

설문지 · 시험지 문항의 신뢰성 분석*

허 명 회**

<요 약>

본 강의에서는 심리학 및 교육학에서 흔히 활용되는 설문지 또는 시험지 문항들의 신뢰성 분석에 대한 이론을 간단히 소개한다. 아울러 SPSS/PC+의 RELIABILITY 모듈과 SAS의 PROC CORR 등을 이용하여 실제 사례를 분석하고 몇가지 관련된 사항을 토의한다.

1. 서 론

측정이론은 전통적으로 심리학과 교육학 분야에서 설문지와 시험지의 구성과 관련하여 연구되어 왔으나 근래에 들어서는 사회학, 정치학, 인류학, 저널리즘 등 여러분야에서 관심을 끌고 있다. 측정(measurement)은 추상적 개념을 경험적 지표로 구현시키는 과정이라고 할 수 있겠는데, 이 때 문제가 되는 것이 측정의 신뢰성(reliability)과 타당성(validity)이다. 그 중 전자의 측도인 신뢰도(영어로는 신뢰성과 동일한 reliability)는 반복측정이 얼마나 동일한 결과를 표출하는 지에 관한 것이며, 타당성은 경험적 지표가 그것이 측정하고자 한 추상적 개념을 얼마나 충실히 반영하고 있는지에 관한 것이다.

예를 들어 '대입학력고사'를 생각하여 보자. 이 시험의 목적이 대입지원자들이 대학에서 어느 정도의 '학업능력(academic ability)'을 갖고 있는지를 측정하려는 데 있다고 볼 때, 과연 의도대로 학업능력이라는 추상적 개념이 제대로 대입학력고사라는 경험적 지표로 나타내어지고 있는지를 평가하기 위해서는 신뢰성과 타당성의 두 관점에서의 분석이 요구된다. 신뢰도가 낮은 시험은 마치 잭 때마다 다른 수치를 가르키는 저울과 같은 것으로, 이 경우 신뢰성이란 대입지원자가 서로 동등한 여러 대입학력고사 문제지로 시험을 치루었을 경우 과연 채점결과들이 어느 정도 유사하게 나타나려는 지에 관한 것이다. 한편 '대입학력고사'에 대하여 타당성의 여러 면모를 예로 들어 설명한다면

가. 기준관련 타당성(criterion-related validity) : 대학진학 후 지원자들이 취득한 평점(GPA)과 대입학력고사 점수의 연관성, 즉 예측타당성(predictive validity),

나. 내용 타당성(content validity) : 시험지의 각 문항이 대학에서의 학업을 수행하는 데

* 1990년 12월 8일 통계교육·상담 연구회의 정기모임에서 발표되었음.

** (136-701) 서울시 성북구 안암동 고려대학교 통계학과 교수.

필요한 모든 사항을 점검하고 있는지에 관한 적절성,

다. 구인 타당성(construct validity) : 학력고사 점수가 높은 학생들을 선발한 A 대학 학생들이 대학교과 성취에서, 그보다 못한 학생들을 선발한 B 대학의 학생들에 비하여 보이는 비교우위도

라고 할 수 있다 (단순히 설명만을 목적으로 하였으므로 학력고사에 대하여 지나치게 일부 관점만이 기술되었음을 독자들께서는 양해하기 바란다).

본 강의의 범위는 측정의 신뢰성으로 잡았으므로 측정의 타당성에 관하여는 더이상 언급하지 않기로 하겠다. 신뢰성과 타당성에 관하여 관심이 있는 독자는 Carmines and Zeller (1979), Cronbach (1971)와 Stanley (1971)를 참조하기 바란다.

