

다변량

7장 과제

과목	다변량데이터분석
담당교수	임태진 교수님
전공	산업정보시스템공학과
학번	20201368
이름	한채원
제출일	2022.11.30

데이터: bfi (psych 패키지) 사용

성별에 따라 데이터를 분할하여, 각각에 대해 다음의 요인분석을 수행하고

전체 데이터를 모두 사용했을 때의 결과와 비교하시오.

-성별 데이터 분리

```
> library(psych)
```

```
> str(bfi)
```

```
'data.frame': 2800 obs. of 28 variables:
```

```
$ A1      : int  2 2 5 4 2 6 2 4 4 2 ...
$ A2      : int  4 4 4 4 3 6 5 3 3 5 ...
$ A3      : int  3 5 5 6 3 5 5 1 6 6 ...
$ A4      : int  4 2 4 5 4 6 3 5 3 6 ...
$ A5      : int  4 5 4 5 5 5 5 1 3 5 ...
$ C1      : int  2 5 4 4 4 6 5 3 6 6 ...
$ C2      : int  3 4 5 4 4 6 4 2 6 5 ...
$ C3      : int  3 4 4 3 5 6 4 4 3 6 ...
$ C4      : int  4 3 2 5 3 1 2 2 4 2 ...
$ C5      : int  4 4 5 5 2 3 3 4 5 1 ...
$ E1      : int  3 1 2 5 2 2 4 3 5 2 ...
$ E2      : int  3 1 4 3 2 1 3 6 3 2 ...
$ E3      : int  3 6 4 4 5 6 4 4 NA 4 ...
$ E4      : int  4 4 4 4 4 5 5 2 4 5 ...
$ E5      : int  4 3 5 4 5 6 5 1 3 5 ...
$ N1      : int  3 3 4 2 2 3 1 6 5 5 ...
$ N2      : int  4 3 5 5 3 5 2 3 5 5 ...
$ N3      : int  2 3 4 2 4 2 2 2 2 5 ...
$ N4      : int  2 5 2 4 4 2 1 6 3 2 ...
$ N5      : int  3 5 3 1 3 3 1 4 3 4 ...
$ O1      : int  3 4 4 3 3 4 5 3 6 5 ...
$ O2      : int  6 2 2 3 3 3 2 2 6 1 ...
$ O3      : int  3 4 5 4 4 5 5 4 6 5 ...
$ O4      : int  4 3 5 3 3 6 6 5 6 5 ...
$ O5      : int  3 3 2 5 3 1 1 3 1 2 ...
$ gender   : int  1 2 2 2 1 2 1 1 1 2 ...
$ education: int  NA NA NA NA NA 3 NA 2 1 NA ...
$ age      : int  16 18 17 17 17 21 18 19 19 17 ...
```

```
> bfi2 = bfi[bfi$gender==1,]
```

```
> bfi3 = bfi[bfi$gender==2,]
```

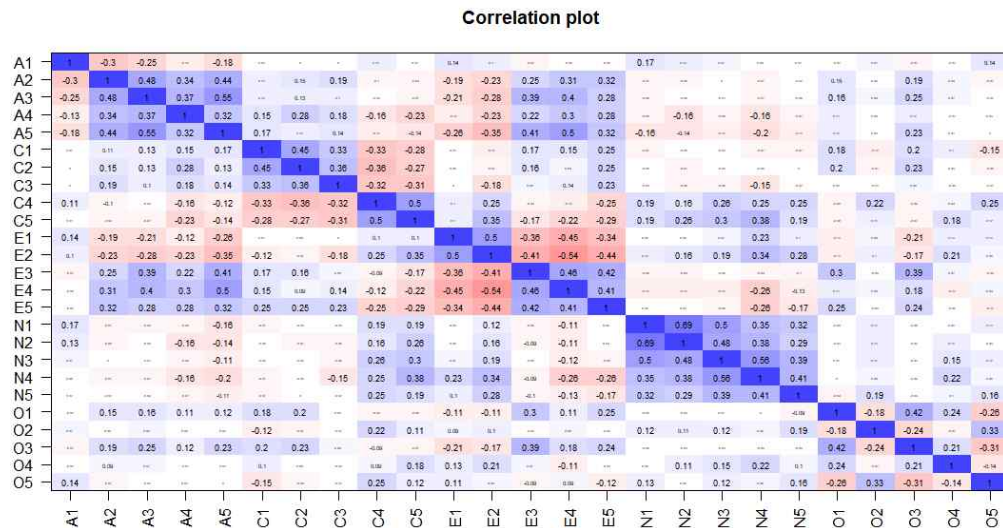
```
> df2 = bfi2[1:25];df3 = bfi3[1:25]
```

(1) 요인분석 타당성 평가: 상관관계수, 구형성, KMO, reliability 등

1. 상관관계수

-남성일 때(gender=1일 때)

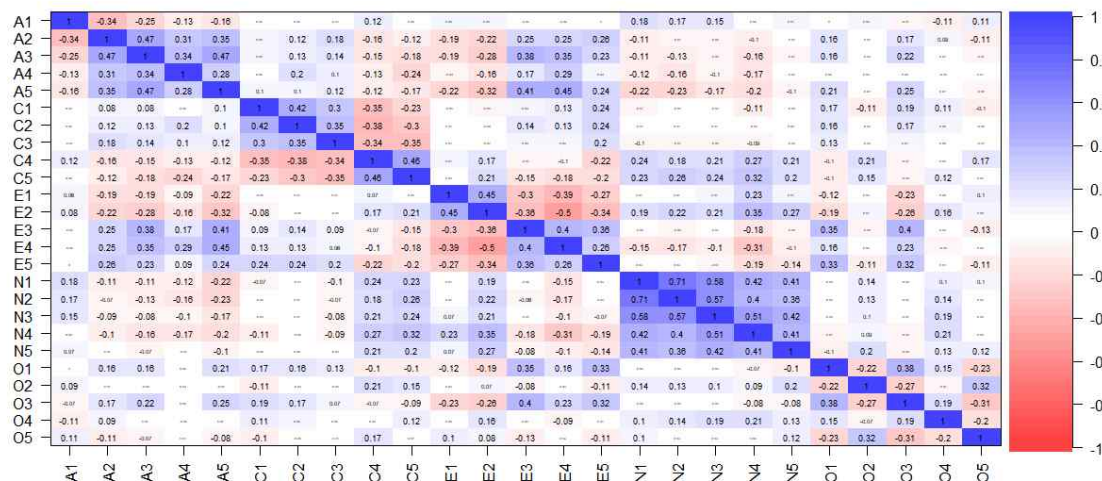
> corPlot(df2,las=2)



N1,N2,N3,N4,N5는 양의 상관관계를 보이고 있고, A1은 음의 상관관계를 보이고 있고 O1,O2,O3,O4,O5은 음과 양의 상관관계가 섞여있음을 알 수 있다.

