



叙述性偏好法

城市分析方法系列课程

苏州大学 王灿

1、SP简介

什么是SP？为什么SP？

2、SP基本原理

要素水平，正交实验设计

3、SP应用案例

养老机构选择偏好研究





1、SP简介

什么是SP? 为什么SP?

2、SP基本原理

要素水平, 正交实验设计

3、SP应用案例

养老机构选择偏好研究



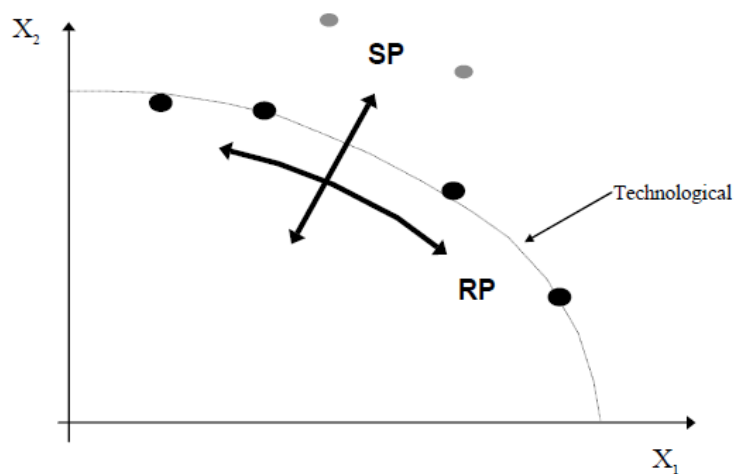
什么是RP? 什么是SP?

- 选择行为的背后是偏好 (preference)
- **显示性偏好**数据 (revealed preference, **RP**) : 反映人们在真实世界中的实际选择的数据被称显示性偏好数据
- **叙述性偏好**数据 (stated preference, **SP**) : 通过选择实验的方式收集的数据。在实验中, 研究者为受访者提供一些经过设计的虚拟选择情景, 由受访者判断并叙述自己将会做出的选择。

叙述性偏好法简介

为什么需要SP? SP的优势在哪里?

- 现实不存在: 需要评估新的选项 (产品)、新的属性、新的解释变量, 或者产品本身不具有交易属性。
- 现实区别度低: 在真实环境下, 选项之间在某些解释变量上可能缺少差异性



叙述性偏好法简介

为什么需要SP? SP的优势在哪里?

- 统计污点: RP数据中的**解释变量常常相关**, 使单个变量的**效应无法独立估计**, 影响结果的准确性, 这可能是RP数据最普遍、最严重的问题。
- 成本: SP法简单、快捷, 收集SP数据的成本远小于同等规模的RP数据。



参观选择: 排队是正效应吗?

吸引力**高**, 且排队**短**



吸引力**高**, 且排队**长**



吸引力**低**, 且排队**短**



吸引力**低**, 且排队**长**



叙述性偏好法简介

为什么需要SP? SP的优势在哪里?

RP :the world as it is 如你所见的世界

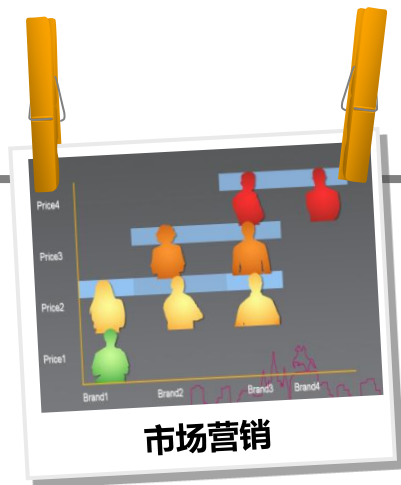
- 选项及属性必须真实存在;
- 选项间很可能在某些属性差异过小;
- **属性间在难以避免的相关性, 各属性的效应不能独立的估计出来, 结果一般存在偏误;**
- 可能较为费时费力。

SP: the world as it could be 如你所愿的世界

- 选项或属性不必真实存在;
- 可以通过实验设计保证选项间差异;
- **可以通过实验设计规避模型界定的问题, 从而保证所需各项系数能被独立且无偏地估计;**
- 一般较为简单易行。

RP相比于SP, 最大的优点是真实; 而SP是在很短的时间内做出的, 可能“言行不一”。
RP与SP总体上是互补的, 一种数据的弱点可以由另一种数据的优点所弥补。

叙述性偏好法简介



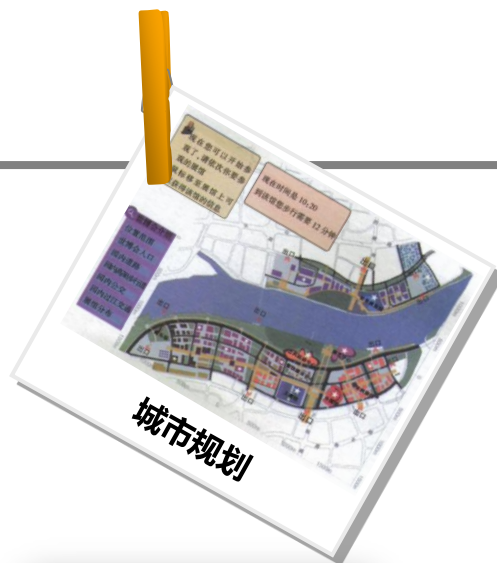
□最原始和应用最广的领域：选择不同的商品，选择品牌、服务、.....



□交通需求预测、路径选择、交通方式选择、停车收费政策.....



□非市场化商品的支付意愿评估：环境、文化、游憩、景观、健康.....



