

1. 概述

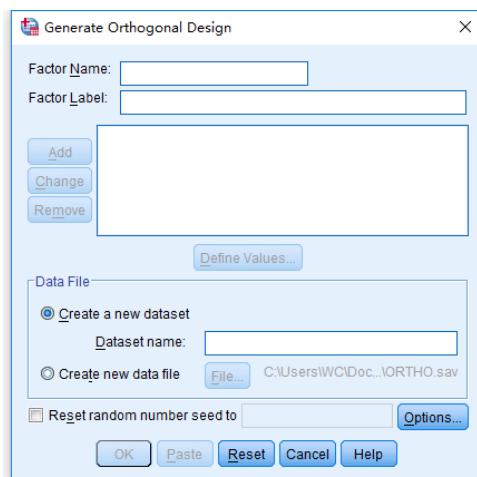
本节以老年人选择养老院为例，演示叙述性偏好法（以下简称 SP 法）的操作过程。SP 法的操作主要包括两个部分：实验设计和选择模型估计，由于后者已经在离散选择模型中介绍，因此本节主要关注实验设计，具体而言是正交设计。

有许多统计软件都可以实现正交设计，SPSS 是最常用的软件之一。需要注意的是，这些软件并不是专门面向 SP 选择实验的，有时会出现一些小问题，不过在不严格要求的情况下，其结果也能基本满足我们的需要。如果希望进一步提高实验设计的质量，可以使用 Ngene 软件。该软件是 NLogit 属于同一系列，是专门面向 SP 选择的实验设计软件，除了正交设计以外，还提供高效设计等更前沿的实验设计功能。本节将同时演示 SPSS 与 Ngene 中正交设计的方法。后者需要用到的代码文件为“sp_seq.ngs”和“sp_sim.ngs”。

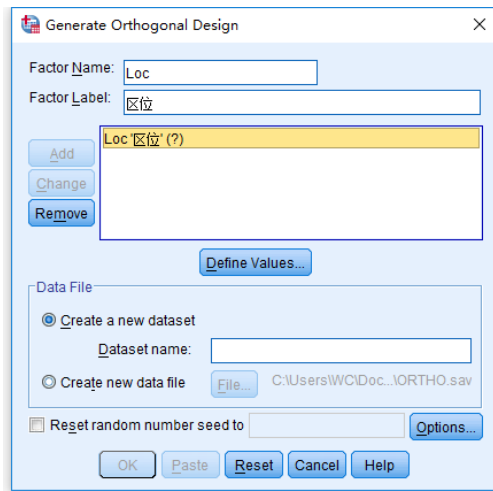
在实验设计完成后需要用 SP 问卷收集选择数据。本节的数据文件为“sp_data.csv”。之后，我们将使用 NLogit 软件对该数据估计 Logit 模型，使用的代码文件为“sp_logit.lim”。

2. 用 SPSS 进行分步正交设计

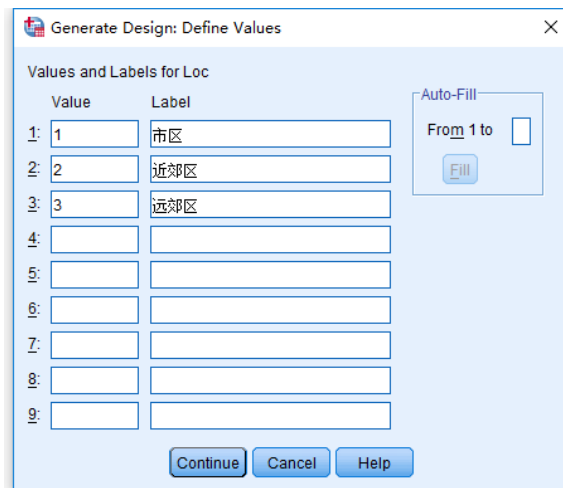
打开 SPSS 软件，在菜单栏中选择“Data – Orthogonal Design – Generate”，打开正交设计生成对话框。



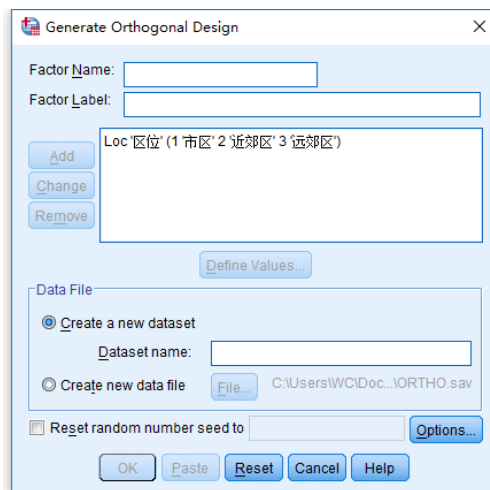
下面添加第一个要素——区位。在“Factor Name”中输入“Loc”，在“Factor Label”中输入“区位”，然后点击“Add”，将其选入下方的要素列表中，此时呈现的是“Loc '区位' (?)”，如下所示。后面的“(?)”代表该要素还未定义水平。



在要素列表中选中“Loc '区位' (?)”，点击下方的“Define Values”，为区位要素定义水平，如下图所示。“Value”和“Label”分别是各水平的取值编码和标签，这里共定义了3个水平：“1：市区”、“2：近郊区”、“3：远郊区”。



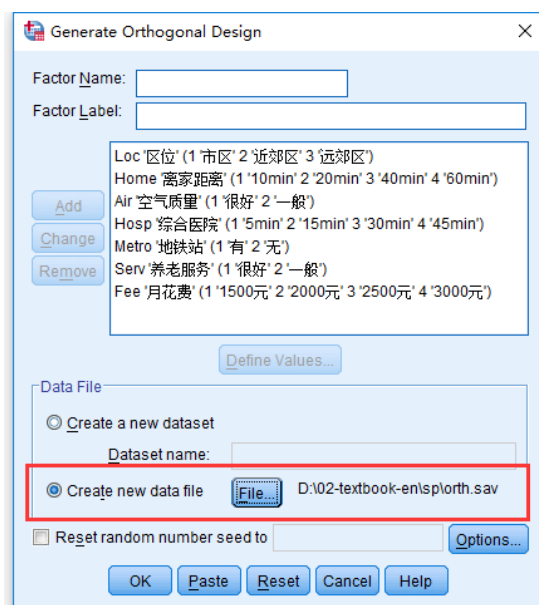
点击“Continue”回到要素列表对话框中，可以看到，此时的区位要素已变为“Loc '区位' (1 '市区' 2 '近郊区' 3 '远郊区)’”。



类似地输入其他要素和相应的水平，要素水平设置参照课件。所有要素设置完后的效果如下图所示。



在“Data File”中选中“Create new data file”，并点击“File”按钮，选择数据文件保存的位置，该文件的内容即是正交设计的结果。下图中，文件被保存至“D:\02-textbook-en\sp\orth.sav”。



点击“OK”运行分析，在 SPSS 的结果窗口中显示如下信息。这里的“16 cards”即生成了 16 个备选项轮廓。

Warnings

A plan is successfully generated with 16 cards.