-여성일 때(gender=2일 때)

> corPlot(df3,las=2)



마찬가지로 N1,N2,N3,N4,N5는 양의 상관관계를 보이고 있고, A1은 음의 상관관계를 보이고 있고 O1,O2,O3,O4,O5은 음과 양의 상관관계가 섞여있음을 알 수 있다.

남성과 여성 모두 비슷한 양상을 보이고 있음을 파악할 수 있다.

2. 구형성

-남성일 때

```
> cortest.bartlett(df2)
```

R was not square, finding R from data

\$chisq

[1] 6986.281

\$p.value

[1] 0

\$df

[1] 300

-여성일 때

```
> cortest.bartlett(df3)
```

R was not square, finding R from data

\$chisq

[1] 13312.67

\$p.value

[1] 0

\$df

[1] 300

남성과 여성 모두 p값이 매우 작게 나왔다. 따라서 상관행렬을 항등행렬이 아니므로 변수들간의 상관관계가 존재함을 알 수 있다.

3. KMO

-남성일 때

```
> KMO(df2)
```

Kaiser-Meyer-Olkin factor adequacy

Call: KMO(r = df2)

Overall MSA = 0.84

MSA for each item =

	A1	A2	A3	A4	A5	C1	C2	C3	C4	C5
	0.73	0.85	0.86	0.89	0.89	0.84	0.79	0.84	0.82	0.85
	E1	E2	E3	E4	E5	N1	N2	N3	N4	N5
	0.83	0.88	0.88	0.88	0.90	0.73	0.74	0.84	0.84	0.84

O1 O2 O3 O4 O5
0.83 0.77 0.83 0.77 0.74

-여성일 때

> KMO(df3)

Kaiser-Meyer-Olkin factor adequacy

Call: KMO(r = df3)

Overall MSA = 0.84

MSA for each item =

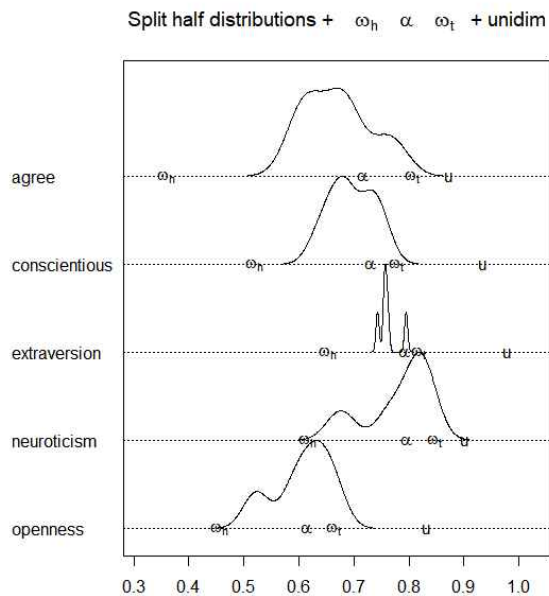
A1	A2	A3	A4	A5	C1	C2	C3	C4	C5
0.74	0.82	0.87	0.84	0.90	0.82	0.78	0.83	0.82	0.86
E1	E2	E3	E4	E5	N1	N2	N3	N4	N5
0.83	0.88	0.89	0.86	0.88	0.80	0.80	0.86	0.89	0.88
O1	O2	O3	O4	O5					
0.86	0.77	0.85	0.75	0.75					

남성과 여성 모두 Overall MSA 값이 0.84로 1에 가까우므로 표본이 적합하다고 판단할 수 있다.

4. reliability

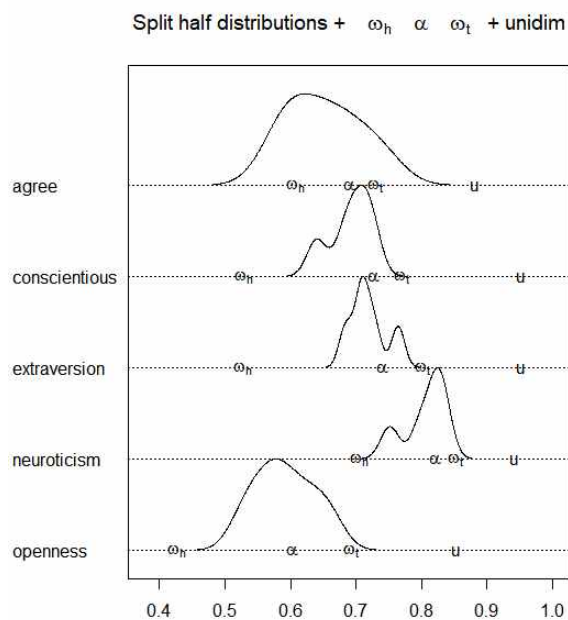
-남성일 때

> bfrel2 = reliability(bfi.keys,df2);plot(bfrel2)



-여성일 때

> bfrel3 = reliability(bfi.keys,df3);plot(bfrel3)

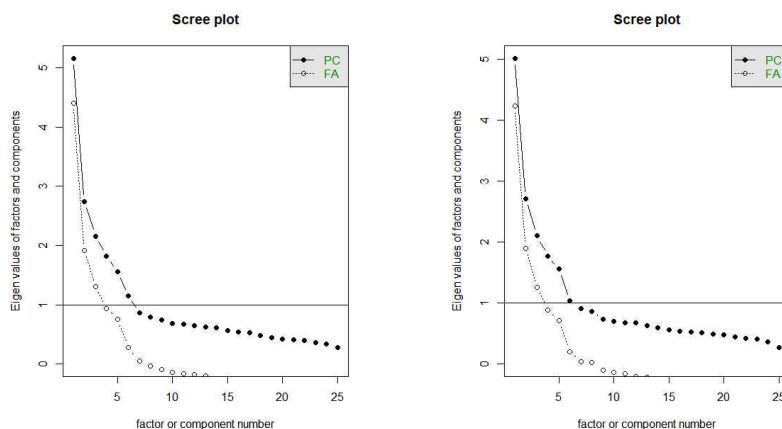


남성과 여성 모두 agree, conscientious, extraversion, neuroticism의 신뢰성을 보면 모두 0.7을 넘어가 응답자의 일치성이 높게 나타나는 것을 파악할 수 있었지만 openness에서는 상관관계 plot에서도 확인을 했지만 음과 양이 섞여 있어 다소 일치성이 낮게 나왔음을 파악할 수 있다.

