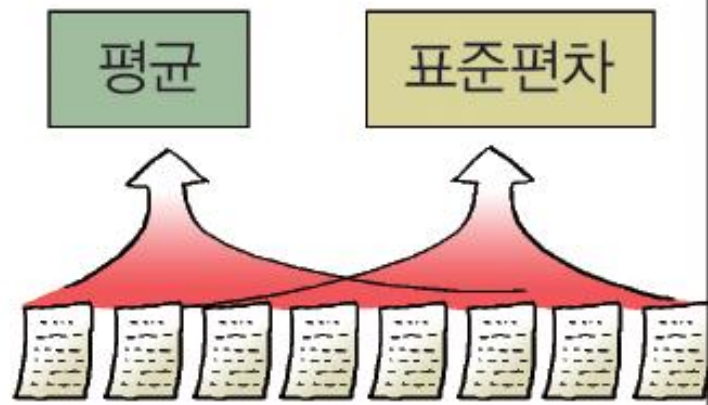




그 자료들이 어떤 특성을 가지고 있는지
쉽게 알 수 있도록 해주는 방법을
말하는 것입니다.



다음 예제들을 따라
해보면 여러분도 쉽게
이해할 수 있습니다.



01 신뢰도 분석

1) 신뢰도 분석 예제



예제

직업재활을 전공한 학부생이 직업재활 현장실습을 마친 후의 전문성 수준을 파악하기 위한 설문 문항을 설정하였다. 전문성 측정문항은 5개의 하부영역으로 문항 수는 각 5개이며, 설문문항들이 전문성을 측정하기 위해 어느 정도의 신뢰도가 있는지 분석하고자 한다.

변수	하부영역	문항수	변수번호	
전문성	전문조직의 준거성	5	전문성	_1, _4, _11, _15*, _17*
	공적서비스에 대한 신념	5		_2*, _5, _8*, _12*, _22
	자율성	5		_3, _10*, _19*, _21, _25*
	자체규제에 대한 신념	5		_6, _13*, _16*, _20*, _23
	직업에 대한 소명의식	5		_7, _9, _14, _18, _24
	소계	25		

* 표시항목은 부적 설문문항이므로 역점 처리함

(1) 신뢰도 분석 대화상자 열기

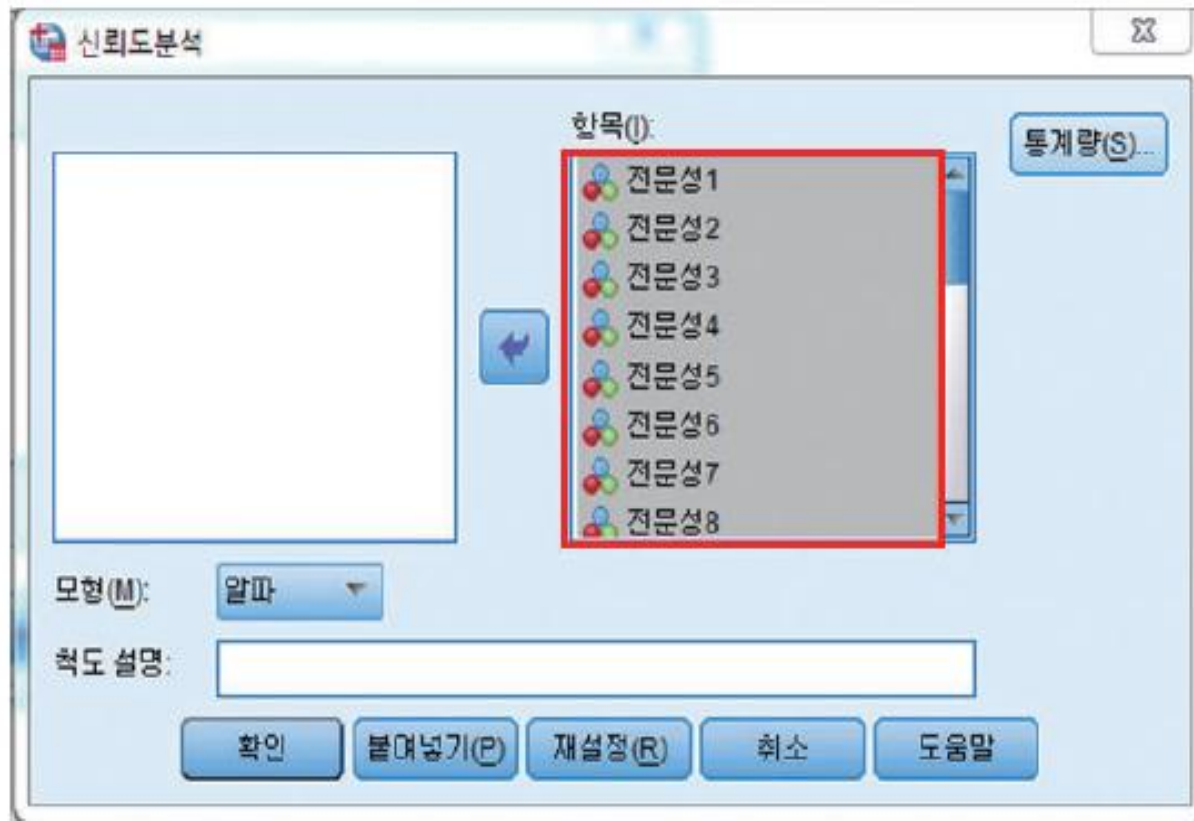
메뉴에서 ① [분석(A)] → ② [척도(A)] → ③ [신뢰도분석(R)]을 클릭한다.

IBM SPSS Statistics Data Editor window showing the menu path: **1** [분석(A)] → **2** [척도(A)] → **3** [신뢰도분석(R)]을 클릭한다.

