

제6장 카이제곱검정

1. 기본개념

1) 범주형 자료

- 범주로만 분류될 수 있는 관측 값(수치적으로 측정되지 않는 자료)

Ex. 종교 문제의 연구(불교, 기독교, 회교 등)

- 직업의 만족도(만족하는가, 보통인가, 만족하지 않는가)

2) 분할표(Contingency table)

- 두 가지 혹은 그 이상의 속성에 따라 분류된 표본 관측 값으로부터 얻어지는 도수로 구성되는 데이터.

1. 기본개념

3) 범주형 자료 분석 - 피어슨의 검정

- 각 모집단이 두 가지 이상의 서로 다른 속성을 갖는 개체들로 나뉘는 경우에 여러 모집단을 비교하는 방법으로 다음과 같은 검정이 있다.

- ① 여러 범주로 분류되는 단일표본 - 적합도 검정
- ② 여러 범주로 분류되는 독립표본 - 동질성 검정
- ③ 두 특성에 따라 동시에 분류되는 단일표본 - 독립성 검정

1. 기본개념

4) 모집단 비교방법

	적합성	독립성	동질성	
범주형	$\chi^2 - test$	$\chi^2 - test$	$\chi^2 - test$	명목척도 서열척도
수치형	t-test	t-test	t-test	등간척도 비율척도
	보통 1집단	2집단	여러집단	

2. 적합도 검정

1. 적합도에 대한 피어슨의 χ^2 검정

- 대표본 (n : 각 칸의 기대도수가 5이상)에서만 유효한 근사 검정

1) 적합도 검정에서의 통계적 모형

칸	1	2	...	k	합계
관측도수(O)	n_1	n_2	...	n_k	n
H_0 하의 확률	P_{10}	P_{20}	...	P_{k0}	1
H_0 하의 기대도수(E)	nP_{10}	nP_{20}	...	nP_{k0}	n

2. 적합도 검정

2) 적합도에 대한 피어슨의 χ^2 검정 (대표본의 경우)

① 귀무가설 : $H_0 : P_1 = P_{10} = \dots = P_k = P_{k0}$

② 검정통계량 : $\chi^2 = \sum_{i=1}^k \frac{(n_i - nP_{i0})^2}{nP_{i0}} = \sum_{\text{관}} \frac{(O-E)^2}{E}$

(χ^2 : 적합도에 대한 피어슨의 χ^2 통계량)

③ 기각역 : $\chi^2 \geq \chi^2_{\alpha}$, $d.f. = k - 1$ (칸의 수 - 1)

2. 적합도 검정

예제 p106 완두콩 교배 실험에서 나타나는 잡종은 A(둥글고 노란완두), B(둥글고 녹색), C(모나고 노랑), D(모나고 녹색)

- ▶ 멘델의 유전 이론적 모형에 의하면 A, B, C, D형태의 잡종은 9:3:3:1의 비율
- ▶ 실험으로 이것을 확인하기 위하여 두 가지 형태의 식물의 교잡으로 나타난 잡종은 다음표와 같다.

2. 적합도 검정

▶ 잡종의 분류

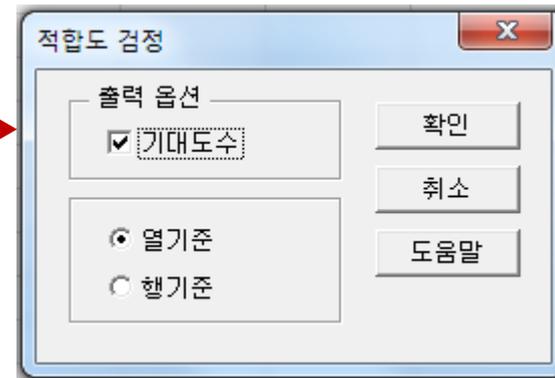
인자형	A	B	C	D	합 계
관 측 도 수	315	108	101	32	100

- ▶ 이 데이터는 유전의 이론적 모형에 부합된다고 할 수 있는가?(유의수준:0.05)

2. 적합도 검정

- ▶ 분석방법
- ▶ 통계분석 - 범주형자료분석 - 적합도검정

	A	B	C	D
1	A	B	C	D
2	322	108	98	32
3	9	3	3	1



2. 적합도 검정

- ▶ 가설 $H_0 : p_A = \frac{9}{16}, p_B = \frac{3}{16}, p_C = \frac{3}{16}, p_D = \frac{1}{16},$
- ▶ $H_1 : H_0$ 가 아니다.

