

SPSS による傾向スコアとマッチングの手順

Procedure of Propensity Score and Matching Using SPSS

石村 貞夫・石村 友二郎

Sadao ISHIMURA and Yujiro ISHIMURA

「鶴見大学紀要」第 52 号 第 4 部

人文・社会・自然科学編 (平成 27 年 3 月) 別刷

SPSSによる傾向スコアとマッチングの手順

Procedure of Propensity Score and Matching Using SPSS

石村 貞夫・石村 友二郎

Sadao ISHIMURA and Yujiro ISHIMURA

1. 序文

いろいろな研究分野において、2つのグループの比較をおこなうことは、よくある。このようなとき、統計処理としては、2つの母平均の差の検定を行うのが一般的であるが、その際に、2つのグループのデータ数が、同じであるとは限らない。

比較をしたいグループの共変量の条件を揃えたり、選択バイアスを除く方法として、最近、傾向スコアとマッチングがよく利用されている。

この掌編では、傾向スコアとマッチングの手順について、具体例を用い、解説する。

2. 2つのグループの比較をおこなう研究例

研究目的

脳卒中なしのグループと脳卒中ありの2つのグループにおいて、飲酒量に差があるかどうか分析したい。

統計処理

2つの母平均の差の検定を利用する。

データ

表 脳卒中なしのグループ

脳卒中なし

被験者	飲酒量	性別	喫煙	血圧
1	2.5	1	2	142
2	1.3	0	1	133
3	1.6	0	3	147
4	4.3	1	3	135
5	1.2	1	0	142
6	2.1	1	0	158
7	3.9	1	1	165
8	3.4	1	2	145
9	4.6	1	1	148
10	1.1	0	0	135
11	4.8	1	0	149
12	1.7	0	1	176
13	3.1	1	1	141
14	2.5	1	0	136
15	4.7	1	1	148

表 脳卒中ありのグループ

脳卒中あり

被験者	飲酒量	性別	喫煙	血圧
1	3.5	0	2	154
2	6.4	1	1	135
3	4.5	1	2	159
4	5.2	1	3	165
5	2.8	0	2	147

分析の方針

1. データ数が15人と5人なので、傾向スコアとマッチングをおこなう。
2. そのマッチングをもとに、2つのグループのデータ数を同程度にする。
3. 最後に、同程度にした2つのグループについて、母平均の差の検定を行う。

傾向スコアの計算方法

脳卒中の有無を従属変数、性別、喫煙、血圧を共変量にし、ロジステック回帰分析を用いて、予測確率を計算する。

手順1.

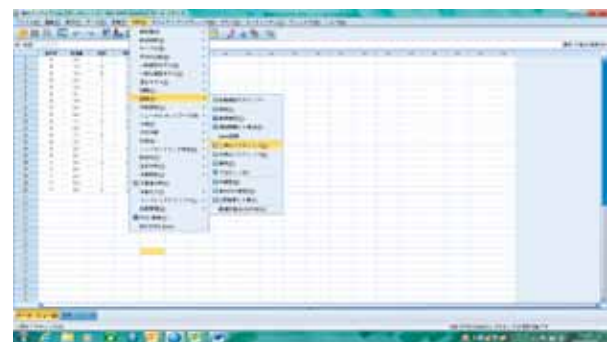


図1 SPSSによるロジステック回帰分析

SPSS による傾向スコアとマッチングの手順

手順2.

