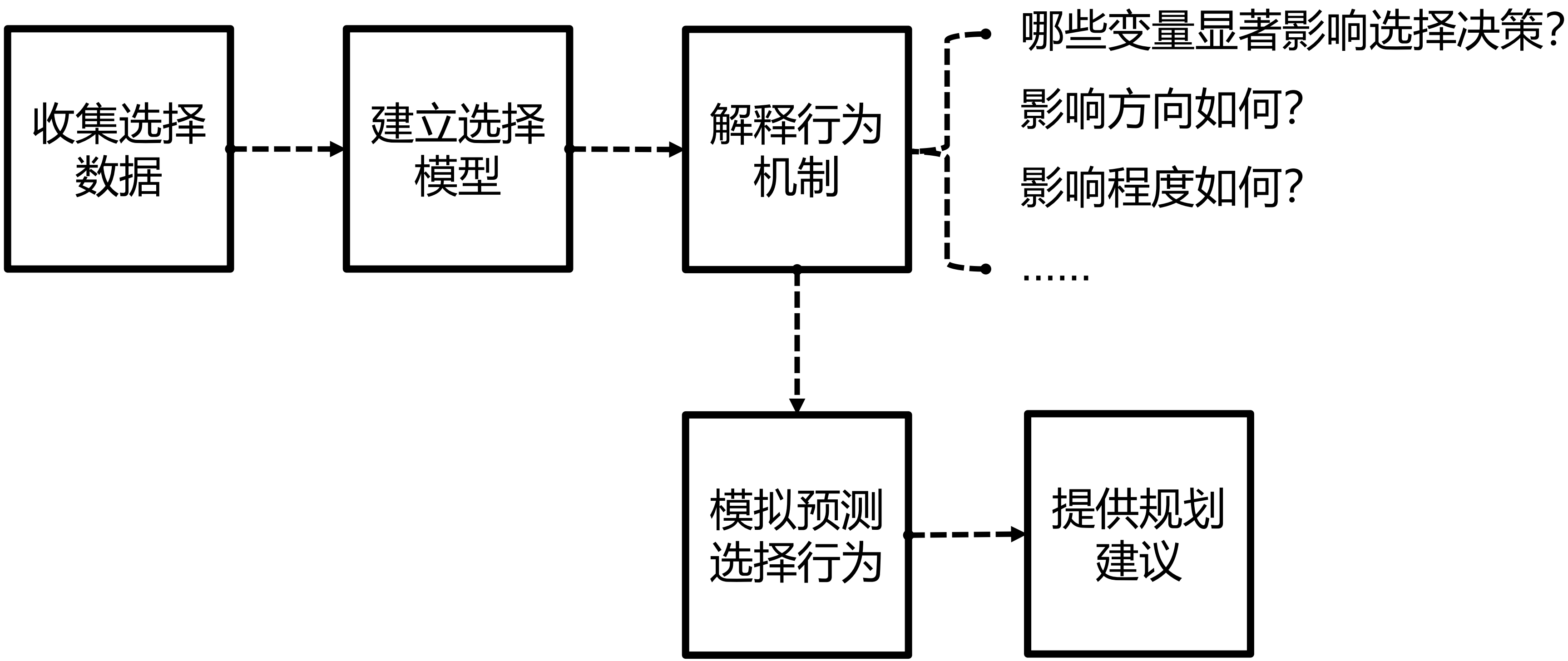
离散选择模型简介

行为模型：

Life is all about CHOICES!!!
研究个体如何在给定的有限个选项中，选择其中一个作为其最佳选择的决策过程。

离散选择模型简介

离散选择模型的分析流程



离散选择模型简介

分类 vs. 选择

分类

这块地会更新成什么功能?

这张图像里是猫还是狗?

解释变量：只有分类对象的属性

数据格式：一行是一个样本

选择

平江路、山塘街、观前街，我选择去哪里?

地铁、打车、自行车，我选择怎么去?

解释变量：选择者的属性 + **被选项的属性**

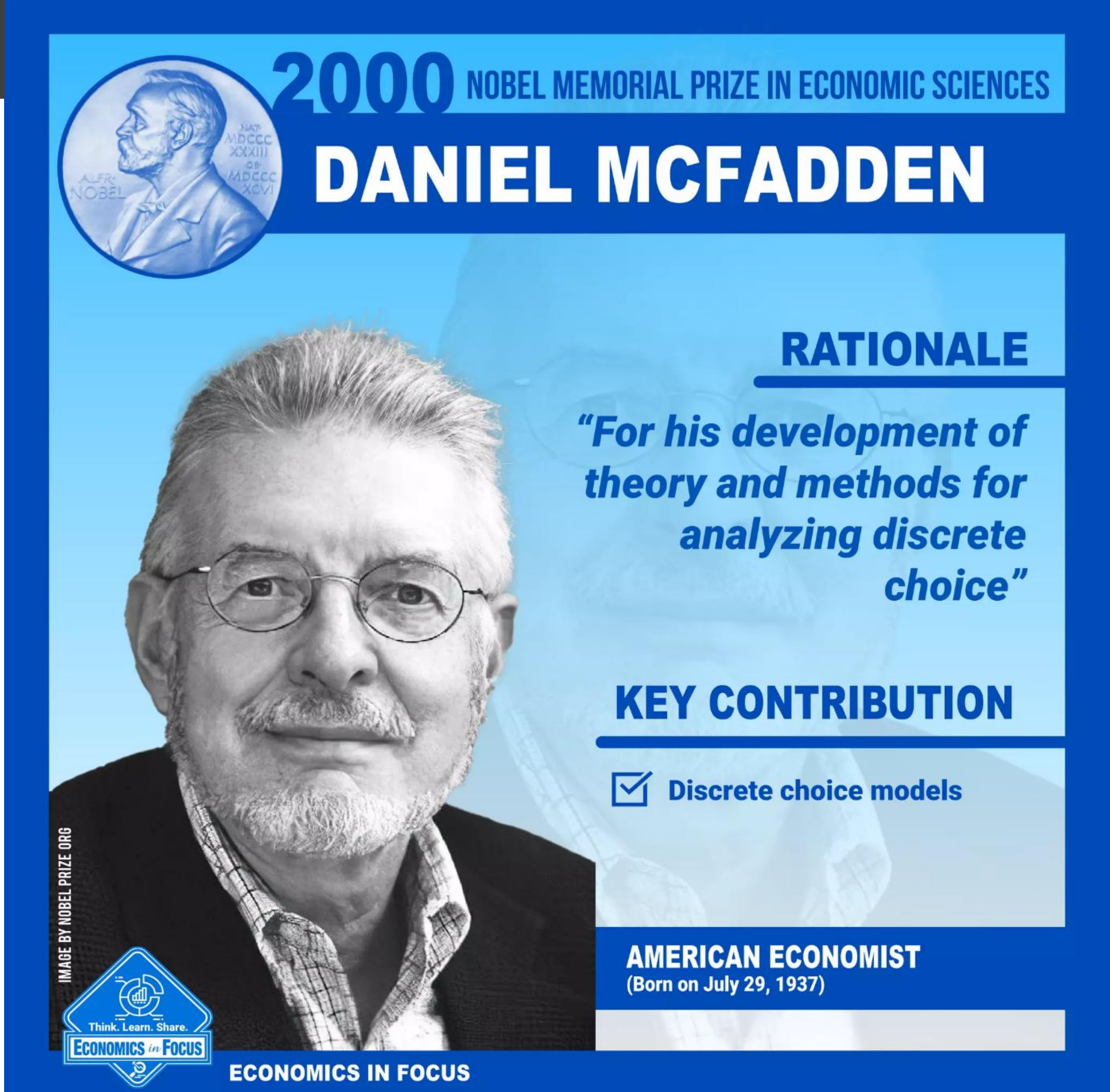
数据格式：**?**

Name (1)	NF (2)	LA (3)	FA (4)	IC (5)	Loc (6)	Dist (7)	Road (8)	SO (9)
××电子管厂	1	1.77	3.34	5.02	1	0.82	2	1
××治药厂	1	0.83	1.87	3.89	1	0.79	2	1
××印刷厂	1	1.22	1.83	4.09	3	0.64	2	1

离散选择模型简介

离散选择模型的意义

- 2000年诺贝尔经济学奖



离散选择模型的意义

- UrbanSim中的选择模型
  - 家庭：要不要迁居？搬到哪？
  - 就业：要不要空间迁移？迁移到哪？
  - 开发商：要不要开发？在哪开发？开发什么功能？

https://urbansim.com/urbansim

TABLE 1. Comparison of operational model characteristics.

