

作业 4

辛柏赢 2020111753

2023-05-08

第一题

对十项全能运动得分样本相关矩阵进行因子分析

```
# Load data
path1 <- "/Users/xinby/Desktop/Sufe/Multivariate-Stat-Analysis/Hw&Proj/hw4/exec8.4.xlsx"
dat1 <- read_xlsx(path1) %>% select(-event)

# Factor Analysis with no rotations
print("-----")
print("|               Factor Analysis with No Rotations:               |")
print("-----")
factor <- fa( dat1, nfactors = 4, rotate = "none",
              residuals = T, fm='ml') # Maximum Likelihood Factor Analysis
print(factor)

# Factor Analysis with rotations
print("-----")
print("|               Factor Analysis with Max-Variance Rotations:               |")
print("-----")
factor.maxvar <- fa( dat1, nfactors = 4, rotate = "varimax",
                    residuals = T, fm='ml') # Maximum Likelihood Factor Analysis
print(factor.maxvar)

## [1] "-----"
## [1] "|               Factor Analysis with No Rotations:               |"
## [1] "-----"
## Factor Analysis using method = ml
## Call: fa(r = dat1, nfactors = 4, rotate = "none", residuals = T, fm = "ml")
## Standardized loadings (pattern matrix) based upon correlation matrix
##      ML2   ML3   ML1   ML4   h2    u2 com
## x1  0.21  0.82  0.30 -0.17 0.84 0.159 1.5
## x2  0.38  0.59  0.25  0.27 0.62 0.379 2.6
## x3  0.64 -0.02  0.76  0.00 1.00 0.005 1.9
## x4  0.41  0.33  0.16  0.44 0.50 0.499 3.1
## x5  0.45  0.66 -0.11 -0.14 0.67 0.328 1.9
## x6  0.26  0.42  0.26  0.39 0.46 0.539 3.4
## x7  0.50  0.01  0.54  0.01 0.54 0.462 2.0
## x8  0.31  0.22  0.06  0.39 0.30 0.699 2.6
## x9  0.31 -0.02  0.31  0.09 0.21 0.795 2.2
## x10 0.71 -0.01 -0.70  0.00 1.00 0.005 2.0
##
##
##      ML2   ML3   ML1   ML4
## SS loadings      1.98 1.79 1.72 0.63
## Proportion Var    0.20 0.18 0.17 0.06
## Cumulative Var    0.20 0.38 0.55 0.61
## Proportion Explained 0.32 0.29 0.28 0.10
```

```

## Cumulative Proportion 0.32 0.62 0.90 1.00
##
## Mean item complexity = 2.3
## Test of the hypothesis that 4 factors are sufficient.
##
## The degrees of freedom for the null model are 45 and the objective function was 3.7
## The degrees of freedom for the model are 11 and the objective function was 0.07
##
## The root mean square of the residuals (RMSR) is 0.02
## The df corrected root mean square of the residuals is 0.04
##
## Fit based upon off diagonal values = 1
## Measures of factor score adequacy
##
## Correlation of (regression) scores with factors      ML2  ML3  ML1  ML4
## Multiple R square of scores with factors            1.00 0.94 1.00 0.76
## Minimum correlation of possible factor scores        0.99 0.88 1.00 0.57
## [1] "-----"
## [1] "|                Factor Analysis with Max-Variance Rotations:                |"
## [1] "-----"
## Factor Analysis using method = ml
## Call: fa(r = dat1, nfactors = 4, rotate = "varimax", residuals = T,
##       fm = "ml")
## Standardized loadings (pattern matrix) based upon correlation matrix
##      ML1 ML3 ML4 ML2 h2 u2 com
## x1  0.17 0.86 0.25 -0.14 0.84 0.159 1.3
## x2  0.24 0.48 0.58  0.01 0.62 0.379 2.3
## x3  0.96 0.15 0.20 -0.06 1.00 0.005 1.1
## x4  0.24 0.17 0.63  0.11 0.50 0.499 1.5
## x5  0.06 0.71 0.24  0.33 0.67 0.328 1.7
## x6  0.21 0.26 0.59 -0.07 0.46 0.539 1.7
## x7  0.70 0.13 0.18 -0.01 0.54 0.462 1.2
## x8  0.14 0.08 0.51  0.12 0.30 0.699 1.3
## x9  0.42 0.02 0.17  0.00 0.21 0.795 1.3
## x10 -0.05 0.06 0.11  0.99 1.00 0.005 1.0
##
##
##      ML1 ML3 ML4 ML2
## SS loadings      1.80 1.61 1.58 1.14
## Proportion Var    0.18 0.16 0.16 0.11
## Cumulative Var    0.18 0.34 0.50 0.61
## Proportion Explained 0.29 0.26 0.26 0.19
## Cumulative Proportion 0.29 0.56 0.81 1.00
##
## Mean item complexity = 1.5
## Test of the hypothesis that 4 factors are sufficient.
##
## The degrees of freedom for the null model are 45 and the objective function was 3.7
## The degrees of freedom for the model are 11 and the objective function was 0.07
##
## The root mean square of the residuals (RMSR) is 0.02
## The df corrected root mean square of the residuals is 0.04
##
## Fit based upon off diagonal values = 1
## Measures of factor score adequacy
##
##      ML1 ML3 ML4 ML2