在刚才指定的路径找到并打开新生成的数据文件“orth.sav”，如下所示。这结果就是一个备选项（设为养老机构 A）的所有轮廓。

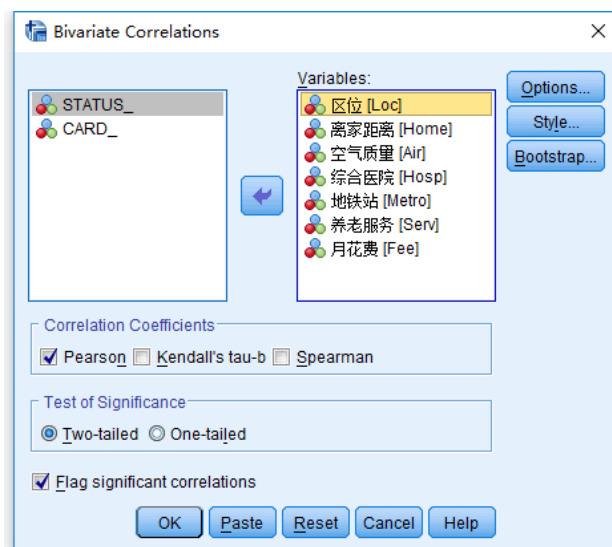
Loc	Home	Air	Hosp	Metro	Serv	Fee	STATUS_	CARD_
近郊区	10min	很好	15min	无	一般	2500元	Design	1
近郊区	60min	一般	30min	无	很好	1500元	Design	2
远郊区	40min	一般	5min	无	很好	2500元	Design	3
市区	10min	很好	5min	有	很好	1500元	Design	4
市区	10min	一般	45min	无	很好	2000元	Design	5
远郊区	60min	很好	15min	有	很好	2000元	Design	6
远郊区	10min	一般	30min	有	一般	3000元	Design	7
市区	60min	很好	5min	无	一般	3000元	Design	8
市区	60min	一般	45min	有	一般	2500元	Design	9
市区	40min	很好	30min	无	一般	2000元	Design	10
市区	20min	一般	15min	无	很好	3000元	Design	11
近郊区	20min	一般	5min	有	一般	2000元	Design	12
近郊区	40min	很好	45min	有	很好	3000元	Design	13
远郊区	20min	很好	45min	无	一般	1500元	Design	14
市区	20min	很好	30min	有	很好	2500元	Design	15
市区	40min	一般	15min	有	一般	1500元	Design	16

可以看到：SPSS 共生了 16 个备选项轮廓，前面 7 列均是实验设计的要素，例如，第 1 行的备选项即为：“位于近郊区+距离家 10min+空气质量很好+距离综合医院 15min+附近无地铁站+养老服务一般+月花费 2500 元”。“STATUS_”列提示的是备选项的状态，可以忽略，“CARD_”列是备选项的编号。

对于分步正交设计而言，在获得了养老机构 A 的轮廓表后，下一步将其复制一份并随机打乱，即得到养老机构 B 的轮廓表，然后将 A、B 两表每一行对应的轮廓进行比较，即得到 16 个选择情景。

然而，读者应当注意到，课件中给出的正交设计矩阵中共有 24 行，而非 SPSS 得到的 16 行。事实上，SPSS 的结果不满足要素水平均衡，不是一个完美的结果，下面我们进行验证。

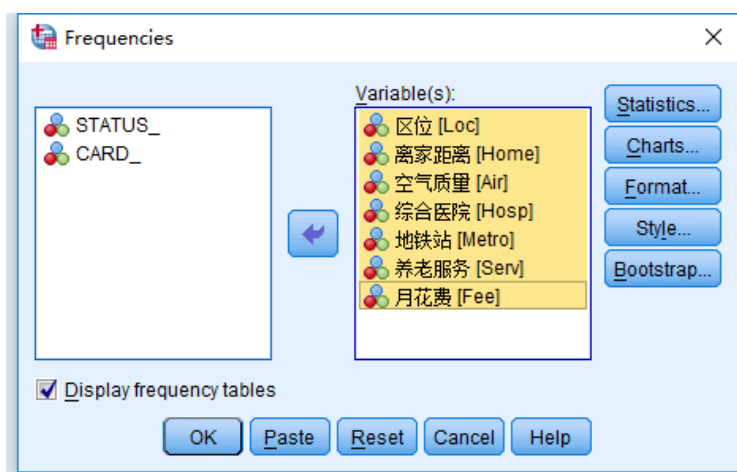
首先，在“orth.sav”数据文件的菜单栏中选择“Analyze – Correlate – Bivariate”，打开相关分析对话框。将 7 个要素选入“Variables”中，如下所示。



点击“OK”运行相关分析，结果如下所示。可以看到，7个要素两两之间的Pearson相关系数均为0，满足互不相关的要求。

		区位	离家距离	空气质量	综合医院	地铁站	养老服务	月花费
区位	Pearson Correlation	1	.000	.000	.000	.000	.000	.000
	Sig. (2-tailed)		1.000	1.000	1.000	1.000	1.000	1.000
	N	16	16	16	16	16	16	16
离家距离	Pearson Correlation	.000	1	.000	.000	.000	.000	.000
	Sig. (2-tailed)	1.000		1.000	1.000	1.000	1.000	1.000
	N	16	16	16	16	16	16	16
空气质量	Pearson Correlation	.000	.000	1	.000	.000	.000	.000
	Sig. (2-tailed)	1.000	1.000		1.000	1.000	1.000	1.000
	N	16	16	16	16	16	16	16
综合医院	Pearson Correlation	.000	.000	.000	1	.000	.000	.000
	Sig. (2-tailed)	1.000	1.000	1.000		1.000	1.000	1.000
	N	16	16	16	16	16	16	16
地铁站	Pearson Correlation	.000	.000	.000	.000	1	.000	.000
	Sig. (2-tailed)	1.000	1.000	1.000	1.000		1.000	1.000
	N	16	16	16	16	16	16	16
养老服务	Pearson Correlation	.000	.000	.000	.000	.000	1	.000
	Sig. (2-tailed)	1.000	1.000	1.000	1.000	1.000		1.000
	N	16	16	16	16	16	16	16
月花费	Pearson Correlation	.000	.000	.000	.000	.000	.000	1
	Sig. (2-tailed)	1.000	1.000	1.000	1.000	1.000	1.000	
	N	16	16	16	16	16	16	16

我们再检查要素水平均衡的要求。在菜单栏中选择“Analyze – Descriptive Statistics – Frequencies”，打开频数统计对话框，将7个要素选入“Variable(s)”中，如下所示。



点击“OK”运行分析，区位要素的统计结果如下。

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 市区	8	50.0	50.0	50.0
近郊区	4	25.0	25.0	75.0
远郊区	4	25.0	25.0	100.0
Total	16	100.0	100.0	