□各种各样的空间行为研究

叙述性偏好法简介

SP长什么样？

虚拟选择场景: #2

你的当前住宅

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	30人/ha
人口收入构成	低收入人口 = 10%

选择留下来

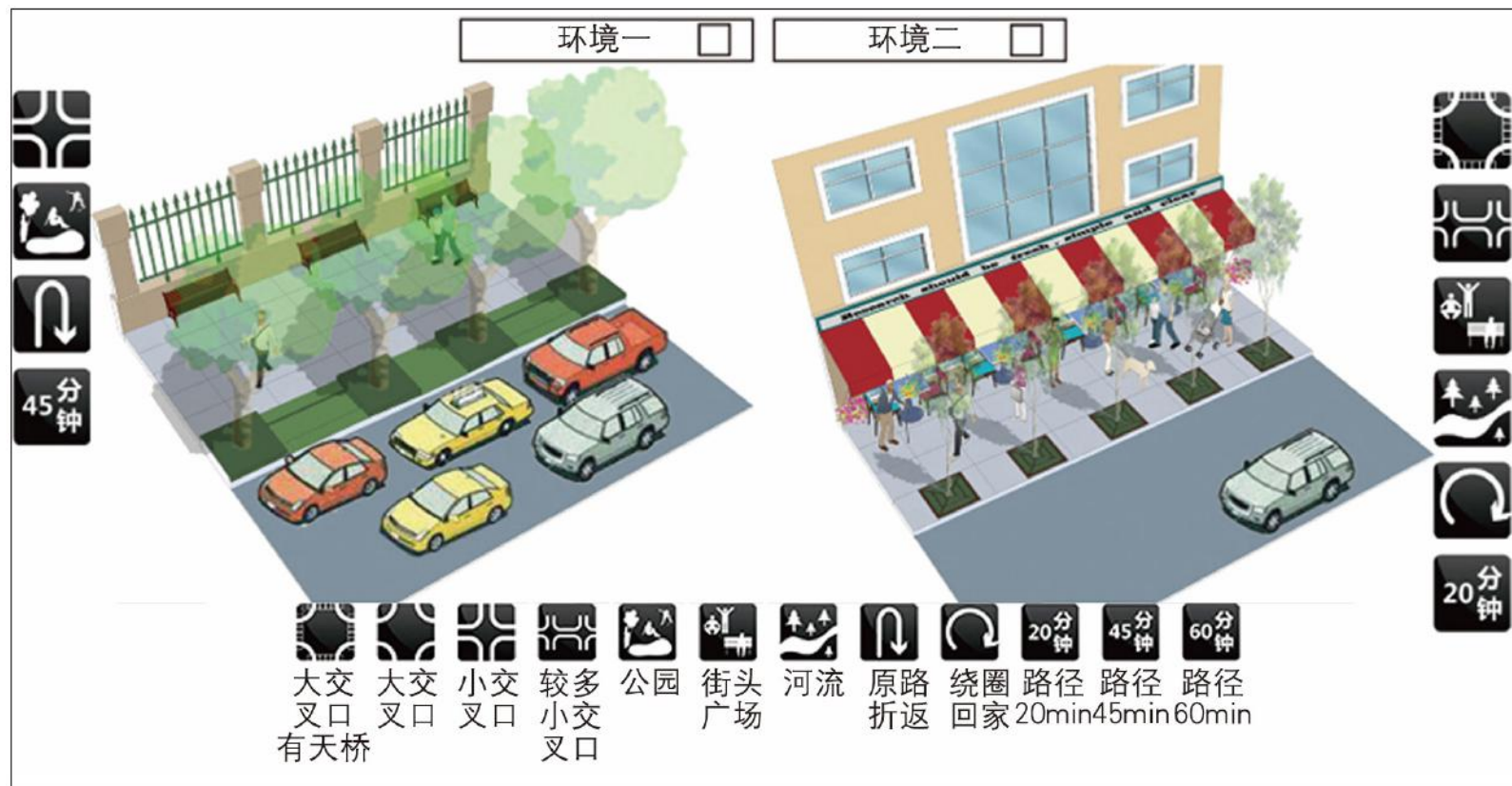
一个可能的备选住宅

属性	描述
通勤距离	10km
房间大小	小房间 (15m2)
月租金	1500元
人口密度	30人/ha
人口收入构成	低收入人口 = 50%

选择搬到这里

叙述性偏好法简介

SP长什么样?



1、SP简介

什么是SP? 为什么SP?



2、SP基本原理

要素水平, 正交实验设计

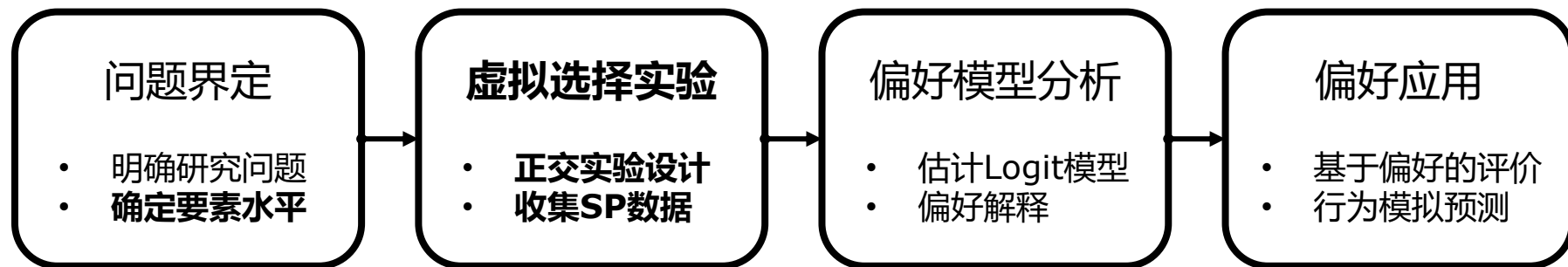
3、SP应用案例

养老机构选择偏好研究



叙述性偏好法基本原理

SP 的工作流程



本章重点：确定要素和水平、正交实验设计、收集SP数据。

其余部分参见“离散选择模型”章节。

叙述性偏好法基本原理

确定要素及其水平

- 要素 (attributes) : 选择实验中, 每个虚拟情景中的选项由若干个要素所描述, 这些要素将成为 Logit 模型中的解释变量。
 - 实际中有影响的要素很多, 但纳入模型的要素不宜过多, 否则不易回答。
 - 一般要根据文献、访谈、预调查等前期研究的结果, 挑选重要性高、符合研究兴趣的要素。
- 要素的水平 (levels) : 要素在实验中可能的取值。
 - 实际中有无限可能, 但实验中的水平数有限, 以2-4个最为常见。
 - 应该具有现实意义。

叙述性偏好法基本原理

实验设计

- 把不同要素的特定水平组合起来，就得到了选项和选择情景。
- **实验设计** (design of experiments, **DOE**) ：如何进行要素水平组合的方法。

确定要素和水平

要素	水平
行程时间 (x_1)	水平1: 60min
	水平2: 90min
费用 (x_2)	水平1: 15元
	水平2: 30元

生成问卷，收集SP数据

在下列情景中，您愿意选择哪种交通方式？

1.	巴士	火车	都不选
行程时间	60min	60min	
费用	15元	15元	
您的选择	☐	☐	☐

.....