```

```
## Correlation of (regression) scores with factors    0.99 0.91 0.80 1.00
## Multiple R square of scores with factors          0.98 0.82 0.65 0.99
## Minimum correlation of possible factor scores      0.96 0.65 0.29 0.98
```

因子分析结果如上所示。分别展示了不进行旋转和进行了最大方差旋转，以 MLE 为估计准则的结果。具体而言，主要的分析结果摘录如下（最大方差旋转结果）：

	ML1	ML3	ML4	ML2	h2	u2	com
x1	0.17	0.86	0.25	-0.14	0.84	0.159	1.3
x2	0.24	0.48	0.58	0.01	0.62	0.379	2.3
x3	0.96	0.15	0.20	-0.06	1.00	0.005	1.1
x4	0.24	0.17	0.63	0.11	0.50	0.499	1.5
x5	0.06	0.71	0.24	0.33	0.67	0.328	1.7
x6	0.21	0.26	0.59	-0.07	0.46	0.539	1.7
x7	0.70	0.13	0.18	-0.01	0.54	0.462	1.2
x8	0.14	0.08	0.51	0.12	0.30	0.699	1.3
x9	0.42	0.02	0.17	0.00	0.21	0.795	1.3
x10	-0.05	0.06	0.11	0.99	1.00	0.005	1.0

其中 ML1~ML4 表示的是最终得到的四个因子。h2 表示共性方差，表示的是这些公共因子对各个变量 x 的方差贡献；u 表示特殊方差，表示的是特殊因子对于 x 的方差贡献。

对于该因子分析，其相应的效果评估为：

	ML1	ML3	ML4	ML2
SS loadings	1.80	1.61	1.58	1.14
Proportion Var	0.18	0.16	0.16	0.11
Cumulative Var	0.18	0.34	0.50	0.61
Proportion Explained	0.29	0.26	0.26	0.19
Cumulative Proportion	0.29	0.56	0.81	1.00