可以看到：在 16 个备选项轮廓中，市区出现了 8 次，近郊区与远郊区各自只出现了 4 次，因此不满足要素水平均衡的要求。除了区位以外，其他要素则均满足要求。

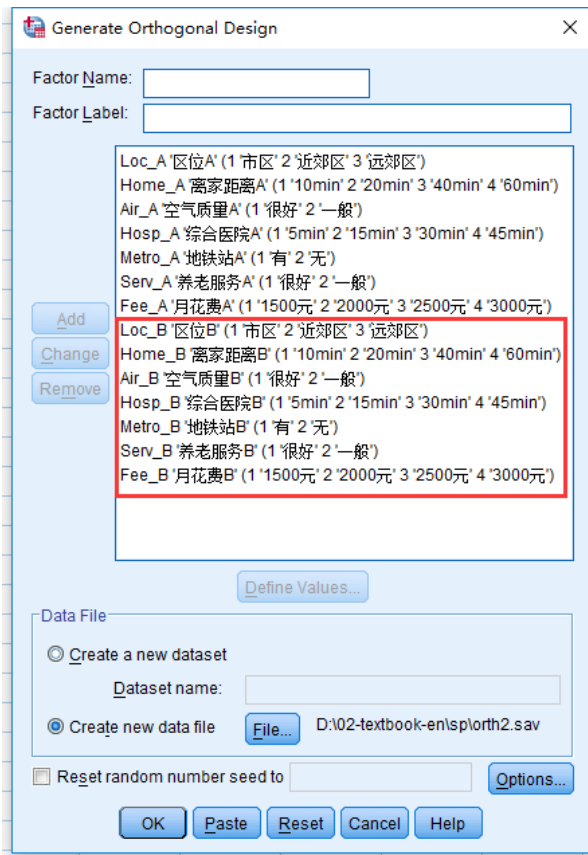
由此可见，SPSS 在进行正交实验设计时，可能会出现要素水平不均衡的问题，特别是同时存在 3 水平要素和 4 水平要素时，该问题尤为常见。在实际应用中，研究者可以尽量避免 3 水平要素的使用，统一使用 2 水平和 4 水平要素，并注意对 SPSS 生成结果的检验。另一方面，即使出现了这一问题，在不严格要求的情况下也可以使用该结果收集数据，毕竟在整个 SP 的全过程中有许多环节都可能影响正交性，与没有任何控制的 RP 数据相比，这样的数据仍然能很好地改善解释变量相关的问题。

3. 用 SPSS 进行同步正交设计

同步正交设计的操作与分步正交设计类似，只是将要素的数据增加一倍，以区分养老机构 A 和 B 的要素。在 SPSS 的菜单栏中选择“Data – Orthogonal Design – Generate”，打开正交设计生成对话框，用相同的方法设置 7 个要素及相应的水平，只是在要素名称和标签中加入后缀“A”，表示其属于养老机构 A，如下所示。



用相同的方法重复设置这 7 个要素及相应的水平，并在要素名称和标签中加入后缀“B”，表示这些要素属于养老机构 B，如下所示。同时，在“Data File”中选“Create new data file”，将生成的结果保存至指定的数据文件中。



点击“OK”，SPSS 结果窗口显示生成了 64 组备选项轮廓。

Warnings

A plan is successfully generated with 64 cards.

打开生成的数据文件，如下所示，这里仅显示了前 20 行。可以看到，第 1-7 列（Loc_A ~ Fee_A）是养老机构 A 的要素水平，第 8-14 列（Loc_B ~ Fee_B）是养老机构 B 的要素水平，每一行从中间断开，自然形成了 A、B 两个备选项，因此 SPSS 共生了 64 个选择情景。

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	STATUS	CARD
远郊区	10min	很好	15min	无	很好	2500元	远郊区	10min	一般	5min	有	很好	3000元	Design	1
市区	40min	很好	45min	无	一般	3000元	近郊区	60min	很好	15min	无	很好	3000元	Design	2
市区	20min	一般	30min	无	很好	1500元	市区	20min	很好	15min	无	一般	3000元	Design	3
市区	20min	很好	5min	有	很好	2500元	近郊区	60min	很好	5min	有	很好	2000元	Design	4
市区	60min	一般	15min	有	很好	2500元	市区	10min	很好	15min	有	很好	3000元	Design	5
市区	40min	一般	15min	有	很好	1500元	远郊区	60min	很好	15min	有	很好	2500元	Design	6
市区	10min	一般	30min	无	一般	1500元	远郊区	20min	一般	45min	有	很好	3000元	Design	7
近郊区	60min	很好	30min	有	很好	2000元	近郊区	60min	很好	45min	有	一般	1500元	Design	8
市区	20min	一般	30min	无	一般	2500元	市区	40min	一般	45min	有	很好	2500元	Design	9
市区	60min	一般	15min	有	一般	3000元	近郊区	40min	很好	5min	有	一般	1500元	Design	10
市区	10min	一般	30min	无	很好	2000元	市区	60min	一般	30min	有	一般	1500元	Design	11
远郊区	10min	很好	15min	无	一般	1500元	近郊区	60min	很好	30min	无	一般	2500元	Design	12
市区	10min	一般	30min	无	很好	2500元	近郊区	40min	很好	15min	无	一般	2500元	Design	13
近郊区	40min	一般	5min	无	很好	2500元	市区	60min	一般	15min	有	很好	2000元	Design	14
近郊区	40min	一般	5min	无	一般	1500元	市区	10min	很好	45min	无	一般	1500元	Design	15
远郊区	60min	一般	5min	无	一般	2500元	远郊区	60min	很好	45min	无	一般	2000元	Design	16
近郊区	60min	很好	30min	有	很好	2500元	市区	40min	一般	5min	无	一般	2500元	Design	17
市区	20min	很好	5min	有	一般	1500元	远郊区	10min	一般	30min	无	一般	1500元	Design	18
远郊区	60min	一般	5min	无	一般	2000元	市区	40min	一般	5min	有	一般	3000元	Design	19
远郊区	40min	很好	30min	有	一般	2500元	远郊区	40min	很好	30min	有	很好	2500元	Design	20

