

범주형 자료분석

목 차

| | | |
|------------|----------------------|-----------|
| 제1장 | 소 개 | 3 |
| | - 관계의 이해 | 3 |
| | - 범주형 자료 | 7 |
| | | |
| 제2장 | 두 모비율 차이의 검정 | 10 |
| | - 2X2 교차표 : 독립표본의 경우 | 10 |
| | - 2X2 교차표 : 독립표본(예) | 11 |
| | - 2X2 교차표 : 대응표본의 경우 | 15 |
| | - 2X2 교차표 : 대응표본(예) | 16 |
| | | |
| 제3장 | 교차표 분석 | 19 |
| | - 2X2 에서 확장 | 19 |
| | - 관찰도수와 기대도수 | 22 |

| | | |
|------------|-------------------|-----------|
| 제3장 | 교차표 분석 | 19 |
| | - 카이제곱 검정 | 23 |
| | - 동일성 검정 예 | 24 |
| | - 열이 순서형일 때 | 34 |
| | - 독립성 검정 예 | 38 |
| | - 행과 열이 모두 순서형일 때 | 40 |
| | - 정확 검정 | 43 |
| | - 연관성 측도 | 45 |
| | - 연관성 측도 : 람다 | 46 |
| | - 연관성 측도 : 감마 | 51 |
| | - 심슨의 파라독스 | 55 |
| | - 코호트 연구 | 57 |
| | - 사례-대조 연구 | 58 |

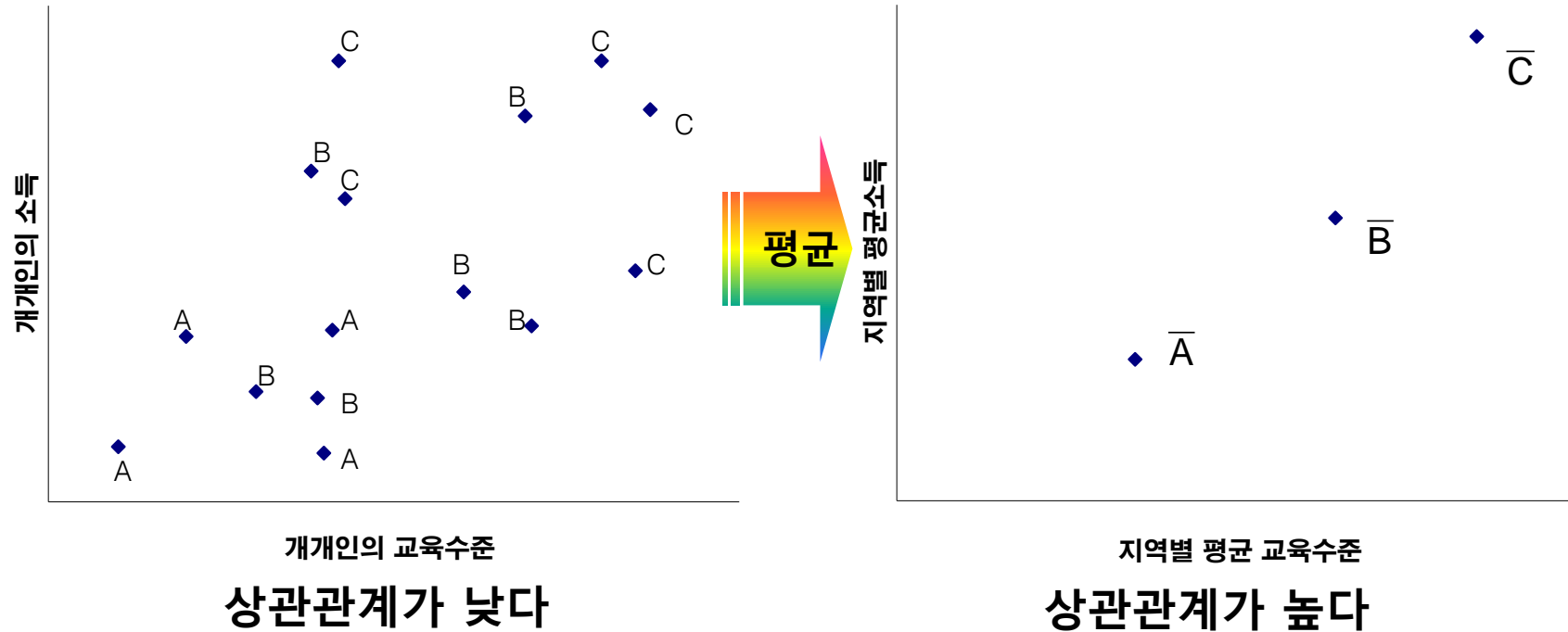
제1장. 소개 (관계의 이해)



