CHAPTER 7

요인 분석과 신뢰도 계수

7.1. 요인 분석 개념

요인 분석(FA: Factor Analysis 혹은 인자 분석이라고도 함)은 사람의 지적 능력을 측정하고 이에 연관된 변수들을 이해하려는 노력의 일환으로 Galton(회귀 분석 창시자)에 (1888) 의해 제안되었고 수학적인 모형은 Spearman(1904 상관 계수 제안자)에 의해 발전되었다. 요인 분석은 변수들의 내재된 상관 관계를 이용하여 요인을 구하고 이를 이용하여 (1)변수들을 분류하고 (변수 그룹에는 원 변수 일부만 포함되어 있다) (2)그룹에 적절한 의미를 부여하는(그룹 이름 부여) 분석 방법이다.

요인 분석의 예를 보면, 설문 조사에서 동일한 개념을 측정하기 위해 설계된 리커드(Likert) 척도 문항들이 정말 그런지 알아보기 위한 분석 방법으로 요인 분석이 사용된다. 물론 그 문항들의 신뢰도(혹은 내적 일치도)는 Cronbach α로 측정된다. 예를 들어 학생들의 학교 만족도를 측정하기 위하여 교수 강의, 조교, 행정 인력, 강의실, 도서실, 전산실습실, 체육 시설, 건물 만족도를 조사하였다고 하자. 8 개의 만족도 항목을 그룹화 할 수 있을까? 이에 대한 해답을 요인 분석이 제공한다. A 기업 지원자 48 명의 능력에 대해 측정한 15 개 항목 점수들을 분류(그룹)하고자 할 때 사용된다. 또한 기업 관련 지표에 관해 20 개의 항목을 (매출액, 종업원 수, 부채비율, …) 유사한 항목끼리 분류하고 할 때 사용한다.

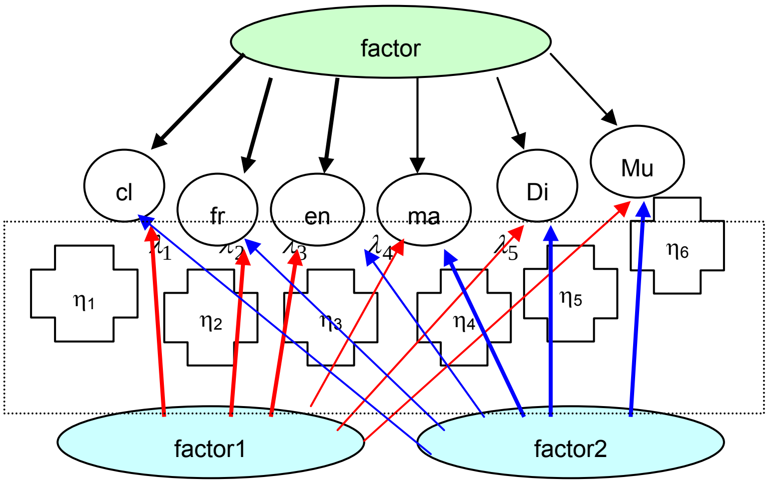
7.1.1. Spearman (1904)

Spearman 은 학생들의 6 과목 성적에 대한 상관 계수를 구한 후 상관 계수 값을 이용하여 각 학생들의 과목 성적은 다음과 같이 두 부분으로 나눌 수 있을 것이라 생각했다.(언어, 수리) 그러나 상관 계수 값으로 과목을 분류하는데 한계에 부딪히게 된다.

	Classic	French	English	Math	Discover	Music
Classic	1	.83	.78	.7	.66	.63
French		1	.67	.67	.65	.57
English			1	.64	.54	.51
Math				1	.45	.51
Discover					1	.4
Music						1

각 시험 점수(변수)는 그림과 같이 변수 간에 내재된 공통 개념(*f*: 이를 factor 라 함) 부분과 랜덤 부분에 해당하는(η) 부분으로 나눌 수 있을 것이다. 물론 f 와 η_j 들은 서로 독립이라고 가정하였다. 또한 그는 학생들의 과목 성적은 일반적 재능으로 해석되는 인자 f 와 과목에 대한 특별 재능으로 나눌 수 있다고 믿었다. 각 점수가 설문 조사 에서는 척도 문항에 해당 된다.

아래 그림 위의 까만 색 부분은 공통 개념이 하나(fact: 초록 부분)인 경우를 도식화 한 것이다. 이전 페이지 관계식을 도식화 한 것이다. 아래 부분은 6 개 과목에 2 개의 공통 개념(factor1: 빨간색 factor2: 파란색)이 존재할 때 도식화 한 그림이다. 굵은 선은 영향을 많이 미치는 것을 의미하므로 무엇인지는 모르지만 공통 개념이 영향을 주는 정도가 같은 과목끼리(변수끼리) 묶으면 될 것이다. 즉 고전 (classic), 불어(French), 영어(English)를 하나로 묶고 수학(Math), 과학(Discovery), 음악(Music) 하나로 묶을 수 있을 것이다.



$$classic = \lambda_1 f + \eta_1$$
$$french = \lambda_2 f + \eta_2$$
$$english = \lambda_3 f + \eta_3$$
$$math = \lambda_4 f + \eta_4$$
$$discover = \lambda_5 f + \eta_5$$
$$music = \lambda_6 f + \eta_6$$

7.1.2. 상관 계수

상관 계수(correlation coefficient)의 의미는 두 변수간의 선형 관계(linear relationship) 정도를 계산한 값으로 –1 에서 1 사이의 값을 갖는다. 두 변수간의 선형 상관 관계 정도를 나타내는 상관 계수 계산식은 다음과 같다.