2. 신뢰성의 평가

어떤 지표의 신뢰성을 평가하기 위한 방법들과 신뢰성에 관한 측도들을 소개하기로 하겠다. 편의상 지표를 변수 X 로 나타낸다고 하자. 예를 들자면 어느 설문지의 문항 11부터 문항 20까지가 어느 동일한 개념을 측정하기 위한 것으로 각각이 5점 척도일 때, 이들 열 개 문항의 점수합을 지표 X 로 사용한다고 하자. 이런 경우 전통적 시험이론(classical test theory)에서는 관측점수(observed score) X 에 대하여 다음과 같은 모형을 상정한다.

$$X = t + e$$

여기서 t 는 참점수(true score)이고 e 는 확률오차이며 서로 독립이라고 가정된다. 참점수 t 는 일종의 가상적인 변량으로서 어느 개인이 동일한 변수에 대해 무한번 측정되었을 때 얻어지는 평균값이다.

만약 X 와 완벽하게 동등한 측정 X' 이 있다면 다음과 같은 표현이 가능할 것이다.

$$X' = t + e'$$

여기서 e' 은 t 와 독립이며 또한 e 와 독립으로 가정된다. 이 경우

$$\text{Cov}(X, X') = \text{Var}(t).$$

$$\text{Var}(X) = \text{Var}(X') = \text{Var}(t) + \text{Var}(e)$$

이므로 X 와 X' 사이의 상관계수는

$$\rho = \text{Var}(t)/\text{Var}(X) = \text{Var}(t)/\{\text{Var}(t) + \text{Var}(e)\}$$

인데, 이것을 측정 X 의 신뢰도 또는 신뢰계수(reliability coefficient)로 정의한다. 즉 신뢰도는 서로 동등한 척도(scale) 측정치들 사이의 상관계수라고 할 수 있다. 신뢰도에 대한 또 하나의 해석은 그것이 관측점수 X 와 참점수 t 사이의 상관계수의 제곱이라는 것이다. 왜냐하면

$$\text{Cov}(X, t) = \text{Cov}(t+e, t) = \text{Var}(t)$$

로부터

$$\text{Corr}^2(X, t) = \text{Cov}^2(X, t) / \{\text{Var}(X) \cdot \text{Var}(t)\} = \text{Var}(t) / \text{Var}(X) = \rho$$

가 되기 때문이다.

이제 문제는 신뢰계수 ρ 를 추정하는 것인데 이에 대하여는 다음과 같이 네 가지 방법이 있다(Carmines and Zeller, 1979).

2. 1 재시험법(retest method)

이 방법은 동일한 시험을 일정 시간 후에 단순히 반복하여 신뢰도를 추정하는 방법이다. 만약 응답자가 두 시험에서 동일한 결과를 보인다면 신뢰도는 당연히 1이어야 하나 실제로는 서로 다른 두 시점에서의 응답자 반응은 여러 이유에서 다를 수 있다. 이 방법은 매우 단순하다는 이점이 있으나, 응답자의 기억효과(memory effect)로 인하여 신뢰도가 과대추정될 소지가 있으며 충분한 간격의 두 시점에서 동일한 응답자로부터 측정이 이루어져야 하므로 관리 상의 애로점이 있다.

2. 2 동형 방법(alternative-form method)

이 방법은 서로 매우 비슷한 유형의 시험지를 2부 개발하여 각 응답자에게 적용하여 신뢰도를 추정하는 방법이다. 이 방법에서는 두 시험지의 각 문항들은 서로 유사하나 같지는 않으므로 재시험법에서 나타나는 기억효과가 여기서는 문제가 되지 않는다. 그러나 동형(alternative form)의 개발이 경우에 따라서는 상당히 어려울 수 있다는 사실이 현실적 문제로 남는다.

