

20. 因子分析

因子分析の手法は、多様ですので、どのようなデータをどのような目的で因子分析を施すのかを、別途検討する必要があります。ここでは、最も基本的なヴァリマックス直交回転をさせた主成分分析のシンタックスについて紹介します。

20.1 主成分分析（ヴァリマックス直交回転）

一般的には、

```
factor /variables 変数リスト /extraction pc /rotation varimax.
```

主成分得点を SPSS データファイルに保存するには、

```
factor /variables 変数リスト /extraction pc /rotation varimax  
/save regression (all ルート名).
```

となります。ここで、all はすべての主成分得点について保存することを意味しており、ルート名とは、主成分解に対して任意に付ける名称で、主成分得点の変数が順次、ルート名 1、ルート名 2 というように、後ろに番号がついて保存されます。

具体例として、1995 年国勢調査の東京都市区町村別アグリゲートデータを使って、因子生態学分析をする場合を取り上げます。使用する変数は、人口成長率 (popgr)、年少人口比率 (child95r)、老年人口指数 (aged95r)、女性雇用者比率 (femp95r)、上級ホワイトカラー比率 (upwht95r)、ブルーカラー比率 (blue95r)、人口密度 (dens95) です。ルート名は、tk95a とします。

```
factor /variables popgr child95r aged95r femp95r upwht95r blue95r dens95  
/extraction pc /rotation varimax /save regression (all tk95a).
```

実行結果は、図 20.1 のようになります。固有値 1 以上の主成分が 3 つ抽出されました (固有値 1 以上は、デフォルト)。

図 20.1 の画面には出ていませんが、「記録」欄 (図 20.2) に 3 つの主成分得点が、それぞれ TK95A1、TK95A2、TK95A3 として、保存されていることが示されています。