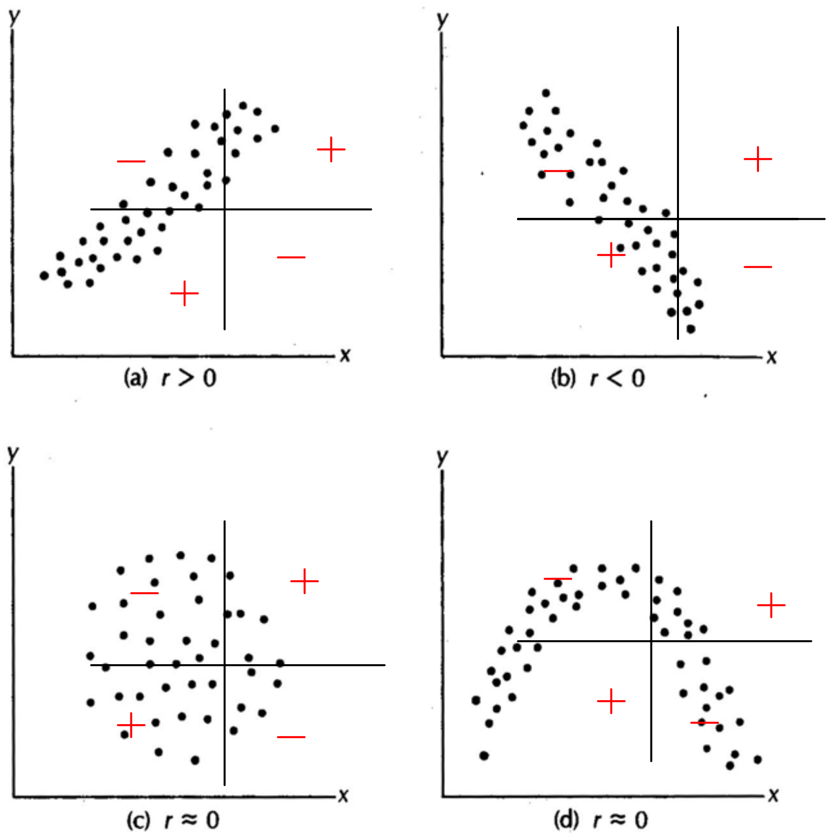
$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

상관 계수가 완벽한 1 이면 양의 상관 관계가 있는데 한 변수의 값이 증가(감소)하면 다른 변수의 값도 증가(감소)할 뿐 아니라 직선을 따라 움직인다. –1 이면 완벽한 음의 상관 관계가 있는데 한 변수 값이 증가(감소)하면 다른 변수의 값이 감소(증가)하고 직선을 따라 움직인다. 0 이면 상관 관계가 전혀 없다. 그럼 상관 계수의 값이 얼마이면 상관 관계가 존재한다고 할 수 있나? 이는 표본의 크기, 변수의 분산에 따라 다라지므로 다음 방법에 의해 모집단 상관 계수에 대한 t-검정을 하면 된다.

▪귀무가설: 두 변수간의 상관 관계는 없다. 상관 계수는 0 이다. $\rho = 0$

▪대립가설: 두 변수간의 상관 관계가 존재한다. $\rho \neq 0$

▪검정통계량 $t = \dfrac{r}{\sqrt{(1-r^2)/(n-2)}} \sim t(df = n-2)$



7.1.3. 요인 분석의 목적 및 예제

요인분석은 p 개의 변수들이 상호 어떤 관계가 있는지 결정하여 변수들을 m 개의 변수 그룹으로 (subset) 나누는데 목적이 있다.

▪변수에 내재된 관계를 설명할 상호 독립인 인자를 얻고 몇 개의 인자로 내재된 관계를 충분히 설명할 수 있는가 살펴본다.

▪인자의 부하(loading) 값에 의해 변수를 그룹화 하고 (다소 subjective) 변수 그룹에 적절한 이름을 붙인다.

요인분석의 가장 전형적인 예는 설문 분석에서 리커드 척도 문항(변수)들을 유사한 문항으로 분류할 때 사용한다. 예를 들어 학생들의 학교 만족도를 측정하기 위하여 교수 강의, 조교, 행정 인력, 강의실, 도서실, 전산실습실, 체육 시설, 건물 만족도를 조사하였다고 하자. 8개의 만족도 항목을 그룹화 할 수 있을까?

A 기업 지원자 48 명의 능력에 대해 측정한 15 개 항목 점수들을 분류(그룹)하고자 할 때 사용된다. 또한 기업 관련 지표에 관해 20 개의 항목을 (매출액, 종업원 수, 부채비율, ...) 유사한 항목끼리 분류하고 할 때 사용한다.