2. 3 반분법(split-halves method)

이 방법은 한 시험지의 문항들을 반 썩으로 나누어 구성된 두 파트의 상관계수로부터 신뢰도의 추정치를 얻는 방법이다. 예를 들어 홀수 문항의 응답치 합계를 X_1 , 짝수 문항의 응답치 합계를 X_2 라고 할 때, X_1 과 X_2 의 상관계수 ρ_{X_1, X_2} 를 가지고 신뢰계수 ρ 에 관한 추정을 하는 것이다. 그러나 ρ_{X_1, X_2} 는 문항수가 절반인 시험지의 신뢰도에 관한 측도이므로 전체 시험지의 신뢰도 ρ 에 대한 것으로 쓰기 위하여는 다음과 같이 보정되어야 한다.

$$\rho = 2\rho_{X_1, X_2} / (1 + \rho_{X_1, X_2}).$$

그 이유는 X 와 동등한 X' 을 가상하고 $\rho = \text{Corr}(X_1 + X_2, X_1' + X_2')$ 를 계산하여 보면 명확하다. 이 식은 스피어만-브라운 공식(Spearman-Brown prophecy formula)으로 알려져 있다.

2. 4 내적 합치도 방법(internal consistency method)

앞에서 설명한 반분법의 결점은 실제의 자료분석에 있어서는 어떻게 시험지를 두 부분으로 나누느냐에 따라 신뢰도 ρ 의 추정치가 달라진다는 사실이다. 이에 대한 대안으로서 다음과 같은 생각을 할 수 있다. X_1, \dots, X_N 을 각 문항 점수라고 하고 X'_1, \dots, X'_N 를 동형의 문항점수라고 한다면(즉, 문항 1, ..., N, 1', ..., N' 이 문항들의 전체집단으로부터의 확률표본이라면), 등분산성(equal variances)과 등공분산성(equal covariances)의 가정하에서

$$\text{Cov}(X_1 + \dots + X_N, X'_1 + \dots + X'_N) = N^2 \sigma_{12},$$

$$\text{Var}(X_1 + \dots + X_N) = \text{Var}(X'_1 + \dots + X'_N) = N\sigma^2 + N(N-1)\sigma_{12}$$

이다. 여기서 $\sigma_{12} = \text{Cov}(X_i, X_j)$, $\sigma^2 = \text{Var}(X_i)$ 이다. 한편

$$\text{Cov}(X_1 + \dots + X_N, X'_1 + \dots + X'_N) = \{\text{Var}(X_1 + \dots + X_N) - \sum \text{Var}(X_i)\} \cdot N / (N-1)$$

이 된다. 따라서 신뢰도는

$$\begin{aligned} \rho &= \text{Corr}(X_1 + \dots + X_N, X'_1 + \dots + X'_N) \\ &= [\{\text{Var}(X_1 + \dots + X_N) - \sum \text{Var}(X_i)\} \cdot N / (N-1)] / \text{Var}(X_1 + \dots + X_N) \\ &= N / (N-1) \cdot \{1 - \sum \text{Var}(X_i) / \text{Var}(\sum X_i)\} \end{aligned}$$

으로 표현된다. 이 공식을 Cronbach의 Alpha 혹은 그냥 **알파 계수**라고 한다. 만약 각 문항의 동등성에 관한 가정이 깨지게 되면 알파는 신뢰도 ρ 를 과소 측정하게 된다는 사실이 알려져 있으므로, 실제로는 알파계수는 신뢰도의 하한(lower bound)으로 생각될 수 있다. 알파계수는 각 문항이 이항형(0 또는 1)인 경우 KR-20(Kuder-Richardson-20) 계수로도 알려져 있다.

만약 모든 문항점수를 표준화되어 $\text{Var}(X_i) = 1$ 이라면

$$\rho = N \rho_{12} / \{1 + \rho_{12}(N-1)\}$$

가 된다. 이 식에서 N 이 2인 경우는 2.3 절의 스피어만-브라운 공식과 일치하며, 또 이 식으로부터 신뢰도 ρ 는 문항간의 상관계수 ρ_{12} 에 뿐만 아니라 문항수 N 에도 의존하고 있음을 알 수 있다. 신뢰도에 관한 다음 표에서 볼 수 있는 분명한 사실은 비록 문항간의 상관계수가 작다고 하더라도 문항수가 충분히 많으면 신뢰도 혹은 알파계수는 1에 가깝게 커진다는 것이다 (이 점이 설문지의 문항 수가 커지는 경향이 있게 되는 부분적 이유이다. 그렇지만 응답자의 인내 한계와 반응유류를 고려한다면 설문지의 문항수는 작을수록 좋으므로, 연구자는 절약의 원칙을 지켜야 할 것이다).