	Operational urban model			
Characteristic	DRAM/EMPAL	MEPLAN and TRANUS	CUF-2	UrbanSim
Model structure	Spatial interaction	Spatial input-output	Discrete choice	Discrete choice
Household location choice	Modeled	Modeled	Not modeled	Modeled
Household classification	Aggregate, 8 categories	Aggregate, user-defined	Not represented	Disaggregate, income, persons, workers, child
Employment location choice	Modeled	Modeled	Not modeled	Modeled
Employment classification	Aggregate, 8 categories	Aggregate, user-defined	Not modeled	Disaggregate, 10–20 sectors

Waddell. UrbanSim: Modeling Urban Development for Land Use, Transportation and Environmental Planning (2002)

Table 1. Agents, choices and models in UrbanSim

Agent	Choice	Model
Household	In- and out-migration	Demographic transition model
Household	Residential moves	Household relocation model
Household	Residential location	Household location choice model
Person	Work at home	Work at home model
Person	Job choice	Workplace choice model
Business	Birth and death	Economic transition model
Business	Business relocation	Business relocation model
Business	Business location	Business location choice model
Developer	Parcel development	Real estate development model
Market	Real estate prices	Real estate price model

Waddell. Integrated Land Use and Transportation Planning and Modelling: Addressing Challenges in Research and Practice (2010)

离散选择模型简介


Census Tract #179
Cycling 13
Walking 58
and predict how the design
might impact mobility choices
in nearby neighborhoods...

离散选择模型简介

离散选择模型

多项Logit模型（Multinomial Logit Model，**MNL**）
或称**条件**Logit模型（Conditional Logit Model，**CL**）
或简称为Logit模型

嵌套Logit模型（Nested Logit Model，**NL**）

混合Logit模型（Mixed Logit Model，**MXL**）

潜在类别模型（Latent Class Model，**LCM**）

Probit模型（Probit Model）

较复杂

离散选择模型简介

软件工具

- 通用软件：Stata、R、Python、SAS 等
- 专用软件：NLogit 和 Biogeme

大纲

- 离散选择模型简介
- **Logit模型的核心原理**
- 应用案例：交通方式选择
- 个体模拟初步

Logit模型的基本原理

随机效用理论（Random Utility Theory）

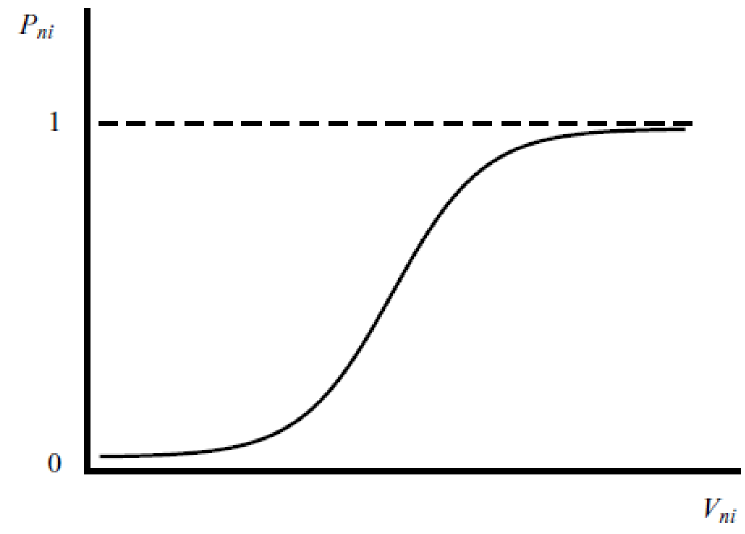
- 效用（Utility）：选择者可以从每个备选项获得一定的效用，反映了备选项满足选择者需求的程度。选择者永远选择效用最大的备选项。
- 如同回归分析一样，效用与解释变量的关系是线性的。
- 但因为误差的存在，我们无法绝对确定哪个备选项的效用最大，只能确定某个备选项效用最大的概率，即被选中的概率 。

$$U_A = \beta_{time} Dist_A + \beta_{price}\, Price_A + \beta_{size} Size_A + \ \dots + \boldsymbol{\varepsilon_A}$$

$$U_B = \underbrace{\beta_{move} + \beta_{time} Dist_B + \beta_{price}\, Price_B + \beta_{size} Size_B + \ \dots}_{\text{可见效用}(observable\ utility):V} + \boldsymbol{\varepsilon_B}$$

MNL模型

- 选择者永远选择效用最大的备选项。
- 但是，因为误差的存在，无法绝对确定哪个备选项的效用最大，只能某个备选项效用最大的概率，即被选中的概率 。
- 效用与选择概率正相关，关系非线性（S型）



$$Prob(A) = Prob(U_A > U_B) = \frac{\exp(V_A)}{\exp(V_A) + \exp(V_B)}$$

$$Prob(B) = Prob(U_B > U_A) = \frac{\exp(V_B)}{\exp(V_A) + \exp(V_B)}$$

Logit模型的基本原理

Logit模型

- 选择者可以从备选项中获得一定的效用。

$$
\begin{aligned}
U_A &= \phantom{\beta_{move} +{}} \beta_{time}\, Time_A + \beta_{price}\, Price_A + \beta_{size}\, Size_A + \ldots + \varepsilon_A \\
U_B &= \beta_{move} + \beta_{time}\, Time_B + \beta_{price}\, Price_B + \beta_{size}\, Size_B + \ldots + \varepsilon_B \\
U_C &= \beta_{move} + \beta_{time}\, Time_C + \beta_{price}\, Price_C + \beta_{size}\, Size_C + \ldots + \varepsilon_C \\
U_D &= \underbrace{\beta_{move} + \beta_{time}\, Time_D + \beta_{price}\, Price_D + \beta_{size}\, Size_D + \ldots}_{observable\ utility:V} + \varepsilon_D
\end{aligned}
$$

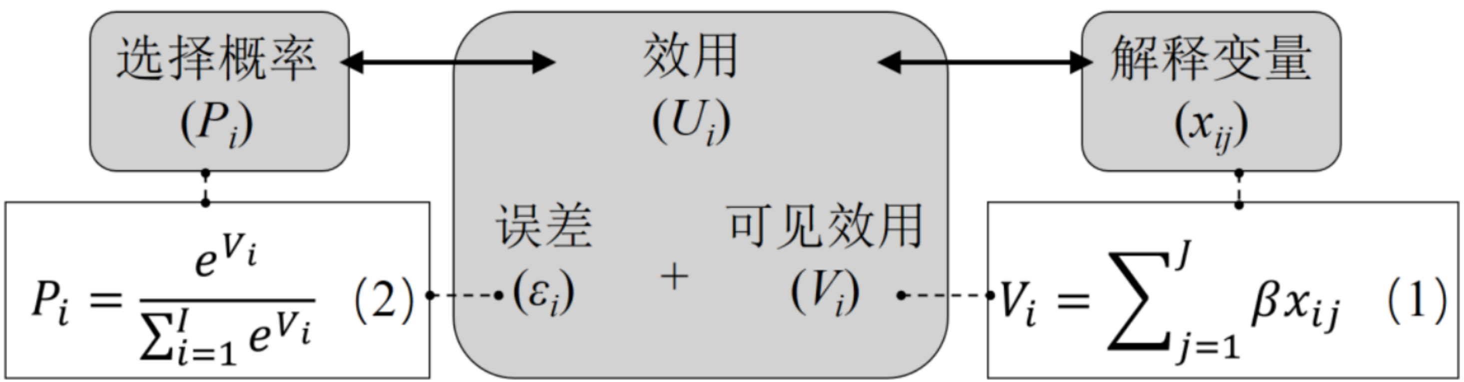
- 效用与选择概率正相关。

$$P_A = \frac{\exp(V_A)}{\exp(V_A) + \exp(V_B) + \exp(V_C) + \exp(V_D)} \qquad P_C = \frac{\exp(V_C)}{\exp(V_A) + \exp(V_B) + \exp(V_C) + \exp(V_D)}$$

$$P_B = \frac{\exp(V_B)}{\exp(V_A) + \exp(V_B) + \exp(V_C) + \exp(V_D)} \qquad P_D = \frac{\exp(V_D)}{\exp(V_A) + \exp(V_B) + \exp(V_C) + \exp(V_D)}$$