- 검정통계량

- $\chi^2 = 0.9651$
- 유의확률 = $0.8097 > 0.05$

이므로 유의수준 0.05에서
귀무가설 기각할 수 없다.

- 그러므로 유전이론모형에 부합한다.

적합도 검정

도수표

	A	B	C	D
관측도수	322	108	98	32
기대도수	315	105	105	35

카이제곱 통계량 : 0.9651
유의확률 : 0.8097

2. 적합도 검정

EX. 아래의 데이터는 주사위를 300번 던졌을 때 관측된 도수이다. 이 데이터가 주사위의 공정성에 의문을 제기한다고 볼 수 있는가?
(유의수준:0.05)

면의 번호	1	2	3	4	5	6	합 계
도 수	33	61	49	65	55	37	300

2. 적합도 검정

EX. 아래의 데이터는 주사위를 300번 던졌을 때 관측된 도수이다. 이 데이터가 주사위의 공정성에 의문을 제기한다고 볼 수 있는가?
(유의수준:0.05)

- ▶ 가설 :
- ▶ 검정통계량:
- ▶ 유의확률:
- ▶ 결론:

2. 적합도 검정

예제 - 실습예제 6-1 p112

3. 교차분석

- ▶ 두 변수들의 관련성을 알아보는 분석.
- ▶ 데이터에서 한 변수의 범주를 다른 변수의 범주에 따라 빈도(frequency)를 교차 분류하는 교차표(cross tabulation) 또는 분할표(contingency table)를 작성하고 변수간의 관련성 알아본다.
- ▶ 교차분석 프로시저를 사용
 - 피어슨(Pearson)의 카이제곱,
 - 우도비 카이제곱,
 - 선형별 결합 검정,
 - 피셔(Fisher)의 정확한 검정,
 - 피어슨(Pearson)의 상관계수,
 - 스피어만(Spearman)의 상관계수 등 많은 통계량과 결합측도가 출력

3. 교차분석 (독립성 검정)

▶ 카이 제곱 검정(Chi-Square Test)

▶ 가설

- 귀무가설 : 두 가지 특성 A 와 B는 서로 독립이다.
- 대립가설 : 두 가지 특성 A 와 B는 서로 독립이 아니다.
- 검정통계량 (대표본인 경우)

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O - E)^2}{E} \sim \chi^2(r - 1)(c - 1),$$

- O : 관측도수 r : 행의수 c : 열의수
- E : 기대도수 n : 총합계
- n_i : 행합계 n_j : 열합계

- 기각역 R: $\chi^2 \geq \chi^2_{\alpha}((r - 1)(c - 1))$

3. 교차분석

- ▶ 기대빈도와 실제빈도의 차이를 이용하여 독립적인가, 관련성이 있느냐를 판단한다
- ▶ 교차표

		성별		Total
		남성	여성	
학력	고졸이하	12	11	n1(행합계) 23
	전문대재졸	13	6	n2(행합계) 19
Total		n1(열합계) 25	n2(열합계) 17	n(총합계) 42

3. 교차분석

- ▶ 두 변수의 형태가 명목, 순서
- ▶ 교차표

		성별	
		남성	여성
흡연 여부	흡연	A	B
	비흡연	C	D

- A : 흡연하는 남성의 빈도
- B : 흡연하는 여성의 빈도
- C : 흡연하지 않는 남성의 빈도
- D : 흡연하지 않는 여성의 빈도

3. 교차분석

▶ 성별, 학력의 교차표

		성별		Total
		남성	여성	
학력	고졸이하	12	11	23
	전문대재졸	13	6	19
	대재졸	65	47	112
	대학원재졸	41	5	46
Total		131	69	200

- 성별과 학력이 서로 관계가 있다고 말할 수 있는가?