读者可以用前述的方法验证，上表中所有的 14 个要素两两之间的 Pearson 相

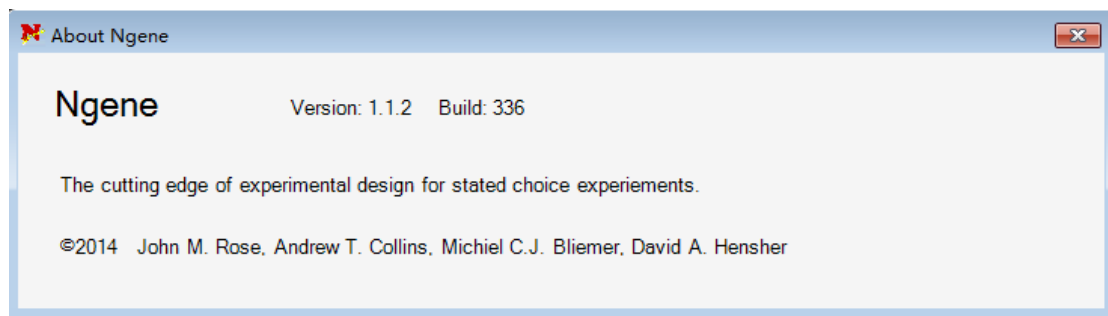
关系数均为 0, Loc_A 与 Loc_B 这两个 3 水平要素不满足要素水平均衡 (如下所示), 其余的 2 水平或 4 水平要素则满足要素水平均衡。另外, 该结果与课件中的结果并不一致, 课件中只需要 36 个选择情景, 表明 SPSS 的结果不够精炼。

区位A

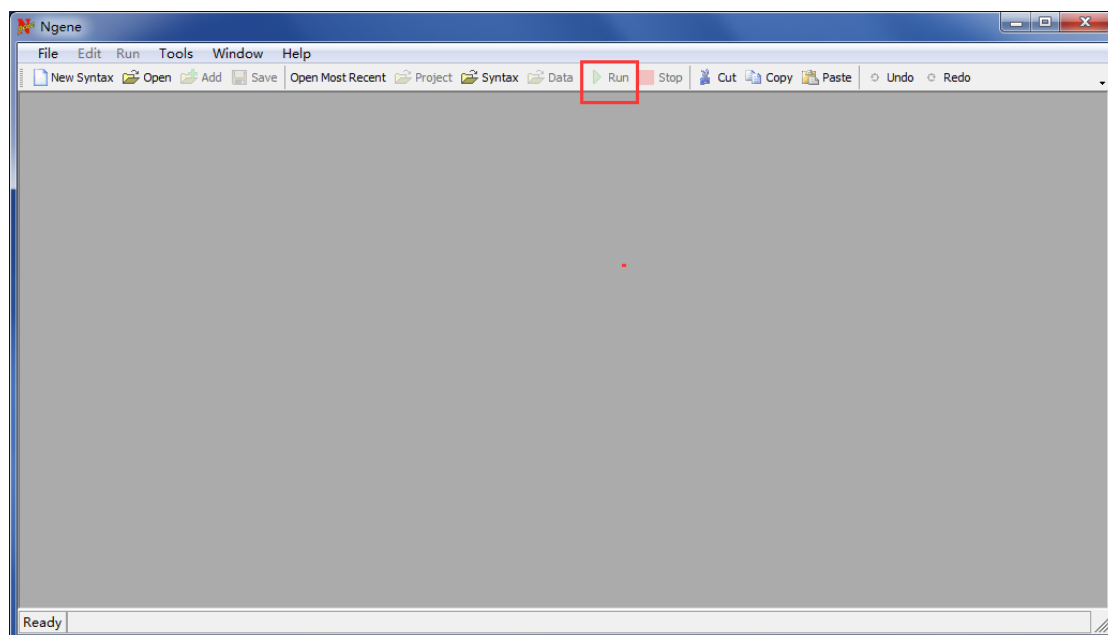
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 市区	32	50.0	50.0	50.0
近郊区	16	25.0	25.0	75.0
远郊区	16	25.0	25.0	100.0
Total	64	100.0	100.0	

4. 用 Ngene 进行分步正交设计

Ngene 是 NLogit 同系列的软件, 专门用于 SP 选择的实验设计。



打开 Ngene, 其界面如下所示。我们通过编写代码文件, 然后点击红框中的“Run”按钮运行代码, 得到实验设计结果。



Ngene 代码文件的后缀名为 “.ngs”, 其本质上是文本文件, 可以通过记事本

等文本工具打开和编辑。本小节用的到代码文件为“sp_seq.ngs”，在工具栏中点击“Open”，然后打开该文件，如下图所示。

```
design
;alts=a, b
;rows=16
;orth=seq
;model:
u(a)=b1*loc[1,2,3] + b2*home[1,2,3,4] + b3*air[1,2] + b4*hosp[1,2,3,4] + b5*metro[1,2] + b6*serv[1,2] + b7*fee[1,2,3,4]/
u(b) = b1*loc + b2*home + b3*air + b4*hosp + b5*metro + b6*serv + b7*fee
$
```

我们也可以在工具栏点击“New Syntax”，在新建代码文件中输入上述代码。

Ngene 的代码格式与 NLogit 非常相似：(1) 以一个命令开始，这里为“design”；(2) 最后以美元符号“\$”结束；(3) 中间是对该命令的一系列选项的设置，每个选项均以“;”开始。

可以看到，上图中“design”命令共有 4 个选项：

第 2 行中的“alts”选项用于设置备选项名称，这里的“;alt=a,b”即设置了 2 个备选项——“a”和“b”，分别代表养老机构 A 和养老机构 B。

第 3 行的“rows”选项用于设置行数，即选择情景的期望数量，Ngene 将会在满足实验设计要求（如正交设计中的正交性）的情况下，尽量靠近该数值，但是不一定完全相等。通常而言，研究者不必担心行数的精确设置，只需要输入一个大概的期望即可，如果行数过少，无法满足最基本的需求，Ngene 将报错，此时将行数调高，再重新运行。

第 4 行的“orth”选项要求 Ngene 使用正交设计，同时，还用于设置正交设计的方法，这里的“seq”代表 sequential，即分步正交设计。

第 5 行的“model”选项用于设置要素及水平。与 SPSS 等软件不同，Ngene 是通过定义选择模型中备选项的效用来设置要素和水平的，其格式如下。请注意在“model”之后需加引号“:”，在两个备选项的效用公式之间需加斜杠“/”。

model:

$u(\text{备选项 1 的名称}) = \text{系数名称} * \text{要素 1 名称}[\text{要素 1 的水平名称}] + \text{系数名称} * \text{要素 2 名称}[\text{要素 2 的水平名称}] + \dots /$

$u(\text{备选项 2 的名称}) = \text{系数名称} * \text{要素 1 名称}[\text{要素 1 的水平名称}] + \text{系数名称} * \text{要素 2 名称}[\text{要素 2 的水平名称}] + \dots /$

.....

