

# 倾向评分匹配的SPSS和R实现方法

作者：张华 赵一鸣 来源：临床流行病学和循证医学

本文原地址：<https://www.iikx.com/news/statistics/2267.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

## 倾向评分匹配的SPSS和R实现方法

。前面我们讲解了倾向评分匹配的原理，由于实现时需要编程，当时并没有讲软件的实现方法。而SPSS在22版和23版加入了倾向评分匹配方法，笔者多次操作，程序界面还算友好，现给大家展示一下，供初次使用者参考。

如下图，一个数据，包括了id(病例的唯一编码)、group(干预方法)、cf1-cf6(六个混杂因子)。



	名称	类型	宽度	小数位数
1	id	数字	12	0
2	group	数字	2	0
3	name	字符串	9	0
4	sex	数字	12	0
5	age	数字	12	0
6	cf1	数字	12	0
7	cf2	数字	12	0
8	cf3	数字	12	0
9	cf4	数字	12	0
10	cf5	数字	12	0
11	cf6	数字	12	0
12				

操作方法:1.点击“数据”-“倾向得分匹配”，如下图：





2.弹出下图对话框，组指示符选择“group”，即干预因素，须为二分类变量;预测变量框里选入所有混杂因素，倾向变量名即每个个体的倾向评分得分变量名，可随意填写(字母或字母加数字)、匹配容差可从较小的数值填写，根据情况填写;匹配id变量名，即可输出一个变量，告诉我们每个case的匹配对象的id;数据集名称可自行填写。点击确定可得到匹配结果。

id	group	name	sex	age	cf1	cf2	cf3	cf4	cf5	cf6	ps	psid
19539	0	姜...	2	65	1	0	0	0	0	0	.24037	.
22028	0	陈...	2	68	0	0	0	0	0	0	.21767	.
224852	0	姜...	2	64	0	1	1	0	0	0	.18645	.
231236	0	王...	1	51	0	1	0	0	0	1	.17047	.
290849	0	冯...	1	73	0	1	0	0	0	0	.18821	.
294414	0	刘...	1	85	0	1	0	1	1	1	.17543	.
309612	0	刘...	1	70	1	1	0	0	0	1	.18944	.
323230	1	杜...	1	67	1	0	0	0	0	1	.21903	556016
329894	1	王...	1	60	1	1	0	0	0	1	.18944	552410
342002	0	郝...	1	60	1	1	0	0	0	1	.18944	.
345619	0	刑...	1	73	1	1	1	0	0	1	.18767	.
358367	0	苏...	1	57	0	1	0	0	0	1	.17047	.
368995	1	周...	2	85	0	1	1	0	0	0	.18645	537698
380846	0	许...	1	76	1	0	0	0	0	0	.24037	.
386206	0	李...	1	50	0	0	0	0	0	1	.19782	.
387451	0	王...	1	66	1	1	0	0	0	0	.20866	.
394022	1	殷...	1	57	1	1	1	0	0	1	.18767	602292
402114	0	李...	1	58	0	1	1	0	0	1	.16884	.
403417	0	李...	1	60	0	0	0	0	0	1	.19782	.
404005	1	颜...	1	48	1	1	0	0	0	1	.18944	561393
410729	0	赵...	2	79	1	1	0	1	0	0	.20033	.
415780	1	王...	2	75	1	1	1	1	0	0	.21144	436872
416840	0	王...	1	50	1	1	0	0	0	1	.18944	.

可以在下图中看到，在我们的变量后面多了两个新变量，ps即每个case的倾向得分，psid即匹配对象，第一个19539未能匹配，第八个323230匹配对象的id是556016，依次类推。

另外SPSS会将所有匹配的对象重新筛选出来生成一个新的数据，即可以用于分析的数据。

在用SPSS做倾向评分匹配时应注意：1.SPSS安装需要同意安装Python Essentials插件，否则无法使用;2.所有用于分析的变量名和界面填写的变量名必须是英文或英文加数字，不能是中文;3.匹配容差需要根据实际情况确定，如两组样本量差异较大(两组差10倍以上)，可以用较小的容差，如0.001，如较小容差不能匹配，再将容差调大后重试;4.完成匹配后应对两组进行均衡性检验;5.如果一个个体有多个匹配对象，程序会从中随机选择，因此每次运行可能得到的匹配结果不同;6.SPSS程序只能进行1：1匹配。

在R程序中，提供了1：n匹配，以及多类匹配方法，现常使用MatchIt程序包进行匹配，实现起来也很容易。现将程序及说明附在下面，使用时修改相关的变量名即可。( #后为说明文字，在R中不运行)

```
installed.packages("MatchIt") #安装MatchIt程序包，用于匹配
```

```
library(MatchIt) #加载MatchIt程序包
```

```
installed.packages("foreign") #安装foreign程序包，用于读取SPSS数据
```

```
library(foreign) #加载foreign程序包
```

---

```
mydata=read.spss("E:/data3.sav") #读取数据到mydata
names(mydata) #查看数据变量名
mydata=data.frame(mydata) #将mydata转换成数据框
attach(mydata) #绑定mydata
m.out = matchit(group ~cf1 + cf2 + cf3 + cf4 + cf5 + cf6,
data = mydata, method = "nearest",
ratio = 4)
#匹配过程，method包括"exact" (exactmatching), "full" (full matching), "genetic" (geneticmatching),
"nearest" (nearest neighbor matching), "optimal"(optimal matching), "subclass" (subclassification) are
available.默认为 "nearest".
#ratio可设置匹配比例
summary(m.out) #查看匹配情况
plot(m.out, type = "jitter") #查看匹配前后货币评分分布图
m.data1 <- match.data(m.out) #导出匹配数据到m.data1
write.csv(m.data1, file="E:/psm20160121.csv") #导出匹配数据成CSV格式，供后续分析使用.
```

更多 统计方法 请访问 <https://www.iikx.com/news/statistics/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发