3. 교차분석

▶ 성별, 학력의 교차표(기대빈도)

		성별		Total
		남성	여성	
학력	고졸이하	$23 \times 131 / 200$ =15.1	$23 \times 69 / 200$ =7.9	23
	전문대재졸	$19 \times 131 / 200$ =12.4	$19 \times 69 / 200$ =6.6	19
	대재졸	$112 \times 131 / 200$ =73.4	$112 \times 69 / 200$ =38.6	112
	대학원재졸	$46 \times 131 / 200$ =30.1	$46 \times 69 / 200$ =15.9	46
Total		131	69	200

◦ 기대빈도: (해당 행의 합) * (해당 열의 합) / (전체 빈도의 합)

3. 교차분석

- ▶ 성별, 학력의 교차표(관찰빈도-기대빈도, 제공값)

		성별	
		남성	여성
학력	고졸이하	$12 - 15.1 = -3.1$ (9.61)	$11 - 7.9 = 3.1$ (9.61)
	전문대재졸	$13 - 12.4 = 0.6$ (0.36)	$6 - 6.6 = -0.6$ (0.36)
	대재졸	$65 - 73.4 = -8.4$ (70.56)	$47 - 38.6 = 8.4$ (70.56)
	대학원재졸	$41 - 30.1 = 10.9$ (118.81)	$5 - 15.9 = -10.9$ (118.81)

3. 교차분석

- ▶ 성별, 학력의 교차표(제곱값/기대빈도)

		성별		Total
		남성	여성	
학력	고졸이하	9.61 / 15.1 =0.64	9.61 / 7.9 =1.22	1.86
	전문대재졸	0.36 / 12.4 =0.03	0.36 / 6.6 =0.05	0.08
	대재졸	70.56 / 73.4 =0.96	70.56 / 38.6 =1.83	2.79
	대학원재졸	118.81 / 30.1 =3.95	118.81 / 15.9 =7.47	11.42
Total		5.58	10.57	16.15

- Chi-square값 : 16.15
- 자유도 : $(2-1)*(4-1) = 3$

3. 교차분석

- ▶ 독립성검정: 카이제곱검정(chi-square test)
- ▶ 귀무가설
 - 두 변수는 독립적인 관계이다.
- ▶ 대립가설
 - 두 변수는 독립적인 관계가 아니다.
- ▶ 독립성분석의 주의사항
 - 기대빈도가 5 미만인 셀이 전체의 20%를 초과할 때에는 독립성분석을 하지 않는 것이 좋다.
 - 빈도가 적은 셀은 제외한 뒤 분석
 - 빈도가 적은 셀을 다른 셀과 합친 뒤 분석
 - 더 많은 표본 확보

3. 교차분석(독립성 검정)

▶ 성별, 학력의 교차표

		성별		Total
		남성	여성	
학력	고졸이하	12	11	23
	전문대재졸	13	6	19
	대재졸	65	47	112
	대학원재졸	41	5	46
Total		131	69	200

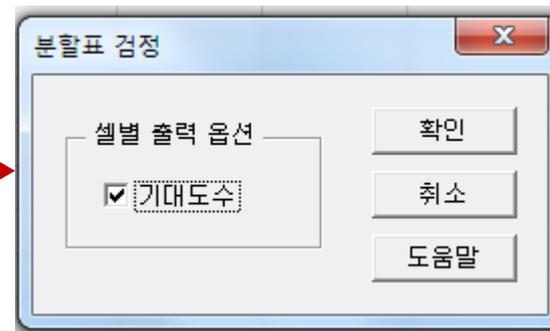
- 성별과 학력이 서로 관계가 있다고 말할 수 있는가?
- 유의수준: 0.05

3. 교차분석(독립성 검정)

▶ 분석방법

- 통계분석 - 범주형자료분석 - 분할표

	A	B	C
1		남성	여성
2	고졸	12	11
3	전문대졸	13	6
4	대재졸	65	47
5	대학원졸	41	5



3. 교차분석(독립성 검정)

▶ 성별과 학력이 서로 관계가 있다고 말할 수 있는가?