本例中，养老机构 A 的效用定义如下式：

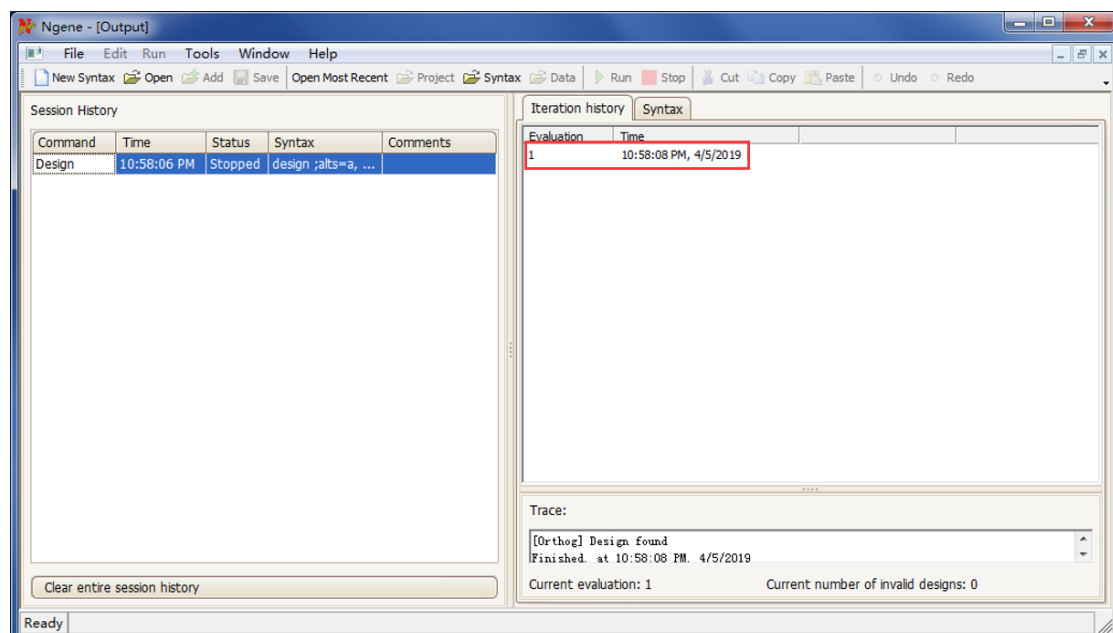
$u(a) = b1 * \text{loc}[1,2,3] + b2 * \text{home}[1,2,3,4] + b3 * \text{air}[1,2] + b4 * \text{hosp}[1,2,3,4] + b5 * \text{metro}[1,2] + b6 * \text{serv}[1,2] + b7 * \text{fee}[1,2,3,4]$

可以看到，式中共有 7 个要素：loc、home、air、hosp、metro、serv、fee，b1~b7 是它们各自的系数。“loc[1,2,3]”表明“loc”要素共有 3 个水平，分别编码为 1、2、3；“home[1,2,3,4]”表明“home”要素共有 4 个水平，分别编码为 1、2、3、4。Ngene 不需要输入每个要素和每个水平编码的标签意义。

需要注意的是，如果已经在前面的备选项中定义了某个要素的水平，那么后面的备选项在使用该要素时，不需要再重复地写出该要素的水平，除非水平的定义发生变化。本例中，养老机构 A 与养老机构 B 使用的要素相同，且水平设置相同，那么既然在“u(a)”中已经定义了 7 个要素的水平，在“u(b)”中就只需要写出要素名称，后面“[]”内的水平设置可以省略，如下式。

$u(b) = b1 * \text{loc} + b2 * \text{home} + b3 * \text{air} + b4 * \text{hosp} + b5 * \text{metro} + b6 * \text{serv} + b7 * \text{fee}$

选中这段代码，点击工具栏中的“Run”按钮执行分析，弹出如下结果窗口。

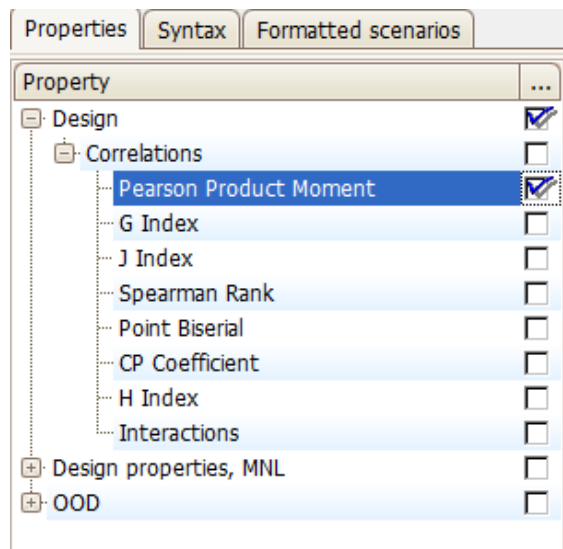


双击上图中红框的位置，打开结果窗口，如下所示。

Choice situation	a.loc	a.home	a.air	a.hosp	a.metro	a.serv	a.fee	b.loc	b.home	b.air	b.hosp	b.metro	b.serv	b.fee
1	1	1	1	1	1	1	1	1	3	2	3	1	2	2
2	1	1	2	4	2	1	3	2	2	1	2	1	2	4
3	1	3	2	3	1	2	2	1	4	1	1	2	2	4
4	1	3	2	2	2	2	3	1	1	2	4	2	1	3
5	2	2	1	2	1	2	4	2	3	1	1	2	1	2
6	2	3	1	1	2	1	2	3	4	1	3	2	2	1
7	2	4	1	4	1	1	1	1	2	1	3	2	1	4
8	2	4	2	2	2	1	4	3	2	2	1	1	2	3
9	3	1	1	4	1	2	4	3	1	2	2	2	1	2
10	3	2	2	3	2	1	2	3	3	1	2	2	2	1
11	3	2	2	1	1	2	3	2	4	1	4	1	1	1
12	3	4	1	3	2	2	1	2	1	1	3	2	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	2	2	4	2	2	1
15	1	2	1	2	1	1	1	2	1	2	1	1	2	2
16	1	2	1	3	2	1	4	1	3	2	2	2	2	3
17	2	3	2	3	1	1	3	2	4	2	2	2	1	4
18	2	2	2	4	2	2	1	3	4	2	1	1	1	3
19	2	1	2	1	1	2	2	1	1	1	1	1	1	1
20	2	1	1	3	2	2	3	2	3	2	3	1	1	3
21	3	4	2	1	1	1	3	3	1	1	4	1	2	4
22	3	3	1	2	2	2	1	1	4	2	4	1	2	2
23	3	3	1	4	1	1	4	1	2	1	2	1	1	1
24	3	1	2	2	2	1	2	3	2	2	3	2	1	2

表中，“Choice situation”表示选择情景，共有 24 个。后面各列中，“a.loc ~ a.fee”是养老机构 A 的要素水平，“b.loc ~ b.fee”是养老机构 B 的要素水平。Ngene 在后台自动完成了以下过程：首先生成了 A 的 24 个轮廓，然后复制一份并随机打乱顺序，得到 B 的轮廓，最后再进行组合比较。该结果即是课件中呈现的结果，只是由于随机原因，每次运行时 B 的轮廓出现的顺序有所不同。