10.	巴士	火车	都不选
行程时间	90min	60min	
费用	15元	30元	
您的选择	☐	☐	☐

.....

叙述性偏好法基本原理

实验设计

- 把不同要素的特定水平组合起来，就得到了选项和选择情景。
- **实验设计** (design of experiments, **DOE**)：如何进行要素水平组合的方法。

确定要素和水平

要素	水平
行程时间 (x_1)	水平1: 60min
	水平2: 90min
费用 (x_2)	水平1: 15元
	水平2: 30元

实验设计：设计矩阵

	巴士 x_1	巴士 x_2	火车 x_1	火车 x_2
1	1	1	1	1
2	1	1	1	2
3	1	1	2	1
4	1	1	2	2
5	1	2	1	1
6	1	2	1	2
7	1	2	2	1
8	1	2	2	2
9	2	1	1	1
10	2	1	1	2
11	2	1	2	1
12	2	1	2	2
13	2	2	1	1
14	2	2	1	2
15	2	2	2	1
16	2	2	2	2

生成问卷，收集SP数据

在下列情景中，您愿意选择哪种交通方式？

1.	巴士	火车	都不选
行程时间	60min	60min	
费用	15元	15元	
您的选择	☐	☐	☐

.....

10.	巴士	火车	都不选
行程时间	90min	60min	
费用	15元	30元	
您的选择	☐	☐	☐

.....

叙述性偏好法基本原理

实验设计

- 完全因子设计 (full factorial design) : 考虑所有可能的组合。
- 4要素: 巴士 x_1 、巴士 x_2 、火车 x_1 、火车 x_2 , 每要素2水平: $2^4=16$ 。

确定要素和水平

要素	水平
行程时间 (x_1)	水平1: 60min
	水平2: 90min
费用 (x_2)	水平1: 15元
	水平2: 30元

实验设计: 设计矩阵

	巴士 x_1	巴士 x_2	火车 x_1	火车 x_2
1	1	1	1	1
2	1	1	1	2
3	1	1	2	1
4	1	1	2	2
5	1	2	1	1
6	1	2	1	2
7	1	2	2	1
8	1	2	2	2
9	2	1	1	1
10	2	1	1	2
11	2	1	2	1
12	2	1	2	2
13	2	2	1	1
14	2	2	1	2
15	2	2	2	1
16	2	2	2	2

生成问卷, 收集SP数据

在下列情景中, 您愿意选择哪种交通方式?

1.	巴士	火车	都不选
行程时间	60min	60min	
费用	15元	15元	
您的选择	☐	☐	☐

.....

10.	巴士	火车	都不选
行程时间	90min	60min	
费用	15元	30元	
您的选择	☐	☐	☐

.....

叙述性偏好法基本原理

部分因子设计

- 随着要素、水平数量的增加，选择情景的数量将几何增长，SP实验难度激增。
- 部分因子设计 (fractional factorial design)：只使用一部分情景。

完全因子设计

	巴士 x_1	巴士 x_2	火车 x_1	火车 x_2
1	1	1	1	1
2	1	1	1	2
3	1	1	2	1
4	1	1	2	2
5	1	2	1	1
6	1	2	1	2
7	1	2	2	1
8	1	2	2	2
9	2	1	1	1
10	2	1	1	2
11	2	1	2	1
12	2	1	2	2
13	2	2	1	1
14	2	2	1	2
15	2	2	2	1
16	2	2	2	2

正交设计

	巴士 x_1	巴士 x_2	火车 x_1	火车 x_2
1	1	1	1	1
2	1	1	1	2
3	1	2	2	1
4	1	2	2	2
5	2	1	2	1
6	2	1	2	2
7	2	2	1	1
8	2	2	1	2

叙述性偏好法基本原理

正交设计

- 部分因子设计 (fractional factorial design) : 只使用一部分情景。
- **正交设计** (orthogonal design) : 大大减少选择情景数量, 且统计性质良好。
 - 要素水平均衡 (attribute level balance) : 均匀分散, 整齐可比。
 - 要素之间互不相关: 各要素的效应可以独立估计。

正交设计

	巴士x ₁	巴士x ₂	火车x ₁	火车x ₂
1	1	1	1	1
2	1	1	1	2
3	1	2	2	1
4	1	2	2	2
5	2	1	2	1
6	2	1	2	2
7	2	2	1	1
8	2	2	1	2

	BUS_x1	BUS_x2	TRAIN_x1	TRAIN_x2
BUS_x1	1.0000			
BUS_x2	0.0000	1.0000		
TRAIN_x1	0.0000	0.0000	1.0000	
TRAIN_x2	0.0000	0.0000	0.0000	1.0000

叙述性偏好法基本原理

正交设计：正交表

■ $L8(2^4)$ 正交表

- 括号内的指数4：4个要素，对应于表中的4列；
- 括号内的底数2：每个要素均有2个水平；
- 括号外的数字8：正交设计共生成了8行，对应于8次选择实验。

正交设计

	巴士 x_1	巴士 x_2	火车 x_1	火车 x_2
1	1	1	1	1
2	1	1	1	2
3	1	2	2	1
4	1	2	2	2
5	2	1	2	1
6	2	1	2	2
7	2	2	1	1
8	2	2	1	2