(2) 요인개수 판단: scree, parallel, nfactors 등

1. Scree plot

> scree(df2);scree(df3)



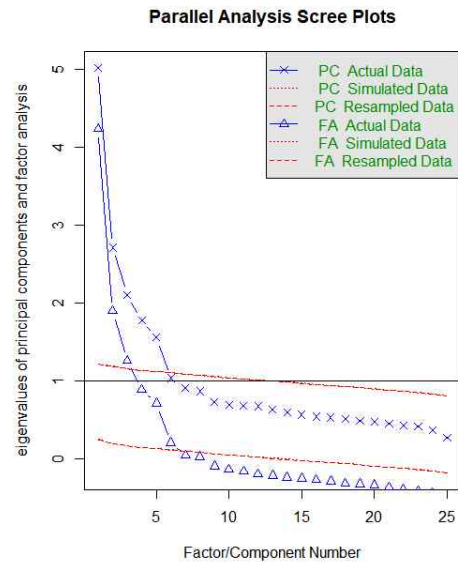
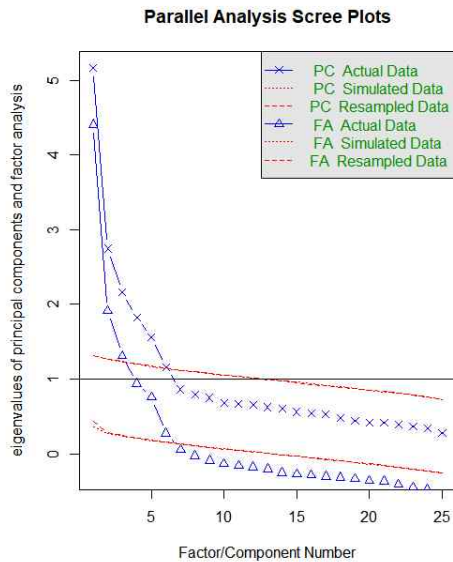
PCA에서는 남녀 모두 6개, FA에서는 3개의 요인 개수가 나왔음을 파악할 수 있다.

2. Parallel

```
> fa.parallel(df2):fa.parallel(df3)
```

Parallel analysis suggests that the number of factors = 6 and the number of components = 5

Parallel analysis suggests that the number of factors = 6 and the number of components = 5



3. nfactors

-남성일 때

```
> nfactors(df2)
```

Number of factors

Call: `vss(x = x, n = n, rotate = rotate, diagonal = diagonal, fm = fm,`

`n.obs = n.obs, plot = FALSE, title = title, use = use, cor = cor)`

VSS complexity 1 achieves a maximum of Although the vss.max shows 4 factors, it is probably more reasonable to think about 2 factors

VSS complexity 2 achieves a maximum of 0.74 with 4 factors

The Velicer MAP achieves a minimum of 0.02 with 5 factors

Empirical BIC achieves a minimum of -852.79 with 6 factors

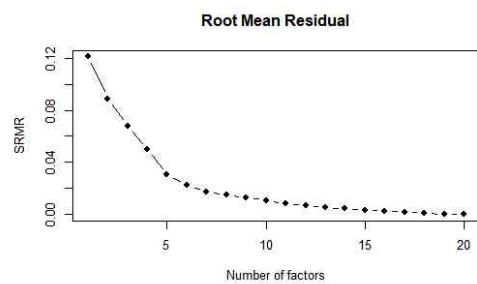
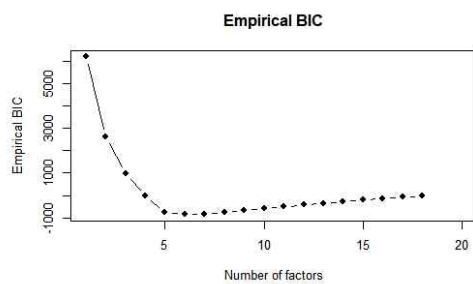
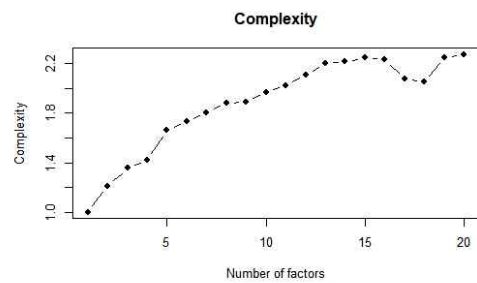
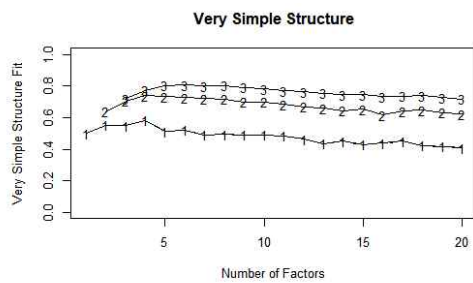
Sample Size adjusted BIC achieves a minimum of -229.8 with 8 factors

Statistics by number of factors

	vss1	vss2	map	dof	chisq	prob	sqresid	fit
1	0.50	0.00	0.025	275	4.1e+03	0.0e+00	26.1	0.50
2	0.55	0.64	0.020	251	2.8e+03	0.0e+00	18.8	0.64

3	0.55	0.70	0.019	228	2.0e+03	9.2e-277	14.5	0.72
4	0.58	0.74	0.017	206	1.4e+03	2.2e-168	11.5	0.78
5	0.51	0.73	0.015	185	7.0e+02	9.2e-61	9.2	0.82
6	0.52	0.73	0.017	165	4.7e+02	1.4e-30	8.0	0.85
7	0.49	0.72	0.019	146	3.1e+02	2.6e-13	7.5	0.86
8	0.49	0.72	0.023	128	2.4e+02	1.6e-08	7.1	0.86
9	0.49	0.70	0.028	111	1.8e+02	2.1e-05	6.6	0.87
10	0.49	0.70	0.034	95	1.4e+02	2.7e-03	6.3	0.88
11	0.48	0.69	0.040	80	8.6e+01	3.1e-01	5.9	0.89
12	0.46	0.67	0.047	66	5.9e+01	7.2e-01	5.7	0.89
13	0.44	0.66	0.056	53	3.8e+01	9.4e-01	5.5	0.90
14	0.45	0.65	0.067	41	2.7e+01	9.6e-01	5.1	0.90
15	0.43	0.65	0.078	30	1.5e+01	9.9e-01	5.0	0.90
16	0.44	0.62	0.093	20	9.4e+00	9.8e-01	4.5	0.91
17	0.45	0.64	0.110	11	5.2e+00	9.2e-01	4.2	0.92
18	0.42	0.65	0.128	3	1.8e+00	6.1e-01	4.0	0.92
19	0.42	0.63	0.156	-4	8.4e-05	NA	4.2	0.92
20	0.41	0.62	0.194	-10	4.7e-06	NA	4.3	0.92
RMSEA BIC SABIC complex eChisq SRMR eCRMS								
1	0.1232	2233	3106.1	1.0	8.1e+03	1.2e-01	0.1264	
2	0.1042	1041	1838.5	1.2	4.4e+03	8.9e-02	0.0972	
3	0.0915	426	1149.9	1.4	2.5e+03	6.8e-02	0.0779	
4	0.0779	-49	605.1	1.4	1.4e+03	5.0e-02	0.0603	
5	0.0550	-563	24.2	1.7	5.2e+02	3.1e-02	0.0392	
6	0.0446	-659	-135.3	1.7	2.7e+02	2.2e-02	0.0300	
7	0.0344	-691	-227.2	1.8	1.7e+02	1.8e-02	0.0251	
8	0.0304	-636	-229.8	1.9	1.3e+02	1.5e-02	0.0232	
9	0.0265	-575	-222.0	1.9	9.4e+01	1.3e-02	0.0215	
10	0.0221	-510	-208.6	2.0	6.6e+01	1.1e-02	0.0195	
11	0.0087	-460	-206.1	2.0	3.9e+01	8.4e-03	0.0163	
12	0.0000	-391	-181.7	2.1	2.5e+01	6.7e-03	0.0144	
13	0.0000	-324	-155.5	2.2	1.5e+01	5.3e-03	0.0126	
14	0.0000	-253	-122.7	2.2	1.1e+01	4.4e-03	0.0118	
15	0.0000	-190	-94.6	2.2	6.4e+00	3.4e-03	0.0108	
16	0.0000	-127	-63.6	2.2	3.6e+00	2.5e-03	0.0099	
17	0.0000	-70	-34.9	2.1	1.9e+00	1.9e-03	0.0097	
18	0.0000	-19	-9.1	2.1	5.9e-01	1.0e-03	0.0104	
19	NA	NA	NA	2.2	2.7e-05	7.0e-06	NA	
20	NA	NA	NA	2.3	1.6e-06	1.7e-06	NA	
eBIC								
1	6201							