其中 Cumulative Var 表示累计方差比例，表示前 n 个因子解释的方差占总方差的比例

第二题

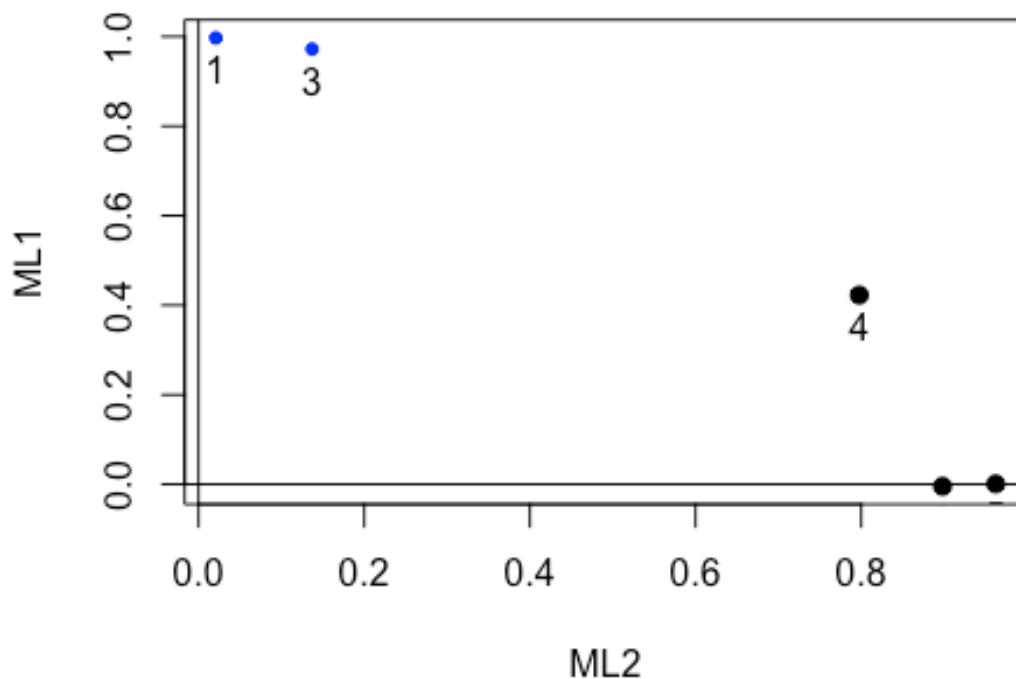
对洛杉矶人口调查进行因子分析；变量 1~5 分别表示：人口数、教育程度/教育年数中位数、佣人总数、服务业人数、房价中位数。

```
# Load data
path2 <- "/Users/xinby/Desktop/Sufe/Multivariate-Stat-Analysis/Hw&Proj/hw4/exec8.5.xlsx"
dat2 <- read_xlsx(path2) %>% select(-i)
dat2.R <- cor(dat2)

# Factor Analysis with rotations
print("-----")
print("|               Factor Analysis with Max-Variance Rotations:               |")
print("-----")
factor.ny<- fa( dat2.R , nfactors = 2, rotate = "varimax",
               residuals = T, fm='ml') # Maximum Likelihood Factor Analysis
```

```
print(factor.ny)
plot(factor.ny)
```

Factor Analysis



```
## [1] "-----"
## [1] "|               Factor Analysis with Max-Variance Rotations:               |"
## [1] "-----"
## Factor Analysis using method = ml
## Call: fa(r = dat2.R, nfactors = 2, rotate = "varimax", residuals = T,
##      fm = "ml")
## Standardized loadings (pattern matrix) based upon correlation matrix
##      ML2  ML1   h2    u2 com
## X1 0.02 1.00 1.00 0.005 1.0
## X2 0.90 0.00 0.81 0.193 1.0
## X3 0.14 0.97 0.96 0.036 1.0
## X4 0.80 0.42 0.81 0.185 1.5
## X5 0.96 0.00 0.93 0.074 1.0
##
##
##      ML2  ML1
## SS loadings      2.39 2.12
## Proportion Var    0.48 0.42
## Cumulative Var    0.48 0.90
## Proportion Explained 0.53 0.47
## Cumulative Proportion 0.53 1.00
##
## Mean item complexity = 1.1
## Test of the hypothesis that 2 factors are sufficient.
##
## The degrees of freedom for the null model are 10 and the objective function was 6.3
```

```

8
## The degrees of freedom for the model are 1 and the objective function was 0.31
##
## The root mean square of the residuals (RMSR) is 0.01
## The df corrected root mean square of the residuals is 0.05
##
## Fit based upon off diagonal values = 1
## Measures of factor score adequacy
##
## Correlation of (regression) scores with factors      ML2  ML1
## Multiple R square of scores with factors            0.98 1.00
## Minimum correlation of possible factor scores       0.95 1.00
## Minimum correlation of possible factor scores       0.91 0.99

```

因子分析结果如上所示，其中对应因子载荷为：

	ML2	ML1	h2	u2	com
X1	0.02	1.00	1.00	0.005	1.0
X2	0.90	0.00	0.81	0.193	1.0
X3	0.14	0.97	0.96	0.036	1.0
X4	0.80	0.42	0.81	0.185	1.5
X5	0.96	0.00	0.93	0.074	1.0