	전문성1	전문성2	전문성4	전문성5	전문성6	전문성7	전문성8
1	4.00	2.00	4.00	4.00	4.00	4.00	4.00
2	3.00	3.00	3.00	3.00	3.00	4.00	4.00
3	4.00	2.00	3.00	4.00	4.00	4.00	4.00
4	4.00	4.00	3.00	5.00	2.00	5.00	5.00
5	2.00	3.00	2.00	3.00	3.00	4.00	4.00
6	5.00	5.00	5.00	5.00	5.00	5.00	5.00
7	5.00	5.00	5.00	5.00	5.00	5.00	5.00
8	3.00	3.00	4.00	5.00	2.00	5.00	5.00
9	4.00	4.00	2.00	4.00	3.00	4.00	4.00
10	3.00	2.00	4.00	5.00	4.00	4.00	4.00
11	5.00	4.00	5.00	4.00	5.00	5.00	5.00
12	5.00	4.00	4.00	4.00	3.00	4.00	4.00
13	4.00	3.00	4.00	4.00	4.00	5.00	5.00
14	5.00	4.00	4.00	4.00	4.00	4.00	4.00
15	4.00	1.00	3.00	3.00	4.00	2.00	2.00
16	4.00	2.00	4.00	5.00	4.00	4.00	4.00

(2) 변수 옮기기

[신뢰도분석(R)] 대화상자 왼쪽에 있는 변수의 목록들 중 측정문항을 문항상자에 옮긴다.



(3) 분석 실행하기

신뢰도 분석 대화상자에서 [통계량(S)]을 클릭한다. 대화상자가 나타나면 다음에 대한 기술통계량에서 항목(I), 척도(S), 항목 제거 시 척도(A)를 선택한다.

신뢰도 분석: 통계량

다음에 대한 기술통계량

- 항목(I)
- 척도(S)
- 항목제거시 척도(A)

항목내

- 상관관계(R)
- 공분산(E)

요약값

- 평균(M)
- 분산(V)
- 공분산(E)
- 상관관계(R)

분산분석표

- 지정없음(N)
- F-검정(F)
- Friedman 카이제곱(Q)
- Cochran 카이제곱(H)

Hotelling의 T 제곱(G) Tukey의 가법성 검정(K)

급내 상관계수(U)

모형(M): 이차원 혼합 유형(Y): 일차

신뢰구간(U): 95 %의 케이스 추출 검정값(A): 0

계속 취소 도움말

(4) 결과 해석하기

각 항목에 대한 평균, 표준편차, 표본에 대한 기술통계량을 확인할 수 있다.

항목통계량

	평균	표준편차	N
전문성1	3.6259	.86424	294
전문성2	2.8878	1.07611	294
전문성3	3.7619	.77797	294
전문성4	3.7517	.79862	294
전문성5	3.9762	.78561	294
전문성6	3.6599	.80524	294
전문성7	3.8844	.84673	294
전문성8	2.8333	1.03329	294
전문성9	3.8912	.74016	294
전문성10	2.9116	1.00119	294
전문성11	3.7041	.78652	294
전문성12	2.7687	1.01905	294

전문성 문항 전체의 **크론바 알파**(Cronbach's a)는 0.878로 신뢰할 만한 수준으로 나타났다.

신뢰도 통계량	
Cronbach의 알파	항목 수
.878	25

사용되는 각 설문문항의 값들은 양의 상관관계를 갖는데, 만약 음의 상관관계가 있는 설문문항들이 포함되면 크론바 알파값은 낮아진다.


5점 척도 또는 7점 척도의 설문문항들에서 음의 상관관계가 나왔다면 설문문항의 내용을 보고, 질문의 의미가 반대인지 확인하고 역코딩하여 신뢰도분석을 다시 합니다.

역코딩 : 1 -> 5, 2 -> 4 ... 5 -> 1



Tip

신뢰도 분석에서 자주 쓰이는 크론바 알파의 기준값이 일반적으로 **0.6 이상이면 '신뢰도가 있다'**라고 하는데, 신뢰도의 레퍼런스로 가장 많이 쓰이는 **Nunnally에 의하면 0.7 이상이면 '신뢰할 만하다'**라고 한다. 이전의 사회과학연구들을 보면 0.6 이상이면 허용하였으나, 요즘 저널들의 추세를 살펴보면 0.7 이상 나와야만 허용하는 쪽으로 방향을 바뀌어졌다 (**$\alpha > 0.8$: Excellent, $\alpha > 0.7$: Good, $\alpha > 0.5$: Miniman**).

 여기서 돋보기! **논문**에 실제로 적용하기

직업재활 실습을 마친 학부생의 전문성을 측정하기 위한 설문문항에 대한 신뢰도를 분석한 결과, 설문문항 25개 전체의 크론바 알파계수는 0.878로 높게 나타났다.

표 7-1. 설문지의 하부영역 별 신뢰도

변수	하부영역	문항수	신뢰도
전문성	전문조직의 준거성	5	.646
	공적서비스에 대한 신념	5	.631
	자율성	5	.614
	자체규제에 대한 신념	5	.676
	직업에 대한 소명의식	5	.724
	소계	25	.878

02 데이터의 검정

1) 정규성 검정

모집단의 형태가 정규분포인지 아닌지를 검증하는 방법이다.

(1) 정규성 검정의 예제



예제

직업재활 실습을 마친 남녀학생을 대상으로 실습 전 학교에서 실시하는 실습교육의 만족도에 대해 조사를 하였다.

이에 앞서 자료의 정규성을 검증하였다.

1 정규성 검정 대화상자 열기

메뉴에서 ① [분석(A)] → ② [기술통계량(E)] → ③ [데이터 탐색(E)]을 클릭한다.