2 2642
 3 987
 4 -28
 5 -741
 6 -853
 7 -827
 8 -747
 9 -663
 10 -582
 11 -507
 12 -425
 13 -346
 14 -269
 15 -198
 16 -133
 17 -73
 18 -20
 19 NA
 20 NA



-여성일 때

> nfactors(df3)

Number of factors

Call: vss(x = x, n = n, rotate = rotate, diagonal = diagonal, fm = fm,

n.obs = n.obs, plot = FALSE, title = title, use = use, cor = cor)

VSS complexity 1 achieves a maximum of 0.56 with 4 factors

VSS complexity 2 achieves a maximum of 0.74 with 5 factors

The Velicer MAP achieves a minimum of 0.01 with 5 factors

Empirical BIC achieves a minimum of -762.88 with 8 factors

Sample Size adjusted BIC achieves a minimum of -196 with 10 factors

Statistics by number of factors

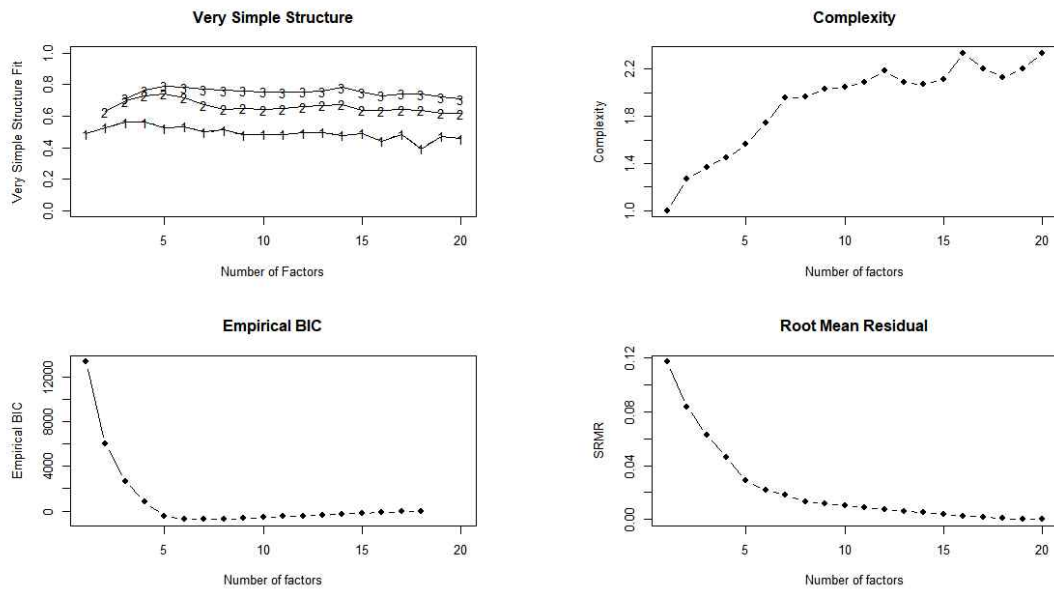
	vss1	vss2	map	dof	chisq	prob	sqresid	fit
1	0.49	0.00	0.024	275	7.8e+03	0.0e+00	25.7	0.49
2	0.53	0.63	0.017	251	4.8e+03	0.0e+00	18.6	0.63
3	0.56	0.70	0.016	228	3.2e+03	0.0e+00	14.5	0.71
4	0.56	0.73	0.015	206	2.2e+03	0.0e+00	11.7	0.77
5	0.53	0.74	0.015	185	1.3e+03	1.6e-159	9.5	0.81
6	0.53	0.72	0.016	165	7.8e+02	4.3e-80	8.6	0.83
7	0.50	0.67	0.019	146	5.6e+02	6.6e-50	8.1	0.84
8	0.51	0.64	0.022	128	3.7e+02	2.9e-25	7.5	0.85
9	0.48	0.65	0.027	111	3.0e+02	8.9e-20	7.1	0.86
10	0.49	0.64	0.033	95	2.2e+02	1.0e-11	6.9	0.86
11	0.49	0.65	0.038	80	1.8e+02	3.1e-09	6.7	0.87
12	0.49	0.66	0.047	66	1.3e+02	7.5e-06	6.4	0.87
13	0.50	0.67	0.056	53	9.6e+01	3.1e-04	6.0	0.88
14	0.48	0.67	0.066	41	6.0e+01	2.7e-02	5.7	0.89
15	0.49	0.64	0.078	30	3.9e+01	1.2e-01	5.4	0.89
16	0.44	0.63	0.090	20	2.3e+01	2.9e-01	5.2	0.90
17	0.48	0.64	0.108	11	8.1e+00	7.0e-01	5.0	0.90
18	0.39	0.64	0.135	3	1.7e+00	6.5e-01	4.7	0.91
19	0.47	0.62	0.167	-4	2.5e-02	NA	4.3	0.91
20	0.46	0.62	0.204	-10	3.0e-05	NA	4.7	0.91