신뢰도 (reliabilty coefficient)

문항수 N	문항간 상관계수 ρ_{12}					
	0.0	0.2	0.4	0.6	0.8	1.0
2	0.0	0.33	0.57	0.75	0.89	1.0
4	0.0	0.50	0.73	0.86	0.94	1.0
6	0.0	0.60	0.80	0.90	0.96	1.0
8	0.0	0.67	0.84	0.92	0.97	1.0
10	0.0	0.71	0.87	0.94	0.98	1.0

3. SPSS의 RELIABILITY 모듈

고려대학교 통계학과 통계상담실에 의뢰된 어떤 설문조사에 있어서의 신뢰성 분석을 하기 위하여 SPSS의 RELIABILITY 모듈을 이용하기로 한다. 이 조사의 일부항목인 B1-B17은 모두 5점 척도 반응으로 구성되어 있으며 실제 설문내용은 다음과 같다(정말 그렇다 = 5, ..., 정말 그렇지 않다 = 1).

- B1. 나는 집안 일 이외에 취업하고 있는 것에 대해 만족한다.
- B2. 나는 교사라는 직업을 가진 것에 대해 만족하고 있다.
- B3. 남편은 내가 직장에 다니고 있는 것에 대해 긍정적이다.
- B4. 학교에 있는 동안에도 집에 있는 자녀들에 대해 자주 생각을 한다.
- B5. 집에 전화를 걸어서 자녀의 안부를 확인하는 일이 많다.
- B6. 자녀가 아프거나 특별한 외출이 필요한 경우 집에 전화를 거는 경우가 많다.
- B7. 학교에 있을 때는 집안일이나 자녀에 대한 생각은 되도록 하지 않고 학교일에 집중하려고 노력한다.
- B8. 여성은 남성과 비교해 볼 때 능력이 부족한 편이다.
- B9. 결혼을 하게 되면 여성은 남편의 지위에 따라 지위가 결정된다.
- B10. 남편이 가정에서 일하고 주부가 가정 밖에서 일하는 가정이 이상적이다.
- B11. 여자가 결혼 후 직업을 갖기를 원해도 남편이나 자녀가 반대하면 포기하는 것이 좋다.
- B12. 여자가 잘되는 길은 자신이 직장생활에서 성공하기 보다는 남편과 자식이 성공하도록 하는 것이다.
- B13. 남편들도 집안일과 자녀교육의 책임을 부인과 똑같이 져야 한다.
- B14. 여자는 남자와 달라서 많은 책임을 져야하는 직책을 맡지 않는 것이 좋다.
- B15. 결혼한 여자는 직장생활보다 가정생활이 우선이어야 한다.
- B16. 아내가 직업이 있더라도 가족부양은 남편의 책임이다.
- B17. 부인의 직업은 남편의 직업과 마찬가지로 중요하다.

상담의뢰인은 B1-B3를 '취업에 대한 태도', B4-B7을 '자녀와의 결속감', B8-B17을 '성역할 태도'의 척도로 하고자 하였으며, 이들 중에서 B7, B13과 B17은 다른 설문과는 다른 방향의 경향을 갖는 것으로 생각되어 5, 4, 3, 2, 1의 반응값을 1, 2, 3, 4, 5로 바꾸어 분석에 들어갔다. 이 예에서 신뢰성 분석을 위해 사용된 SPSS/PC+ 프로그램과 출력결과는 다음과 같다.