7.1.4. 요인 분석 모형 및 가정

$$\underline{x} = L\underline{f} + \underline{\eta} \quad \Leftrightarrow \quad \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_p \end{pmatrix} = \begin{pmatrix} l_{11}\, l_{12} \cdots l_{1m} \\ l_{21}\, l_{22} \cdots l_{2m} \\ \cdots \\ l_{p1}\, l_{p2} \cdots l_{pm} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ \cdots \\ f_m \end{pmatrix} + \begin{pmatrix} \eta_1 \\ \eta_2 \\ \cdots \\ \eta_p \end{pmatrix}$$

- $\underline{x}$ 는 원 변수 벡터이다.
- $f_k, f_2, \ldots, f_m$ 들은 공통 인자 (요인: common factor)
- l_{ij} 들은 인자 부하 (factor loading) ➔ i 번째 변수에 j 번째 요인이 미치는 영향
- $\eta_1, \eta_2, \ldots, \eta_p$ 특정 인자 (specific factor) 한다. ➔ η_j 는 j 번째 변수에 한정된 오차 변동

※방정식 풀기

원변수의 상관계수행렬(R)에 대해 $R = LL' + \psi$ 성립하므로 $R = LIL' + \psi = R = (LP)(LP)' + \psi$ 에 의해 이를 만족하는 L 은 무수히 많다. (무수히 많은 성질을 이용하여 요인의 부하 값을 잘 구별할 수 있도록 요인 회전 방법을 사용할 수 있게 된다. 가장 흔히 사용되는 방법이 VARIMAX 방법이다.) 요인 방정식의 해를 구하는 방법으로 가장 많이 사용되는 것이 principal factoring 방법이다.

※Principal factoring w/ or w/o iteration

변수의 상관 계수 행렬 R 에 대한 eigen-value(고유치)와 eigen-vector(고유 벡터)를 구하여 그것을 각각 $\lambda_1 \ge \lambda_2 \ge ...\lambda_p$, $\underline{e}_1, \underline{e}_2, ..., \underline{e}_p$ 라고 하자. 그리고 여기서부터 구한 주성분을 $y_1, y_2, ..., y_p$ 라 하자. 그러면 요인 방정식은 다음과 같이 쓸 수 있다.

$$\begin{aligned} x_1 &= y_1 e_{11} + y_2 e_{12} + ... + y_p e_{1p} \\ x_2 &= y_1 e_{21} + y_2 e_{22} + ... + y_p e_{2p} \\ &: \\ x_p &= y_1 e_{p1} + y_2 e_{p2} + ... + y_p e_{pp} \end{aligned}$$

요인은 $f_1 = y_1 / \sqrt{\lambda_1}$, $f_2 = y_2 / \sqrt{\lambda_2}$, …, $f_p = y_p / \sqrt{\lambda_p}$ 이므로

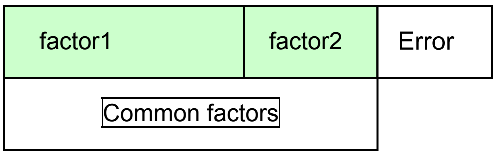
$$\begin{aligned} x_1 &= \sqrt{\lambda_1} e_{11} f_1 + \sqrt{\lambda_2} e_{12} f_2 + ... + \sqrt{\lambda_p} e_{1p} f_p \\ x_2 &= \sqrt{\lambda_{21}} e_{21} f_1 + \sqrt{\lambda_2} e_{22} f_2 + ... + \sqrt{\lambda_p} e_{2p} f_p \\ &: \\ x_p &= \sqrt{\lambda_1} e_{p1} f_1 + \sqrt{\lambda_2} e_{p2} f_2 + ... + \sqrt{\lambda_p} e_{pp} f_p \end{aligned}$$

요인의 부하 값은 $l_{ij} = \sqrt{\lambda_j} e_{ij}$ 이고 특정 인자(오차)는 $\psi_j^2 = \sigma_j^2 - (l_{j1}^2 + l_{j2}^2 + ... + l_{jm}^2)$ 이다.

7.1.5. 요인의 개수와 부하

부하(loading) 값의 의미는 각 요인이 원 변수를 설명하는 정도(크기)를 나타내며 요인은 변수들에 내재된 관계에서 공통 부분에 해당된다.

$$\begin{pmatrix} x_1 \\ x_2 \\ ... \\ x_p \end{pmatrix} = \begin{pmatrix} l_{11}\ l_{12} \dots l_{1m} \\ l_{21}\ l_{22} \dots l_{2m} \\ ... \\ l_{p1}\ l_{p2} \dots l_{pm} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ ... \\ f_m \end{pmatrix} + \begin{pmatrix} \eta_1 \\ \eta_2 \\ ... \\ \eta_p \end{pmatrix}$$

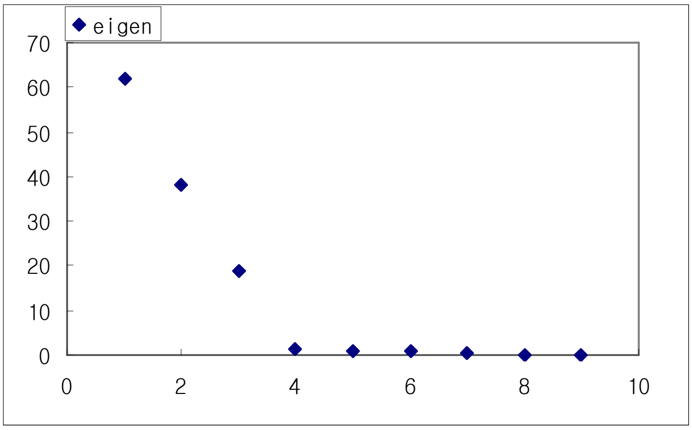


그러므로 각 요인에서 부하 값의 절대값이 큰 것들만 (음의 부호는 동일 개념의 반대 척도) 선택하여 변수들을 그룹화 하면 된다.