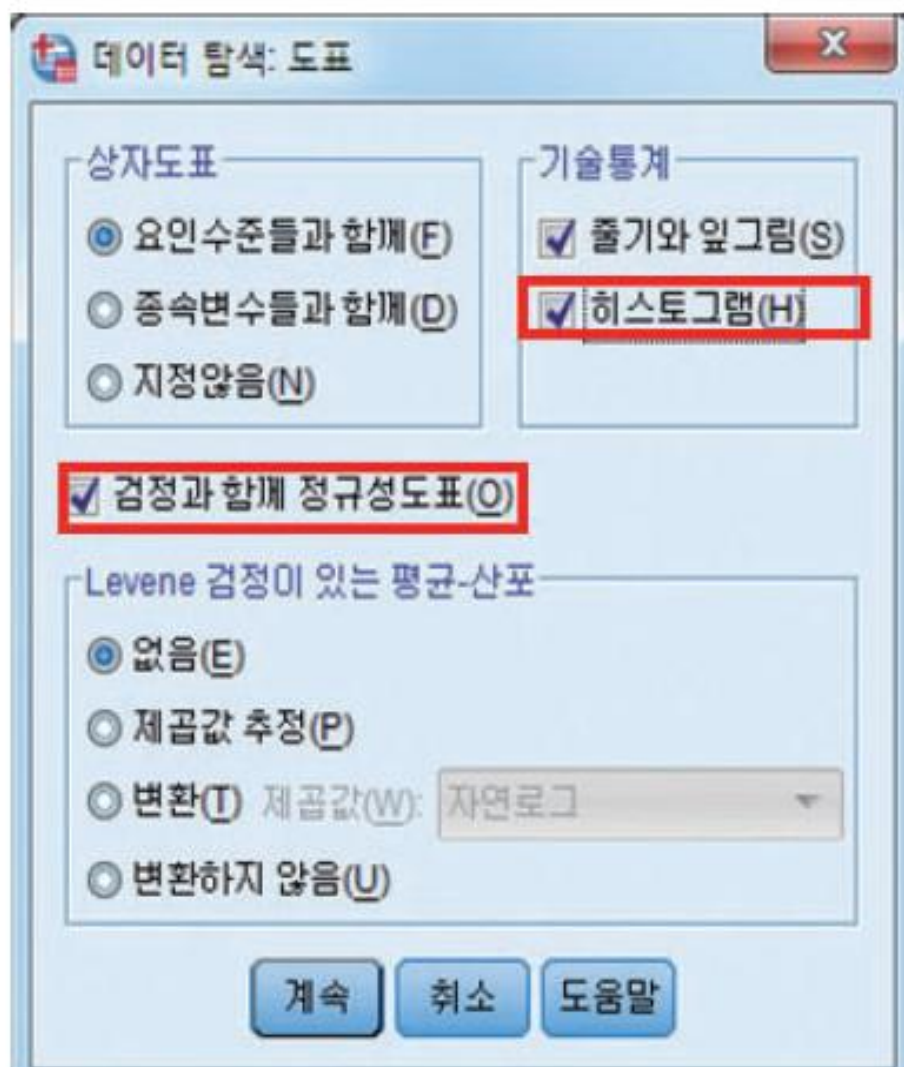
② 변수 옮기기

실습을 끝낸 남녀학생들의 정규분포 가정을 만족하는지를 검증하기 위해 성별을 **요인(F)**으로 이동시키고, school은 **종속변수(D)**로 이동시킨다.



③ 분석 실행하기

도표를 클릭해 히스토그램(H)에 체크하고, 정규성을 검증위해 검정과 함께 정규성도표(O)를 체크한 다음 [계속]을 클릭한다.



[확인]을 눌러 결과를 본다.



4 결과 해석하기

케이스 처리 요약에는 유효케이스와 결측케이스가 표시된다. 남녀별 결측치가 없음이 확인되었다.

케이스 처리 요약

성별		케이스					
		유효		결측		전체	
		N		N	퍼센트	N	퍼센트
school	남자	75	100.0%	0	.0%	75	100.0%
	여자	219	100.0%	0	.0%	219	100.0%

기술통계에는 각 그룹별 평균, 중앙값, 분산, 표준편차, 최소값, 최대값, 범위, 사분위수 범위, 왜도와 첨도를 알려준다.

기술통계

성별		통계량	표준오차
school 남자	평균	73.9867	1.78669
	평균의 95% 신뢰구간	하한 70.4266 상한 77.5467	
	5% 절삭평균	74.4222	
	중위수	74.0000	
	분산	239.419	
	표준편차	15.47316	
	최소값	27.00	
	최대값	105.00	
	범위	78.00	
	사분위수 범위	18.00	
	왜도	-.431	.277
	첨도	.992	.548

왜도가 <0 이면 오른쪽으로 치우친 분포, 첨도<0이면 편평한 분포를 의미하지만 남녀 모두 -2~2 사이에 있어 크게 문제되지 않는다.

|왜도| < 0.2 , |첨도| < 0.4 를 충족하면 크게 정규성을 이탈하지 않는다.

여자	평균		71.6804	1.02793
	평균의 95% 신뢰구간	하한 상한	69.6544 73.7063	
	5% 절삭평균		72.3031	
	중위수		72.0000	
	분산		231.402	
	표준편차		15.21190	
	최소값		21.00	
	최대값		101.00	
	범위		80.00	
	사분위수 범위		19.00	
	왜도		-.568	.164
	첨도		.593	.327

정규성 검정

성별	Klmogorov-Smirnov ^a			Shapiro-Wilk		
	통계량	자유도	유의확률	통계량	자유도	유의확률
school 남자	.092	75	.186	.968	75	.057
여자	.074	219	.005	.975	219	.001

a. Lilliefors 유의확률 수정

(두 종류의 정규성 결과 출력)

콜모고로프-스미노프검정과 샤피로-윌크검정

정규성 검정에 대한 가설은 유의확률을 보고 결론을 내림

H_0 = 집단은 정규분포를 이룬다.

H_A = 집단은 정규분포를 이루지 않는다.

정규성 검정

성별	Klmogorov-Smirnov ^a			Shapiro-Wilk		
	통계량	자유도	유의확률	통계량	자유도	유의확률
school 남자	.092	75	.186	.968	75	.057
여자	.074	219	.005	.975	219	.001

a. Lilliefors 유의확률 수정

- **콜모고로프-스미노프검정** **0.05보다 작으면 대립가설기각**

남자집단의 유의확률은 0.186으로 '정규분포를 이룬다'는 영가설을 기각하지 못하므로 영가설을 받아들여 '집단은 정규분포를 이룬다'

여자집단의 경우 유의확률은 0.005로 '집단은 정규분포를 이룬다'는 대립가설을 받아들여 '집단은 정규분포를 이루지 않는다'

정규성 검정

성별	Klmogorov-Smirnov ^a			Shapiro-Wilk		
	통계량	자유도	유의확률	통계량	자유도	유의확률
school 남자	.092	75	.186	.968	75	.057
여자	.074	219	.005	.975	219	.001

a. Lilliefors 유의확률 수정

• 샤피로-윌크검정

남자집단의 유의확률은 0.057로 '정규분포를 이룬다'는 영가설을 기각하지 못하므로 영가설을 받아들여 '집단은 정규분포를 이룬다'.

여자집단의 경우 유의확률은 0.001로 '집단은 정규분포를 이룬다'는 대립가설을 받아들여 '집단은 정규분포를 이루지 않는다'.