상관관계가 곧바로 인과관계를 말하지는 않습니다.

제1장. 소개 (관계의 이해)

상관계수가 실제의 관계를 과장하는 경우



- 집단별 비율이나 평균에 기초하여 구한 상관계수는 실제의 관계를 과장한다.

제1장. 소개 (관계의 이해)

분기별 음료수 판매 실적과 소아마비 발생률



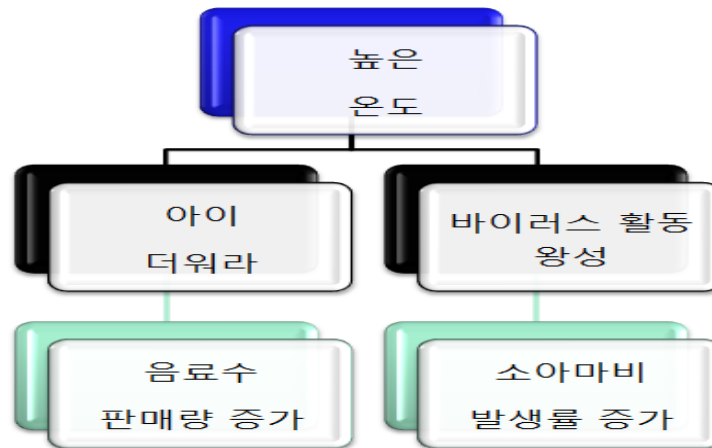
➤ 두 변수 사이에 강한 양의 관련성이 있는 것처럼 보인다.

제1장. 소개 (관계의 이해)

➤ 관련성을 인과성(causation)으로 잘못 판단한다면?

소아마비 발생률 낮추기 위해
음료수 판매 중지!

NO !!!!



제1장. 소개



- 범주형 변수 (Categorical Variable)

- 일반적으로 연산 안됨 (예 : 강원도+여자?)

예 : 성별(gender) - 여(0), 남(1)

출신지역 - 서울·경기(1), 강원(2), 충청(3), 호남(4), 영남(5)

교육수준 - 초등(1), 중(2), 고등(3), 대학(4)

유형: 이항형(binary) 예- 성별

명목형(nominal) 예- 출신지역

순서형(ordered) 예- 교육수준

: 대학생이 고등학교 출신보다 교육을 더 많이 받았다.

가변수 혹은 연속형으로 처리가 가능

- 범주형 자료 (Categorical Data)

- 일반적으로 반응변수가 이산적(discrete)인 자료

⇒ 범주형 변수만 있거나 반응변수가 범주형인 경우 범주형 자료라 함

제1장. 소 개



범주형 자료의 특성



회귀분석  VS 범주형자료분석



표본추출모형
(이항표본추출)


제1장. 소 개 - 범주형 변수의 분포

이항분포 (Binomial Dist.) : $B(n,p)$

$$P(X = x) = f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n.$$

$$\text{여기서, } \binom{n}{x} = {}_n C_x = \frac{n!}{x!(n-x)!}$$

다항분포 (Multinomial Dist.)

$$P(N_1 = n_1, N_2 = n_2, \dots, N_k = n_k) = \frac{n!}{n_1! n_2! \dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}$$