(1)trivial 한 요인은 제외하자. 원 변수 1-2 개에만 부하 값이 큰 요인은 제외하자. 이 요인에 의해 묶을 수 있는 변수는 1-2 개이므로 그룹의 의미가 없기 때문이다.

(2)Kaiser 판단(가장 많이 이용): 변수들의 상관 관계가 0 이면(관계가 없으면) 상관 계수 행렬은 R 은 항등 행렬 I 이다 이 경우 원 변수의 개수와 주성분의 개수가 같아지고 주성분의 분산은 모두 1 이므로 각 주성분이 가지는 분산 평균도 1 이다. 그러므로 상관 계수 행렬로부터 구한 고유치가 평균인 1 이상인 되어야 한다는 판단 하에 고유치가 1 이상인 것만으로 요인의 개수를 정한다. SAS 도 이 방법에 의해 요인의 개수 출력한다. 이 방법에 의해 요인의 개수를 결정하는 것이 가장 보편적인 방법이다.

(3)SCREE 그림 사용: 고유치를 y-축, 개수를 x-축으로 한 산점도인 SCREE 그림을 사용하여 인자의 개수를 예상한다. 총변동 80%에 연연하지 말고 주성분 분산 설명 변동의 크기(고유치)가 갑자기 줄어들기 바로 전까지의 개수로 적절한 인자 개수로 사용하면 된다.



7.1.6. 요인 회전 (Factor Rotate)

요인 분석에서 요인의 부하 값은 요인(공통 개념)과 원 변수의 상관 관계 정도를 나타내는 크기로 해석될 수 있으므로 부하 값에 의해 원 변수를 그룹화 한다. 그러나 (1)요인의 복합성: 하나의 원 변수에 부하 값이 큰 요인이 2 개 이상 존재하거나 (2)인자의 크기가 0 을 중심으로 ±의 작은 값이 있는 경우 부하 값으로 변수를 그룹화 하는 것은 불가능하다. 요인 회전은 각 요인이 상대적으로 큰 부하 값을 갖도록 요인을 회전(rotate)하는 것으로 QUARTIMAX rotation, OBLIQUE rotation, PROMAX rotation 방법이 있는데 가장 많이 사용되는 것은 직교 회전 방법인 VARIMAX 방법이다. VARIMAX 방법은 Kaiser 가 제안한 것으로 간단한 구조의 측정치로 요인 행렬의 각 열 내의 부하 제곱의 분산의 합을 제안하고 이 분산을 최대화 하는 회전 방법이다.

7.2 설문 분석에 요인 분석 이용

7.2.1. 사용 방법

요인 분석이 언제 설문 분석에 이용될 수 있을까? 리커드(Likert) 척도로 조사된 문항들을 그룹화하는데 사용된다. 몇 문항들을 합쳐 하나의 지표(index) 점수로 사용할 수 있느냐를 알아볼 때 요인 분석이 사용된다. 위에서 원 변수 $x_1, x_2, ..., x_p$ 가 설문 조사의 각 리커드 척도 문항에 해당된다. 예제 설문에서 시설물 관련 만족 정도를 묻는 문항이 Q4-Q13 으로 열 문항이다. 이 10 문항을 하나로 혹은 2-3 그룹으로 묶어 어떤 항목을 측정하는 점수로 사용할 수 있느냐가 궁금할 것이다. 만약 하나로 묶어진다면 그 10 개 문항의 (평균) 점수가 응답자들의 시설물 만족도 점수가 되는 것이다. 만약 2 개 이상으로 묶어진다면 각 그룹을 구성하는 문항을 고려하여 조사자가 이름을 부여하면 된다.

문항을 몇 개의 그룹으로 묶을 수 있느냐는 고유치가 1 이상인 **요인의 수**에 의해 결정되고 그룹에 어떤 문항이 묶여지느냐는 **loading(부하) 값**에 의해 결정된다. 설문 분석에서 요인 분석이 가능 하려면 다음 2 조건이 만족되어야 한다.

(1)리커드 척도 문항이어야 한다.

(2)여러 문항들을 몇 개의 그룹으로 묶으려는 목적에서 실시해야 한다.

7.2.2 통계 소프트웨어 사용

[SAS]

예제 설문 시설물에 대한 Q4-Q12 번 문항을 어떻게 그룹화 할지 요인 분석하여 보자. 요인 분석은 그룹으로 나눌 필요가 있는 리커드(Likert) 척도 문항에(4 점, 5 점, 7 점 척도) 대해 가능하다. 다음 프로그램은 리커드 척도 문항에 대한 요인 분석을 실시할 때 사용되는 전형적인 프로그램이다. 다른 부분은 그대로 사용하고 DATA = ~ 부분과 VAR = ~의 ~ 부분만 적절히 고쳐주면 된다.

```
PROC FACTOR DATA=SURVEY ROTATE=VARIMAX REORDER;
    VAR Q4-Q12;
RUN;
```

ROTATE=VARIMAX 옵션은 요인을 직교 변환하는 방법 중 VARIANCE 를 최대한 방법을 사용하라는 것으로 부하 값을 잘 구별할 수 있다.