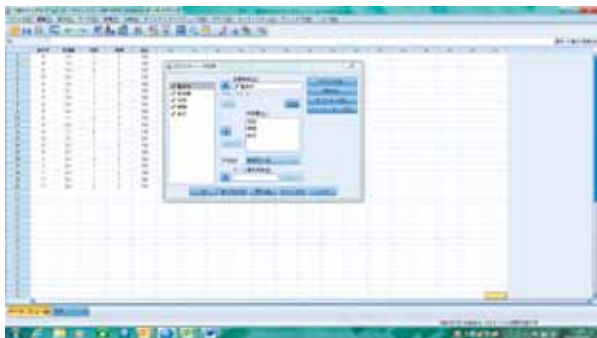


図2 SPSS によるロジステック回帰分析

手順3

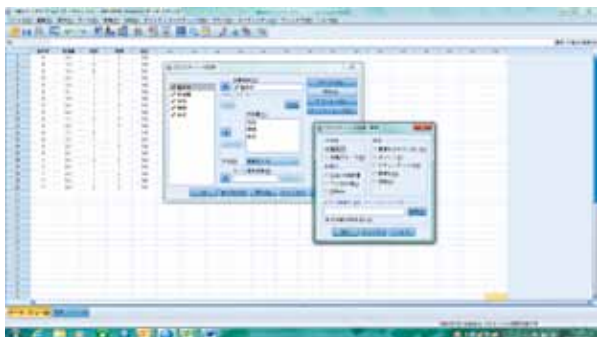


図3 SPSS によるロジステック回帰分析

手順4.

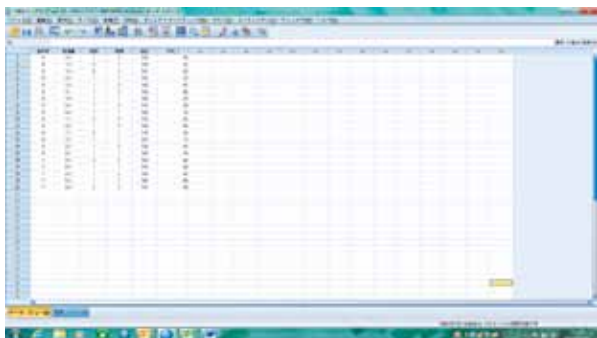


図4 SPSS によるロジステック回帰分析

その予測確率を傾向スコアとする。

傾向スコアが近い調査対象者をマッチングするために、予測確率を昇順に並べ替える。

手順5.

表 脳卒中なしのグループ

被験者	飲酒量	性別	喫煙	血圧	予測確率
14	2.5	1	0	136	0.04
5	1.2	1	0	142	0.05
10	1.1	0	0	135	0.05
11	4.8	1	0	149	0.06
6	2.1	1	0	158	0.08
13	3.1	1	1	141	0.11
2	1.3	0	1	133	0.12
9	4.6	1	1	148	0.14
15	4.7	1	1	148	0.14
7	3.9	1	1	165	0.23
1	2.5	1	2	142	0.26
8	3.4	1	2	145	0.29
12	1.7	0	1	176	0.38
4	4.3	1	3	135	0.43
3	1.6	0	3	147	0.63

表 脳卒中ありのグループ

被験者	飲酒量	性別	喫煙	血圧	予測確率
2	6.4	1	1	135	0.09
5	2.8	0	2	147	0.38
3	4.5	1	2	159	0.40
1	3.5	0	2	154	0.44
4	5.2	1	3	165	0.69

手順6.

2つのグループのデータ数を同程度とするために、それぞれのグループから、予測確率の近い被験者を、次のように4人ずつ選び出す。

表 脳卒中なしのグループ

被験者	飲酒量	性別	喫煙	血圧	予測確率
6	2.1	1	0	158	0.08
12	1.7	0	1	176	0.38
4	4.3	1	3	135	0.43
3	1.6	0	3	147	0.63

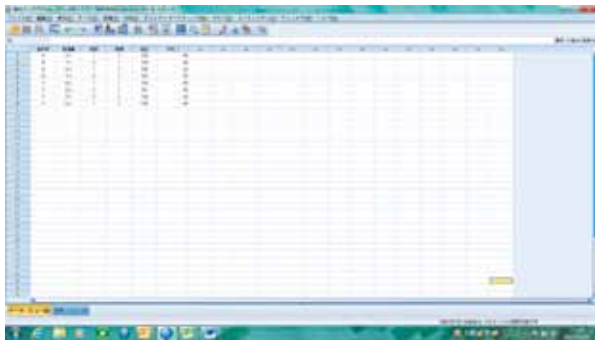
表 脳卒中ありのグループ

被験者	飲酒量	性別	喫煙	血圧	予測確率
2	6.4	1	1	135	0.09
5	2.8	0	2	147	0.38
1	3.5	0	2	154	0.44
4	5.2	1	3	165	0.69