叙述性偏好法基本原理

正交设计：正交表

■ $L8(2^4)$ 正交表

- 括号内的指数4：4个要素
- 括号内的底数2：每个要素2个水平；
- 括号外的数字8：8次选择实验。

正交设计

	巴士 x_1	巴士 x_2	火车 x_1	火车 x_2
1	1	1	1	1
2	1	1	1	2
3	1	2	2	1
4	1	2	2	2
5	2	1	2	1
6	2	1	2	2
7	2	2	1	1
8	2	2	1	2

	Factor Levels			
	2	3	4	5
4	L4(2^3)			
8	L8(2^7)			
9		L9(3^4)		
12	L12(2^{11})			
16	L16(2^{15})		L16(4^5)	
25				L25(5^8)
27		L27(3^{13}) L27(3^{22})		
32	L32(2^{31})			

叙述性偏好法基本原理

正交设计：正交表

■ $L8(2^4)$ 正交表

- 括号内的指数4：4个要素
- 括号内的底数2：每个要素2个水平；
- 括号外的数字8：8次选择实验。

正交设计

	巴士 x_1	巴士 x_2	火车 x_1	火车 x_2
1	1	1	1	1
2	1	1	1	2
3	1	2	2	1
4	1	2	2	2
5	2	1	2	1
6	2	1	2	2
7	2	2	1	1
8	2	2	1	2

		Factor Levels			
		2 and 3	2 and 4	2 and 8	3 and 6
8	8		$L8(2^4 \times 4^1)$		
	16		$L16(2^9 \times 4^2)$ $L16(2^6 \times 4^3)$ $L16(2^3 \times 4^4)$ $L16(2^{12} \times 4^1)$	$L16(2^8 \times 8^1)$	
	18	$L18(2^1 \times 3^7)$ $L18(2^2 \times 3^6)$			$L18(3^6 \times 6^1)$
	32		$L32(2^1 \times 4^9)$		
36	36	$L36(2^{11} \times 3^{12})$ $L36(2^3 \times 3^{13})$			
	54	$L54(2^1 \times 3^{25})$			

叙述性偏好法基本原理

正交设计

- **同步正交设计**：多个备选项同时考虑，一步到位生成实验设计。
- **分步正交设计**：先生成单个备选项，再对备选项进行组合对比。

同步正交设计

	巴士 x_1	巴士 x_2	火车 x_1	火车 x_2
1	1	1	1	1
2	1	1	1	2
3	1	2	2	1
4	1	2	2	2
5	2	1	2	1
6	2	1	2	2
7	2	2	1	1
8	2	2	1	2

分步正交设计

	x_1	x_2		
1	1	1		
2	1	2		
3	2	1		
4	2	2		

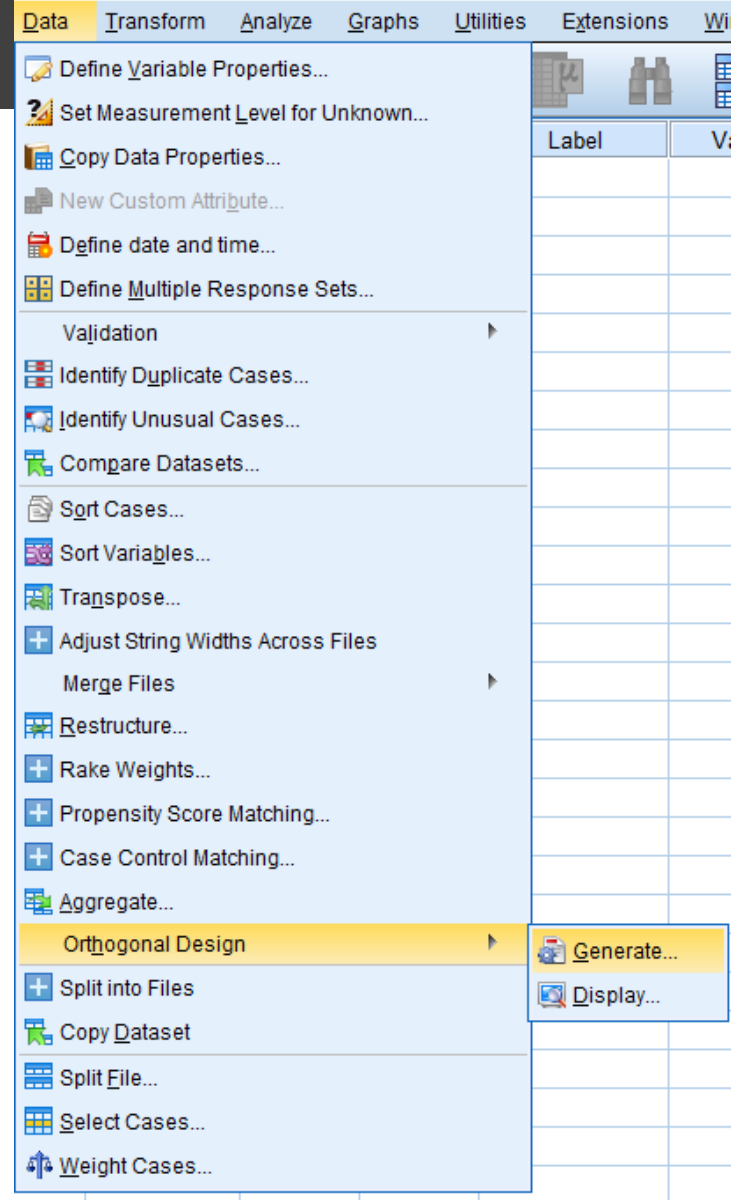


	选项一 x_1	选项一 x_2	选项二 x_1	选项二 x_2
1	1	1	2	2
2	1	2	2	1
3	2	1	1	2
4	2	2	1	1

叙述性偏好法基本原理

正交设计

- 软件：SPSS、NGENE



1、SP简介

什么是SP? 为什么SP?