REORDER 옵션은 부하 값의 크기 순서대로 출력하라는 명령으로 변수(문항)를 그룹화 하는데 편리하다.

COVARIANCE 변수(문항)들의 공분산 행렬을 이용하여 요인을 추정한다. 변수들의 측정 단위가 다를 경우는 상관 행렬(COVARIANCE 를 사용하지 않으면 된다. default)을 사용 해야 하나 설문 분석에서 문항들은 리커드 척도이고 같은 점수 척도이므로 공분산 행렬을 권한다.

요인 추정 방법(METHOD)은 default=주성분 방법을 선택하였다.

Eigenvalues of the Correlation Matrix: Total = 9 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	3.75509467	2.75103818	0.4172	0.4172
2	1.00405650	0.05050753	0.1116	0.5288
3	0.95354897	0.16659613	0.1059	0.6347
4	0.78695284	0.13707952	0.0874	0.7222
5	0.64987332	0.14284924	0.0722	0.7944
6	0.50702408	0.02290537	0.0563	0.8507
7	0.48411871	0.04536948	0.0538	0.9045
8	0.43874923	0.01816755	0.0487	0.9533
9	0.42058168		0.0467	1.0000

2 factors will be retained by the MINEIGEN criterion.

실험실 자료나 측정 자료인 경우 고유치가 1 이상인 경우만 택해도 누적 설명 비율이 80% 이지만 리커드 척도 문항과 같이 1, 2, 3, 4, 5 이산형 데이터인 경우 누적 설명 비율이 매우 낮다.(53%) 그러나 설문 조사에서 요인 분석은 리커드 척도 문항 분류하는데 사용되므로 걱정하지 말자.

Rotated Factor Pattern

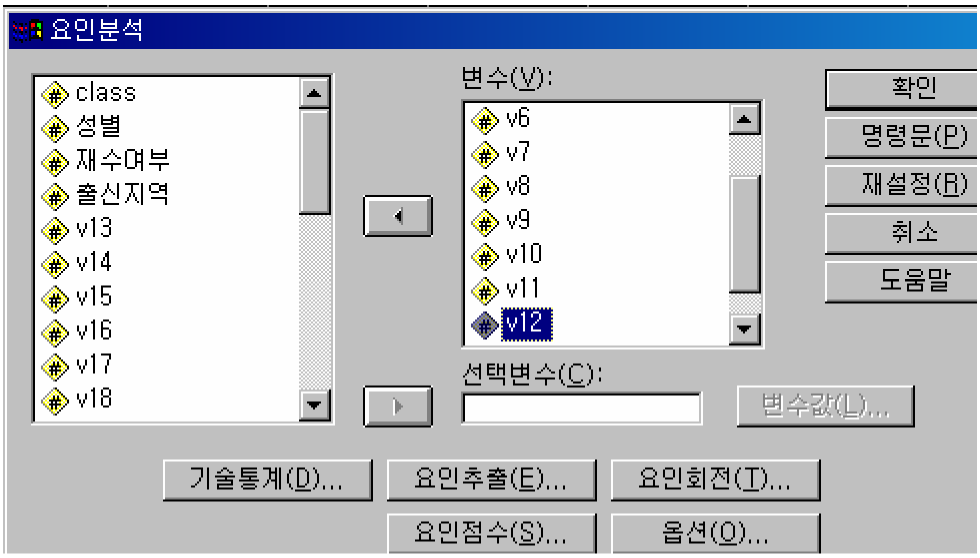
	Factor1	Factor2
Q7	0.73815	0.21534
Q8	0.69564	-0.03666
Q6	0.64055	0.20458
Q5	0.62266	0.37486
Q9	0.56059	0.49437
Q10	0.13259	0.80343
Q11	0.21992	0.75404
Q12	0.08964	0.55255
Q4	0.51461	0.53426

← 이 부분 결과를 이용하면 된다

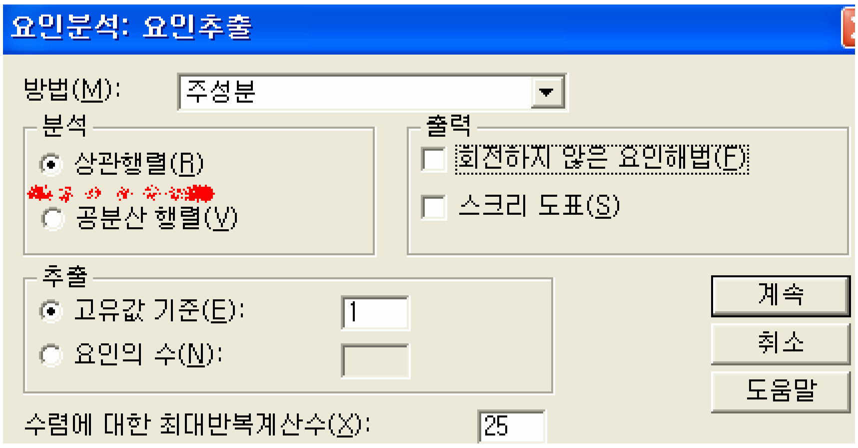
부하의 값의 크기가 0.6 이상인 (크기 값이 유사하고) 변수(문항)를 묶으면(분류하면) 된다. 요인 1(factor 1)이 주로 설명하는 변수는 Q5-Q8 이므로 묶으면 되고, 요인 2 의 부하 값에