在左侧的“Property”中展开并勾选“Design – Correlation – Pearson Product Moment”，如下所示。



右侧结果栏的下方将出现 Pearson 相关系数：

Correlations (Pearson Product Moment)														
Attribute	a.loc	a.home	a.air	a.hosp	a.metro	a.serv	a.fee	b.loc	b.home	b.air	b.hosp	b.metro	b.serv	b.fee
a.loc	1	0	0	0	0	0	0	0.3125	-0.045644	-0.204124	0.136931	0	-0.204124	-0.319505
a.home	0	1	0	0	0	0	0	0.045644	-0.033333	-0.223607	0.333333	0.149071	0.223607	0.3
a.air	0	0	1	0	0	0	0	0.306186	0.074536	-0.166667	-0.074536	-0.166667	-0.166667	0.298142
a.hosp	0	0	0	1	0	0	0	0.045644	0.033333	-0.074536	-0.266667	0.074536	-0.298142	0.466667
a.metro	0	0	0	0	1	0	0	0.204124	0.074536	0.333333	0.149071	0.166667	0.333333	-0.074536
a.serv	0	0	0	0	0	1	0	-0.102062	0.074536	0	0.149071	0	-0.333333	-0.074536
a.fee	0	0	0	0	0	0	1	0.091287	-0.133333	0.074536	-0.066667	0.149071	-0.149071	-0.1
b.loc	0.3125	0.045644	0.306186	0.045644	0.204124	-0.102062	0.091287	1	0	0	0	0	0	0
b.home	-0.045644	-0.033333	0.074536	0.033333	0.074536	0.074536	-0.133333	0	1	0	0	0	0	0
b.air	-0.204124	-0.223607	-0.166667	-0.074536	0.333333	0	0.074536	0	0	1	0	0	0	0
b.hosp	0.136931	0.333333	-0.074536	-0.266667	0.149071	0.149071	-0.066667	0	0	0	1	0	0	0
b.metro	0	0.149071	-0.166667	0.074536	0.166667	0	0.149071	0	0	0	0	1	0	0
b.serv	-0.204124	0.223607	-0.166667	-0.298142	0.333333	-0.333333	-0.149071	0	0	0	0	0	1	0
b.fee	-0.319505	0.3	0.298142	0.466667	-0.074536	-0.074536	-0.1	0	0	0	0	0	0	1

可以看到，养老机构 A 内部的 7 个要素之间，以及养老机构 B 内部的 7 个要素之间的相关系数均为 0，而跨选项要素的相关系数不为 0。这是分步正交设计的特点。对于无标签数据，跨选项的正交性不那么重要。另外，读者可以自行验证，所有的 14 个要素均满足要素水平均衡。

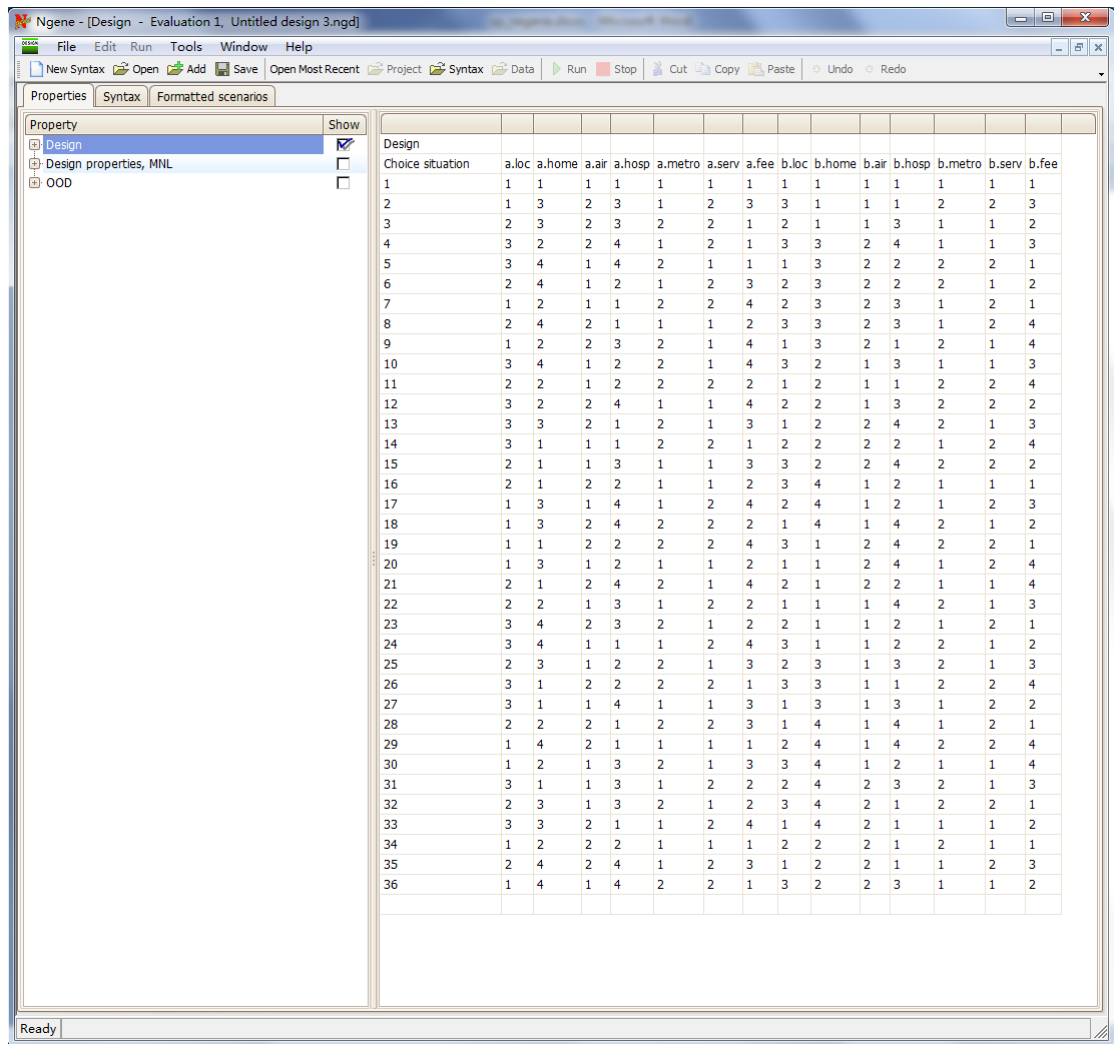
5. 用 Ngene 进行同步正交设计

在 Ngene 打开同步正交设计的代码“sp_sim.ngs”，如下所示。

```

design
:alts=a, b
:rows=16
:orth=sim
:model:
u(a)=b1*loc[1,2,3] + b2*home[1,2,3,4] + b3*air[1,2] + b4*hosp[1,2,3,4] + b5*metro[1,2] + b6*serv[1,2] + b7*fee[1,2,3,4]/
u(b) = b1*loc + b2*home + b3*air + b4*hosp + b5*metro + b6*serv + b7*fee
$
    
```