Logit模型的基本原理

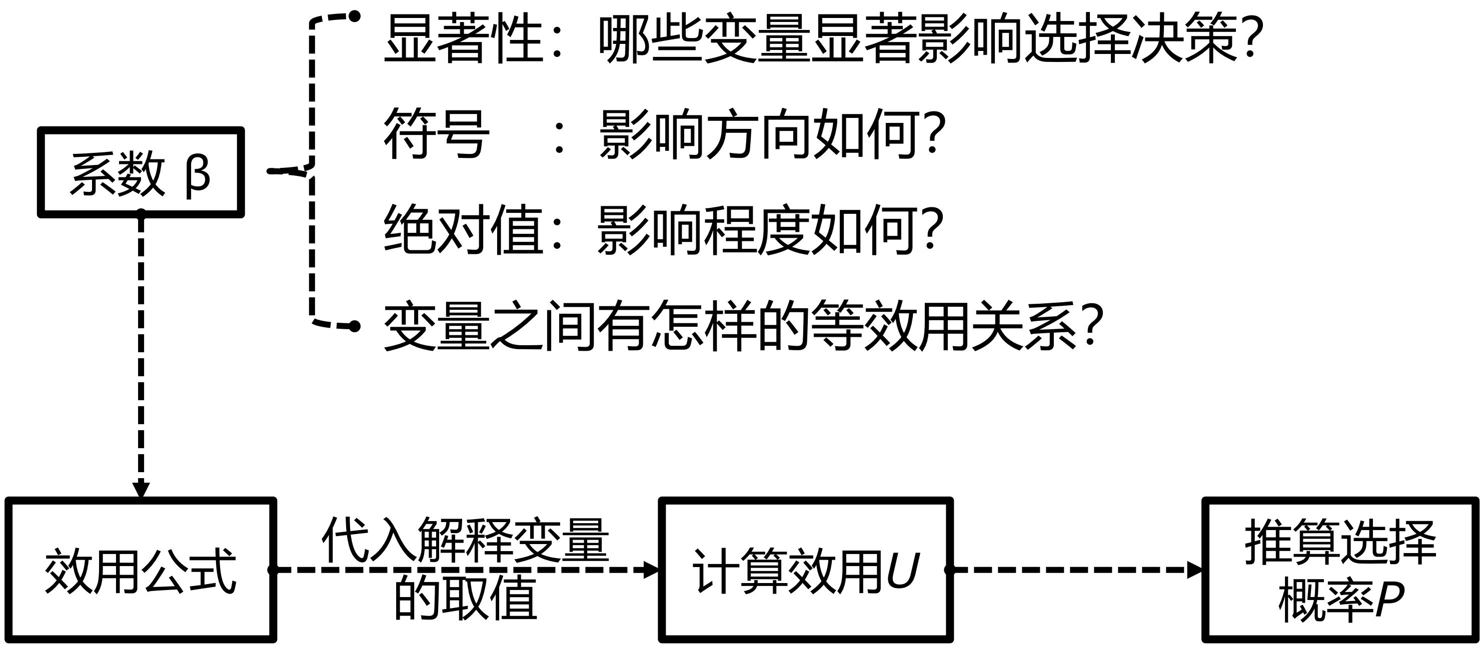
- 离散选择模型建立被解释变量——备选项的选择概率(probability, *P*)与相应解释变量(*x*)的定量关系。其中，备选项的效用(utility, *U*)扮演了中介的角色。
- 效用 *U* = 可见效用 *V* + 误差项 *ε*
- 可见效用***V***与解释变量***x***的关系是**线性的**，与一般的线性回归类似。
- 选择概率***P***与可见效用***V***之间关系是**非线性的**，其具体形式取决于误差项的分布。



备选项：i=1,……,I　　解释变量：j=1,……,J

Logit模型的基本原理

Logit 模型的主要估计结果是解释变量的**系数 β**



大纲

- 离散选择模型简介
- Logit模型的核心原理
- **应用案例：交通方式选择**
- 个体行为模拟初步

Logit模型案例：交通方式选择

数据：

收集了 210 位出行者在某次出行时对航空(Air)、火车(Train)、 巴士(Bus)、私家车(Car)这 4 种**交通方式**的选择结果，以及每种交通方式对应的**站点等候时间**、**行程时间**、**行程费用**。

问题：

1）出行者对交通方式的选择，受到了哪些要素的影响？

2）对于新的选择情景，如何预测出行者对各种交通方式的选择概率？

Logit模型案例：交通方式选择

样本编号 (1)	备选项 (2)	选择结果 (3)	TTME (4)	INVT (5)	INVC (6)
1	Air	0	69	100	59
1	Train	0	34	372	31
1	Bus	0	35	417	25
1	**Car**	**1**	**0**	**180**	**10**
2	Air	0	69	125	115
2	Train	0	34	892	98
2	Bus	0	35	882	53
2	**Car**	**1**	**0**	**720**	**23**
3	Air	0	69	100	59
3	**Train**	**1**	**40**	**345**	**20**
3	Bus	0	35	417	13
3	Car	0	0	284	12
4	**Air**	**1**	**45**	**115**	**148**
4	Train	0	34	945	111
4	Bus	0	35	935	66
4	Car	0	0	821	36
5	Air	0	69	129	103
5	Train	0	34	925	88
5	**Bus**	**1**	**10**	**790**	**30**
5	Car	0	0	887	50

Logit模型案例：交通方式选择

数据格式

- 1次选择情景占据多行，每行对应1个备选项。
- 因变量：即选择结果，1表示选中、0表示未选中，必须且仅能选中1项。
- 解释变量：站点等候时间 (TTME 单位：分钟)、行程时间 (INVT 单位：分钟)、行程费用 (INVC 单位：元)。

样本编号	备选项	选择结果	TTME	INVT	INVC
(1)	(2)	(3)	(4)	(5)	(6)
1	Air	0	69	100	59
1	Train	0	34	372	31
1	Bus	0	35	417	25
1	**Car**	√ **1**	**0**	**180**	**10**
2	Air	0	69	125	115
2	Train	0	34	892	98
2	Bus	0	35	882	53
2	**Car**	√ **1**	**0**	**720**	**23**

选择情景1#

1个备选项

被选中项

Logit模型案例：交通方式选择

定义备选项的效用

$$V_{Air} = \beta_{Air} + \beta_{TTME} TTME_{Air} + \beta_{INVT} INVT_{Air} + \beta_{INVC} INVC_{Air}$$

$$V_{Train} = \beta_{Train} + \beta_{TTME} TTME_{Train} + \beta_{INVT} INVT_{Train} + \beta_{INVC} INVC_{Train}$$

$$V_{Bus} = \beta_{Bus} + \beta_{TTME} TTME_{Bus} + \beta_{INVT} INVT_{Bus} + \beta_{INVC} INVC_{Bus}$$

$$V_{Car} = \beta_{TTME} TTME_{Car} + \beta_{INVT} INVT_{Car} + \beta_{INVC} INVC_{Car}$$

Logit模型案例：交通方式选择

总共有几个系数要估计?

$$V_{Air} = \beta_{Air} + \beta_{TTME} TTME_{Air} + \beta_{INVT} INVT_{Air} + \beta_{INVC} INVC_{Air}$$

$$V_{Train} = \beta_{Train} + \beta_{TTME} TTME_{Train} + \beta_{INVT} INVT_{Train} + \beta_{INVC} INVC_{Train}$$

$$V_{Bus} = \beta_{Bus} + \beta_{TTME} TTME_{Bus} + \beta_{INVT} INVT_{Bus} + \beta_{INVC} INVC_{Bus}$$

$$V_{Car} = \beta_{TTME} TTME_{Car} + \beta_{INVT} INVT_{Car} + \beta_{INVC} INVC_{Car}$$

- β_{TTME}、β_{INVT}、 β_{INVC} 分别为站点等候时间、行程时间、行程费用的系数，通常在各选项中保持一致。
- β_{Air}、β_{Train}、β_{Bus} 均为常数项。
- 选择模型的一般准则：**Only difference matters!**

Logit模型案例：交通方式选择

效用中的常数项

- Logit模型中，每个备选项都对应不同的常数项，因此称为**备选项特定的常数项**(alternative specific constant, **ASC**）。
- 一般地，有***N*个备选项**时，即应当有***N*-1个ASC**常数项，剩下一个备选项作为参照水平。
- 本案例的4个备选项对应了3个ASC，其隐含设定是β_{Car}=0，即以**小汽车为参照水平**，β_{Air}、β_{Train}、β_{Bus} 分别测度了航空、火车、巴士与小汽车的**固有效用差异**，即未被上述 3 个自变量所解释、但又客观存在的平均效用差异（舒适度、灵活性等其他要素）。

Logit模型案例：交通方式选择

一定要有常数项吗?

你的当前住宅

属性	描述
通勤距离	3km
房间大小	小房间（15m2）
月租金	2300元
人口密度	10人/ha
人口收入构成	低收入人口 = 50%

选择留下来

一个可能的备选住宅

属性	描述
通勤距离	5km
房间大小	大房间（30m2）
月租金	1500元
人口密度	80人/ha
人口收入构成	低收入人口 = 75%

选择搬到这里

$$U_A = \beta_{time}\ Time_A + \beta_{price}\ Price_A + \beta_{size}\ Size_A + \ldots + \varepsilon_A$$

$$U_B = \beta_{move} + \beta_{time}\ Time_B + \beta_{price}\ Price_B + \beta_{size}\ Size_B + \ldots + \varepsilon_B$$

Logit模型案例：交通方式选择

一定要有常数项吗?