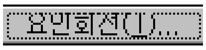
의해서는 Q10-Q11 을 하나로 묶을 수 있을 것이다. Q4, Q12 도 요인 2 에 의해 묶을 수 있을 것 같으나 0.75 에 비해 0.55 면 차이가 많으므로 안 묶는 것이 좋다. 그러므로 문항은 (강의실 만족도: Q5, Q6, Q7, Q8), (정보 시설 만족도: Q10, Q11)을 묶고 나머지 문항들은 개별 문항으로 간주하여 분석한다.

[SPSS]

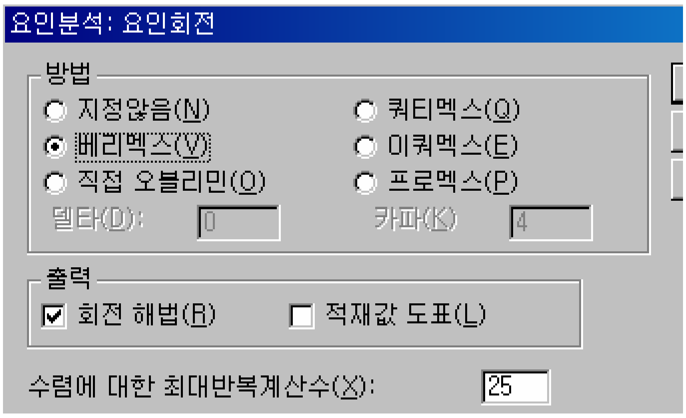


요인추출(E)... 요인들을 구할 때 상관 계수 행렬을 사용해야 변수의 변동을 사용하는 것을 권하다. 요인 수를 제한하는 방법으로 고유 값이 1 이상인 것만 출력하게 한다. 아래와 같이 설정하기 바란다.





요인의 부하 값을 잘 구별하기 위하여 직교 회전 방법을 택하는 것이 좋다. VARIMAX 방법을 권한다.



회전된 성분행렬

	성분	
	1	2
V4	.515	.534
V5	.623	.375
V6	.641	.205
V7	.738	.215
V8	.696	-3.67E-02
V9	.561	.494
V10	.133	.803
V11	.220	.754
V12	3.964E-02	.553

출력 결과는 **SAS** 와 동일하다.

7.2.3. 보고서 작성

리커드 척도 문항들에 대한 요인 분석 결과는 다음과 같이 정리하면 된다.

	Factor1	Factor2
Q7	**0.74**	**0.22**
Q8	**0.70**	**-0.04**
Q6	**0.64**	**0.20**
Q5	**0.62**	**0.37**
Q9	**0.56**	**0.49**
Q10	**0.13**	**0.80**
Q11	**0.22**	**0.75**
Q12	**0.09**	**0.55**
Q4	**0.51**	**0.53**
신뢰도 계수	**0.69**	**0.68**

이제 분류된 리커드 척도 문항 사용에 대해 설명해 보자. 요인 1(factor 1)에 의해 Q5-Q8 을 묶고 강의실 만족도라 하고 요인 2 에 의해 Q10, Q11 을 묶어 정보 시설 만족도라 하였다. 향후 분석(회귀 분석, 분산 분석, 기초 통계량 분석)에서는 묶은 변수 집단(묶은 변수는 합보다는 평균을 이용하는 것이 바람직하다. 이유는 (1)단위 맞추기 (2)결측치 있는 경우)을 하나로 사용하는 것이 좋다. 물론 개별적으로 보기도 하지만…

```
DATA SURVEYO;
    SET SURVEY;
        LECTURE=MEAN(OF Q5-Q8);
        INFORMATION=MEAN(Q10, Q11);
RUN;
```

요인을 2 개 선택하였음에도 불구하고 누적 변동이 55% 밖에 되지 않는 것은 Q8 이 빠져 있고 요인 1 에서 Q4, Q6, Q8 의 부하 값이 다른 문항에 비해 작기 때문이다. 그러나 이 부분에 대해서는 굳이 언급할 필요가 없다. 설문 분석 시 요인 분석을 이용하는 주된 이유는 문항 분류라는 사실을 잊지 말아야 할 것이다. 그리고 리커드 척도 문항과 같이 실험실 측

정 자료가 아니면 누적 변동이 낮을 수 밖에 없다. 마지막 행에는 묶은 문항의 신뢰도 계수(내적 일치도)를 적어 주면 된다. 이 값을 구하는 방법은 다음 절에 있다.