对应的因子分析效果为：

	ML2	ML1
SS loadings	2.39	2.12
Proportion Var	0.48	0.42
Cumulative Var	0.48	0.90
Proportion Explained	0.53	0.47
Cumulative Proportion	0.53	1.00

根据实际情况对比，大约可以认为 ML2 表示的是有关社会生产条件、生活条件等方面相关，ML1 主要与地区人口状况相关。

第三题

对应征者 15 个方面的得分进行因子分析

```

# Load data
path3 <- "/Users/xinby/Desktop/Sufe/Multivariate-Stat-Analysis/Hw&Proj/hw4/exec8.6.xlsx"
dat3 <- read_xlsx(path3) %>% select(-i)
dat3.R <- cor(dat3)

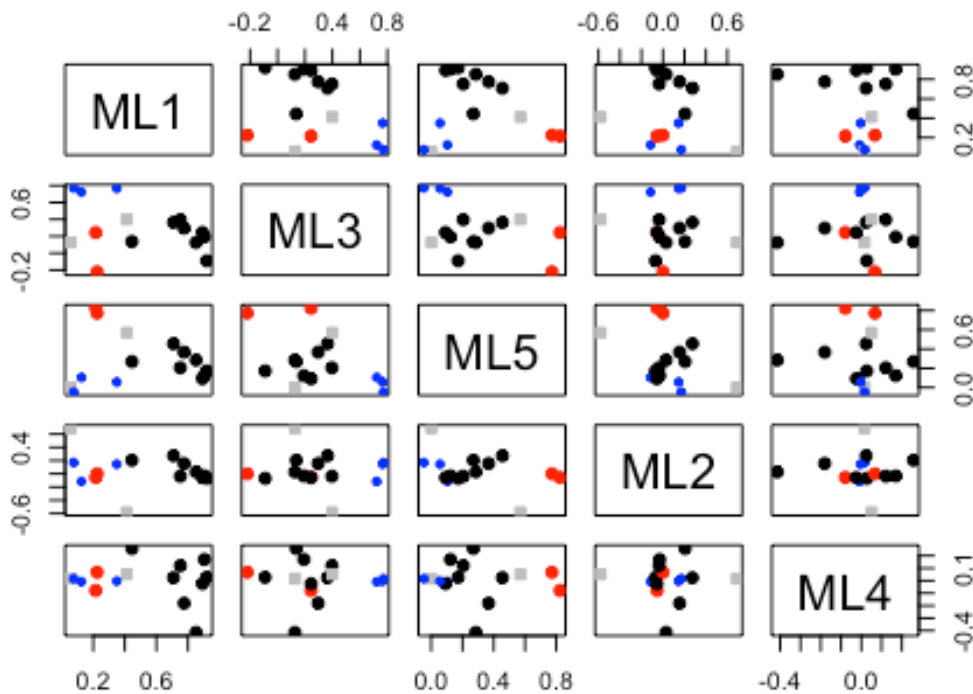
# Factor Analysis with rotations
print("-----")
print("|               Factor Analysis with Max-Variance Rotations:               |")
print("-----")
factor.app<- fa( dat3.R , nfactors = 5, rotate = "varimax",
                 residuals = T, fm='ml') # Maximum Likelihood Factor Analysis

print(factor.app)