备选住宅A

属性	描述
通勤距离	3km
房间大小	小房间（15m2）
月租金	2300元
人口密度	10人/ha
人口收入构成	低收入人口 = 50%

选择搬到这里

备选住宅B

属性	描述
通勤距离	5km
房间大小	大房间（30m2）
月租金	1500元
人口密度	80人/ha
人口收入构成	低收入人口 = 75%

选择搬到这里

$$U_A = \beta_{time}\, Time_A + \beta_{price}\, Price_A + \beta_{size}\, Size_A + \ldots + \varepsilon_A$$

$$U_B = \beta_B + \beta_{time}\, Time_B + \beta_{price}\, Price_B + \beta_{size}\, Size_B + \ldots + \varepsilon_B$$

Logit模型案例：交通方式选择

选择概率

$$
\begin{aligned}
V_{Air} &= \beta_{Air} + \beta_{TTME} TTME_{Air} + \beta_{INVT} INVT_{Air} + \beta_{INVC} INVC_{Air} \\
V_{Train} &= \beta_{Train} + \beta_{TTME} TTME_{Train} + \beta_{INVT} INVT_{Train} + \beta_{INVC} INVC_{Train} \\
V_{Bus} &= \beta_{Bus} + \beta_{TTME} TTME_{Bus} + \beta_{INVT} INVT_{Bus} + \beta_{INVC} INVC_{Bus} \\
V_{Car} &= \beta_{TTME} TTME_{Car} + \beta_{INVT} INVT_{Car} + \beta_{INVC} INVC_{Car}
\end{aligned}
$$

1. 取指数
2. 求和作分母
3. 除法

$$P_{Air} = \frac{e^{V_{Air}}}{e^{V_{Air}} + e^{V_{Train}} + e^{V_{Bus}} + e^{V_{Car}}}$$

$$P_{Bus} = \frac{e^{V_{Bus}}}{e^{V_{Air}} + e^{V_{Train}} + e^{V_{Bus}} + e^{V_{Car}}}$$

$$P_{Train} = \frac{e^{V_{Train}}}{e^{V_{Air}} + e^{V_{Train}} + e^{V_{Bus}} + e^{V_{Car}}}$$

$$P_{Car} = \frac{e^{V_{Car}}}{e^{V_{Air}} + e^{V_{Train}} + e^{V_{Bus}} + e^{V_{Car}}}$$

已知**解释变量**，什么样的β才算预测的准？

Logit模型案例：交通方式选择

最大似然法

Q：已知**解释变量**，什么样的***β***才算预测的准？

A：用***β***预测选择概率，<u>实际被选中的备选项的预测概率</u>应当尽可能高。

似然值（likelihood）

【给定参数下，观测到当前数据样本的概率】

最大似然法

(Maximum Likelihood Estimation,**MLE**)

【选择使得观测到的数据出现概率最大的参数值】

Logit模型案例：交通方式选择

最大似然法

如果只有一个样本：样本1选择了小汽车

样本编号 (1)	备选项 (2)	选择结果 (3)	TTME (4)	INVT (5)	INVC (6)
1	Air	0	69	100	59
1	Train	0	34	372	31
1	Bus	0	35	417	25
1	**Car**	**1**	**0**	**180**	**10**

$$P_{Car,1} = e^{\beta_{TTME}*0+\beta_{INVT}*180+\beta_{INVC}*10} \Bigg/ \left(\begin{array}{l} e^{\beta_{Air}+\beta_{TTME}*69+\beta_{INVT}*100+\beta_{INVC}*59} \\ +e^{\beta_{Train}+\beta_{TTME}*34+\beta_{INVT}*372+\beta_{INVC}*31} \\ +e^{\beta_{Bus}+\beta_{TTME}*69+\beta_{INVT}*100+\beta_{INVC}*59} \\ +e^{\beta_{TTME}*0+\beta_{INVT}*180+\beta_{INVC}*10} \end{array} \right)$$

Task：找到一组$\boldsymbol{\beta}$，让$P_{car,1}$尽可能高（尽可能接近1）。

Logit模型案例：交通方式选择

最大似然法

然而我们有210个样本......

样本1#：选择小汽车 $P_{Car,1}$

样本2#：选择小汽车 $P_{Car,2}$

样本3#：选择火车 $P_{Train,3}$

样本4#：选择航空 $P_{Air,4}$

样本5#：选择巴士 $P_{Bus,5}$

......

样本210#：......

将210个相互独立的概率连乘起来，即为模型整体的**似然值**

最大似然法的目的，即是找到使此似然值最大的一组$\boldsymbol{\beta}$。

Logit模型案例：交通方式选择

最大似然法

- 由于连乘的形式不利于计算，通常将其取自然对数后，变成连加的形式，结果为模型的**对数似然值**（Log Likelihood，**LL**）。

$$LL = ln\left(\prod_{n=1}^{210} P_n\right) = \sum_{n=1}^{210} ln(P_n)$$

P_n 为模型预测的第 n 个样本选择其实际选中的备选项的概率，如$P_1=P_{Car,1}$，$P_3=P_{Train,3}$

- 对数似然值一般**取值为负**：概率的取值范围为 0~1，所以$ln(P_n) \leq 0$。
- 作为**拟合优度**（goodness-of-fit）：对数似然值越大（绝对值后的正数越小），表明模型对数据的拟合程度越高。理论最大值为 0，即模型以100%的概率完美命中了每个样本的实际选中结果（基本不可能）。

Logit模型案例：交通方式选择

估计结果

样本量(N)	210		变量个数(K)	6
对数似然值(LL)	-192.89		r^2	0.3374

变量	系数(Coefficient)	标准误(S.E.)	z 值	相伴概率(p 值)
β_{TTME}	-0.0969	0.0103	-9.37	<0.0001
β_{INVT}	-0.0040	0.0085	-4.70	<0.0001
β_{INVC}	-0.0139	0.0067	-2.09	0.0365
β_{Air}	4.7399	0.8675	5.46	<0.0001
β_{Train}	3.9532	0.4686	8.44	<0.0001
β_{Bus}	3.3062	0.4583	7.21	<0.0001

一般而言：

- 增加**样本量**：有利于减少参数估计的误差（改善显著性）。
- 引入更多**变量**：有利于改善模型拟合效果（改善r^2），但同时会提高复杂度。

Logit模型案例：交通方式选择

估计结果：拟合优度

样本量(N)	210	变量个数(K)	6
对数似然值(LL)	-192.89	r^2	0.3374

- 根据对数似然值，可反推出每次选择的**平均预测准确率**：

$$\overline{P} = e^{\frac{LL}{N}} = e^{\frac{-192.89}{210}} = 40\%$$

- 这一准确率看似不高，但已经较“零模型”有了很大改善。
- **零模型**：通常指没有任何解释变量和参数的模型，可以理解为完全随机选择 。本例中，在4个选项间“瞎猜”的方式只能获得$\overline{P_0} = 1/4 = 25\%$的平均预测准确率。

Logit模型案例：交通方式选择

估计结果：拟合优度

样本量(N)	210	变量个数(K)	6
对数似然值(LL)	-192.89	$\boxed{r^2}$	$\boxed{0.3374}$

- r^2 反映了实际估计的模型与上述零模型相比，对数似然值提升的程度。
- 零模型的对数似然值 LL_0 可以按下式计算：

$$LL_0 = N * ln\left(\overline{P_0}\right) = 210 * ln\left(\frac{1}{4}\right) = -291.12$$

- 进而，Logit 模型的 r^2 由下式计算：

$$r^2 = 1 - \frac{LL}{LL_0} = 1 - \frac{-192.89}{-291.12} = 0.3374$$

Logit模型案例：交通方式选择

估计结果：拟合优度

对数似然值：尺度会受到样本量的影响，很少直接使用。

r^2：

- 消除了样本量的影响，其取值范围被归一化至 0~1 之间，取值越大，拟合优度越高，因此更直观方便。
- Logit模型中的r^2 vs. 普通线性回归中的r^2

伪 r^2（pseudo r^2）　　**样本决定系数**

  - 计算方法不同，不能类似地解释为因变量的变异被模型解释的比例。
  - 判断高低的标准不同。 Logit模型的 r^2 一般远低于线性回归。
  - >0.1可接受，0.2~0.4优秀

Logit模型案例：交通方式选择

估计结果：系数

样本量(N)	210		变量个数(K)	6
对数似然值(LL)	-192.89		r^2	0.3374

变量	系数(Coefficient)	标准误(S.E.)	z 值	相伴概率(p 值)
β_{TTME}	-0.0969	0.0103	-9.37	<0.0001
β_{INVT}	-0.0040	0.0085	-4.70	<0.0001
β_{INVC}	-0.0139	0.0067	-2.09	0.0365
β_{Air}	4.7399	0.8675	5.46	<0.0001
β_{Train}	3.9532	0.4686	8.44	<0.0001
β_{Bus}	3.3062	0.4583	7.21	<0.0001

- **系数**(Coefficient)：控制了模型中的其他自变量以后，各解释变量和常数项对个体选择的平均影响。

估计结果：系数

将系数代回效用公式：

$$
\begin{aligned}
V_{Air} &= 4.7399 - 0.0969 TTME_{Air} - 0.0040 INVT_{Air} - 0.0139 INVC_{Air} \\
V_{Train} &= 3.9532 - 0.0969 TTME_{Train} - 0.0040 INVT_{Train} - 0.0139 INVC_{Train} \\
V_{Bus} &= 3.3062 - 0.0969 TTME_{Bus} - 0.0040 INVT_{Bus} - 0.0139 INVC_{Bus} \\
V_{Car} &= -0.0969 TTME_{Car} - 0.0040 INVT_{Car} - 0.0139 INVC_{Car}
\end{aligned}
$$

Logit模型案例：交通方式选择

估计结果：系数

系数的符号**正负**——反映了该变量的影响**方向**

站点等候时间(β_{TTME})
行程时间(β_{INVT})
行程费用(β_{INVC})

系数均为负，表明它们均对选择具有负效应。

时间和费用增加 → 效用下降 （线性） → 选择概率下降 （非线性）

结果是否符合预期？✔

Logit模型案例：交通方式选择

估计结果：系数

系数的**绝对值大小**——反映了该变量影响的**强度**

变量	系数(Coefficient)
β_{TTME}	-0.0969
β_{INVT}	-0.0040
β_{INVC}	-0.0139
β_{Air}	4.7399
β_{Train}	3.9532
β_{Bus}	3.3062