	RMSEA	BIC	SABIC	complex	eChisq	SRMR	eCRMS
1	0.121	5758	6632	1.0	1.5e+04	1.2e-01	0.1222
2	0.098	2870	3667	1.3	7.9e+03	8.4e-02	0.0914
3	0.084	1508	2232	1.4	4.4e+03	6.2e-02	0.0716
4	0.072	660	1315	1.4	2.4e+03	4.6e-02	0.0555
5	0.056	-133	454	1.6	9.4e+02	2.9e-02	0.0368
6	0.044	-467	58	1.7	5.3e+02	2.2e-02	0.0293
7	0.039	-540	-76	2.0	3.6e+02	1.8e-02	0.0256

8	0.032	-596	-189	2.0	2.0e+02	1.3e-02	0.0205
9	0.030	-534	-181	2.0	1.6e+02	1.2e-02	0.0194
10	0.026	-498	-196	2.0	1.2e+02	1.0e-02	0.0181
11	0.025	-427	-172	2.1	8.8e+01	8.8e-03	0.0171
12	0.022	-370	-160	2.2	6.3e+01	7.5e-03	0.0160
13	0.021	-304	-136	2.1	4.3e+01	6.1e-03	0.0146
14	0.016	-249	-119	2.1	2.8e+01	5.0e-03	0.0134
15	0.013	-187	-92	2.1	1.6e+01	3.8e-03	0.0120
16	0.009	-128	-64	2.3	8.4e+00	2.7e-03	0.0105
17	0.000	-75	-40	2.2	3.9e+00	1.9e-03	0.0097
18	0.000	-21	-11	2.1	7.6e-01	8.2e-04	0.0082
19	NA	NA	NA	2.2	1.1e-02	1.0e-04	NA
20	NA	NA	NA	2.3	1.1e-05	3.1e-06	NA

eBIC

1	13386
2	5997
3	2678
4	833
5	-450
6	-711
7	-741
8	-763
9	-680
10	-600
11	-515
12	-434
13	-357
14	-281
15	-210
16	-142
17	-79
18	-22
19	NA
20	NA



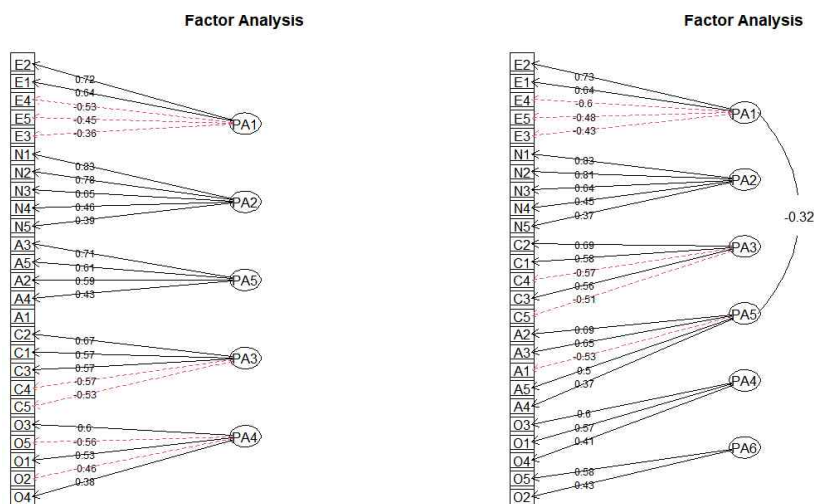
요인의 개수를 위의 그래프들을 참고하여 고려해봤을 때 5개와 6개로 각각 요인개수에 대안을 세워본다.

(3) 2개 대안(요인개수)에 대하여 요인분석 수행: fa diagram 비교, 요인적재 산점행렬도, 공통성, 회전 전/후의 변동 설명 비율 등

1. fa diagram

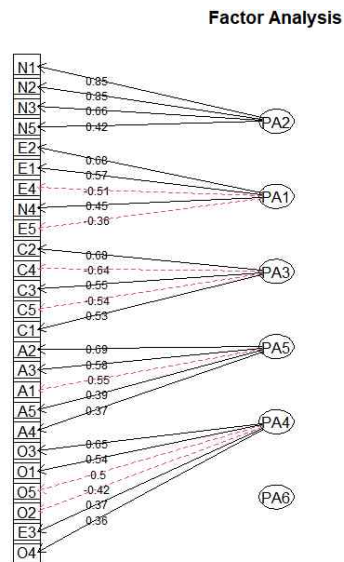
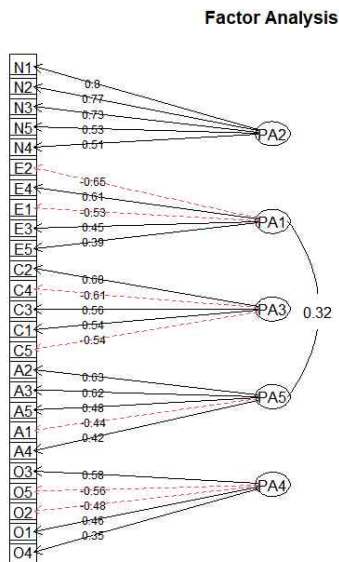
-남성일 때

```
> pa2_5 = fa(df2,nfactors = 5, fm='pa',max.iter = 100,rotate='oblimin');fa.diagram(pa2_5,digits = 2)
> pa2_6 = fa(df2,nfactors = 6, fm='pa',max.iter = 100,rotate='oblimin');fa.diagram(pa2_6,digits = 2)
```



-여성일 때

```
> pa3_5 = fa(df3,nfactors = 5, fm='pa',max.iter = 100,rotate='oblimin');fa.diagram(pa3_5,digits = 2)
> pa3_6 = fa(df3,nfactors = 6, fm='pa',max.iter = 100,rotate='oblimin');fa.diagram(pa3_6,digits = 2)
```

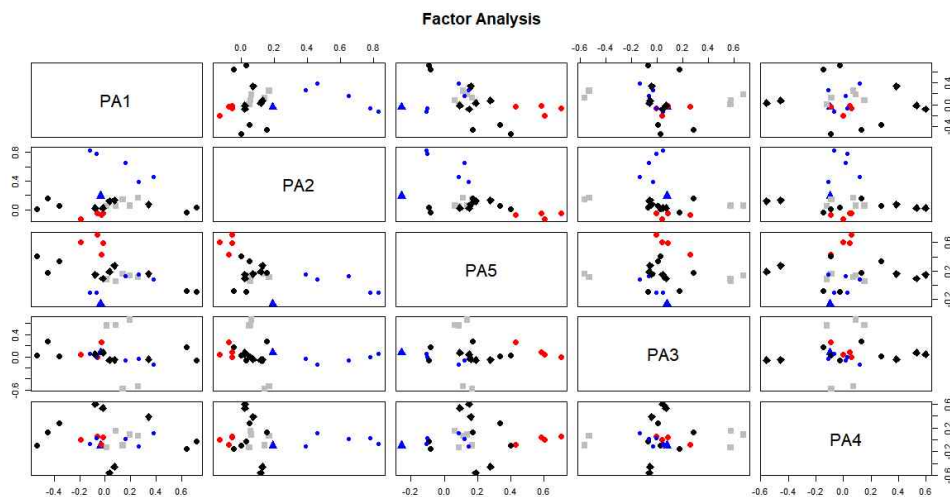


남성에서는 PA1이 가장 중요한 요인으로 나왔지만. 여성에서는 PA2가 가장 중요한 요인으로 나왔음을 알 수 있다. 또한 PA6은 요인으로서 영향력이 부족하다.