여기서 $n_1 + n_2 + \dots + n_k = n$ 이고 $p_1 + p_2 + \dots + p_k = 1$

- 동전 \rightarrow 이항, 주사위 \rightarrow 6항 ... p -값 계산 가능

제2장. 두 모비율 차이 검정-2×2 교차표 : 독립표본의 경우

□ 독립표본의 경우

$$\begin{cases} \hat{p}_1 = x_1/n_1 \\ \hat{p}_2 = x_2/n_2 \end{cases}$$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

검정통계량 :

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})(1/n_1 + 1/n_2)}}$$

기각역 :

| | 귀무가설 | 대립가설 | 기각역 |
|-----|----------------------|----------------------|-------------------------|
| (a) | $H_0 : p_1 \leq p_2$ | $H_1 : p_1 > p_2$ | $Z \geq z_\alpha$ |
| (b) | $H_0 : p_1 \geq p_2$ | $H_1 : p_1 < p_2$ | $Z \leq -z_\alpha$ |
| (c) | $H_0 : p_1 = p_2$ | $H_1 : p_1 \neq p_2$ | $ Z \geq z_{\alpha/2}$ |

- ✓ Z 검정통계량은 표본크기가 충분히 클 때 근사적으로 표준정규분포에 따름.
- ✓ 정확한 유의확률은 이항분포를 이용하여 계산될 수 있음.

제2장. 두 모비율 차이 검정-2×2 교차표 : 독립표본의 경우(예)

- 어느 지역에서 핵폐기물 처리장 설치에 대한 의견이 성인 남녀별로 차이가 있는가를 알아보기 위해 조사를 해본 결과, 임의로 뽑은 250명의 성인 남자 중 110명이 핵폐기물 처리장설치를 찬성하였고, 200명의 성인 여자 중 104명이 찬성한 것으로 나타났다. 성인 남녀별로 핵폐기물 처리장 설치 찬성비율에 차이가 있는지를 유의수준 5%하에서 검정하여 보자.

- 표본비율 : $\hat{p}_1 = \frac{110}{250} = 0.440, \quad \hat{p}_2 = \frac{104}{200} = 0.520$

- 공통비율 : $\hat{p} = \frac{110 + 104}{250 + 200} = \frac{214}{450} = 0.476$

- 검정통계량 :
$$Z = \frac{0.440 - 0.520}{\sqrt{(0.476)(0.524)(1/250 + 1/200)}} = 1.69$$

- 검정통계량의 값($Z=1.69$)이 기각치($z_{0.025}=1.96$)보다 작으므로 귀무가설을 기각할 수 없다. 따라서 정부에 대한 지지율이 남녀별로 차이가 있다고 할 수 없다.

제2장. 두 모비율 차이 검정-2×2 교차표 : 독립표본의 경우(예)

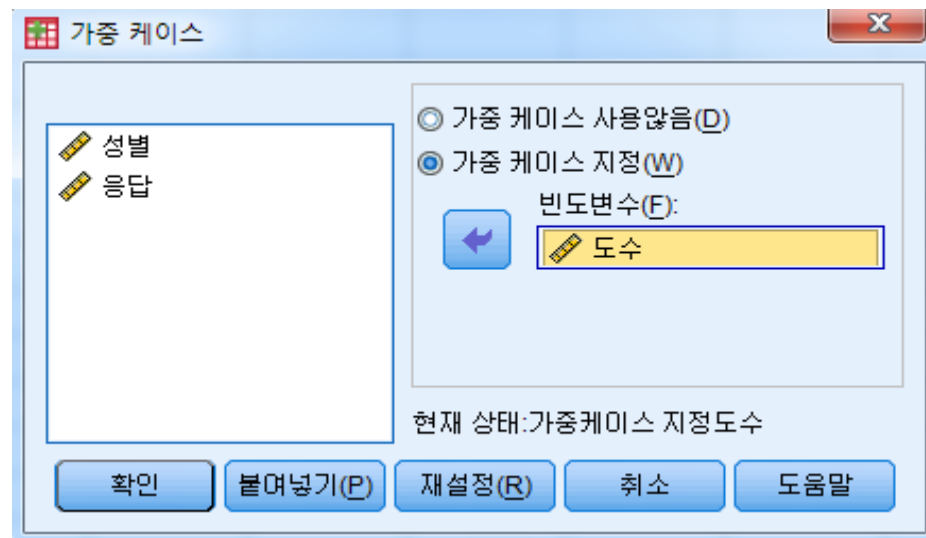
- SPSS 사례분석 - 데이터(CAT01.sav)