2、SP基本原理

要素水平, 正交实验设计



3、SP应用案例

养老机构选择偏好研究



案例：养老院选择偏好

背景

- 养老机构的问题：市中心一床难求，郊区空置率高。
- 原因：结构性供需不平衡——提供的养老院不能满足老年人的需求。
- 研究目的：通过SP+离散选择模型，剖析老年人的养老院选择偏好。

案例：养老院选择偏好

要素及其水平

■ 7个要素，2/3/4个水平。

要素	水平			
	1	2	3	4
Loc：区位	市区	近郊区	远郊区	
Home：离家距离	10min	20min	40min	60min
Air：空气质量	很好	一般		
Hosp：综合医院	5min	15min	30min	45min
Metro：地铁站	有	无		
Serv：养老服务	很好	一般		
Fee：月花费	1500元	2000元	2500元	3000元

■ 备选项：养老机构A、养老机构B、都不选。

案例：养老院选择偏好

分步正交设计

- 首先，让我们看看一个选项长啥样，即选项A的轮廓（profile）。

(7个要素参与正交)

编号	区位	离家距离	空气质量	综合医院	地铁站	养老服务	月花费
1	1	1	1	1	1	1	1
2	1	1	2	4	2	1	3
3	1	3	2	3	1	2	2
4	1	3	2	2	2	2	3
5	2	2	1	2	1	2	4
6	2	3	1	1	2	1	2
7	2	4	1	4	1	1	1
8	2	4	2	2	2	1	4
9	3	1	1	4	1	2	4
10	3	2	2	3	2	1	2
11	3	2	2	1	1	2	3
12	3	4	1	3	2	2	1
13	1	4	2	4	1	2	2
14	1	4	1	1	2	2	4
15	1	2	1	2	1	1	1
16	1	2	1	3	2	1	4
17	2	3	2	3	1	1	3
18	2	2	2	4	2	2	1
19	2	1	2	1	1	2	2
20	2	1	1	3	2	2	3
21	3	4	2	1	1	1	3
22	3	3	1	2	2	2	1
23	3	3	1	4	1	1	4
24	3	1	2	2	2	1	2

案例：养老院选择偏好

分步正交设计

- 首先，让我们看看一个选项长啥样，即生成养老机构A的轮廓（profile）。
- 然后，我们把它们复制一份，随机打散，作为养老机构B的轮廓。
- 将B拼接在A的后面，每一行对应的两个轮廓进行比较，即得到24个选择情景。

(7个要素参与正交)

编号	区位	离家距离	空气质量	综合医院	地铁站	养老服务	月花费
22	3	3	1	2	2	2	1
7	2	4	1	4	1	1	1
9	3	1	1	4	1	2	4
10	3	2	2	3	2	1	2
19	2	1	2	1	1	2	2
21	3	4	2	1	1	1	3
3	1	3	2	3	1	2	2
1	1	1	1	1	1	1	1
4	1	3	2	2	2	2	3
12	3	4	1	3	2	2	1
17	2	3	2	3	1	1	3
11	3	2	2	1	1	2	3
23	3	3	1	4	1	1	4
6	2	3	1	1	2	1	2
16	1	2	1	3	2	1	4
5	2	2	1	2	1	2	4
15	1	2	1	2	1	1	1
13	1	4	2	4	1	2	2
20	2	1	1	3	2	2	3
2	1	1	2	4	2	1	3
18	2	2	2	4	2	2	1
8	2	4	2	2	2	1	4
24	3	1	2	2	2	1	2
14	1	4	1	1	2	2	4

案例：养老院选择偏好

分步正交设计

选择情景	养老机构A							养老机构B						
	Loc	Home	Air	Hosp	Metro	Serv	Fee	Loc	Home	Air	Hosp	Metro	Serv	Fee
1	1	1	1	1	1	1	1	3	3	1	2	2	2	1
2	1	1	2	4	2	1	3	2	4	1	4	1	1	1
3	1	3	2	3	1	2	2	3	1	1	4	1	2	4
4	1	3	2	2	2	2	3	3	2	2	3	2	1	2
5	2	2	1	2	1	2	4	2	1	2	1	1	2	2
6	2	3	1	1	2	1	2	3	4	2	1	1	1	3
7	2	4	1	4	1	1	1	1	3	2	3	1	2	2
8	2	4	2	2	2	1	4	1	1	1	1	1	1	1
9	3	1	1	4	1	2	4	1	3	2	2	2	2	3
10	3	2	2	3	2	1	2	3	4	1	3	2	2	1
11	3	2	2	1	1	2	3	2	3	2	3	1	1	3
12	3	4	1	3	2	2	1	3	2	2	1	1	2	3
13	1	4	2	4	1	2	2	3	3	1	4	1	1	4
14	1	4	1	1	2	2	4	2	3	1	1	2	1	2
15	1	2	1	2	1	1	1	1	2	1	3	2	1	4
16	1	2	1	3	2	1	4	2	2	1	2	1	2	4
17	2	3	2	3	1	1	3	1	2	1	2	1	1	1
18	2	2	2	4	2	2	1	1	4	2	4	1	2	2
19	2	1	2	1	1	2	2	2	1	1	3	2	2	3
20	2	1	1	3	2	2	3	1	1	2	4	2	1	3
21	3	4	2	1	1	1	3	2	2	2	4	2	2	1
22	3	3	1	2	2	2	1	2	4	2	2	2	1	4
23	3	3	1	4	1	1	4	3	1	2	2	2	1	2
24	3	1	2	2	2	1	2	1	4	1	1	2	2	4

案例：养老院选择偏好

分步正交设计

1. 请选择您认为可以接受的养老机构：

	区位	离家距离	空气质量	综合医院	地铁站	养老服务	月花费
A	市区	10min	很好	5min	有	很好	1500 元
B	远郊区	40min	很好	15min	无	一般	1500 元
C	都不选						

您的选择

☐☐☐

5. 请选择您认为可以接受的养老机构：

	区位	离家距离	空气质量	综合医院	地铁站	养老服务	月花费
A	近郊区	20min	很好	15min	有	很好	3000 元
B	近郊区	10min	一般	5min	有	很好	2000 元
C	都不选						

您的选择

☐☐☐

10. 请选择您认为可以接受的养老机构：

	区位	离家距离	空气质量	综合医院	地铁站	养老服务	月花费
A	远郊区	20min	一般	30min	无	很好	2000 元
B	远郊区	60min	很好	30min	无	一般	1500 元
C	都不选						