2. 산점행렬도

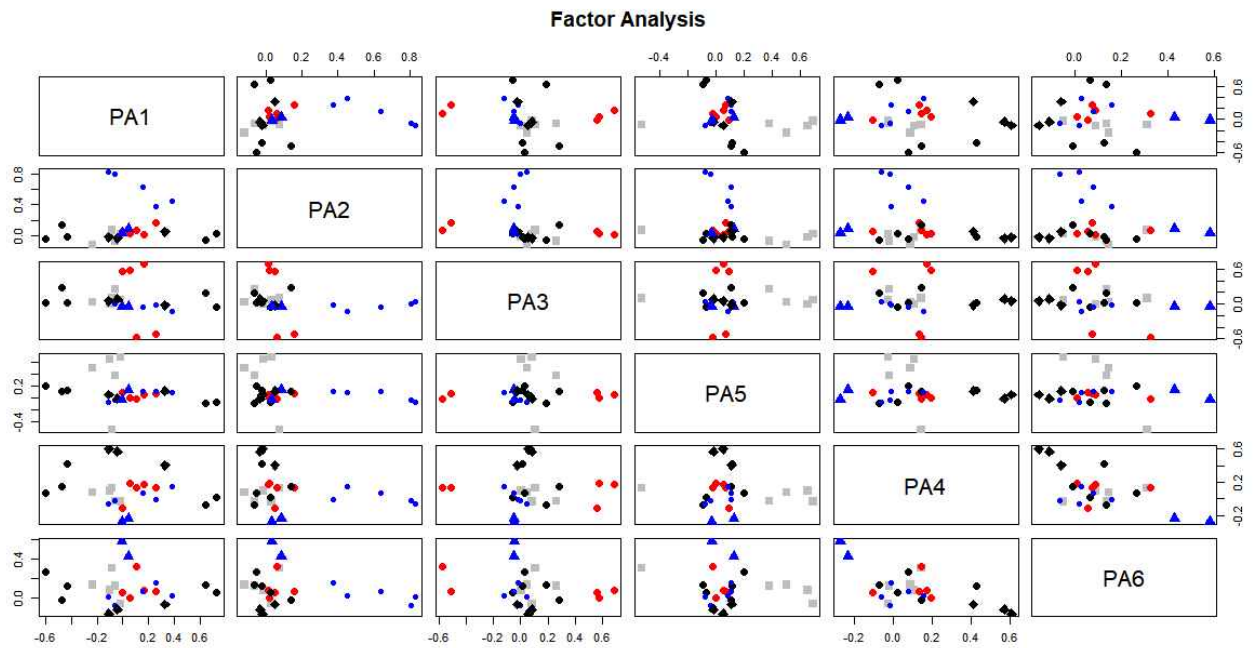
-남성일 때(요인 5개)

```
> plot(pa2_5,cut=0.3,cex=1.5)
```



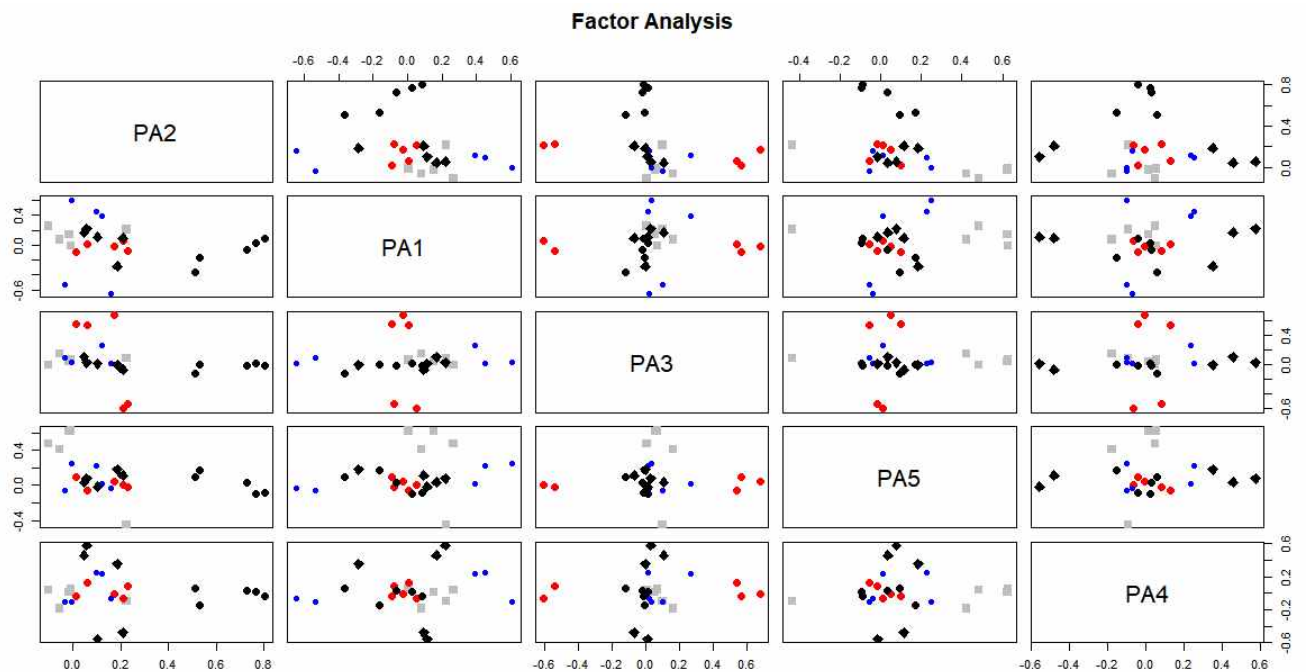
-남성일 때(요인 6개)

> plot(pa2_6,cut=0.3,cex=1.5)