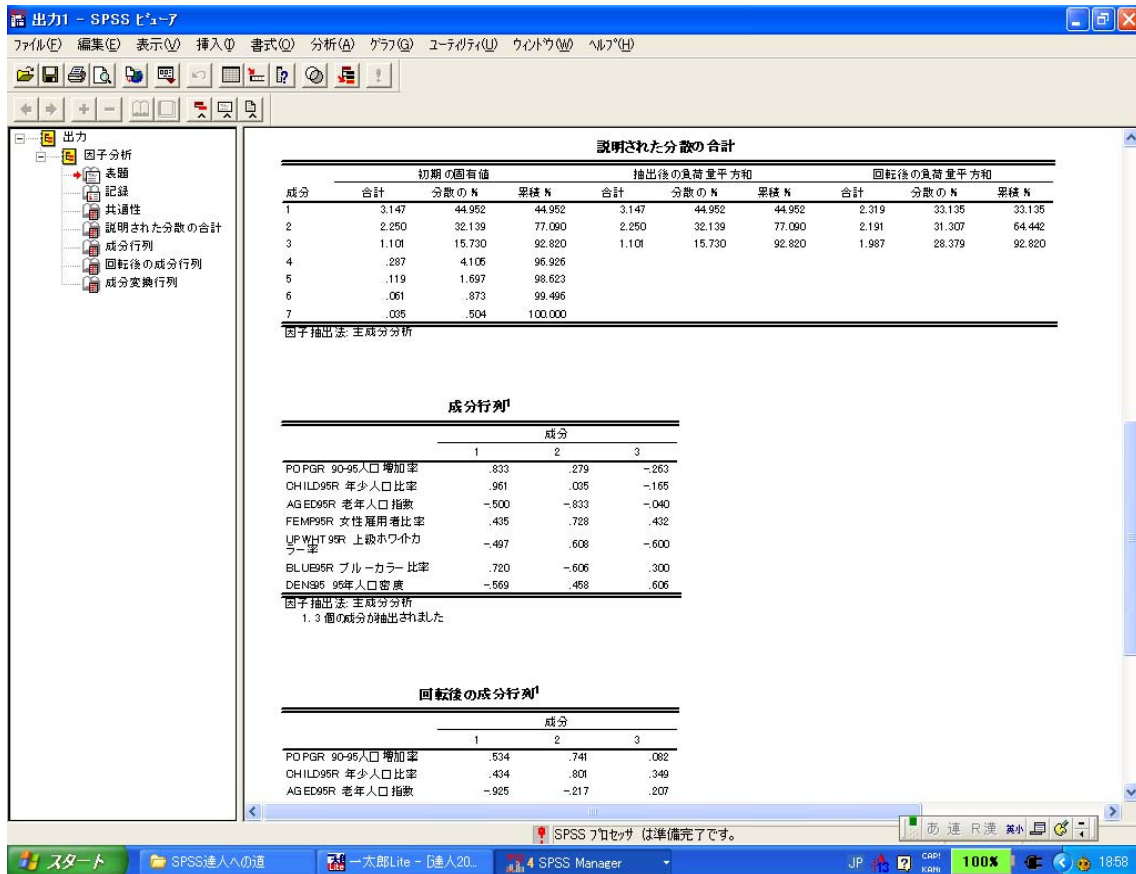


図 20.1 因子分析の出力例

記録

| | | |
|---------|--|--|
| 出力の作成日付 | 02-APR-2006 18:57:10 | |
| コメント | | |
| 入力 | データ | C:\Documents and Settings\松本康\My Documents\東京54市区町村データ\tokyo95.sav |
| | フィルタ | <なし> |
| | 重み付け | <なし> |
| | 分割ファイル | <なし> |
| | 作業データファイル内の行数 | 54 |
| 欠損値処理 | 欠損値の定義 | MISSING=EXCLUDE: ユーザー欠損値は欠損として取り扱われます。 |
| | 使用されたケース | LISTWISE: 統計量はすべての変数に欠損値がないケースに基づいています。 |
| シンタックス | factor /variables popgr child95r aged95r femp95r upwh95r blue95r dens95 /extraction pc /rotation varimax /save regression (all tk95a). | |
| リソース | 経過時間 | 0:00:00.16 |
| | 必要な最大メモリ | 7636 (7.457K) バイト |
| 作成された変数 | TK95A1 | 成分得点 1 |
| | TK95A2 | 成分得点 2 |
| | TK95A3 | 成分得点 3 |

図 20.2 出力された記録。主成分得点が保存されていることが示されている。

説明された分散の合計

| 成分 | 初期の固有値 | | | 抽出後の負荷量平方和 | | | 回転後の負荷量平方和 | | |
|----|--------|--------|---------|------------|--------|--------|------------|--------|--------|
| | 合計 | 分散の% | 累積% | 合計 | 分散の% | 累積% | 合計 | 分散の% | 累積% |
| 1 | 3.147 | 44.952 | 44.952 | 3.147 | 44.952 | 44.952 | 2.319 | 33.135 | 33.135 |
| 2 | 2.250 | 32.139 | 77.090 | 2.250 | 32.139 | 77.090 | 2.191 | 31.307 | 64.442 |
| 3 | 1.101 | 15.730 | 92.820 | 1.101 | 15.730 | 92.820 | 1.987 | 28.379 | 92.820 |
| 4 | .287 | 4.105 | 96.926 | | | | | | |
| 5 | .119 | 1.697 | 98.623 | | | | | | |
| 6 | .061 | .873 | 99.496 | | | | | | |
| 7 | .035 | .504 | 100.000 | | | | | | |

因子抽出法: 主成分分析

図 20.3 出力された各主成分による分散の説明率

回転後の成分行列¹

| | 成分 | | |
|---------------------|-------|-------|-------|
| | 1 | 2 | 3 |
| POPGR 90-95人口増加率 | .534 | .741 | .082 |
| CHILD95R 年少人口比率 | .434 | .801 | .349 |
| AGED95R 老年人口指数 | -.925 | -.217 | .207 |
| FEMP95R 女性雇用者比率 | .945 | -.081 | .076 |
| UPWHT95R 上級ホワイトカラー率 | .041 | -.039 | -.987 |
| BLUE95R ブルーカラー比率 | -.036 | .403 | .901 |
| DENS95 95年人口密度 | .305 | -.885 | -.153 |

因子抽出法: 主成分分析

回転法: Kaiser の正規化を伴わないハリマックス法

1. 5 回の反復で回転が収束しました。

図 20.4 出力された回転後の主成分行列

図 20.4 から、第一成分は、女性雇用者比率 (+) と老年人口指数 (-) に負荷量が高いことがわかります。第二成分は、年少人口比率 (+) と人口増加率 (+) と人口密度 (-) に負荷量が高いことがわかります。第三成分は、上級ホワイトカラー比率 (-) とブルーカラー比率 (+) に負荷量が高いことがわかります。また、図 20.3 から、これら 3 つの主成分によって、分散の 92.9 % が説明できることがわかります。

20.2 ファイルの達人的処理法

生成された主成分得点は、SPSS データ・ファイル上に保存されています。プログラムを閉じるときには、シンタックスを保存して、SPSS データ・ファイルは捨ててしまいます。つぎにこの得点を利用するには、保存されているシンタックスファイルを実行してから、その続きを始めます。因子分析にかかる時間は、わずか 0.16 秒です。