7.3. 문항의 내적 일치도(신뢰도 계수)

7.3.1. 내적 일치도 개념

문항이 요인 분석에 문항이 그룹화 되면 문항들이 하나의 개념(index)을 얼마나 잘 표현하는지를 알아보는 것을 내적 일치도(internal consistency)를 알아본다고 하는데 이 개념을 계산한 값이 Cronbach alpha(α)라 한다. 이를 문항의 신뢰도라 하기도 한다. 응답자로부터 얻은 설문 응답 결과(측정치: observed value)는 실제 응답자의 만족 점수와 측정 오차(measurement error)로 구성되어 있다. $Y = T + E,\ \ \mathrm{cov}(T, E) = 0$. 그러므로 측정치의 신뢰 계수(reliability coefficient)는 다음과 같이 정의된다.

$$\begin{aligned}\sigma^2(Y,T) &= \frac{\mathrm{cov}(Y,T)^2}{\mathrm{var}(Y)\,\mathrm{var}(T)} \\ &= \frac{\mathrm{var}(T^2)}{\mathrm{var}(Y)\,\mathrm{var}(T)} \\ &= \frac{\mathrm{var}(T)}{\mathrm{var}(Y)}\end{aligned}$$

위의 측정치 신뢰 계수는 변수가 하나인 경우인데, 이를 변수가 여러 개인 경우(문항이 여러 개)로 일반화 시킨 것이 Cronbach α 값이다. p 개 문항이 있을 경우 $Y_j = T_j + E_j\,(j = 1,2,\dots,p)$ 이고 $Y_O = \sum Y_j, T_0 = \sum T_j$ 라고 놓으면 다음이 성립한다.

$$\begin{aligned}\alpha &= \left(\frac{p}{p-1}\right)\frac{\sum_{i\neq j}\mathrm{cov}(Y_i, Y_j)}{\mathrm{var}(Y_O)} \\ &= \left(\frac{p}{p-1}\right)\left(1 - \frac{\sum_j \mathrm{var}(Y_j)}{\mathrm{var}(Y_O)}\right)\end{aligned}$$

Cronbach α는 0 과 1 사이의 값이고 1 에 가까울수록 내적 일치도가 높다. 얼마면 높다고 할 수 있는가? 0.6 이상? 0.7 이상? 그러나 이런 기준에는 나는 수긍할 수 없다. 왜냐하면 Cronbach α 값은 문항의 수가 많을수록, 응답자 수가 많을수록 높아지는 경향이 있기 때문이다. 그러므로 값의 크기가 판단의 근거가 되는 것이 아니라 한 문항을 제외했을 때 Cronbach α 값이 적어지느냐, 커지느냐를 를 보고 그 문항을 제외하느냐 그대로 두느냐를 판단하기 바란다. 그러나 보고서나 논문 작성과 같이 내적 일치도 값을 제시해야 하는 경우에는 전체 내적 일치도 값(Cronbach α)을 제시할 수 밖에는 없다. 다시 강조하지만 이 값의 크기가 중요한 것이 아니라 문항을 제외하였을 때 CRONBACH 값의 변화가 더 중요하다.

문항의 보기가 2 개(binary, dichotomous (0,1)) 한 경우 Cronbach α 신뢰 계수는 Kuder-Richardson 20 (KR-20) 신뢰 계수가 된다.

7.3.2. 통계 소프트웨어 사용

[SAS]

요인 분석에 의해 묶은 리커드 척도 문항에 대해서만 내적 일치도 Cronbach α를 구하면 된다.

```
PROC CORR DATA=SURVEY NOSIMPLE NOCORR ALPHA;
    VAR Q5-Q8;
RUN;
```

```
PROC CORR DATA=SURVEY NOSIMPLE NOCORR ALPHA;
    VAR Q10 Q11;
RUN;
```

▪NOCORR 은 변수(문항)들의 상관 계수 값을 출력하지 말라는 옵션이다.

▪NOSIMPLE 은 변수들의 기초 통계량(평균, 표준 편차)을 출력하지 말라는 옵션이다.

▪ALPHA 는 CRONBACH 값을 계산하라는 옵션이다.

Cronbach의 α 계수

변수	α 계수
원데이터	0.686085
표준화	0.690973

변수를 제외했을때의 Cronbach 계수

삭제한 변수	데이터 변수 합계와의 상관 계수	데이터 변수 α 계수	표준화된 변수 합계와의 상관	표준화된 변수 α 계수
Q5	0.513210	0.596535	0.515436	0.599651
Q6	0.446627	0.634924	0.452792	0.639815
Q7	0.555856	0.574055	0.552425	0.575144
Q8	0.386110	0.670726	0.381783	0.683350

일반적으로 변수를 표준화 시킨 후 구한 신뢰도 계수 값이 크므로 이를 이용한다. Q5-Q8 4 개 문항 모두 사용할 경우 신뢰도 계수는 0.69 이다. Q5 를 제외하고 Q6-Q8 만 사용하면 신뢰도 계수가 0.59 로 떨어진다.(생략) Q8 를 제외하고 Q5-Q7 만 사용하면 신뢰도 계수가 0.68 로 떨어진다. 그러므로 4 개 변수(문항)를 묶는 것이 옳으며 신뢰도 계수(내적 일치도)는 0.68 이다.