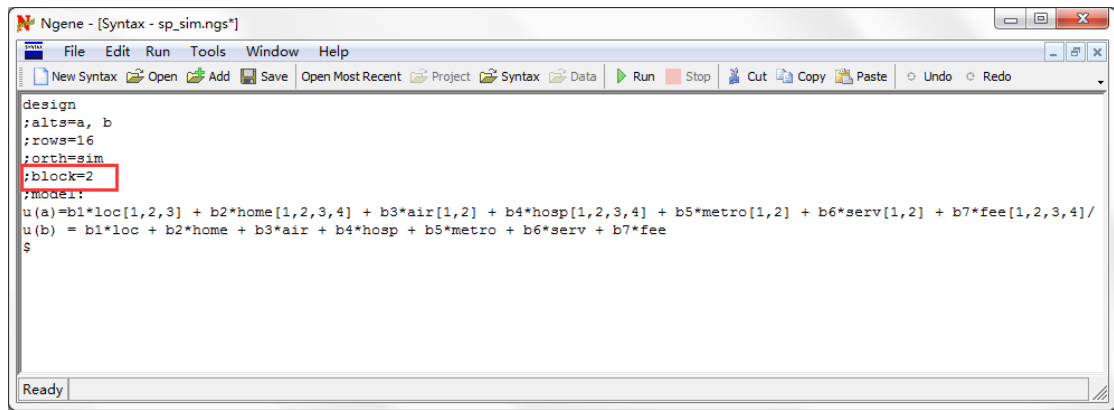
可以看到，与分步正交设计相比，同步正交设计只需要将“orth”选项由“seq”改为“sim”即可，后者是 simultaneous 的缩写。运行代码后，Ngene 得到如下所示的结果。该结果共有 36 个选择情景，与课件中相同，其解读方式与分步正交设计完全相同。



用类似的方法检查全部 Pearson 相关系数，如下所示。可以看到，14 个要素两两之间的相关性为 0。另外，读者可自行验证，所有要素均满足要素水平均衡。

Correlations (Pearson Product Moment)														
Attribute	a.loc	a.home	a.air	a.hosp	a.metro	a.serv	a.fee	b.loc	b.home	b.air	b.hosp	b.metro	b.serv	b.fee
a.loc	1	0	0	0	0	0	0	0	0	0	0	0	0	0
a.home	0	1	0	0	0	0	0	0	0	0	0	0	0	0
a.air	0	0	1	0	0	0	0	0	0	0	0	0	0	0
a.hosp	0	0	0	1	0	0	0	0	0	0	0	0	0	0
a.metro	0	0	0	0	1	0	0	0	0	0	0	0	0	0
a.serv	0	0	0	0	0	1	0	0	0	0	0	0	0	0
a.fee	0	0	0	0	0	0	1	0	0	0	0	0	0	0
b.loc	0	0	0	0	0	0	0	1	0	0	0	0	0	0
b.home	0	0	0	0	0	0	0	0	1	0	0	0	0	0
b.air	0	0	0	0	0	0	0	0	0	1	0	0	0	0
b.hosp	0	0	0	0	0	0	0	0	0	0	1	0	0	0
b.metro	0	0	0	0	0	0	0	0	0	0	0	1	0	0
b.serv	0	0	0	0	0	0	0	0	0	0	0	0	1	0
b.fee	0	0	0	0	0	0	0	0	0	0	0	0	0	1

最后，当选择情景过多时，可以采用分块的方法拆分问卷。在 Ngene 中，可以通过增加“block”选项的方式实现分块，如下所示。



这里的“;block=2”表示将拆分成 2 份问卷。此时得到的结果如下。最右侧的“Block”列即为分块列，我们可以将取值为 1 和取值为 2 的选择情景分别放入两个版本的问卷中。

Choice situation	a.loc	a.home	a.air	a.hosp	a.metro	a.serv	a.fee	b.loc	b.home	b.air	b.hosp	b.metro	b.serv	b.fee	Block
1	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	1	2	3	2
3	2	3	2	3	2	2	1	2	1	1	1	2	1	2	1
4	3	2	2	4	1	2	1	3	4	2	1	2	1	1	2
5	3	4	1	4	2	1	1	1	4	2	1	1	2	3	1
6	2	4	1	2	1	2	3	2	4	2	1	1	2	2	1
7	1	2	1	1	2	2	4	2	3	2	3	2	1	3	1
8	2	4	2	1	1	1	2	3	3	2	3	2	1	4	2
9	1	2	2	3	2	1	4	1	3	2	3	1	2	2	2
10	3	4	1	2	2	1	4	3	1	1	2	2	1	1	2
11	2	2	1	2	2	2	2	1	1	1	2	1	2	4	2
12	3	2	2	4	1	1	4	2	1	1	2	2	2	4	1
13	3	3	2	1	2	1	3	1	2	2	4	2	2	1	2
14	3	1	1	1	2	2	1	2	2	2	4	1	1	4	2
15	2	1	1	3	1	1	3	3	2	2	4	2	2	4	1
16	2	1	2	2	1	1	2	3	4	1	4	1	1	1	1
17	1	3	1	4	1	2	4	2	4	1	4	1	1	3	2
18	1	3	2	4	2	2	2	1	4	1	4	2	2	2	1
19	1	1	2	2	2	2	4	3	2	2	1	2	2	3	1
20	1	3	1	2	1	1	2	1	2	2	1	2	1	4	2
21	2	1	2	4	2	1	4	2	2	2	1	1	1	2	2
22	2	2	1	3	1	2	2	1	2	1	3	2	2	1	2
23	3	4	2	3	2	1	2	2	2	1	3	1	1	3	1
24	3	4	1	1	1	2	4	3	2	1	3	1	2	2	1
25	2	3	1	2	2	1	3	2	3	1	3	2	2	1	2
26	3	1	2	2	2	2	1	3	3	1	3	1	2	4	2
27	3	1	1	4	1	1	3	1	3	1	3	2	1	4	1
28	2	2	2	1	2	2	3	1	4	1	2	2	1	3	1
29	1	4	2	1	1	1	1	2	4	1	2	2	2	4	2
30	1	2	1	3	2	1	3	3	4	1	2	1	1	2	2
31	3	1	1	3	1	2	2	2	3	2	2	2	2	1	2
32	2	3	1	3	2	1	2	3	3	2	2	1	2	3	1
33	3	3	2	1	1	2	4	1	3	2	2	1	1	2	1
34	1	2	2	2	1	1	1	2	1	2	4	1	2	1	1
35	2	4	2	4	1	2	3	1	1	2	4	1	1	3	2
36	1	4	1	4	2	2	1	3	1	2	4	2	1	2	1