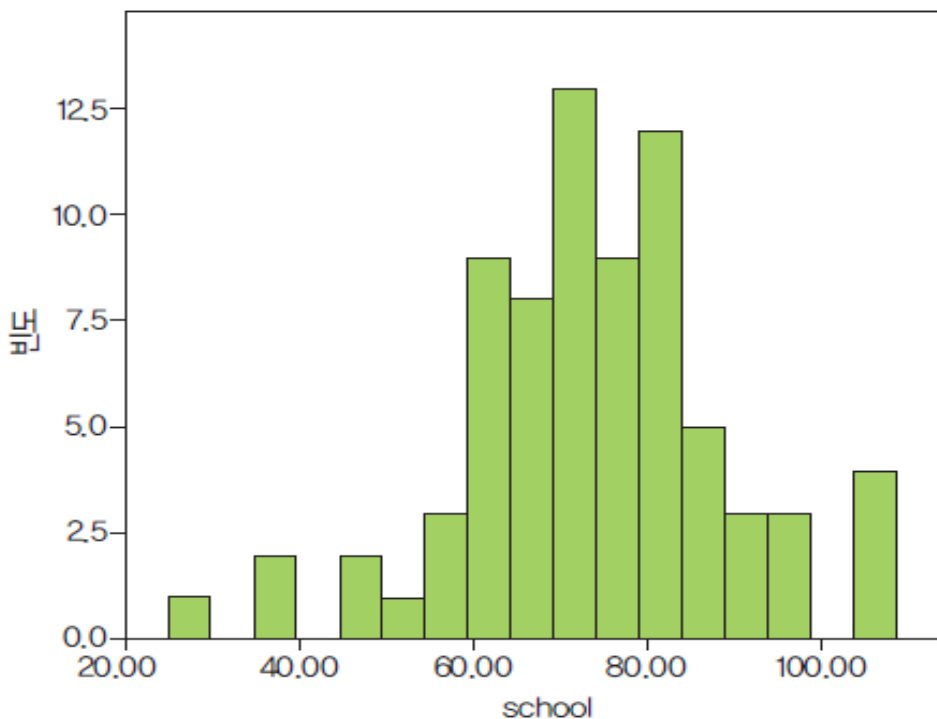
이 방법은 보수적인 방법으로 데이터가 적을 때 사용한다.

결론적으로 두 남녀그룹 중 남자그룹만이 정규성을 띤다고 가정

히스토그램

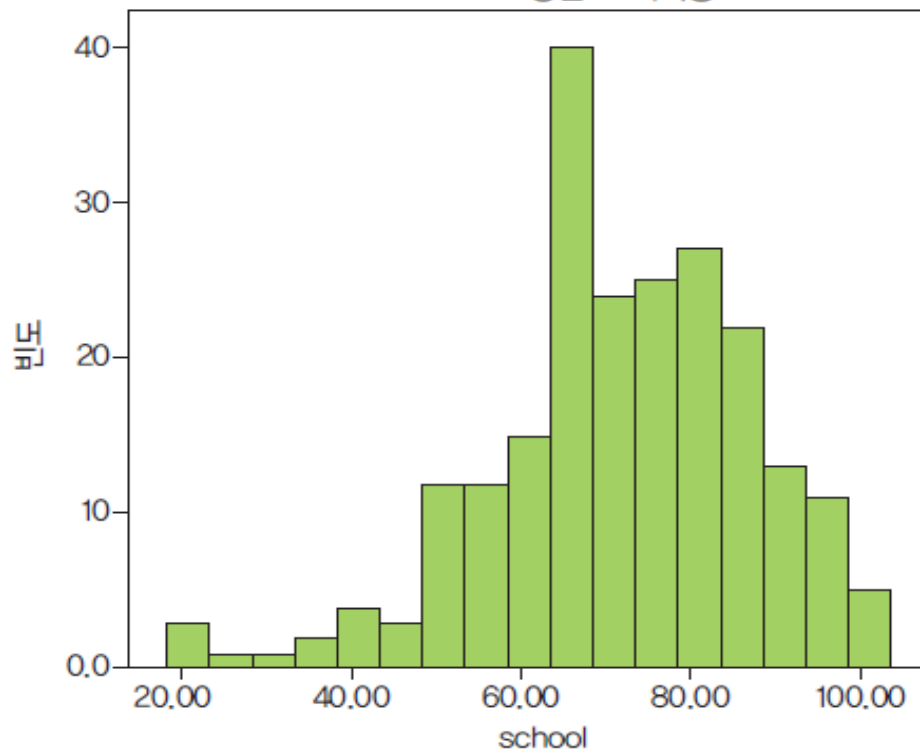
도수 분포의 상태를 기둥 모양의 그래프로 나타 낸 것으로 연속형 자료의 분포를 살펴볼 때 일반적으로 사용되는 방법

히스토그램
성별 = 남자용



평균 = 73.99
표준편차 = 15.473
N = 75

히스토그램
성별 = 여자용



평균 = 71.68
표준편차 = 15.212
N = 219



여기서 돋보기! 논문에 실제로 적용하기

본 연구에서 사용될 변수의 정규성을 검증하기 위하여 샤피로-윌크검정값, 왜도 및 첨도를 조사하였으며, 분석결과 남자그룹의 경우 유의확률이 0.057로 유의수준 $p < 0.05$ 에서 정규성 가정을 충족하였고, 왜도와 첨도의 절대값이 기준치($|\text{왜도}| < 0.2$, $|\text{첨도}| < 0.4$)를 충족하고 있어 자료가 정규성을 크게 이탈하지 않는다고 볼 수 있다(Hong et al. 2003). 그러나 여자그룹의 경우 유의확률이 0.001로 유의수준 0.05에서 정규성 가정이 만족하지 않아 비모수적 분석을 이용하였다.

표 7-2 정규성 검정 결과

변수	평균	표준편차	왜도 (표준오차)	첨도 (표준오차)	Shapiro-Wilks	p값
School(남)	73.98	15.47	-.431(.277)	.992(.548)	.092	.057
School(여)	71.68	15.21	-.568(.164)	.593(.327)	.074	.001

$p < 0.01$

1) 왜 하필 정규성 검정을 하고 정규분포를 이용하는가요?

ANS 통계학에서 기준이 되는 분포가 정규분포가 되어 일반적으로 정규 분포 만족여부에 따른 이론적 전개가 이루어집니다. 그러므로 정규성 검정을 만족하지 않으면, 다른 이론적 접근 방법을 이용하여야 합니다.

2) 임의의 모집단일 경우는 왜 표본의 크기를 따져야 하는가요?

ANS 모집단이 정규분포일 경우에는 랜덤표본의 크기에 상관없이 표본평균들이 정규분포를 이룰 수 있으나, 표본의 크기가 30이하이면 표본평균들이 반드시 정규분포를 이룬다고는 말할 수 없기 때문에 표본분포의 정규성 검정을 해줘야 합니다.