SPSS 프로그램

```
reliability /variables = b1 to b17
            /statistics descriptives correlations scale
            /summary = means variances correlations total
            /scale(job) = b1 to b3
            /scale(child) = b4 to b7
            /scale(sexrole) = b8 to b17
            /model alpha.
```

SPSS 출력결과(부분)

가) 취업에 대한 태도

RELIABILITY ANALYSIS - SCALE (JOB)

		MEAN	STD DEV	CASES
1.	B1	4.1204	.8171	108.0
2.	B2	4.1389	.7164	108.0
3.	B3	4.1111	.7893	108.0

CORRELATION MATRIX

	B1	B2	B3
B1	1.0000		
B2	.6577	1.0000	
B3	.2979	.3360	1.0000

OF CASES = 108.0

ITEM-TOTAL STATISTICS

	SCALE MEAN IF ITEM DELETED	SCALE VARIANCE IF ITEM DELETED	CORRECTED ITEM- TOTAL CORRELATION	SQUARED MULTIPLE CORRELATION	ALPHA IF ITEM DELETED
B1	8.2500	1.5164	.5736	.4392	.5013
B2	8.2315	1.6749	.6202	.4541	.4588
B3	8.2593	1.9508	.3466	.1233	.7894

RELIABILITY COEFFICIENTS 3 ITEMS

ALPHA = .6894 STANDARDIZED ITEM ALPHA = .6940

나) 자녀와의 결속감

RELIABILITY ANALYSIS - SCALE (CHILD)

		MEAN	STD DEV	CASES
1.	B4	3.4167	1.2390	108.0
2.	B5	2.8519	1.3100	108.0
3.	B6	3.7963	1.2953	108.0
4.	B7	2.2500	.9484	108.0

CORRELATION MATRIX

	B4	B5	B6	B7
B4	1.0000			
B5	.6603	1.0000		
B6	.5309	.6650	1.0000	
B7	.2128	.1580	.0723	1.0000

OF CASES = 108.0

ITEM-TOTAL STATISTICS

	SCALE MEAN IF ITEM DELETED	SCALE VARIANCE IF ITEM DELETED	CORRECTED ITEM- TOTAL CORRELATION	SQUARED MULTIPLE CORRELATION	ALPHA IF ITEM DELETED
B4	8.8981	7.1204	.6575	.4644	.5955
B5	9.4630	6.4939	.7178	.5750	.5501
B6	8.5185	7.1866	.5959	.4599	.6336
B6	10.0648	11.0331	.1698	.0505	.8299

RELIABILITY COEFFICIENTS 4 ITEMS

ALPHA = .7357 STANDARDIZED ITEM ALPHA = .7131

다) 성역할 태도

RELIABILITY ANALYSIS - SCALE (SEXROLE)

		MEAN	STD DEV	CASES
1.	B8	2.1481	.8182	108.0
2.	B9	3.2870	1.1361	108.0
3.	B10	1.7870	.6420	108.0
4.	B11	2.8981	1.2147	108.0
5.	B12	3.2593	1.1138	108.0

6.	B13	1.5370	.7289	108.0
7.	B14	2.4537	.9895	108.0
8.	B15	2.9815	1.1105	108.0
9.	B16	3.2315	1.1730	108.0
10.	B17	1.6111	.7213	108.0

CORRELATION MATRIX

	B8	B9	B10	B11	B12
B8	1.0000				
B9	.1650	1.0000			
B10	.1496	.0718	1.0000		
B11	.2598	.2787	.0558	1.0000	
B12	.0908	.1548	-.0789	.2338	1.0000
B13	-.0406	-.1427	.0869	-.1593	-.0465
B14	.2048	.1491	.0211	.2721	.2230
B15	.1573	.1894	-.0318	.2341	.1248
B16	-.0361	.2512	-.0208	.3119	.3113
B17	.1302	.0805	.1828	.1890	.1150

CORRELATION MATRIX

	B13	B14	B15	B16	B17
B13	1.0000				
B14	-.1207	1.0000			
B15	-.0107	.3224	1.0000		
B16	-.1249	.2710	.3405	1.0000	
B17	.2232	.1317	.1076	.1516	1.0000