您的选择

☐☐☐

案例：养老院选择偏好

分步正交设计

假设有两个“篮子”分别装有这24个选项，每次分别从两个“篮子”中无放回地随机抽出一个选项，组成一个选择情景

区位	离家距离	空气质量	综合医院	地铁站
1	1	1	1	1
2	2	1	2	1
2	1	2	1	2
2	1	2	2	1
2	2	1	1	2
1	2	2	2	2
1	1	1	2	2
1	2	2	1	1

区位	离家距离	空气质量	综合医院	地铁站
1	1	1	1	1
2	2	1	2	1
2	1	2	1	2
2	1	2	2	1
2	2	1	1	2
1	2	2	2	2
1	1	1	2	2
1	2	2	1	1

选择集	养老机构A	养老机构B
1	1,1,1,1,1	2,1,2,2,1

1 4 2 5 3 7 8 6
4 5 1 7 8 3 6 2

案例：养老院选择偏好

分步正交设计

假设有两个“篮子”分别装有这24个选项，每次分别从两个“篮子”中无放回地随机抽出一个选项，组成一个选择情景

区位	离家距离	空气质量	综合医院	地铁站
1	1	1	1	1
2	2	1	2	1
2	1	2	1	2
2	1	2	2	1
2	2	1	1	2
1	2	2	2	2
1	1	1	2	2
1	2	2	1	1

区位	离家距离	空气质量	综合医院	地铁站
1	1	1	1	1
2	2	1	2	1
2	1	2	1	2
2	1	2	2	1
2	2	1	1	2
1	2	2	2	2
1	1	1	2	2
1	2	2	1	1

选择集	养老机构A	养老机构B
1	1,1,1,1,1	2,1,2,2,1
2	2,1,2,2,1	2,2,1,1,2
3	2,2,1,2,1	1,1,1,1,1
4	2,2,1,1,2	1,1,1,2,2

14253786

45178362

案例：养老院选择偏好

同步正交设计

■ 1个要素 + 2个备选项 → 2个 “不同” 的要素。 (14个要素参与正交)

选择情景	养老机构A							养老机构B						
	Loc	Home	Air	Hosp	Metro	Serv	Fee	Loc	Home	Air	Hosp	Metro	Serv	Fee
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	3	2	3	1	2	3	3	1	1	1	2	2	3
3	2	3	2	3	2	2	1	2	1	1	3	1	1	2
4	3	2	2	4	1	2	1	3	3	2	4	1	1	3
5	3	4	1	4	2	1	1	1	3	2	2	2	2	1
6	2	4	1	2	1	2	3	2	3	2	2	2	1	2
7	1	2	1	1	2	2	4	2	3	2	3	1	2	1
8	2	4	2	1	1	1	2	3	3	2	3	1	2	4
9	1	2	2	3	2	1	4	1	3	2	1	2	1	4
10	3	4	1	2	2	1	4	3	2	1	3	1	1	3
.....														
.....														
.....														
24	3	4	1	1	1	2	4	3	1	1	2	2	1	2
25	2	3	1	2	2	1	3	2	3	1	3	2	1	3
26	3	1	2	2	2	2	1	3	3	1	1	2	2	4
27	3	1	1	4	1	1	3	1	3	1	3	1	2	2
28	2	2	2	1	2	2	3	1	4	1	4	1	2	1
29	1	4	2	1	1	1	1	2	4	1	4	2	2	4
30	1	2	1	3	2	1	3	3	4	1	2	1	1	4
31	3	1	1	3	1	2	2	2	4	2	3	2	1	3
32	2	3	1	3	2	1	2	3	4	2	1	2	2	1
33	3	3	2	1	1	2	4	1	4	2	1	1	1	2
34	1	2	2	2	1	1	1	2	2	2	1	2	1	1
35	2	4	2	4	1	2	3	1	2	2	1	1	2	3
36	1	4	1	4	2	2	1	3	2	2	3	1	1	2

案例：养老院选择偏好

同步正交设计

■ 1个要素 + 2个备选项 → 2个 “不同” 的要素。 (14个要素参与正交)

选择情景	养老机构A							养老机构B						
	Loc A	Home A	Air A	Hosp A	Metro A	Serv A	Fee A	Loc B	Home B	Air B	Hosp B	Metro B	Serv B	Fee B
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	3	2	3	1	2	3	3	1	1	1	2	2	3
3	2	3	2	3	2	2	1	2	1	1	3	1	1	2
4	3	2	2	4	1	2	1	3	3	2	4	1	1	3
5	3	4	1	4	2	1	1	1	3	2	2	2	2	1
6	2	4	1	2	1	2	3	2	3	2	2	2	1	2
7	1	2	1	1	2	2	4	2	3	2	3	1	2	1
8	2	4	2	1	1	1	2	3	3	2	3	1	2	4
9	1	2	2	3	2	1	4	1	3	2	1	2	1	4
10	3	4	1	2	2	1	4	3	2	1	3	1	1	3
.....														
.....														
.....														
24	3	4	1	1	1	2	4	3	1	1	2	2	1	2
25	2	3	1	2	2	1	3	2	3	1	3	2	1	3
26	3	1	2	2	2	2	1	3	3	1	1	2	2	4
27	3	1	1	4	1	1	3	1	3	1	3	1	2	2
28	2	2	2	1	2	2	3	1	4	1	4	1	2	1
29	1	4	2	1	1	1	1	2	4	1	4	2	2	4
30	1	2	1	3	2	1	3	3	4	1	2	1	1	4
31	3	1	1	3	1	2	2	2	4	2	3	2	1	3
32	2	3	1	3	2	1	2	3	4	2	1	2	2	1
33	3	3	2	1	1	2	4	1	4	2	1	1	1	2
34	1	2	2	2	1	1	1	2	2	2	1	2	1	1
35	2	4	2	4	1	2	3	1	2	2	1	1	2	3
36	1	4	1	4	2	2	1	3	2	2	3	1	1	2