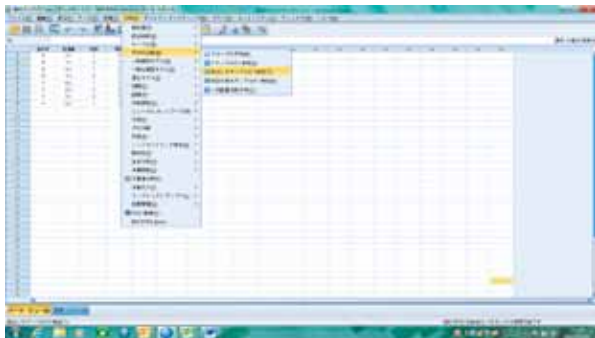
このとき、共変量の条件がそろっているかどうか、チェックを行う。

手順7.

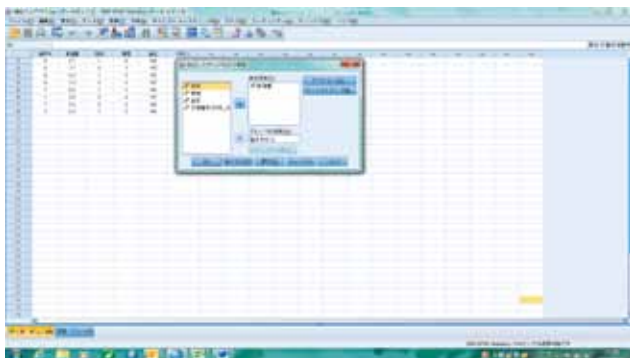
この2つのグループについて、2つの母平均の差の検定を行う。



手順8.



手順9.



手順10.

t 検定

グループ統計量

	脳卒中	N	平均値	標準偏差	平均値の標準誤差
飲酒量 なし		4	2.425	1.2685	.6343
あり		4	4.475	1.6317	.8159

独立サンプルの検定

		飲酒量	
		等分散を仮定する。	等分散を仮定しない。
等分散性のため	F 値	.791	
のLeveneの検定	有意確率	.408	
2つの母平均の	t 値	-1.984	-1.984
差の検定	自由度	6	5.656
	有意確率 (両側)	.095	.097

注1 2群間の差を表す「Standardized Difference」は 0.1 (10%) 未満が推奨されている。

Standardized Difference の算出

■ スケールデータの場合

(Embedded image moved to file: pic48003.jpg)

■ 2値データの場合

(Embedded image moved to file: pic16238.jpg)

注2 傾向スコアを利用しない場合、2つの母平均の差の検定を行うと、次のような出力結果となる。

SPSSによる傾向スコアとマッチングの手順

t検定

グループ統計量

脳卒中	N	平均値	標準偏差	平均値の標準誤差
飲酒量 なし	15	2.853	1.3559	.3501
あり	5	4.480	1.4132	.6320

独立サンプルの検定

		飲酒量	
		等分散を仮定する。	等分散を仮定しない。
等分散性のため のLeveneの検定	F 値	.118	
	有意確率	.735	
2つの母平均の 差の検定	t 値	-2.301	-2.252
	自由度	18	6.653
	有意確率 (両側)	.034	.061

参考文献

1. 調査観察データの統計科学 星野崇宏著 岩波書店 2009
2. 統計的因果推論 宮川雅巳著 朝倉書店 2004
3. 傾向スコアマッチングの手順 IBM SPSS

SPSSによる傾向スコアとマッチングの手順

Procedure of Propensity Score and Matching Using SPSS

歯学部 准教授 石村貞夫
早稲田大学大学院 基幹理工学研究科 応用数学科
石村友二郎