1) Sample size가 커도 정규성 검정을 해야 하나요??

ANS sample size > 30이면 정규성 검정을 하지 않아도 된다는 이론적 근거가 있기 때문에 하지 않아도 됩니다. 그러나 자료의 값 간에 차이가 너무 클 경우에는 transformation(로그변환, 제곱근변환) 으로 정규성 만족을 고려합니다.

2) Sample size가 7보다 작는데, 정규성 검정을 만족하면 정규성 검정을 만족한다는 가정으로 이루어지는 통계분석방법을 이용해야 하나요?

ANS 이 경우는 모수적, 비모수적 접근방법을 모두 이용하는 것이 좋습니다. 대부분 sample size가 7미만이면 sample size가 너무 적어 분석이 어렵습니다.



기억해요

▷ 정규성 검정에서 봐야 할 항목

- ① 중앙값, 평균값이 비슷한지, 5%절단평균
- ② 왜도와 첨도가 $-2 \sim +2$ 사이인지,
- ③ 값을 찍은 점이 normal probability plot과 비슷한 분포를 보이는지,
- ④ Kolmogorov-Smirnov-Test, Shapiro-Wilk Test에서 p-value가 0.05 이상인지 확인한다.

D 대학병원에서 천식 환자 50명에게 Histamine으로 Provocation test를 시행하고 test 받은 환자들 중 15명을 단순임의추출로 뽑아 그들의 iFEV1값을 조사하였더니 아래와 같을 때, Provocation test 받은 환자들의 iFEV1값이 4L와 같다고 할 수 있는가?

index	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
iFEV1	4.1	3.9	4.3	3.8	3.9	4.3	4.0	4.2	4.5	4.5	4.3	3.9	3.9	3.9	4.0

1) 연구 설계

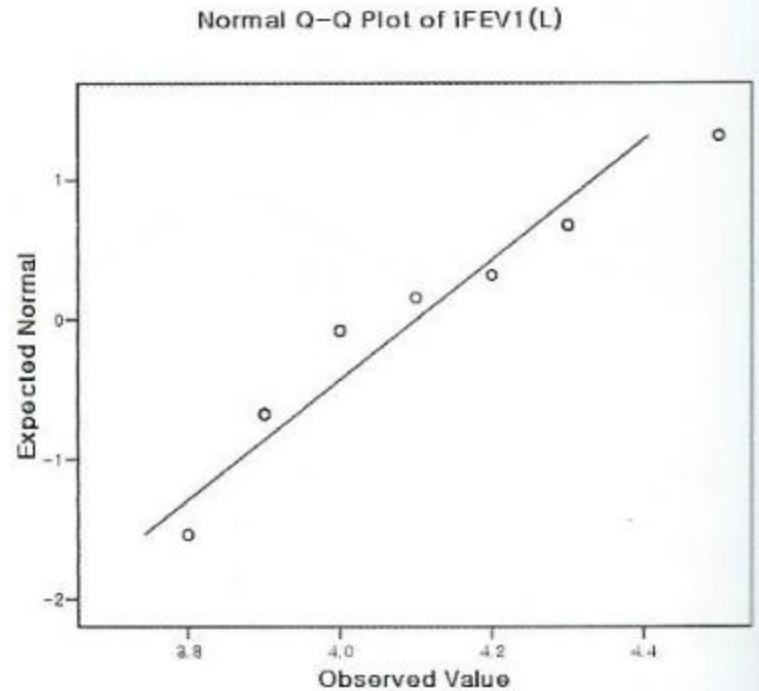
- 연구 목적 : histamine으로 provocation test 받은 환자들의 FEV값이 정상인지 알아본다.
- 연구 대상 및 표본 : 천식 환자로 histamine으로 provocation test를 받은 환자 15명의 iFEV1값
- 연구 방법 : observational study
- Data의 속성 : iFEV1- 양적, 연속, 비율척도

2) 표본의 크기 : $n=15 \leq 30$

3) 표본의 크기가 30이하이므로 정규성 검정이 필요하다.

<SPSS output>

통계량	통계치	표준오차
평균	4.100	0.060
5% 절단 평균	4.094	
중앙값	4.000	
분산	0.054	
표준오차	0.233	
최소값	3.800	
최대값	4.500	
범위	0.700	
사분위수범위	0.400	
왜도	0.547	0.580
첨도	-1.069	1.121



- 평균과 중위수가 각각 4.100, 4.000으로 유사하고
- 왜도, 첨도가 각각 -2와 2사이에 존재한다.
- Normal Q-Q plot을 봤을 때 \circ 가 대부분 선위에 있다.
(동그라미가 관찰값에 따라 계산된 값이다.)

d) Kolmogorov-Smirnov, Shapiro-Wilk test에서 각각 p-value가 0.091, 0.051으로 유의 수준 0.05보다 크다.

⇒ 위의 모든 조건들이 표본이 정규성을 만족함을 보여준다.

Tests of Normality

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
IFEV1(L)	.205	15	.091	.882	15	.051

4) 정규성을 만족하므로 모수적 접근을 한다.

2) 동질성 검정

조사대상 모집단을 여러 개의 부차모집단으로 분할하였을 때
부차모집단의 분포의 형태가 동일한지를 검증

(1) 동질성 검정의 예제



예제

전공실습을 이수한 학생(실험집단)과 아직 이수하지 못한 학생(통제집단)의
진로선택에 대한 생각을 묻는 사전-사후비교로 실험하였다.

실습을 이수한 집단과 이수하지 않은 집단은 동질집단이며, 실습이수 여부가
학생들의 진로선택에 영향을 주었는가를 통계적으로 검증하라

H_0 : 두 집단은 서로 동일한 집단이다.

H_A : 두 집단은 다른 집단이다.

0.05이하면 영가설 기각, 대립가설 채택

본 예제는 독립표본 t검정(실험집단과 통제집단의 동질성 검사)를 통해 동질성검정을 실시하였다.