*CAT01.sav [데이터집합1] - PASW Statistics Data Editor

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 다이렉트 마커

1: 성별 1.00

| | 성별 | 응답 | 도수 | 변수 |
|---|------|------|--------|----|
| 1 | 1.00 | 1.00 | 110.00 | |
| 2 | 1.00 | 2.00 | 140.00 | |
| 3 | 2.00 | 1.00 | 104.00 | |
| 4 | 2.00 | 2.00 | 96.00 | |
| 5 | | | | |



제2장. 두 모비율 차이 검정-2×2 교차표 : 독립표본의 경우(예)

□ SPSS 사례분석 (CAT01.sav)

*CAT01.sav [데이터집합1] - PASW Statistics Data Editor

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 창(W) 도움말(H)

| | 성별 | 응답 | 도수 | 변수 | 변수 | 변수 | 변수 | 변수 | 변수 | 변수 |
|---|------|------|--------|----|----|----|----|----|----|----|
| 1 | 1.00 | 1.00 | 110.00 | | | | | | | |
| 2 | 1.00 | 2.00 | 140.00 | | | | | | | |
| 3 | 2.00 | 1.00 | 104.00 | | | | | | | |
| 4 | 2.00 | 2.00 | 96.00 | | | | | | | |
| 5 | | | | | | | | | | |

교차분석: 통계량

☒ 카이제곱(H) ☐ 상관관계(R)

명목 데이터

☐ 분할계수(O) ☐ 순서

☐ 파이 및 크레이머의 V(P) ☐ 감마(G)

☐ 람다(L) ☐ Somers의 d(S)

☐ 불확실성 계수(U) ☐ Kendall의 타우-b

명목 대 등간 척도

☐ 에타(E) ☐ Kendall의 타우-c

☐ 카파(K)

☐ 위험도(I)

☐ McNemar

☐ Cochran 및 Mantel-Haenszel 통계량

검정 공통승산비가 동일(T): 1

계속 취소 도움말

교차분석

행(W):

도수

성별

열(C):

응답

레이어1대상1

이전(V) 다음(N)

☐ 수평누적 막대도표 출력(B)

☐ 교차표 출력값(T)

확인 붙여넣기(P) 재설정(R) 취소 도움말

정확(X)... 통계량(S)... 셀(E)... 형식(F)... 붓스트랩(A)...

제2장. 두 모비율 차이 검정-2×2 교차표 : 독립표본의 경우(예)

• 출력결과

케이스 처리 요약

| | 케이스 | | | | | |
|---------|-----|--------|----|-----|-----|--------|
| | 유효 | | 결측 | | 전체 | |
| | N | 퍼센트 | N | 퍼센트 | N | 퍼센트 |
| 성별 * 응답 | 450 | 100.0% | 0 | .0% | 450 | 100.0% |

성별 * 응답 교차표

빈도

| | | 응답 | | 전체 |
|----|----|-----|-----|-----|
| | | yes | no | |
| 성별 | 남자 | 110 | 140 | 250 |
| | 여자 | 104 | 96 | 200 |
| 전체 | | 214 | 236 | 450 |

카이제곱 검정

| | 값 | 자유도 | 점근 유의확률 (양측검정) | 정확한 유의확률 (양측검정) | 정확한 유의확률 (단측검정) |
|-------------------|--------------------|-----|-------------------|-----------------------|-----------------------|
| Pearson 카이제곱 | 2.851 ^a | 1 | .091 | .106 | .055 |
| 연속수정 ^b | 2.540 | 1 | .111 | | |
| 우도비 | 2.853 | 1 | .091 | | |
| Fisher의 정확한 검정 | | | | | |
| 선형 대 선형결합 | 2.845 | 1 | .092 | | |
| 유효 케이스 수 | 450 | | | | |

a. 0 셀 (.0%)은(는) 5보다 작은 기대 빈도를 가지는 셀입니다. 최소 기대빈도는 95.11입니다.