6. 选择模型估计

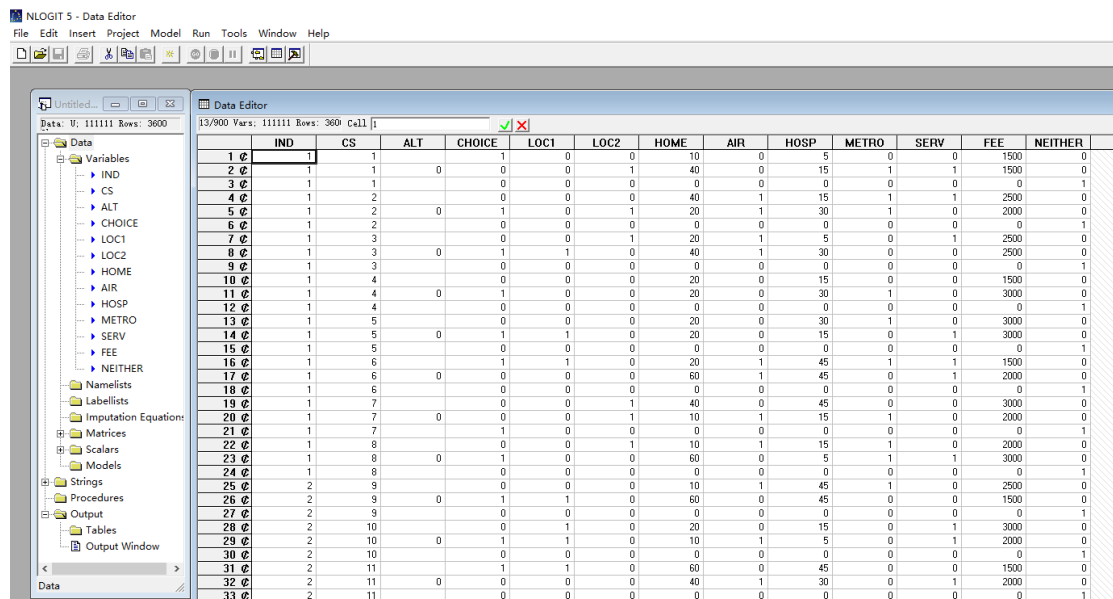
利用 SP 问卷收集的选择数据文件为“sp_data.csv”。打开该文件，如下所示。

“Ind”列为受访者编号，“CS”列为选择情景编号，“Alt”列为备选项，“Choice”列为选择结果，“Loc1”~“Fee”列为养老机构 A 和 B 的要素水平设置，“Neither”

列为专门面向 C 选项（都不选）的特别要素。

Ind	CS	Alt	Choice	Loc1	Loc2	Home	Air	Hosp	Metro	Serv	Fee	Neither
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
1	2 A		0	0	0	40	1	15	1	1	2500	0
1	2 B		1	0	1	20	1	30	1	0	2000	0
1	2 C		0	0	0	0	0	0	0	0	0	1
1	3 A		0	0	1	20	1	5	0	1	2500	0
1	3 B		1	1	0	40	1	30	0	0	2500	0
1	3 C		0	0	0	0	0	0	0	0	0	1
1	4 A		0	0	0	20	0	15	0	0	1500	0
1	4 B		1	0	0	20	0	30	1	0	3000	0
1	4 C		0	0	0	0	0	0	0	0	0	1
1	5 A		0	0	0	20	0	30	1	0	3000	0
1	5 B		1	1	0	20	0	15	0	1	3000	0
1	5 C		0	0	0	0	0	0	0	0	0	1
1	6 A		1	1	0	20	1	45	1	1	1500	0
1	6 B		0	0	0	60	1	45	0	1	2000	0
1	6 C		0	0	0	0	0	0	0	0	0	1
1	7 A		0	0	1	40	0	45	0	0	3000	0
1	7 B		0	0	1	10	1	15	1	0	2000	0
1	7 C		1	0	0	0	0	0	0	0	0	1
1	8 A		0	0	1	10	1	15	1	0	2000	0
1	8 B		1	0	0	60	0	5	1	1	3000	0
1	8 C		0	0	0	0	0	0	0	0	0	1

打开 NLogit 软件，在菜单栏中选择“Project – Import – Variables”，在弹出的窗口中将“sp_data.csv”导入，如下所示。



在 NLogit 中打开代码文件“sp_logit.lim”，如下所示。

```

sp_logit.lim *
Insert
nlogit
;lhs = choice
;choices = a, b, c
;rhs = loc1, loc2, home, air, hosp, metro, serv, fee, neither
$
    
```

我们已经在之前的章节介绍了 NLogit 中估计 Logit 模型的代码语法，这里不

再赘述。需要注意的是，这段代码中没有“rh2”选项，这是因为本案例属于无标签数据——备选的 2 个养老机构的区别完全在于要素水平的设置，而名称的不同（“A”与“B”）没有实际意义，这一点与之前介绍的交通方式选择是不同的。对于这样的无标签数据，我们应当认为决策者对于备选项本身（不考虑其要素水平）没有特别的偏好，因此不设置 ASC 常数项，因此不需要设置“rh2”选项。

运行该段代码，结果如下。

```
-----
Discrete choice (multinomial logit) model
Dependent variable      Choice
Log likelihood function  -1061.66913
Estimation based on N = 1200, K = 9
Inf.Cr.AIC = 2141.3 AIC/N = 1.784
Model estimated: Apr 12, 2019, 00:31:48
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only -1315.4181 .1929 .1899
Response data are given as ind. choices
Number of obs. = 1200, skipped 0 obs
-----
+-----+
CHOICE | Coefficient      Standard      Prob.      95% Confidence
        |                 Error        z          |z|>Z*      Interval
-----+-----+
LOC1   | -.55980***       .11388       -4.92      .0000      -.78300    -.33661
LOC2   | -1.88313***      .13477      -13.97     .0000     -2.14726   -1.61899
HOME   | -.01073***       .00280       -3.83      .0001     -.01622    -.00525
AIR    | -.29840***       .10132       -2.95      .0032     -.49697    -.09982
HOSP   | -.02951***       .00376       -7.84      .0000     -.03688    -.02213
METRO  | -.47690***       .10649       -4.48      .0000     -.68562    -.26817
SERV   | -.88014***       .10219       -8.61      .0000     -1.08043   -.67985
FEE    | -.00084***       .9503D-04    -8.81      .0000     -.00102    -.00065
NEITHER| -4.67755***      .31146      -15.02     .0000     -5.28800   -4.06709
-----
+-----+
Note: mmmn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * ==> Significance at 1%, 5%, 10% level.
-----
```

需要特别注意：由于本数据是无标签数据，没有 ASC 常数项，因此在计算 McFadden R^2 时，作为比较基准的零模型也不应该带有常数项。然后 NLogit 报告的 McFadden R^2 使用了“Constants only”作为零模型，因此是不准确的。我们可以计算，真正意义的零模型的对数似然函数值为： $1200 \cdot \log(1/3) = -1318.3$ ，本模型的对数似然值为 -1061.7，由此可计算真正的 $McFadden R^2 = 1 - (-1061.7) / (-1318.3) = 0.1947$ 。