```

```
plot(factor.app)
```

Factor Analysis



```
## [1] "-----"
## [1] "|               Factor Analysis with Max-Variance Rotations:               |"
## [1] "-----"
## Factor Analysis using method = ml
## Call: fa(r = dat3.R, nfactors = 5, rotate = "varimax", residuals = T,
##      fm = "ml")
## Standardized loadings (pattern matrix) based upon correlation matrix
##      ML1  ML3  ML5  ML2  ML4  h2   u2 com
## x1  0.13  0.72  0.10 -0.12 -0.01 0.56 0.438 1.2
## x2  0.45  0.14  0.27  0.20  0.26 0.40 0.602 3.1
## x3  0.06  0.13  0.00  0.69  0.02 0.49 0.508 1.1
## x4  0.22  0.24  0.82 -0.06 -0.08 0.80 0.205 1.4
## x5  0.92 -0.09  0.17 -0.07  0.03 0.89 0.113 1.1
## x6  0.85  0.13  0.29  0.03 -0.41 1.00 0.005 1.8
## x7  0.23 -0.22  0.77  0.00  0.07 0.70 0.300 1.4
## x8  0.89  0.25  0.09 -0.06 -0.02 0.87 0.133 1.2
## x9  0.08  0.77 -0.05  0.17  0.02 0.63 0.367 1.1
## x10 0.75  0.40  0.20 -0.04  0.12 0.78 0.220 1.8
## x11 0.90  0.19  0.12 -0.03  0.17 0.90 0.101 1.2
## x12 0.78  0.30  0.37  0.15 -0.18 0.88 0.119 2.0
## x13 0.71  0.36  0.46  0.27  0.02 0.92 0.082 2.6
## x14 0.41  0.40  0.57 -0.58  0.05 1.00 0.005 3.6
## x15 0.35  0.77  0.06  0.15  0.00 0.73 0.266 1.5
##
##
##      ML1  ML3  ML5  ML2  ML4
## SS loadings      5.46 2.51 2.21 1.03 0.33
```

```

## Proportion Var          0.36 0.17 0.15 0.07 0.02
## Cumulative Var          0.36 0.53 0.68 0.75 0.77
## Proportion Explained    0.47 0.22 0.19 0.09 0.03
## Cumulative Proportion   0.47 0.69 0.88 0.97 1.00
##
## Mean item complexity = 1.7
## Test of the hypothesis that 5 factors are sufficient.
##
## The degrees of freedom for the null model are 105 and the objective function was 15.75
## The degrees of freedom for the model are 40 and the objective function was 1.66
##
## The root mean square of the residuals (RMSR) is 0.03
## The df corrected root mean square of the residuals is 0.05
##
## Fit based upon off diagonal values = 1
## Measures of factor score adequacy
##
## Correlation of (regression) scores with factors    ML1  ML3  ML5  ML2  ML4
## Multiple R square of scores with factors          0.98 0.94 0.94 0.96 0.95
## Minimum correlation of possible factor scores      0.96 0.88 0.88 0.93 0.89
## Minimum correlation of possible factor scores      0.93 0.76 0.77 0.85 0.79

```

因子分析结果如下:

	ML1	ML3	ML5	ML2	ML4	h2	u2	com
x1	0.13	0.72	0.10	-0.12	-0.01	0.56	0.438	1.2
x2	0.45	0.14	0.27	0.20	0.26	0.40	0.602	3.1
x3	0.06	0.13	0.00	0.69	0.02	0.49	0.508	1.1
x4	0.22	0.24	0.82	-0.06	-0.08	0.80	0.205	1.4
x5	0.92	-0.09	0.17	-0.07	0.03	0.89	0.113	1.1
x6	0.85	0.13	0.29	0.03	-0.41	1.00	0.005	1.8
x7	0.23	-0.22	0.77	0.00	0.07	0.70	0.300	1.4
x8	0.89	0.25	0.09	-0.06	-0.02	0.87	0.133	1.2
x9	0.08	0.77	-0.05	0.17	0.02	0.63	0.367	1.1
x10	0.75	0.40	0.20	-0.04	0.12	0.78	0.220	1.8
x11	0.90	0.19	0.12	-0.03	0.17	0.90	0.101	1.2
x12	0.78	0.30	0.37	0.15	-0.18	0.88	0.119	2.0
x13	0.71	0.36	0.46	0.27	0.02	0.92	0.082	2.6
x14	0.41	0.40	0.57	-0.58	0.05	1.00	0.005	3.6
x15	0.35	0.77	0.06	0.15	0.00	0.73	0.266	1.5

	ML1	ML3	ML5	ML2	ML4
SS loadings	5.46	2.51	2.21	1.03	0.33
Proportion Var	0.36	0.17	0.15	0.07	0.02
Cumulative Var	0.36	0.53	0.68	0.75	0.77
Proportion Explained	0.47	0.22	0.19	0.09	0.03
Cumulative Proportion	0.47	0.69	0.88	0.97	1.00

结合各项的具体含义，**ML1** 主要体现了应征者的能力、进取心，**ML3** 主要体现了应征者的相关经验，**ML5** 主要体现了应征者的人际关系等，**ML2** 主要体现了应征者的专业能力，**ML4** 主要反映了应征者的外在形象等。