① 독립표본 t검정 대화상자 열기

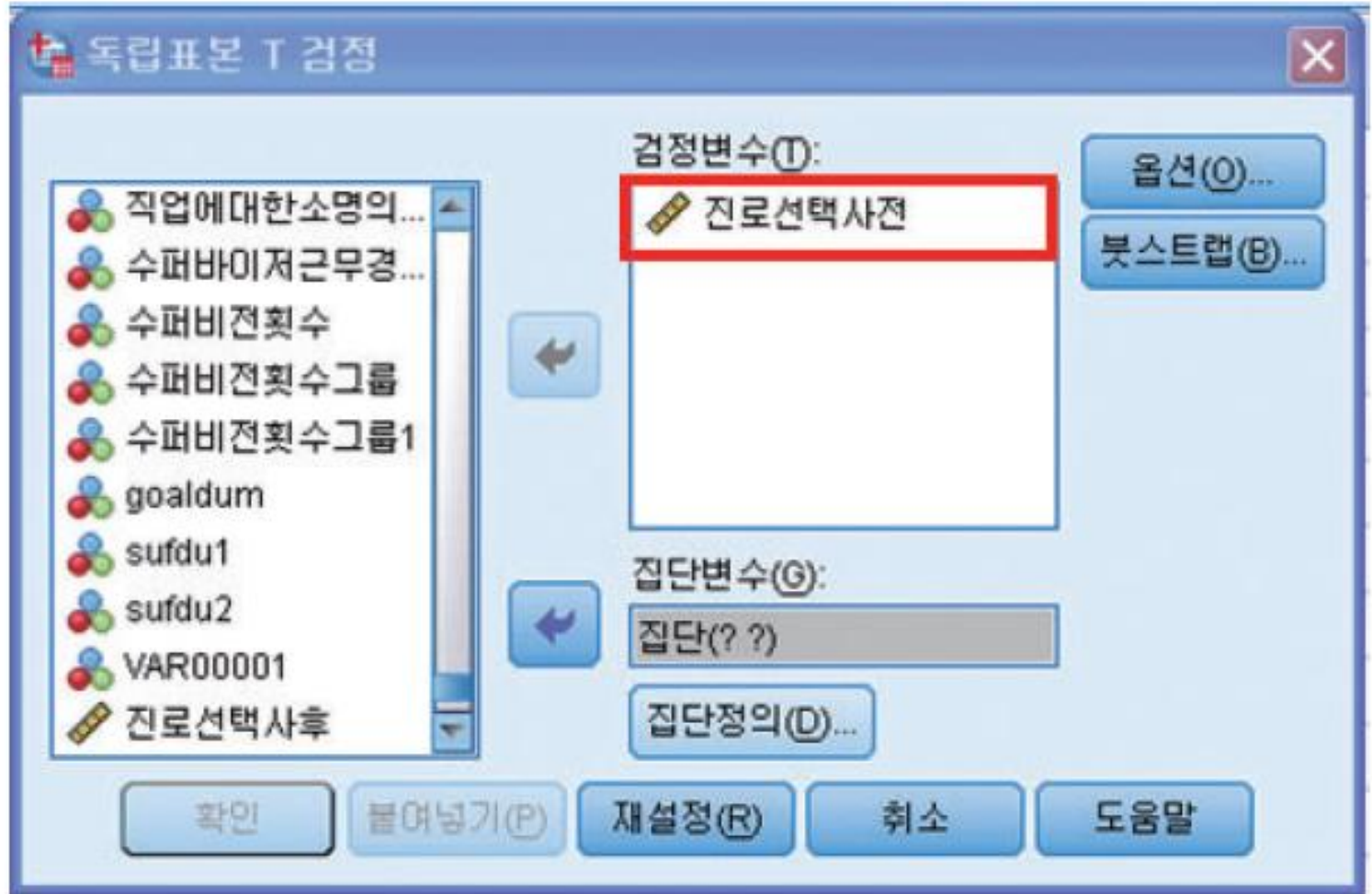
메뉴에서 ① [분석(A)] → ② [평균비교(M)] → ③ [독립표본 T검정(T)]을 클릭한다.

The screenshot shows the IBM SPSS Statistics Data Editor window with the following data table:

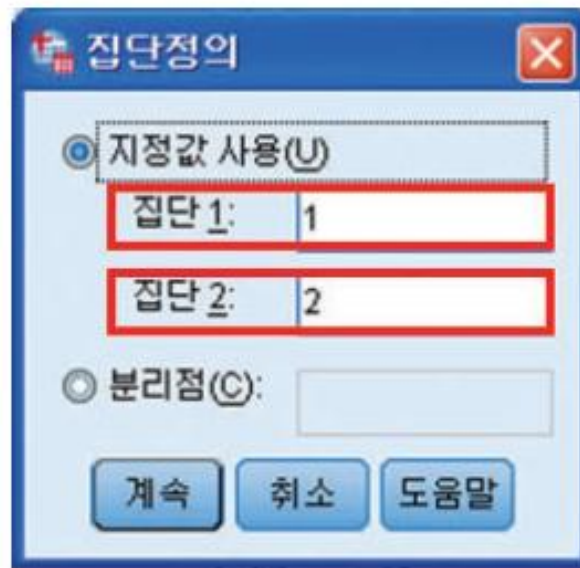
Case Number	Group	Pre-Experiment (사전_실험)	Post-Experiment (사후_통제집단)
1	1	3.61	2.61
2	0	3.50	2.50
3	4	2.44	1.44
4	0	3.50	2.50
5	4	2.44	1.44
6	3	4.33	3.33
7	3	2.39	1.39
8	4	2.94	1.94
9	3	3.33	2.33
10	3	4.28	3.28
11	3	2.28	1.28
12	4	4.44	3.44
13	0	3.67	3.00
14	0	3.61	3.00
15	0	4.17	3.00
16	0	3.28	2.00
17	0	2.83	1.00
18	0	4.67	3.00
19	0	2.39	1.00

② 변수 옮기기

검정변수(T)에는 진로선택사전점수를 이동시킨다.



③ 분석 실행하기



[집단정의(D)]를 클릭하여 집단 1에는 1을, 집단 2에는 2를 입력하고 [확인]을 누른다.

4 결과 해석하기

실험집단과 통제집단의 표본수(N)와 평균이 출력되었다. 실습을 이수하기 전 진로선택 수준에 대한 실험집단의 평균은 2.5541, 통제집단의 평균은 2.6370이다.

집단통계량

집단	N	평균	표준편차	평균의 표준오차
진로선택사전_실험집단	294	2.5541	.75757	.04418
실험집단 통제집단	294	2.6370	.77814	.04538

실험집단과 통제집단의 진로선택수준에 대한 검사를 실시하여 차이를 살펴본 결과 실험 집단과 통제집단 간 진로선택수준의 차이는 $p < 0.05$ 의 수준에서 통계적으로 유의하지 않았다. 따라서 실험집단과 통제집단은 동질로 추론할 수 있다.