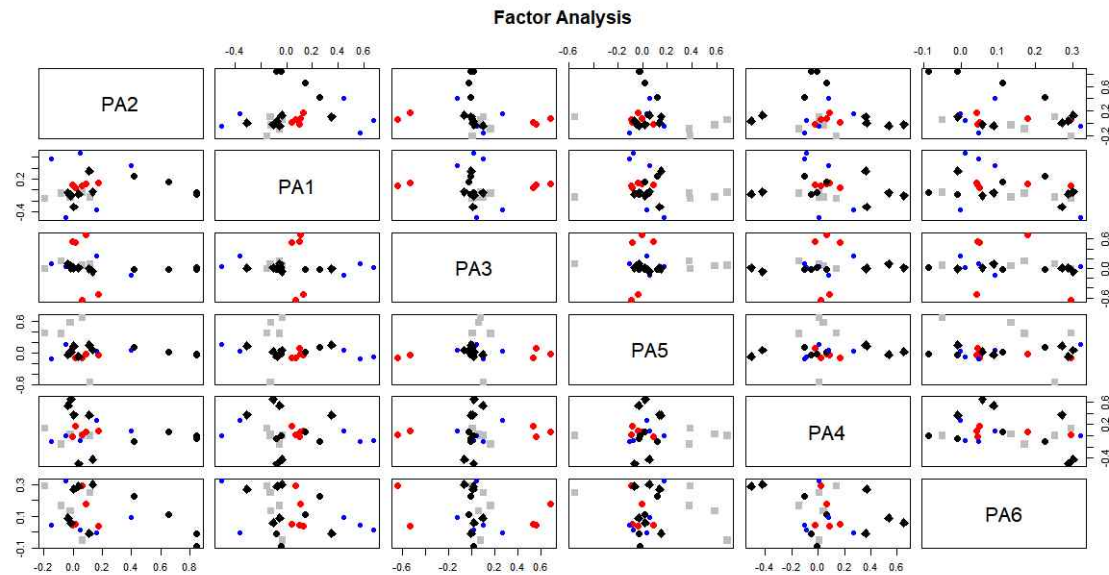
-여성일 때(요인 5개)

> plot(pa3_5,cut=0.3,cex=1.5)



-여성일 때(요인 6개)

> plot(pa3_6,cut=0.3,cex=1.5)



3. 공통성

-남성일 때(요인 5개)

> round(pa2_5\$communality,3)

	A1	A2	A3	A4	A5	C1	C2	C3	C4	C5	E1	E2	E3
E4	0.106	0.392	0.548	0.301	0.510	0.373	0.483	0.324	0.451	0.468	0.443	0.606	0.437
N1	0.460	0.655	0.590	0.516									
N2													
N3													
N4	0.490	0.292	0.332	0.277	0.448	0.272	0.346						
O1													
O2													
O3													
O4													
O5													

-남성일 때(요인 6개)

> round(pa2_6\$communality,3)

	A1	A2	A3	A4	A5	C1	C2	C3	C4	C5	E1	E2	E3
E4	0.296	0.487	0.552	0.298	0.500	0.382	0.506	0.319	0.535	0.464	0.447	0.604	0.503
N1	0.457	0.664	0.624	0.514									
N2													
N3													
N4	0.488	0.289	0.367	0.288	0.468	0.274	0.449						
N5													
O1													
O2													
O3													
O4													
O5													

-여성일 때(요인 5개)

```
> round(pa3_5$communality,3)
  A1   A2   A3   A4   A5   C1   C2   C3   C4   C5   E1   E2   E3
  E4   E5   N1   N2   N3
0.240 0.432 0.496 0.261 0.431 0.319 0.457 0.322 0.458 0.409 0.314 0.515 0.437 0.518
0.371 0.650 0.603 0.555
  N4   N5   O1   O2   O3   O4   O5
0.488 0.357 0.310 0.275 0.464 0.239 0.311
```

-여성일 때(요인 6개)

```
> round(pa3_6$communality,3)
  A1   A2   A3   A4   A5   C1   C2   C3   C4   C5   E1   E2   E3
  E4   E5   N1   N2   N3
0.332 0.491 0.485 0.257 0.464 0.322 0.487 0.315 0.556 0.407 0.356 0.542 0.472 0.530
0.373 0.699 0.684 0.544
  N4   N5   O1   O2   O3   O4   O5
0.495 0.367 0.340 0.290 0.486 0.243 0.329
```

남성 여성 모두 대체적으로 openness에서 공통성이 좀 낮게 나왔음을 파악할 수 있다.

4. 회전 전/후의 변동 설명 비율

-남성일 때(요인 5개)

```
> b.ev1 = pa2_5$e.values
> pv2_5b = 100*b.ev1[1:5]/sum(b.ev1):round(pv2_5b,3)
[1] 20.651 10.984 8.642 7.297 6.234
> a.ev1 = pa2_5$values
> pv2_5a = 100*a.ev1[1:5]/sum(a.ev1):round(pv2_5a,3)
[1] 43.382 20.991 14.760 11.380 9.496
```

-남성일 때(요인 6개)

```
> b.ev2 = pa2_6$e.values
> pv2_6b = 100*b.ev2[1:5]/sum(b.ev2):round(pv2_6b,3)
[1] 20.651 10.984 8.642 7.297 6.234
> a.ev2 = pa2_6$values
> pv2_6a = 100*a.ev2[1:5]/sum(a.ev2):round(pv2_6a,3)
[1] 40.920 19.866 14.120 10.954 9.127
```

요인 개수가 5개일 때가 6개일 때보다 더 설명력이 좋다고 판단할 수 있다.

-여성일 때(요인 5개)

```
> b.ev3 = pa3_5$e.values
> pv3_5b = 100*b.ev3[1:5]/sum(b.ev3);round(pv3_5b,3)
[1] 20.072 10.843 8.411 7.091 6.231
> a.ev3 = pa3_5$values
> pv3_5a = 100*a.ev3[1:5]/sum(a.ev3);round(pv3_5a,3)
[1] 43.679 21.536 14.704 10.995 9.091
```

-여성일 때(요인 6개)

```
> b.ev4 = pa3_6$e.values
> pv3_6b = 100*b.ev4[1:5]/sum(b.ev4);round(pv3_6b,3)
[1] 20.072 10.843 8.411 7.091 6.231
> a.ev4 = pa3_6$values
> pv3_6a = 100*a.ev4[1:5]/sum(a.ev4);round(pv3_6a,3)
[1] 41.372 20.550 14.113 10.503 8.915
```

마찬가지로 여성에서도 요인 개수가 5개일 때가 6개일 때보다 더 설명력이 좋다고 판단할 수 있다. 따라서 요인 5개로 분석을 해본다.