◦ 가설

- 귀무가설
 - 두 변수는 독립적인 관계이다.
- 대립가설
 - 두 변수는 독립적인 관계가 아니다.

• 유의수준 0.05

- 카이제곱통계량 = 16.0075
 - 유의확률 = 0.00113 < 0.05
- 유의수준 0.05에서 귀무가설을 기각 할 수 있다.
즉, 성별과 학력은 서로 관계가 있다고 할 수 있다.

분할표 검정

분할표

	남성	여성	계
고졸			
관측도수	12	11	23
기대도수	15,065	7,935	
전문대졸			
관측도수	13	6	19
기대도수	12,445	6,555	
대재졸			
관측도수	65	47	112
기대도수	73,36	38,64	
대학원졸			
관측도수	41	5	46
기대도수	30,13	15,87	
계	131	69	200

카이제곱 통계량 : 16,0075
유의확률 : 0,00113

3.교차분석(독립성 검정)

예제 - 실습예제 6-4 p118

3. 교차분석(동질성 검정)

- ▶ 범주형 자료에서 여러 집단을 비교 할 때 사용하는 방법
 - 성별, 학력의 교차표

		성별		Total
		남성	여성	
학력	고졸이하	12	11	23
	전문대재졸	13	6	19
	대재졸	65	47	112
	대학원재졸	41	5	46
Total		131	69	200

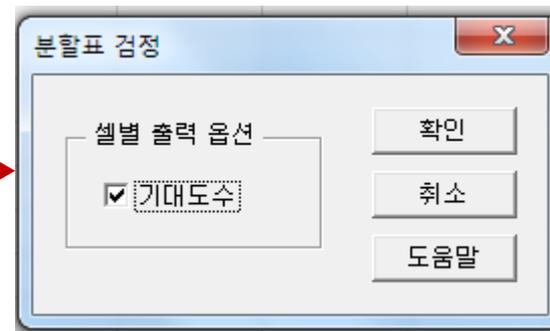
- 학력에 따라 성별은 차이가 있다고 말할 수 있는가?
- 유의수준: 0.05

3. 교차분석(동질성 검정)

▶ 분석방법

- 통계분석 - 범주형자료분석 - 분할표

	A	B	C
1		남성	여성
2	고졸	12	11
3	전문대졸	13	6
4	대재졸	65	47
5	대학원졸	41	5



3. 교차분석(동질성 검정)

▶ 학력에 따라서 성별은 차이가 있다고 말할 수 있는가?

◦ 가설

- 귀무가설
 - 학력에 따른 성별비율의 차이는 없다.
- 대립가설
 - 학력에 따른 성별비율의 차이가 있다.

분할표 검정

분할표

- 유의수준 0.05
 - 카이제곱통계량 = 16.0075
 - 유의확률 = 0.00113 < 0.05
- 결론 :
 - 유의수준 0.05에서 귀무가설을 기각 할 수 있다.
 - 즉, 학력에 따른 성별 비율의 차이가 있다고 할 수 있다.

	남성	여성	계
고졸			
관측도수	12	11	23
기대도수	15,065	7,935	
전문대졸			
관측도수	13	6	19
기대도수	12,445	6,555	
대재졸			
관측도수	65	47	112
기대도수	73,36	38,64	
대학원졸			
관측도수	41	5	46
기대도수	30,13	15,87	
계	131	69	200

카이제곱 통계량 : 16,0075
 유의확률 : 0,00113

과제9

- ▶ 실습예제 6-2
- ▶ 실습예제 6-3
- ▶ 연습문제 6 p120
 - 생활만족도에 따른 성별 비율의 차이가 있다고 할 수 있는지 유의수준 0.05에서 검정하여라.
- ▶ 결과 분석의 내용은 한글파일
제목 :과제9_학번_이름.hwp로 작성하여 제출합니다.