OF CASES = 108.0

ITEM-TOTAL STATISTICS

	SCALE MEAN IF ITEM DELETED	SCALE VARIANCE IF ITEM DELETED	CORRECTED ITEM- TOTAL CORRELATION	SQUARED MULTIPLE CORRELATION	ALPHA IF ITEM DELETED
B8	23.0463	19.1287	.2514	.1555	.5922
B9	21.9074	17.3558	.3117	.1400	.5781
B10	23.4074	20.7484	.0747	.0734	.6199
B11	22.2963	15.8927	.4367	.2459	.5408
B12	21.9352	17.5565	.3000	.1485	.5811
B13	23.6574	21.7974	-.1075	.1271	.6503
B14	22.7407	17.4275	.3862	.1992	.5602
B15	22.2130	16.9542	.3728	.2048	.5613
B16	21.9630	16.3911	.4033	.2868	.5517
B17	23.5833	19.2547	.2879	.1519	.5870

RELIABILITY COEFFICIENTS

10 ITEMS

ALPHA = .6106

STANDARDIZED ITEM ALPHA = .5784

‘취업에 관한 태도’ (B1-B3)에 대하여는 알파계수값이 0.6894 이며 만약 각 문항을 표준화하여 척도로 하는 경우에는 0.6940 임을 출력결과로부터 볼 수 있다. 그런데 B3을 빼고 B1과 B2만으로 척도로 하는 경우 알파 값이 0.7894가 되므로 신뢰도를 높이기 위해서는 오히려 B3가 척도구성에서 제외되는 것이 바람직하다고 볼 수 있다. 사실 이들 설문들을 내용으로 보더라도 B1과 B2는 취업에 관한 ‘나’의 태도에 관한 것인데 B3은 나의 취업에 관한 ‘남편’의 태도에 관한 것이므로 서로 이질적일 수 있음을 알 수 있다.

‘자녀와의 결속감’ (B4-B7)에 대해서는 알파계수 값이 0.7357인 것을 볼 수 있는데 B7을 빼는 경우에는 알파 값이 0.8299로 상당히 증가하는 것을 알 수 있다. 이 설문은 자녀와의 결속감에 대한 것이라기 보다는 자녀양육과 직장생활 사이의 ‘갈등’이라는 좀 다른 개념에 관한 것으로 보아야 할 것으로 생각된다.

‘성 역할 태도’ (B8-B17)에 대하여는 알파계수 값이 0.6106으로 문항수에 비하면 약간 작은 감이 있으며 특히 상관계수행렬을 보면 음의 값을 갖는 것도 다수 산재하고 있음을 알 수 있다. 상관계수 값들의 부호로 보아 B10과 B13이 문제가 된다는 것을 확실히 알 수 있다. 특히 문항 B10을 잘 보면 1, 2, 3, 4, 5의 반응점수가 5, 4, 3, 2, 1로 그 순서가 바뀌어 분석되었어야 한다는 것을 알 수 있다.

앞에서의 알파 계수를 통한 신뢰도 계산외에 SPSS의 RELIABILITY 모듈에서는, /MODEL SPLIT이라는 옵션을 쓰면 2.3절에서 소개한 반분법에 의한 신뢰성 분석이 가능하다. 그리고 /MODEL STRICTPARALLEL과 /MODEL PARALLEL과 같은 전통적 시험이론에서의 동등성 모형을 통하여 신뢰도의 최대우도추정도 가능하다. 그 밖의 자세한 사항에 관하여는 SPSS/PC+ Advanced Statistics 매뉴얼을 참조하기 바란다(Norusis, 1988).