β_{TTME}、β_{INVT} 单位：分钟

β_{INVC} 单位：元

$|\beta_{TTME}|$ 比 $|\beta_{INVT}|$ 高得多

由于两变量的量纲一致，因此可以判断个体对站点等候时间的敏感性比行程时间大得多 。

Logit模型案例：交通方式选择

估计结果：系数

系数的**绝对值大小**——反映了该变量影响的**强度**

变量	系数(Coefficient)
β_{TTME}	-0.0969
β_{INVT}	-0.0040
$\boxed{\beta_{INVC}}$	$\boxed{-0.0139}$
β_{Air}	4.7399
β_{Train}	3.9532
β_{Bus}	3.3062

β_{TTME}、β_{INVT} 单位：分钟

β_{INVC} 单位：元

$|\beta_{TTME}|$ 比 $|\beta_{INVT}|$ 高得多

由于两变量的量纲一致，因此可以判断个体对站点等候时间的敏感性比行程时间大得多 。

行程费用 $|\beta_{INVC}|$ 与它们的量纲不同，不宜简单地得出哪一方更重要的结论。

Logit模型案例：交通方式选择

估计结果：系数（支付意愿）

- **支付意愿**(willingness to pay, **WTP**)：个体为 x 每一个单位的变化所愿意支付的价格，基于**等效用性**计算。

$$WTP = \frac{变量x的系数}{价格的系数}$$

- 行程时间的支付意愿为0.29，意味着个体平均愿意为行程时间每1分钟的减少额外支付 0.29 元。

$$WTP_{INVT} = \frac{\beta_{INVT}}{\beta_{INVC}} = \frac{-0.0040}{-0.0139} = 0.29$$

- 推广到非价格要素：个体平均愿意为站点等候时间每1分钟的减少，而忍受行程时间24分钟的增加。

$$\frac{\beta_{TTME}}{\beta_{INVT}} = \frac{-0.0969}{-0.0040} = 24.23$$

Logit模型案例：交通方式选择

估计结果：系数（ASC常数项）

变量	系数(Coefficient)
β_{TTME}	-0.0969
β_{INVT}	-0.0040
β_{INVC}	-0.0139
β_{Air}	4.7399
β_{Train}	3.9532
β_{Bus}	3.3062

航空、火车、巴士的ASC均为正

表明在控制了站点等候时间、行程时间、费用后，

（即：假定4种交通方式这三方面完全相同），

上述3种交通方式与小汽车这一参照交通方式相比，

具有更高的固有效用。

- 固有效用排序：航空>火车>巴士>小汽车。
- 小汽车最低，可能是由于小汽车需要自己驾驶、较为劳累造成的。

Logit模型案例：交通方式选择

估计结果：显著性

样本量(N)	210		变量个数(K)	6
对数似然值(LL)	-192.89		r^2	0.3374

变量	系数(Coefficient)	标准误(S.E.)	z 值	相伴概率(p 值)
β_{TTME}	-0.0969	0.0103	-9.37	<0.0001
β_{INVT}	-0.0040	0.0085	-4.70	<0.0001
β_{INVC}	-0.0139	0.0067	-2.09	0.0365
β_{Air}	4.7399	0.8675	5.46	<0.0001
β_{Train}	3.9532	0.4686	8.44	<0.0001
β_{Bus}	3.3062	0.4583	7.21	<0.0001

标准误(standard error, S.E.)：由于抽样的随机性所带来的系数估计误差。

Logit模型案例：交通方式选择

估计结果：显著性

样本量(N)	210		变量个数(K)	6
对数似然值(LL)	-192.89		r^2	0.3374

变量	系数(Coefficient)	标准误(S.E.)	z 值	相伴概率(p 值)
β_{TTME}	-0.0969	0.0103	-9.37	<0.0001
β_{INVT}	-0.0040	0.0085	-4.70	<0.0001
β_{INVC}	-0.0139	0.0067	-2.09	0.0365
β_{Air}	4.7399	0.8675	5.46	<0.0001
β_{Train}	3.9532	0.4686	8.44	<0.0001
β_{Bus}	3.3062	0.4583	7.21	<0.0001

z值：假设检验统计量，零假设为：$\boldsymbol{\beta}$与0没有显著差异，即某变量对选择无影响。

$$\mathrm{z} = \frac{\beta - 0}{S.E.} = \frac{\beta}{S.E.}$$

Logit模型案例：交通方式选择

估计结果：显著性

样本量(N)	210		变量个数(K)	6
对数似然值(LL)	-192.89		r^2	0.3374

变量	系数(Coefficient)	标准误(S.E.)	z 值	相伴概率(p 值)
β_{TTME}	-0.0969	0.0103	-9.37	<0.0001
β_{INVT}	-0.0040	0.0085	-4.70	<0.0001
β_{INVC}	-0.0139	0.0067	-2.09	0.0365
β_{Air}	4.7399	0.8675	5.46	<0.0001
β_{Train}	3.9532	0.4686	8.44	<0.0001
β_{Bus}	3.3062	0.4583	7.21	<0.0001

P值：比z值更方便地判断显著性。一般以 <0.05 作为显著的标准。

Logit模型案例：交通方式选择

预测选择概率（以样本1为例）

样本编号 (1)	备选项 (2)	选择结果 (3)	TTME (4)	INVT (5)	INVC (6)
1	Air	0	69	100	59
1	Train	0	34	372	31
1	Bus	0	35	417	25
1	**Car**	**1**	**0**	**180**	**10**

$$V_{Air} = 4.7399 - 0.0969 TTME_{Air} - 0.0040 INVT_{Air} - 0.0139 INVC_{Air}$$

$$V_{Train} = 3.9532 - 0.0969 TTME_{Train} - 0.0040 INVT_{Train} - 0.0139 INVC_{Train}$$

$$V_{Bus} = 3.3062 - 0.0969 TTME_{Bus} - 0.0040 INVT_{Bus} - 0.0139 INVC_{Bus}$$

$$V_{Car} = -0.0969 TTME_{Car} - 0.0040 INVT_{Car} - 0.0139 INVC_{Car}$$

Logit模型案例：交通方式选择

预测选择概率（以样本1为例）

样本编号 (1)	备选项 (2)	选择结果 (3)	TTME (4)	INVT (5)	INVC (6)
1	Air	0	69	100	59
1	Train	0	34	372	31
1	Bus	0	35	417	25
1	**Car**	**1**	**0**	**180**	**10**

$$V_{Air} = 4.7399 - 0.0969 \times 69 - 0.0040 \times 100 - 0.0139 \times 59 = -3.1656$$

$$V_{Train} = 3.9532 - 0.0969 \times 34 - 0.0040 \times 372 - 0.0139 \times 31 = -1.2582$$

$$V_{Bus} = 3.3062 - 0.0969 \times 35 - 0.0040 \times 417 - 0.0139 \times 25 = -2.0984$$

$$V_{Car} = -0.0969 \times 0 - 0.0040 \times 180 - 0.0139 \times 10 = -0.8582$$

Logit模型案例：交通方式选择

预测选择概率（以样本1为例）

$$P_{Air} = \frac{e^{-3.1656}}{e^{-3.1656}+e^{-1.2582}+e^{-2.0984}+e^{-0.8582}} = 4.83\%$$

$$P_{Train} = \frac{e^{-1.2582}}{e^{-3.1656}+e^{-1.2582}+e^{-2.0984}+e^{-0.8582}} = 32.55\%$$

$$P_{Bus} = \frac{e^{-2.0984}}{e^{-3.1656}+e^{-1.2582}+e^{-2.0984}+e^{-0.8582}} = 14.05\%$$

$$P_{Car} = \frac{e^{-0.8582}}{e^{-3.1656}+e^{-1.2582}+e^{-2.0984}+e^{-0.8582}} = \boxed{48.56\%}$$

- 该样本实际选择了小汽车。
- 模型预测的选择小汽车的概率为48.56%，在 4 个概率结果中最高。

大纲

- 离散选择模型简介
- Logit模型的核心原理
- 应用案例：交通方式选择
- **个体行为模拟初步**

个体行为模拟

- 利用Logit模型计算的选择概率，我们可以对个体的选择行为进行模拟预测。
- 个体行为模拟的优势：
  - 个体行为模拟可以充分利用更加具体的个体细节信息，而集合模拟则一般只能使用在集合单元上被汇总后的平均信息。
  - 如果研究对象是“行为过程”——该过程由一次一次的连续行为衔接而成，后一次行为需要建立在前面行为的基础上——以每一位行动者为单元进行个体行为模拟是最为直接、自然、灵活的方式。
  - 个体行为模拟可以更好地捕捉个体之间的局部互动和网络效应。
  - ……

个体行为模拟

哪个概率最高，就一定选哪个吗?

99% vs. **1%**

51% vs. **49%**

选择结果应该一样吗?