독립표본 검정

	Levene의 등분산 검정		평균의 동일성에 대한 t검정						
	F	유의확률	t	자유도	유의확률 (양쪽)	평균차	차이의 표준오차	차이의 95% 신뢰구간	
								하한	상한
진로선택사전_실험집단	.572	.450	-1.309	586	.191	-.08288	.06334	-.20728	.04151
진로선택사전_통제집단	.572	.450	-1.309	585.580	.191	-.08288	.06334	-.20728	.04151



여기서 돋보기!

논문에 실제로 적용하기

실습 전, 실험집단과 통제집단의 진로선택수준에 대한 동질성 검정 결과, 진로선택($t=-1.309$, $p=0.191$)은 유의미한 차이가 없어 두 집단이 동질한 것으로 나타났다.

표 7-3. 실험집단과 통제집단의 진로선택에 대한 동질성 검증

변수	빈도	평균	표준편차	t	p값
실험집단	294	2.55	.75	-1.309	.191
통제집단	294	2.63	.77		

3) 독립성 검정

분류기준(요인)이 둘인 경우 즉 범주형 변수가 2개인 경우 이들 변수
간에 관련성이 있는지를 검증

(1) 독립성 검정의 예제



예제

최근 입법 예고된 의료법안에 대하여 A치료사군과 B치료사군 간의 찬성에 대한 인식을 설문조사한 결과 다음과 같은 결과를 얻었다. 이때 동일한 법안에 대한 두 치료사군의 법안에 대한 찬성의 인식이 동일한지를 분석하기 위하여 SPSS 프로그램을 이용하여 실습해보도록 하자.

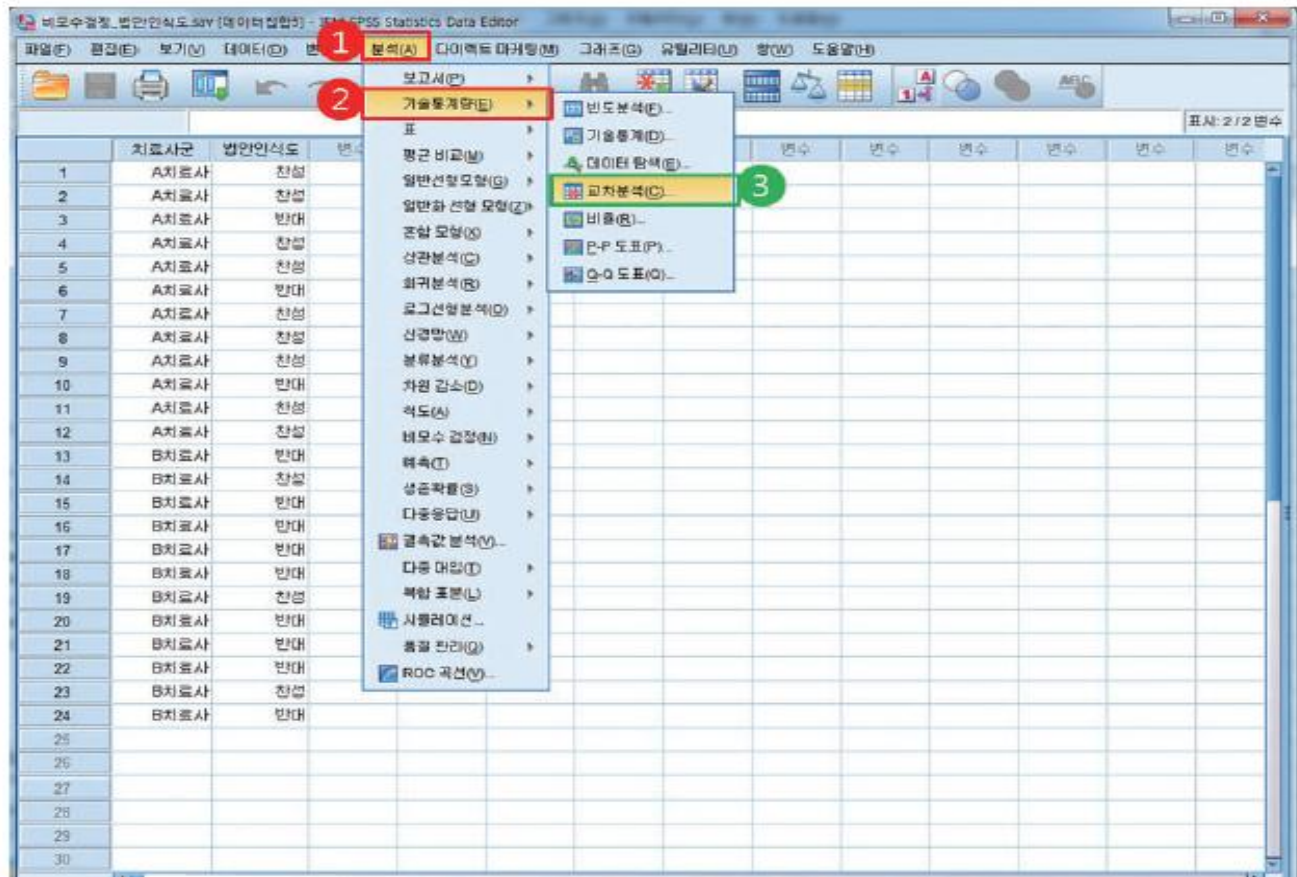
① 통계학적 가설

H_0 : 그룹과 법안에 대한 인식이라는 자료는 서로 독립적이다.

H_A : 그룹과 법안에 대한 인식이라는 자료는 서로 독립적이지 않다.