叙述性偏好法基本原理

同步正交设计

- 同时考虑不同备选项，一步到位生成正交设计：假设选择情景中有 M 个选项，每个选项均有 A 个要素，每个要素均有 K 个水平，则同步正交设计应当采用 $L_n(K^{MA})$ 正交表， n 为正交表的行数，每一行可以直接转译为一个选择情景。

同步正交设计

	巴士 x_1	巴士 x_2	火车 x_1	火车 x_2
1	1	1	1	1
2	1	1	1	2
3	1	2	2	1
4	1	2	2	2
5	2	1	2	1
6	2	1	2	2
7	2	2	1	1
8	2	2	1	2

叙述性偏好法基本原理

分步正交设计

- 先用相对较小的正交表生成一个备选项的各种要素水平组合（轮廓）；
- 把上述正交表复制($M - 1$)份，每一份中将行的顺序随机打乱，作为其他选项的轮廓，再拼接在原正交表的后方，使不同选项随机比较。

分步正交设计

	x_1	x_2		
1	1	1		
2	1	2		
3	2	1		
4	2	2		
	选项一 x_1	选项一 x_2	选项二 x_1	选项二 x_2
1	1	1	2	2
2	1	2	2	1
3	2	1	1	2
4	2	2	1	1

叙述性偏好法基本原理

同步 or 分步？

- 选择题的个数：分步正交设计更少（本例：24 vs. 36）。
- 正交性：同步正交设计更好。
- 不同选项可以有不同要素吗：分步法必须相同。
- Silly choices：同步设计更容易产生。

选择 情景	养老机构A							养老机构B						
	Loc	Home	Air	Hosp	Metro	Serv	Fee	Loc	Home	Air	Hosp	Metro	Serv	Fee
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
21	2	1	2	4	2	1	4	2	1	2	2	1	1	4

- 建议：
 - 无标签（养老机构A vs. 养老机构B）：两者皆可，分步法更便捷
 - 有标签（巴士 vs. 火车）：同步法

叙述性偏好法基本原理

有标签 vs. 无标签

■ ASC常数项

- Logit模型中的常数项被称为**备选项特定的常数项**(alternative specific constant, **ASC**)。有 **N 个备选项**时，即应当有 **$N-1$ 个ASC**常数项。
- 这些常数项衡量了选项之间的**固有效用差异**，即不同的备选项在所有解释变量都相同的情况下，依然存在的效用差异。
- **无标签数据** (unlabeled data) 如“养老机构A vs. 养老机构B”，选项名称无意义，选项的效用全部来自解释变量，一般不应该加入ASC常数项。
- **有标签数据** (labeled data)：如“巴士 vs. 火车”，选项名称有意义，选项的效用既来自于解释变量，也来自于标签，一般应该加入ASC常数项。

案例：养老院选择偏好

SP问卷

1. 请选择您认为可以接受的养老机构:

	区位	离家距离	空气质量	综合医院	地铁站	养老服务	月花费	您的选择
A	市区	10min	很好	5min	有	很好	1500 元	☐
B	远郊区	40min	很好	15min	无	一般	1500 元	☐
C	都不选							☐

5. 请选择您认为可以接受的养老机构:

	区位	离家距离	空气质量	综合医院	地铁站	养老服务	月花费	您的选择
A	近郊区	20min	很好	15min	有	很好	3000 元	☐
B	近郊区	10min	一般	5min	有	很好	2000 元	☐
C	都不选							☐

10. 请选择您认为可以接受的养老机构:

	区位	离家距离	空气质量	综合医院	地铁站	养老服务	月花费	您的选择	
A	远郊区	20min	一般	30min	无	很好	2000 元		☐
B	远郊区	60min	很好	30min	无	一般	1500 元		☐
C	都不选								☐

案例：养老院选择偏好

收集和整理选择数据

[illegible]

案例：养老院选择偏好

“都不选”选项

受访者	选择情景	备选项	选择结果	Loc1	Loc2	Home	Air	Hosp	Metro	Serv	Fee	Neither
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1	1	A	1	0	0	10	0	5	0	0	1500	0
1	1	B	0	0	1	40	0	15	1	1	1500	0
1	1	C	0	0	0	0	0	0	0	0	0	1

EXAMPLE	Alternative A	Alternative B	Alternative C	Alternative D
Brand	McCain's	Old South	Generic	NONE OF THESE PRODUCTS
Grade	A	C	C	
Sweetness	Unsweetened	Sweetened	Sweetened	
Sale Offer	Package of 4	Unit	Unit	
Total Price/Package	\$3.75	\$1.75	\$1.00	
I Would Purchase.... (✓Check Only One)	☐ ⇓	☐ ⇓	☐ ⇓	☐
How Many of the Selected Product?	_____ packages of 4	_____ cans	_____ cans	

Features of hunting area	Site A	Site B	
Distance from home to hunting area	50 kilometres	50 kilometres	
Quality of road from home to hunting area	Mostly gravel or dirt, some paved	Mostly paved, some gravel or dirt	
Access within hunting area	Newer trails, cutlines or seismic lines, passable with a 2WD vehicle	Newer trails, cutlines or seismic lines passable with a 4WD truck	
Encounters with other hunters	No hunters, other than those in my hunting party, are encountered	Other hunters, on ATVs, are encountered	
Forestry activity	Some evidence of recent logging found in the area	No evidence of logging	
Moose population	Evidence of less than 1 moose per day	Evidence of less than 1 moose per day	
Tick ONE box only:	☐ A	☐ B	☐ Neither

案例：养老院选择偏好

“都不选”选项

受访者	选择情景	备选项	选择结果	Loc1	Loc2	Home	Air	Hosp	Metro	Serv	Fee	Neither
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1	1	A	1	0	0	10	0	5	0	0	1500	0
1	1	B	0	0	1	40	0	15	1	1	1500	0
1	1	C	0	0	0	0	0	0	0	0	0	1

■ Neither: 在正常选项取0，在 “都不选” 选项上取1。

案例：养老院选择偏好

“都不选”选项

受访者	选择情景	备选项	选择结果	Loc1	Loc2	Home	Air	Hosp	Metro	Serv	Fee	Neither
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1	1	A	1	0	0	10	0	5	0	0	1500	0
1	1	B	0	0	1	40	0	15	1	1	1500	0
1	1	C	0	0	0	0	0	0	0	0	0	1