个体行为模拟

基于概率的两种模拟方式

最大概率模拟

- 在每一次选择中，个体将选择概率最大的备选项。
- 全有全无（all or nothing）。
- 确定性：每次模拟结果都一样。

蒙特卡洛模拟

- 在每一次选择中，通过“**扔骰子**”的方式决定个体选择哪一个备选项。
- 依概率抽样：概率小也可能被选中。
- 随机性：每次模拟，结果都不一样。
- 由于随机性，可能需要多次模拟。

个体行为模拟

来扔骰子吧！

- 已知：选择Air的概率为1/6

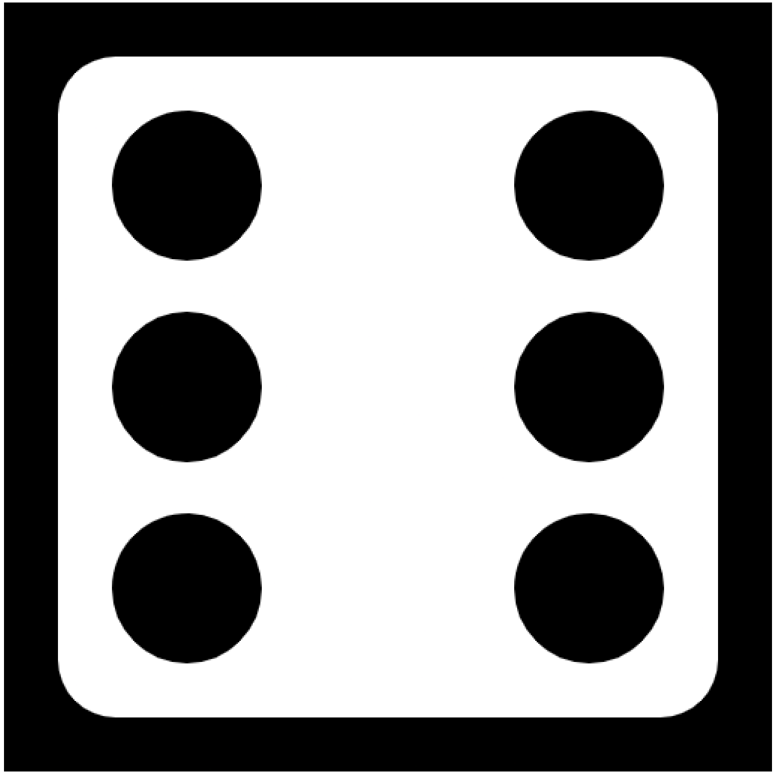
  扔到1点，选Air

- 已知：选择Air的概率为1/2

  扔到1点、2点、3点，选Air

- 已知：Air=1/2, Train=Bus=Car=1/6

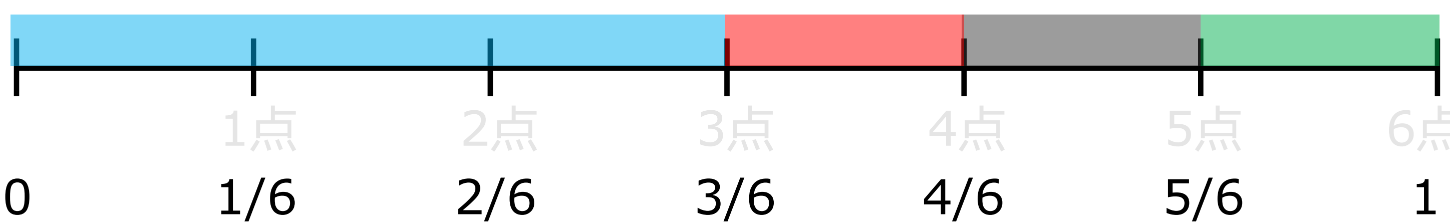
  扔到1点、2点、3点，选Air
  扔到4点，选Train
  扔到5点，选Bus
  扔到6点，选Car



个体行为模拟

蒙特卡洛模拟 (Monte Carlo Simulation)

Air=1/2, Trian=1/6, Bus=1/6, Car=1/6

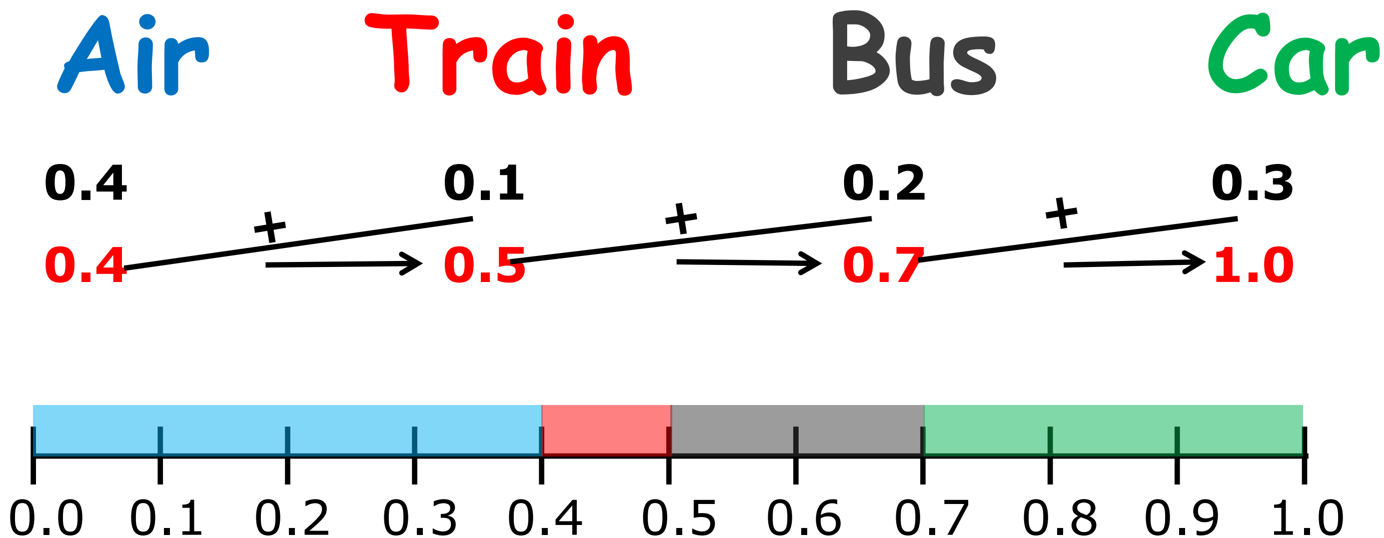


蒙特卡洛模拟 (Monte Carlo Simulation)

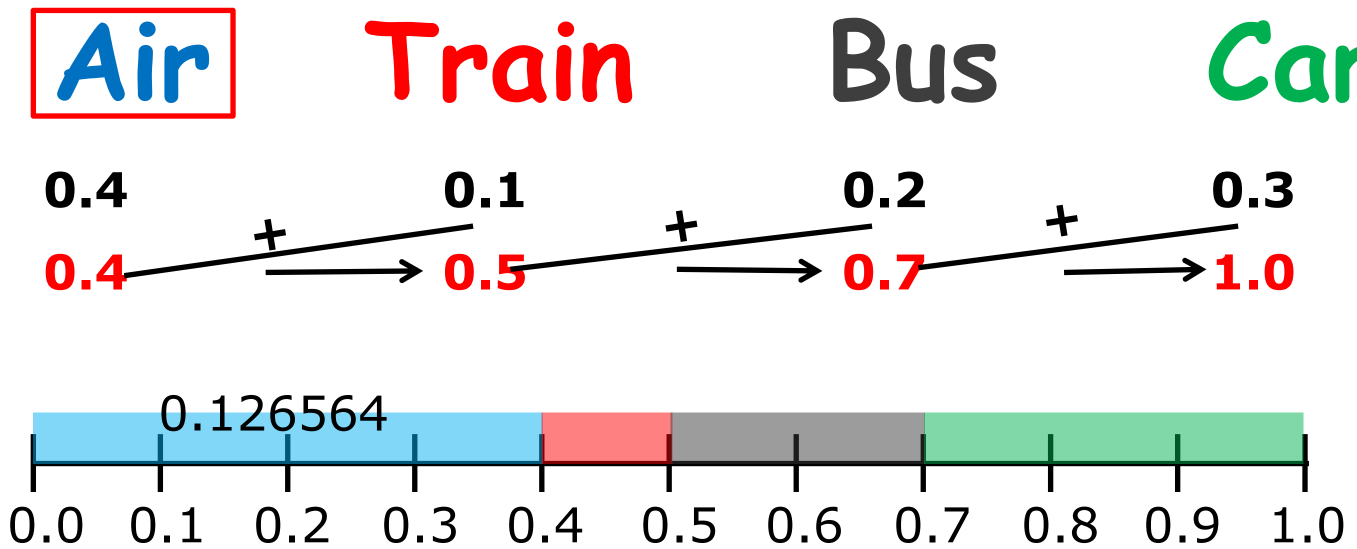
Air	Train	Bus	Car
0.4	0.1	0.2	0.3

个体行为模拟

蒙特卡洛模拟 (Monte Carlo Simulation)