Cronbach의 α 계수

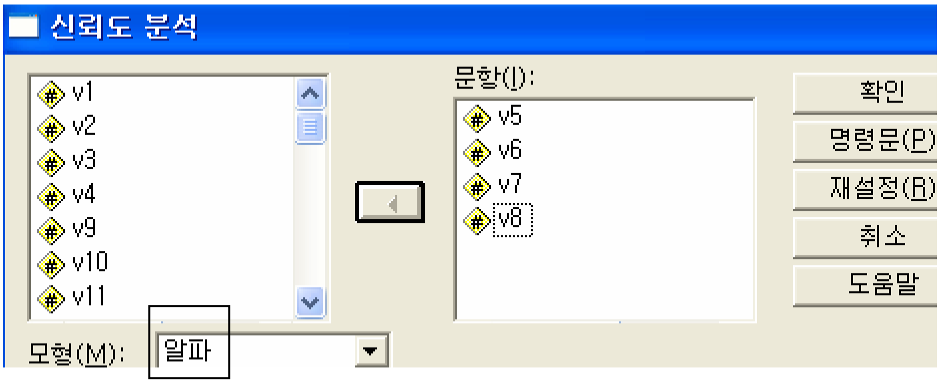
변수	α 계수
원데이터	0.670557
표준화	0.684502

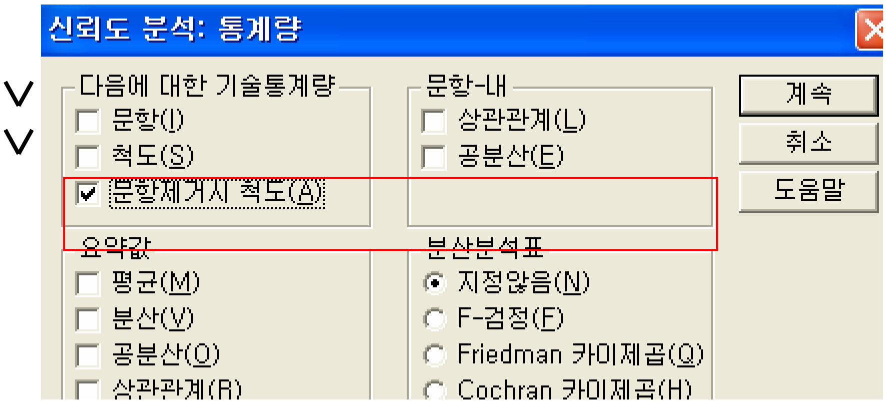
변수를 제외했을때의 Cronbach 계수

삭제한 변수	데이터 변수 합계와의 상관 계수	데이터 변수 α 계수	표준화된 변수 합계와의 상관	표준화된 변수 α 계수
Q10	0.512626	.	0.520337	.
Q11	0.512626	.	0.520337	.

변수가 2 개인 경우는 제외 신뢰도 계수가 계산될 수 없다. (Q10, Q11) 문항의 신뢰도 계수는 0.68 이다.

[SPSS]





7.3.3. 보고서 작성

신뢰도 계수에 대한 보고서 작성은 요인 분석 결과와 같이 하면 된다. 그러므로 7.1.9 절에 있는 표 하나로 요인 분석과 신뢰도 분석 정리는 충분하다. 어떤 책에는 신뢰도 계수가 0.7 이상 되어야 문항들을 신뢰할만하다고 하여 0.7 이하가 나오면 숫자를 고치는 경향이 있는데 그럴 필요 없다. 묶이는 문항 수가 많고 응답자 수가 많으면 신뢰도 계수는 올라가므로 상대적인 것이다. 걱정할 필요 없다.

앞 절에서 언급하였듯이 향후 분석에서는 각 리커드 척도 문항을 따로 분석하는 것이 아니라 요인 분석 결과 묶인 문항을 하나로 하여 분석을 실시하면 된다. 리커드 척도 문항에 대한 개별 정리가 필요하면 아래와 같이 정리하면 된다.

```
DATA SURVEY0;
     SET SURVEY;
          LECTURE=MEAN(OF Q5-Q8);
          INFORMATION=MEAN(Q10, Q11);
RUN;

PROC MEANS DATA=SURVEY0 MEAN STD;
     VAR Q5-Q8 LECTURE;
RUN;
```

변수	평균값	표준편차	100점 만점(*)
휴식공간	2.92	1.57	32.05
강의실 공간	2.76	1.26	29.36
강의실 시설	2.51	1.18	25.13
보조 기자재	2.80	1.27	30.00
LECTURE	2.75	0.95	0.95
(*)(평균-1)*100/6 (6.3.1.절 참고)			

강의실(lecture) 만족도는 2.75(100 점 만점에 30 점)이고 강의실 관련 하위 만족도 가운데 강의실 시설 만족도가 가장 낮으므로 학생들의 강의실 만족도를 높이기 위하여 강의실 시설에 많은 투자가 있어야 할 것이다. 강의실 하위 4 개 문항간 만족도 차이 검정을 굳이 하려면 분산 분석이 아니라 6.4 절에서 언급한 방법을 사용해야 한다.

[연습 문제]

(1)예제 설문 Q18-Q21 문항(교양 과목 만족도)에 대한 요인 분석을 실시하시오.

(2)요인 분석 결과 묶인 문항에 대해 내적 일치도 계산하시오.

(3)(2)에서 사용된 자료를 복사하여(동일 자료 두 번 사용) 내적 일치도 계산하고 결과에 대해 논하시오.

```
DATA SURVEY2;
    SET SURVEY SURVEY;
RUN;
```

(4)팀 프로젝트 설문지에서 리커드 척도 문항(동일 개념을 설문한 문항들끼리)에 대해 요인 분석을 실시하시오.

(5)(4)의 결과 묶인 문항에 대해 내적 일치도(Cronbach α)를 계산하시오.