- Neither：在正常选项取0，在“都不选”选项上取1。
- 其他解释变量：在“都不选”选项上一律取0。

案例：养老院选择偏好

“都不选” 选项

受访者	选择情景	备选项	选择结果	Loc1	Loc2	Home	Air	Hosp	Metro	Serv	Fee	Neither
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1	1	A	1	0	0	10	0	5	0	0	1500	0
1	1	B	0	0	1	40	0	15	1	1	1500	0
1	1	C	0	0	0	0	0	0	0	0	0	1

- Neither: 在正常选项取0，在 “都不选” 选项上取1。
- 其他解释变量：在 “都不选” 选项上一律取0。

$$V = \beta_{Loc1}Loc1 + \beta_{Loc2}Loc2 + \beta_{Home}Home + \cdots + \beta_{Fee}Fee + \beta_{Neither}Neither$$

$$V_A = \beta_{Loc1}Loc1_A + \beta_{Loc2}Loc2_A + \beta_{Home}Home_A + \cdots + \beta_{Fee}Fee_A + \beta_{Neither}Neither_A$$

$$V_C = \beta_{Loc1}Loc1_C + \beta_{Loc2}Loc2_C + \beta_{Home}Home_C + \cdots + \beta_{Fee}Fee_C + \beta_{Neither}Neither_C$$

案例：养老院选择偏好

“都不选”选项

受访者 (1)	选择情景 (2)	备选项 (3)	选择结果 (4)	Loc1 (5)	Loc2 (6)	Home (7)	Air (8)	Hosp (9)	Metro (10)	Serv (11)	Fee (12)	Neither (13)
1	1	A	1	0	0	10	0	5	0	0	1500	0
1	1	B	0	0	1	40	0	15	1	1	1500	0
1	1	C	0	0	0	0	0	0	0	0	0	1

- Neither: 在正常选项取0，在“都不选”选项上取1。
- 其他解释变量：在“都不选”选项上一律取0。

$$V = \beta_{Loc1}Loc1 + \beta_{Loc2}Loc2 + \beta_{Home}Home + \cdots + \beta_{Fee}Fee + \beta_{Neither}Neither$$

$$V_A = \beta_{Loc1}Loc1_A + \beta_{Loc2}Loc2_A + \beta_{Home}Home_A + \cdots + \beta_{Fee}Fee_A + \beta_{Neither} \times 0$$

$$V_C = \beta_{Loc1} \times 0 + \beta_{Loc2} \times 0 + \beta_{Home} \times 0 + \cdots + \beta_{Fee} \times 0 + \beta_{Neither} \times 1$$

案例：养老院选择偏好

Logit模型估计

样本量(N)	1200	变量个数(K)	9	
对数似然值(LL)	-1061.67	r^2	0.1947	
变量	系数(Coefficient)	标准误(S.E.)	z 值	相伴概率(p 值)
β_{Loc1}	-0.5598	0.1139	-4.9158	<0.0001
β_{Loc2}	-1.8831	0.1348	-13.9731	<0.0001
β_{Home}	-0.0107	0.0028	-3.8344	<0.0001
β_{Air}	-0.2984	0.1013	-2.9450	0.0032
β_{Hosp}	-0.0295	0.0038	-7.8408	<0.0001
β_{Metro}	-0.4769	0.1065	-4.4781	<0.0001
β_{Serv}	-0.8801	0.1022	-8.6125	<0.0001
β_{Fee}	-0.0008	0.0001	-8.8116	<0.0001
$\beta_{Neither}$	-4.6776	0.3115	-15.018	<0.0001

案例：养老院选择偏好

Logit模型解读

- 到最近综合医院的距离每减少1分钟，老年人愿意增加的月花费为 $(-0.0295) \div (-0.0008) = 35.24$ 元。
- 当养老机构的区位由市区变为近郊区时，如果离家距离也相应地增加30分钟，要使养老机构的总效用保持不变：
 - 需要月花费降低 $(-0.0107 \times 30 - 0.5598) \div (-0.0008) = 1053$ 元。
 - 到综合医院的距离减少 $(-0.0107 \times 30 - 0.5598) \div (-0.0295) = 29.88$ 分钟。
- 最后，“都不选”的效用为-4.6774，其作用相当于可接受阈值，当养老机构的效用水平低于该阈值时，老年人倾向于拒绝该养老机构。

案例：养老院选择偏好

Logit模型解读

要素		系数	养老机构		
			市区	近郊	远郊
区位	近郊	-0.5598	0	1	0
	远郊	-1.8831	0	0	1
离家距离		-0.0107	10min	40min	60min
空气质量		-0.2984	一般	一般	很好
综合医院		-0.0295	10min	20min	30min
地铁站		-0.4769	有	无	无
养老服务		-0.8801	一般	很好	很好
月花费		-0.0008	3000元	2500元	2000元

案例：养老院选择偏好

Logit模型解读

要素		系数	养老机构		
			市区	近郊	远郊
区位	近郊	-0.5598	0	1	0
	远郊	-1.8831	0	0	1
离家距离		-0.0107	10min	40min	60min
空气质量		-0.2984	一般	一般	很好
综合医院		-0.0295	10min	20min	30min
地铁站		-0.4769	有	无	无
养老服务		-0.8801	一般	很好	很好
月花费		-0.0008	3000元	2500元	2000元
效用			-4.0930	-4.4479	-5.5639
选择概率	三者相互比较		52%	36%	12%
	各自与“都不选”比较		64%	56%	29%

小结

- SP的优势：突破现实、统计性质好（要素不相关）、**方便快捷**。
- 研究问题 → 预调研 → **确定要素和水平** → **正交设计** → 发问卷 → 录入数据
 - 正交设计：均匀分散、整齐可比 ⇒ 比较充分+题数少
 - 具体技术：分步式 or 同步式
 - 标签效应（ASC常数项）
 - “都不选” 选项的处理
- SP调查（数据） + Logit模型（解释） + 个体行为模拟（预测）