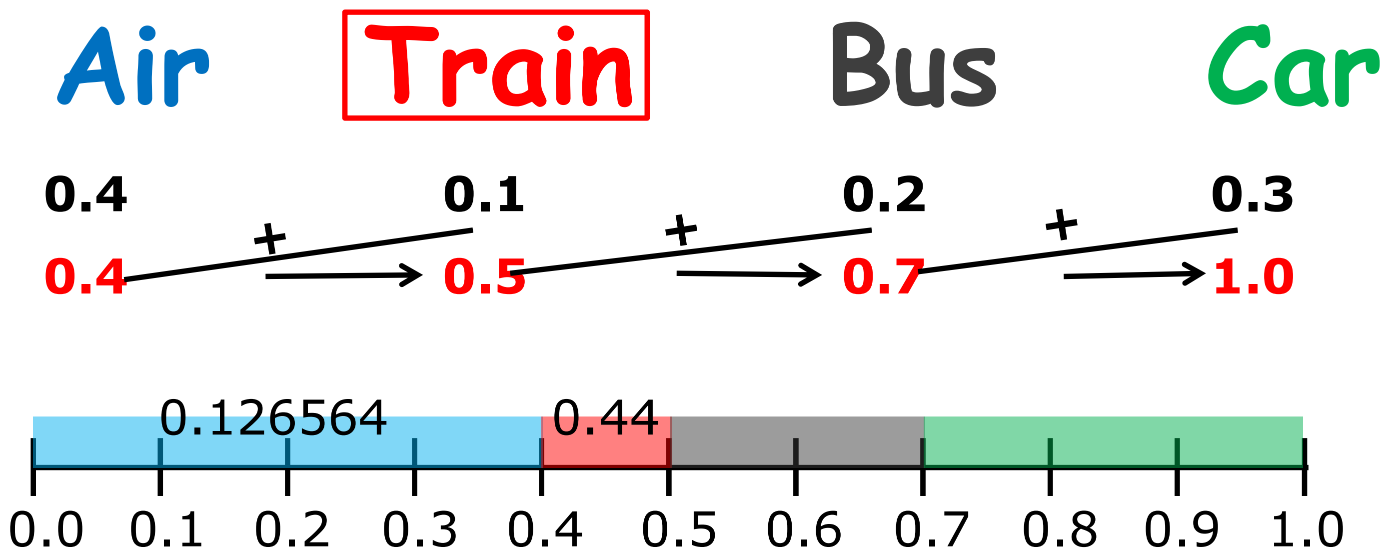


蒙特卡洛模拟 (Monte Carlo Simulation)

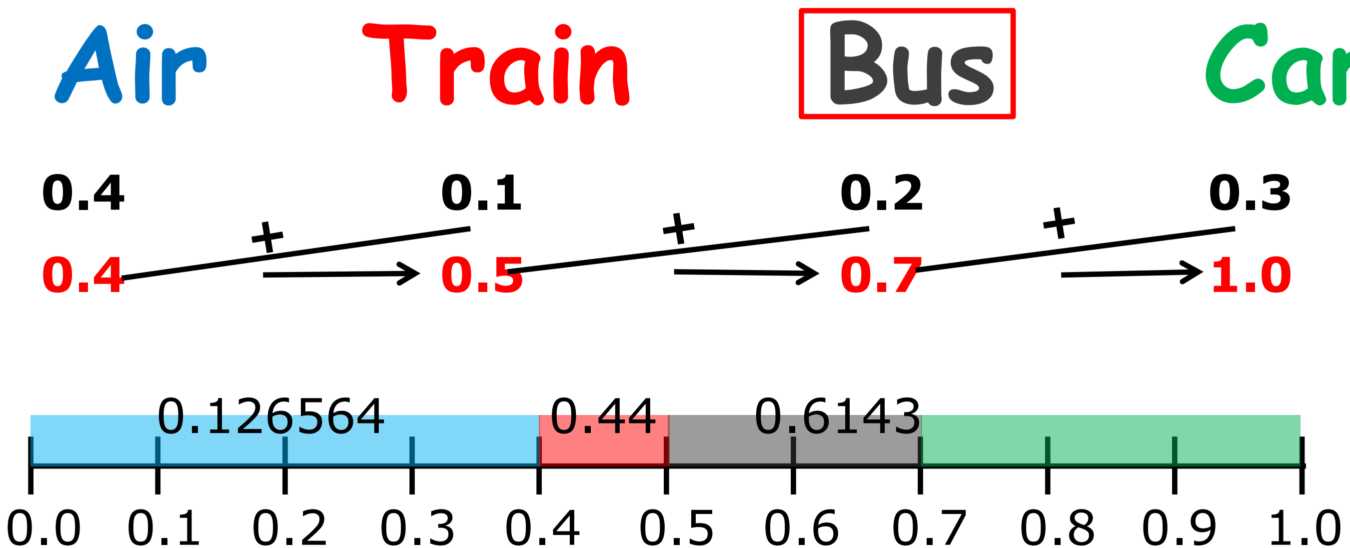


个体行为模拟

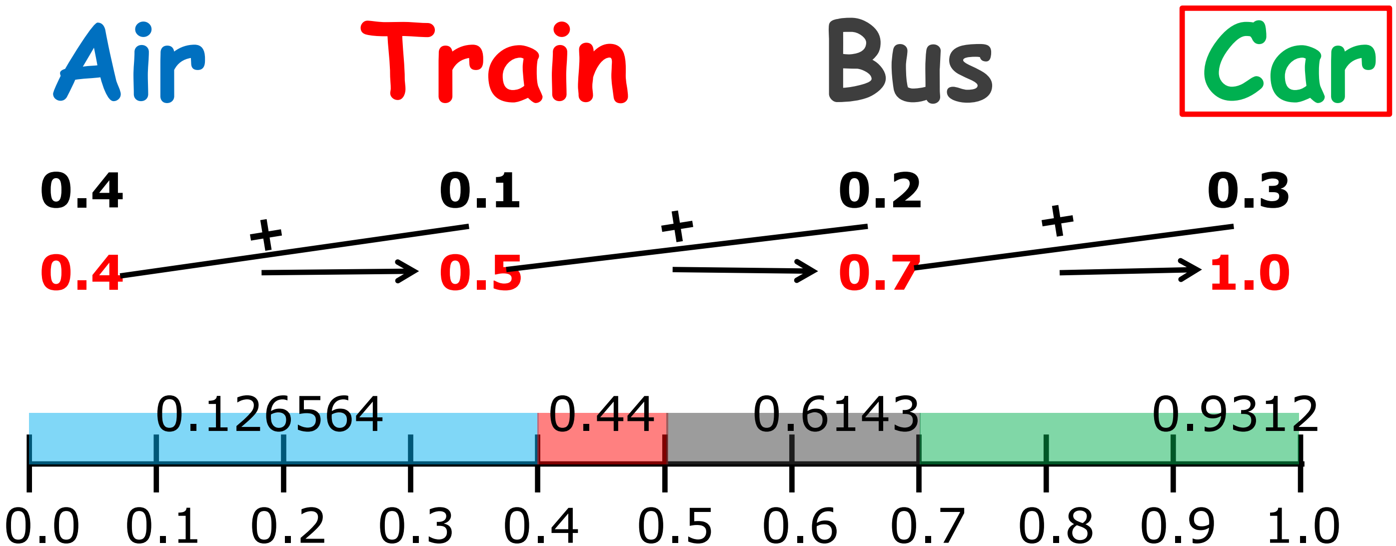
蒙特卡洛模拟 (Monte Carlo Simulation)



蒙特卡洛模拟 (Monte Carlo Simulation)



蒙特卡洛模拟 (Monte Carlo Simulation)



蒙特卡洛模拟 (Monte Carlo Simulation)

- 在一次选择中，可以通过概率累加的方式将[0, 1]分为N个区间：[0, P_1], $(P_1, P_1+P_2]$，$(P_1+P_2, \sum_{k=1}^{3} P_k]$…, $(\sum_{k=1}^{n-2} P_k, \sum_{k=1}^{n-1} P_k]$, $(\sum_{k=1}^{n-1} P_k, 1]$，其中第i个区间为$(\sum_{k=1}^{i-1} P_k, \sum_{k=1}^{i} P_k]$，由此保证这$N$个区段的长度分别对应于$N$个备选项的选择概率，进而生成一个0-1的均匀分布随机数，根据其落入的区段确定哪个对应的备选项被选中。

- 蒙洛卡洛模拟中的“骰子”是不公平的，而是根据预设的选择概率分布设定的，概率越高的备选项越有可能被选中，同时，小概率备选项也相应的机会。

个体行为模拟

样本1：蒙特卡洛模拟

$$P_{Air} = \frac{e^{-3.1656}}{e^{-3.1656} + e^{-1.2582} + e^{-2.0984} + e^{-0.8582}} = 4.83\%$$

选择**Air**： 0 - 0.0483

$$P_{Train} = \frac{e^{-1.2582}}{e^{-3.1656} + e^{-1.2582} + e^{-2.0984} + e^{-0.8582}} = 32.55\%$$

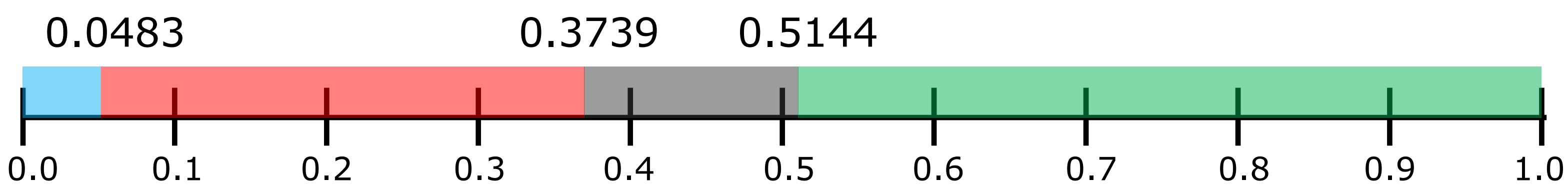
选择**Train**：0.0483 - 0.3739

$$P_{Bus} = \frac{e^{-2.0984}}{e^{-3.1656} + e^{-1.2582} + e^{-2.0984} + e^{-0.8582}} = 14.05\%$$