(4) 요인점수 추출 후 education, age에 대해 각각 회귀분석 유의성 검정

-남성일 때

```
> sc2_5 = factor.scores(df2,pa2_5);str(sc2_5)
List of 5
 $ scores : num [1:919, 1:5] -0.389 -0.754 -0.294 1.112 NA ...
 ..- attr(*, "dimnames")=List of 2
 .. ..$ : chr [1:919] "61617" "61622" "61624" "61629" ...
 .. ..$ : chr [1:5] "PA1" "PA2" "PA5" "PA3" ...
 $ weights : num [1:25, 1:5] -0.0182 0.042 0.0378 0.0218 -0.0311 ...
 ..- attr(*, "dimnames")=List of 2
 .. ..$ : chr [1:25] "A1" "A2" "A3" "A4" ...
 .. ..$ : chr [1:5] "PA1" "PA2" "PA5" "PA3" ...
 $ r.scores: num [1:5, 1:5] 1 0.222 -0.277 -0.246 -0.1 ...
 ..- attr(*, "dimnames")=List of 2
 .. ..$ : chr [1:5] "PA1" "PA2" "PA5" "PA3" ...
 .. ..$ : chr [1:5] "PA1" "PA2" "PA5" "PA3" ...
 $ missing : Named num [1:919] 0 0 0 0 1 0 1 0 0 0 ...
 ..- attr(*, "names")= chr [1:919] "61617" "61622" "61624" "61629" ...
 $ R2      : Named num [1:5] 0.9 0.917 0.892 0.881 0.844
 ..- attr(*, "names")= chr [1:5] "PA1" "PA2" "PA5" "PA3" ...
```

```

> sc2_5 = sc2_5$scores
> mylm2_1 = function(x){reg=lm(x~bfi2$education);rr=summary(reg)$coef;rr[2,3:4]}
> mylm2_2 = function(x){reg=lm(x~bfi2$age);rr=summary(reg)$coef;rr[2,3:4]}
> apply(sc2_5,2,mylm2_1)
          PA1          PA2          PA5          PA3          PA4
t value  0.2457986 -0.3124004 -0.7405849 -0.1980569 2.02792006
Pr(>|t|) 0.8059069 0.7548251 0.4591823 0.8430555 0.04292962
> apply(sc2_5,2,mylm2_2)
          PA1          PA2          PA5          PA3          PA4
t value -0.09133293 -1.0878786 -0.2175376 4.574785e+00 3.914749e+00
Pr(>|t|) 0.92725083 0.2769751 0.8278447 5.520225e-06 9.817819e-05

```

남성에서의 교육 수준이 올라갈수록 해당 요인들과의 유의한 변화가 있다고 하기에는 힘들다고 파악이 된다. 또한 나이가 들수록 conscientious와 openness가 증가함을 알 수 있다.

-여성일 때

```

> sc3_5 = factor.scores(df3,pa3_5):str(sc3_5)
List of 5
 $ scores : num [1:1881, 1:5] 0.0842 0.6147 -0.1685 0.1798 0.812 ...
 ..- attr(*, "dimnames")=List of 2
 .. ..$ : chr [1:1881] "61618" "61620" "61621" "61623" ...
 .. ..$ : chr [1:5] "PA2" "PA1" "PA3" "PA5" ...
 $ weights : num [1:25, 1:5] 0.047 0.00484 -0.00472 -0.00942 -0.00815 ...
 ..- attr(*, "dimnames")=List of 2
 .. ..$ : chr [1:25] "A1" "A2" "A3" "A4" ...
 .. ..$ : chr [1:5] "PA2" "PA1" "PA3" "PA5" ...
 $ r.scores: num [1:5, 1:5] 1 -0.191 -0.174 -0.114 -0.018 ...
 ..- attr(*, "dimnames")=List of 2
 .. ..$ : chr [1:5] "PA2" "PA1" "PA3" "PA5" ...
 .. ..$ : chr [1:5] "PA2" "PA1" "PA3" "PA5" ...
 $ missing : Named num [1:1881] 0 0 0 0 0 0 0 0 0 0 ...
 ..- attr(*, "names")= chr [1:1881] "61618" "61620" "61621" "61623" ...
 $ R2      : Named num [1:5] 0.925 0.883 0.874 0.859 0.835
 ..- attr(*, "names")= chr [1:5] "PA2" "PA1" "PA3" "PA5" ...
> sc3_5 = sc3_5$scores
> mylm3_1 = function(x){reg=lm(x~bfi3$education);rr=summary(reg)$coef;rr[2,3:4]}
> mylm3_2 = function(x){reg=lm(x~bfi3$age);rr=summary(reg)$coef;rr[2,3:4]}
> apply(sc3_5,2,mylm3_1)
          PA2          PA1          PA3          PA5          PA4
t value -2.34929515 -1.1949947 1.5568199 2.3096800 5.876289e+00
Pr(>|t|) 0.01893818 0.2322781 0.1197243 0.0210412 5.161159e-09

```

```
> apply(sc3_5,2,mylm3_2)
```

	PA2	PA1	PA3	PA5	PA4
t value	-7.046783e+00	2.841960433	3.4540229059	6.809303e+00	2.699441771
Pr(> t)	2.693014e-12	0.004539551	0.0005664394	1.374714e-11	0.007017417

여성에서의 교육 수준이 올라갈수록 neuroticism은 떨어짐을 알 수 있고, conscientious와 openness가 증가함을 알 수 있다. 또한 여성이 나이가 들수록 neuroticism은 떨어짐을 알 수 있고 agree, conscientious, extraversion, openness는 증가함을 알 수 있다.

(5) 결과 해석

다음은 전체 데이터였을때의 회귀분석 유의성 검정이다.

```
apply(sc5, 2, mylm1)
```

	PA2	PA1	PA3	PA5	PA4
t value	-6.241595e+00	2.4017049	5.860964e+00	5.711623e+00	4.058459e+00
Pr(> t)	5.093494e-10	0.0163934	5.225008e-09	1.255316e-08	5.095523e-05

```
apply(sc5, 2, mylm2)
```

	PA2	PA1	PA3	PA5	PA4
t value	-2.12146848	-1.0247996	1.1960157	1.0907551	5.695084e+00
Pr(> t)	0.03399222	0.3055687	0.2318173	0.2754983	1.395501e-08

결과를 보면 나이가 들수록 neuroticism이 줄어듦을 파악할 수 있고, 나머지 요인들은 나이가 들수록 증가함을 알 수 있었다. 하지만 남성이 나이가 들수록 conscientious와 openness가 증가함을 알 수 있었고 여성에서는 전체 데이터의 경향과 비슷하게 neuroticism이 줄어듦을 파악할 수 있고, 나머지 요인들은 나이가 들수록 증가함을 알 수 있었다.

교육 수준에서 살펴보면 전체 데이터에서는 학력이 높아질수록 neuroticism이 떨어짐을 알 수 있었고, openness는 증가함을 파악할 수 있었다. 하지만 남성에서의 교육수준이 올라갈수록 해당 요인들과의 유의한 변화가 있다고 하기에는 힘들다고 파악이 되고 여성에서는 교육수준이 올라갈수록 neuroticism은 떨어짐을 알 수 있고, conscientious와 openness가 증가함을 알 수 있다.

전체 데이터와 여성은 비슷한 경향을 보이고 있음을 파악할 수 있지만 남성은 전체 데이터에 크게 반영이 되지 못함을 파악할 수 있었다. 남성과 여성 간에 유의미한 차이를 확인할 수 있었다.