4. SAS에서의 신뢰성 분석

SAS/PC의 새 버전(version 6.3 이후)에서는 PROC CORR을 통하여 Cronbach의 알파계수를 계산할 수 있게 되었다. 그 이전의 SAS에서는 여러 procedure들을 결합하여서 겨우 신뢰도의 계산이 가능했던 것에 비하면 일이 매우 수월해졌다고 하겠다. 만약 구버전의 SAS로 신뢰도를 계산하기 원하는 독자가 있다면 신충렬(1990)을 참고하기 바란다(신충렬의 SAS루틴은 결측치가 없는 경우에 한하여 SPSS/PC+의 RELIABILITY 모듈과 같은 결과를 제공한다. 왜냐하면 SPSS 분석에서는 결측치를 포함하는 data line 전체가 계수 계산에서 제외되지만 그 SAS루틴에서는 결측치가 해당하는 통계치 계산에만 영향을 주기 때문이다). 3절의 사례분석을 위한 SAS 프로그램과 출력결과는 다음과 같다.

신뢰도 계산을 위한 SAS의 PROC CORR

```

proc corr alpha ;
  var b1-b3 ;
run ;
proc corr alpha ;
  var b4-b7 ;
run ;
proc corr alpha ;
  var b8-b17 ;
run ;

```

SAS의 출력결과 (부분)

CORRELATION ANALYSIS

3 'VAR' Variables : B1 B2 B3

Simple Statistics

Variable	N	Mean	Std Dev	Sum
B1	108	4.12037	0.81708	445.00000
B2	108	4.13889	0.71641	447.00000
B3	108	4.11111	0.78934	444.00000

Cronbach Coefficient Alpha

for RAW variables : 0.689418
 for STANDARDIZED variables : 0.694016

Raw Variables

Std. Variables

Deleted Variable	Correlation with Total	Alpha	Correlation with Total	Alpha
B1	0.573573	0.501284	0.584569	0.503048
B2	0.620204	0.458796	0.616806	0.459007
B3	0.346615	0.789390	0.348146	0.793512

Pearson Correlation Coefficients / Prob > |R| under H_0 : $Rho = 0$ / $N = 108$

	B1	B2	B3
B1	1.00000	0.65770	0.29786
	0.0	0.0001	0.0017
B2	0.65770	1.00000	0.33605
	0.0001	0.0	0.0004
B3	0.29786	0.33605	1.00000
	0.0017	0.0004	0.0

5. 토의 및 기타

신뢰성 분석은 심리학과 교육학에서의 측정이론(measurement theory)에서 기인하는 일련의 통계적 방법론이라고 할 수 있는데, 이에 관한 관련문헌은 실로 방대하다. 국내에 소개된 이 분야의 전문서적으로는 이종성 편역(1985)의 『측정이론의 기초』가 있으므로 관심이 있는 독자는 참조하기 바란다. 다음으로는 실제 통계분석에 있어서 관심이 될 수 있는 몇몇 사항을 토의함으로써 본 강의를 마치기로 한다.

5. 1 신뢰도의 크기에 관한 지침

일반적으로 척도(scale)의 신뢰도가 얼마 이상이어야 하는가 하는 질문에 대하여 그 기준을 타당한 논리의 바탕하에 제시하는 것은 불가능한 것으로 생각된다. 알파값이 0.8 이상이어야 한다는 등의 출처를 알 수 없는 지침으로 조사연구자들이 위축되는 사례가 간혹 있으나, 알파 값에 관한 기준설정은 회귀분석에서의 결정계수의 크기에 관한 문제와 같은 맥락에서 이해해도 될 것이다. 또 유의해야 할 점은 회귀분석에서 설명변수의 수가 늘어남에 따라 결정계수가 커지는 경향이 있듯이, 신뢰도의 경우에는 문항수가 많아짐에 따라 신뢰도가 자연스레 증가한다는 사실이다(2.4절 참조). 따라서 신뢰도를 높이기 위해서는 문항수를 단순히 늘리는 것이 가장 쉽게 생각되는 방법이지만 설문조사의 제반요소를 동시에 고려한다면 그다지 좋은 방법은 아니다. 원칙적이며 이상적으로는, 한정된 수의 문항을 다듬고 다듬어 가능한 최고의 신뢰도를 얻는 데에 정성을 다하여야 할 것이다.