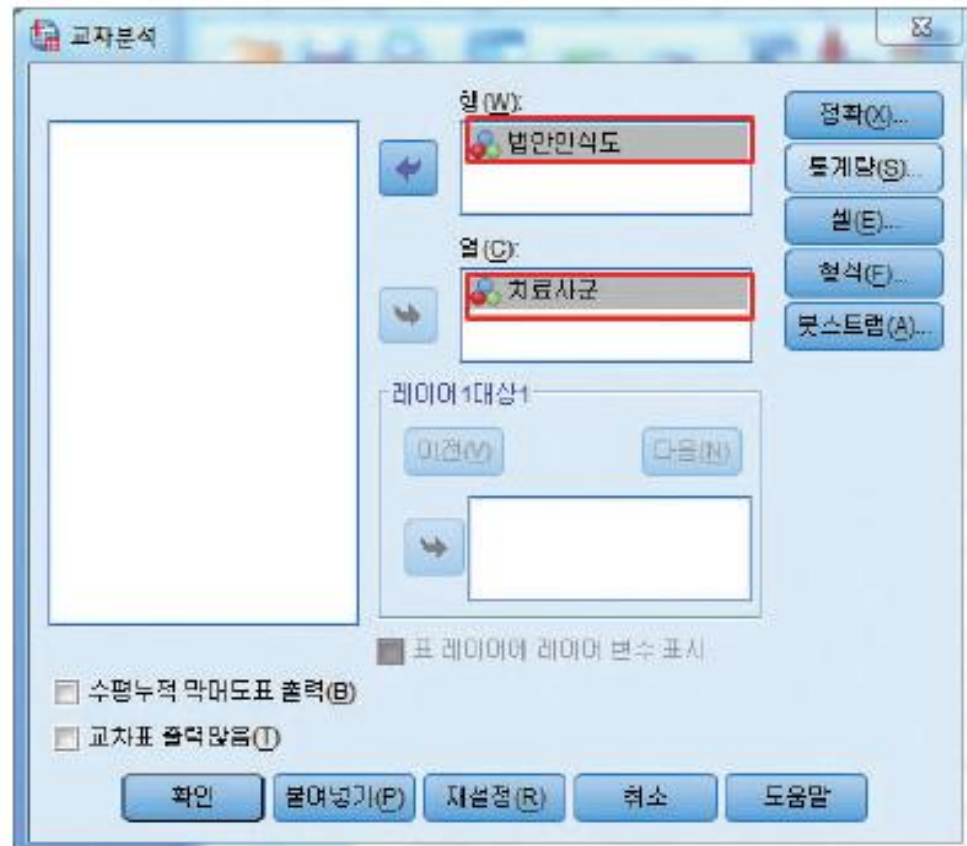
② 독립성 검정 대화상자 열기

메뉴에서 [분석(A)] → [기술통계량(E)] → [교차분석(C)]을 클릭한다.



③ 변수 옮기기

행(W)에는 법안인식도를 이동시키고 열(C)에는 치료사군을 이동시킨다.



4 분석 실행하기

[통계량(S)] 클릭하여 카이제곱(H) 체크하고 [계속]을 클릭한다.



[셀(E)] 클릭하여 관측빈도(O)와 퍼센트의 전체(T)에 체크하고 [계속], [확인]을 클릭한다.

교차분석: 셀 출력

빈도

- 관측빈도(O)
- 기대빈도(E)
- 낮은 빈도 숨기기

미만 5

Z 검정

- 열 비율 비교
- p 값 조정(본페로니법)

퍼센트

- 행(R)
- 열(C)
- 전체(T)

잔차

- 비표준화(U)
- 표준화(S)
- 수정된 표준화(A)

정수가 아닌 가중값

- 셀 수 반올림(N)
- 케이스 가중값 반올림(W)
- 셀 수 절삭(L)
- 케이스 가중값 절삭(H)
- 조정 없음(M)

계속 취소 도움말

5 결과 해석하기

카이제곱검정

	값	자유도	점근 유의확률 (양측검정)	정확한 유의확률 (양측검정)	정확한 유의확률 (단측검정)
Pearson 카이제곱	6.000 ^a	1	.014		
연속수정 ^b	4.167	1	.041		
우도비	6.279	1	.012		
Fisher의 정확한 검정				.039	.020

분석 결과에서 **Pearson 카이제곱의 값**은 6.000이고 이때 점근 유의확률이 0.014이지만 예제처럼 2X2 교차표인 경우, 즉 자유도가 1일 때는 연속수정을 읽어준다. 따라서 카이 제곱 값은 4.167, **유의확률은 0.041로 유의확률이 0.05보다 작기 때문에, 영가설은 기각 하고 대립가설을 채택**하여 그룹과 법안에 대한 인식이라는 자료는 서로 독립적이지 않은 것을 알 수 있다



여기서 돋보기!

논문^에 실제로 적용하기

A치료사와 B치료사 간 법안 인식도의 차이를 분석한 결과 A치료사의 경우는 37.5%가 찬성하였으나 B치료사의 경우는 12.5%만 찬성을 하여 두 치료사군 사이의 법안 인식도에는 유의한 차이가 있었다.

표 7-4. 치료사군 간의 법안인식도

N(%)

법안인식도	치료사 군		전체	p값
	A치료사	B치료사		
찬성	9(37.5)	3(12.5)	12(50)	.041
반대	3(12.5)	9(37.5)	12(50)	
전체	12(50)	12(50)	24(100.0)	

스피어만 상관계수 Spearman

자료가 서열척도로서 구성된 경우 자료의 원래 값 대신 순위를 이용하여 계산하는 상관계수. 두 변수의 상관관계가 선형이든 비선형이든 관계없이 상관관계만 밝혀줌.

두 변수의 값이 $-1 < r < 1$ 의 값을 갖으며 두 변수의 데이터 값들의 순위가 완전히 일치하면 그 값은 1이며 반대로 일치하지 않으면 -1이 된다.

Pearson (피어슨 상관계수)

두 변수간의 관련성을 구하기 위해 보편적으로 이용되는 척도.

두 변수 간의 값이 $-1 < r < 1$ 의 값을 갖으며 여기에서 상관계수가 0이라는 것은 선형의 상관관계가 없다는 의미이다.

보통 상관계수가 0.7~0.8이면 강한 양의 상관관계라 보고 0.6이며 보통의 양의 상관관계를 갖고 있다고 볼 수 있다.



Tip

교차분석 시 유의할 점

1. 최소기대빈도를 먼저 확인하여 5보다 커야 하며 전체 25가 넘으면 안 된다.
2. 기대빈도는 25가 넘으면 안 된다(넘을 경우 p 값을 볼 수 없다).
 - ① 기대빈도를 5보다 크게 하라.
 - ② 표본수를 늘리거나, 변수를 합하여 분석하라(변수의 하위 항목이 여러 개인 경우)
3. 해석할 때
 - ① 2×2 table에서는 Fisher's 정확한 검정을 읽어라.
 - ② 2×2 table 이상이면서 서열변수일 경우는 trend로 해석하되 이때는 선형 대 선형결합을 읽고, 명목변수는 차이를 해석하라.