选择**Bus**：0.3739 – 0.5144

$$P_{Car} = \frac{e^{-0.8582}}{e^{-3.1656} + e^{-1.2582} + e^{-2.0984} + e^{-0.8582}} = 48.56\%$$

选择**Car**：0.5144 – 1



个体行为模拟

样本1-5

样本编号	备选项	选择结果	TTME	INVT	INVC
(1)	(2)	(3)	(4)	(5)	(6)
1	Air	0	69	100	59
1	Train	0	34	372	31
1	Bus	0	35	417	25
1	**Car**	**1**	**0**	**180**	**10**
2	Air	0	69	125	115
2	Train	0	34	892	98
2	Bus	0	35	882	53
2	**Car**	**1**	**0**	**720**	**23**
3	Air	0	69	100	59
3	**Train**	**1**	**40**	**345**	**20**
3	Bus	0	35	417	13
3	Car	0	0	284	12
4	**Air**	**1**	**45**	**115**	**148**
4	Train	0	34	945	111
4	Bus	0	35	935	66
4	Car	0	0	821	36
5	Air	0	69	129	103
5	Train	0	34	925	88
5	**Bus**	**1**	**10**	**790**	**30**
5	Car	0	0	887	50

个体行为模拟

样本1-5：最大概率模拟

样本	交通方式			
	Air	Train	Bus	Car
1	4.83%	32.55%	14.05%	48.56%*+
2	20.51%	16.41%	15.18%	47.90%*+
3	6.34%	30.99%*	21.78%	40.89%+
4	74.17%*+	5.96%	5.51%	14.36%
5	6.00%	4.16%	85.59%*+	4.25%

注：*表示实际选择结果，+表示预测选择结果

- 模型正确预测了 1、2、4、5 号样本的选择结果。
- 但在 3 号样本上，模型预测其选择小汽车， 而实际选择结果为火车。

个体行为模拟

样本1-5：蒙特卡洛模拟

模拟轮次	随机数	样本	交通方式			
			Air	Train	Bus	Car
I	0.6134	1	[0, 0.0483]	(0.0483, 0.3739]	(0.3739, 0.5144]	(0.5144,1]*+
	0.6555	2	[0, 0.2051]	(0.2051, 0.3692]	(0.3692, 0.5210]	(0.5210,1]*+
	0.1712	3	[0, 0.0634]	(0.0634,0.3733]*+	(0.3733, 0.5911]	(0.5911, 1]
	0.7083	4	[0, 0.7417]*+	(0.7417, 0.8013]	(0.8013, 0.8594]	(0.8594, 1]
	0.1318	5	[0, 0.0600]	(0.0600, 0.1016]	(0.1016, 0.9575]*+	(0.9575, 1]
II	0.2291	1	[0, 0.0483]	(0.0483, 0.3739]+	(0.3739, 0.5144]	(0.5144, 1]*
	0.6787	2	[0, 0.2051]	(0.2051, 0.3692]	(0.3692, 0.5210]	(0.5210, 1]*+
	0.0462	3	[0, 0.0634]+	(0.0634, 0.3733]*	(0.3733, 0.5911]	(0.5911, 1]
	0.0971	4	[0, 0.7417]*+	(0.7417, 0.8013]	(0.8013, 0.8594]	(0.8594, 1]
	0.8235	5	[0, 0.0600]	(0.0600, 0.1016]	(0.1016, 0.9575]*+	(0.9575, 1]

注：*表示实际选择结果，+表示预测选择结果

个体行为模拟

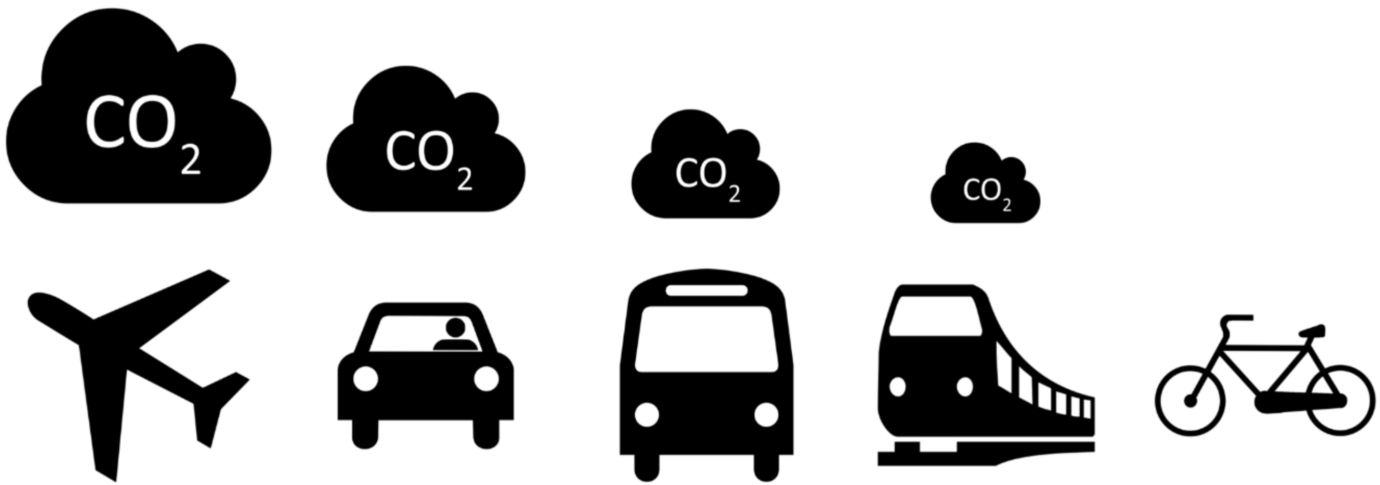
全部样本：最大概率模拟 & 蒙特卡洛模拟

结果	交通方式				均方误差
	Air	Train	Bus	Car	
实际	58 (28%)	63 (30%)	30 (14%)	59 (28%)	-
最大概率模拟	55 (26%)	64 (30%)	24 (11%)	67 (32%)	5.24
蒙特卡洛模拟：1次	60 (29%)	56 (27%)	31 (15%)	63 (30%)	4.18
蒙特卡洛模拟：5次	59 (28%)	64 (30%)	28 (13%)	58 (28%)	1.32
蒙特卡洛模拟：10次	58 (28%)	64 (30%)	29 (14%)	60 (29%)	0.87
蒙特卡洛模拟：20次	58 (28%)	63 (30%)	30 (14%)	59 (28%)	0
蒙特卡洛模拟：50次	58 (28%)	63 (30%)	30 (14%)	59 (28%)	0

- 相比于最大概率模拟，蒙特卡洛模拟通常具有**更小的系统误差**。
- 由于天然的随机性，蒙特卡洛模拟可能需要运行多次，以获得稳定的结果。

个体行为模拟

What If?



（1）提高发车频率，将巴士站点等候时间降低20%。

（2）提高速度，将巴士行程时间减少20%。

（3）降低成本，将巴士行程费用降低20%。

效果如何?

个体行为模拟

What If?

情景	结果	交通方式			
		Air	Train	Bus	Car
-	实际	58 (28%)	63 (30%)	**30 (14%)**	59 (28%)
现实情景	最大概率模拟	55 (26%)	64 (30%)	24 (11%)	67 (32%)
	蒙特卡洛模拟	58 (28%)	63 (30%)	30 (14%)	59 (28%)
1）Bus：站点等候时间减少20%	最大概率模拟	53 (25%)	53 (25%)	44 (21%)	60 (29%)
	蒙特卡洛模拟	54 (26%)	58 (28%)	**46 (22%)**	53 (25%)
2）Bus：行程时间减少20%	最大概率模拟	55 (26%)	55 (26%)	37 (18%)	63 (30%)
	蒙特卡洛模拟	55 (26%)	60 (29%)	**39 (19%)**	56 (27%)
3）Bus：票价减少20%	最大概率模拟	55 (26%)	63 (30%)	25 (12%)	67 (32%)
	蒙特卡洛模拟	58 (28%)	62 (30%)	**32 (15%)**	58 (28%)

以提高巴士分担率、减少小汽车分担率为目标：减少站点等候时间的效果最好，其次为减少行程时间，相对而言，出行者对行程费用的降低最不敏感。

小结

- 离散选择模型用于解释和预测个体的选择决策行为，Logit模型（MNL模型）是其中最简单、常用的模型。
- 选择者选择效用最大的备选项，由于效用中有一部分是随机的，所以选择也具有随机性，我们利用可见效用计算选择概率。
- Logit模型的选择概率： $P_i = \exp(V_i) \,/ \sum_j \exp(V_j)$
- 拟合优度（对数似然值、伪r^2）、系数的方向和大小、支付意愿、显著性
- 个体行为模拟：选择模型计算概率 → 蒙特卡洛模拟 / 最大概率模拟