5. 2 신뢰성 분석과 인자분석

여러 문항 점수들의 합계를 어떤 개념의 지표로 쓸 때에는 해당 문항들이 그 한 개념을 동등하게 측정한다는 전제가 미리 깔려 있는 것이다. 실제의 자료분석에서는 이 가정이 만족되지 않을 수도 있는데 그런 경우에는 다음의 두 방향으로 생각할 수 있다. 첫째는 문항들이 각기 다른 비중으로 개념을 측정하고 있지 않나 하는 점이다. 둘째는 문항들이 어떤 한 개념을 측정하고 있는 것이 아니라 둘 이상의 여러 개념을 동시에 측정하고 있을 가능성이 있다. 인자분석(factor analysis)은 바로 이런 가능성들을 탐색하거나 혹은 제시된 척도의 타당성을 입증하기 위하여는 활용될 수 있다. Carmines and Zeller (1979)는 인자분석을 이용하여 자존심(self-esteem)에 관한 열 문항을 두 부분으로 분리한 사례를 제시한 바 있다.

5. 3 대입학력고사의 신뢰도

매년 국가적 행사로 치루어지는 대입학력고사의 신뢰도는 얼마나 될 것인가? 이 시험이 워낙 중요한 시험이기 때문에 높은 신뢰도가 요구되는 것은 너무나 당연하지만, 그 중요성에 비하여 신뢰도의 검증에 대한 연구는 너무 미미한 것이 아닌가 생각된다. 김상겸(1988)

의 석사학위논문이 학력고사의 신뢰도에 대한 약간의 짐작치를 제공하는데 그는 서울의 D 고교 3년생 101 명이 5 차례 치른 모의고사 점수로부터 다음의 상관계수 행렬을 얻었다. 이들 상관계수들의 평균은 0.8865 이다.

	1회	2회	3회	4회	5회
1회	1				
2회	0.906	1			
3회	0.858	0.868	1		
4회	0.874	0.861	0.892	1	
5회	0.920	0.885	0.877	0.924	1

◇ 참 고 문 헌 ◇

- [1] 김상겸(1988) 『선지원제와 후지원제 하에서의 입학생들의 학력차에 관한 연구』. 고려대학교 대학원 통계학과 석사학위 논문.
- [2] 신충렬(1990) “SAS 언어일반”, 『SAS 를 이용한 통계 Workshop』. 고려대학교 통계연구소 편.
- [3] 이종성(1985) 『행동과학연구를 위한 측정이론의 기초』. 중앙적성출판사. (Mary J. Allen and Wendy M. Yen 의 *Introduction to Measurement Theory* 의 번역)
- [4] Carmines, E. G. and Zeller, R. A. (1979) *Reliability and Validity Assessment*. Series : Quantitative Applications in the Social Sciences No. 17. Sage Publications, Beverly Hills.
- [5] Cronbach, L. J. (1971) “Test Validation,” in *Educational Measurement*, 2nd Edition (Edited by R. L. Thorndike). American Council on Education, Washington, D. C. (Chapter 14)
- [6] Norusis, M. J. (1988) *SPSS/PC+ Advanced Statistics*, Version 2 Edition. SPSS Inc., Chicago.
- [7] Stanley, J. C. (1971) “Reliability,” in *Educational Measurement*, 2nd Edition (Edited by R. L. Thorndike). American Council on Education, Washington, D. C. (Chapter 13)

Reliability Analysis of Questionnaire/Test Items

Myung-Hoe Huh*

<Abstract>

In this tutorial lecture, basic theory underlying reliability analysis of questionnaire /test items is explained. Also, we analyze a real case with SPSS/PC+ reliability module and SAS correlation procedure, and discuss several relevant issues.

* Professor, Department of Statistics, Korea University.
Anam-dong 5-1, Sungbuk-ku, Seoul 136-701, Korea.