b. 2x2 표에 대해서만 계산됨

$$\begin{aligned}
 Z &= \sqrt{\text{Pearson 카이제곱}} \\
 &= \sqrt{2.851} \\
 &= 1.688
 \end{aligned}$$

제2장. 두 모비율 차이 검정-2×2 교차표 : 대응표본의 경우



| | | B | | 합계 |
|---|---------|----------|-----------|-----------|
| | | 예 (1) | 아니오 (0) | |
| A | 예 (1) | n_{11} | n_{10} | n_A |
| | 아니오 (0) | n_{01} | n_{00} | $n - n_A$ |
| | 합계 | n_B | $n - n_B$ | n |

두 표본비율 $p_A = n_A/n$ 과 $p_B = n_B/n$ 는 서로 독립이 아니므로 앞에서 설명한 방법을 적용할 수 없으며, 대응표본의 경우 귀무가설 ' $H_0 : p_A = p_B$ ' 하에서 n_{01} 이 이항분포 $B(n_{01} + n_{10}, 1/2)$ 를 따른다는 사실을 이용하여 검정을 수행할 수 있다. 또한 다음과 같은 정규근사를 이용할 수도 있는데,

$$Z = \frac{n_{01} - (n_{01} + n_{10})/2}{\sqrt{(n_{01} + n_{10})/4}} = \frac{n_{01} - n_{10}}{\sqrt{n_{01} + n_{10}}}$$

검정통계량 Z 은 귀무가설 하에서 표준정규분포를 따르므로 Z^2 은 자유도 1인 카이제곱분포를 따르며, 이를 맥니머(McNemer)의 카이제곱 통계량이라고도 한다.

제2장. 두 모비율 차이 검정-2×2 교차표 : 대응표본의 경우(예)

- 환경부에서 1회용품 사용을 줄이고자 홍보영상을 만들었다. 100명의 사람을 대상으로 일회용품 사용 줄이기를 실천여부를 조사하여 홍보영상을 보기 전과 홍보영상을 본 이후의 응답을 얻었다. 1회용 줄이기 홍보영상이 사람들 인식에 변화를 주었다고 할 수 있는지 유의수준 5% 하에서 검정해 보아라.

| 이후 이전 | 예 | 아니오 | 합계 |
|----------|----|-----|-----|
| | 예 | 4 | 67 |
| 아니오 | 21 | 12 | 33 |
| 합계 | 84 | 16 | 100 |

제2장. 두 모비율 차이 검정-2×2 교차표 : 대응표본의 경우(예)

□ SPSS 사례분석 - 데이터(CAT02.sav)

*CAT02.sav [데이터집합2] - PASW Statistics Data Editor

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 창(W) 도움말(H)

1: 이전 1.00

| | 이전 | 이후 | 도수 |
|---|------|------|-------|
| 1 | 1.00 | 1.00 | 63.00 |
| 2 | 1.00 | 2.00 | 4.00 |
| 3 | 2.00 | 1.00 | 21.00 |
| 4 | 2.00 | 2.00 | 12.00 |

교차분석: 통계량

☒ 카이제곱(H) ☐ 상관관계(R)

명목 데이터

- ☐ 분할계수(O)
- ☐ 파이 및 크레이머의 V(P)
- ☐ 람다(L)
- ☐ 불확실성 계수(U)

명목 대 등간 척도

- ☐ 에타(E)

순서

- ☐ 감마(G)
- ☐ Somers의 d(S)
- ☐ Kendall의 타우-b
- ☐ Kendall의 타우-c

☐ 카파

☐ 위험도(I)

☒ McNemar

☐ Cochran 및 Mantel-Haenszel 통계량

검정 공통승산비가 동일(T): 1

계속 취소 도움말

교차분석

도수

행(W): 이전

열(C): 이후

레이어1대상1

이전(V) 다음(N)

☐ 수평누적 막대도표 출력(B)

☐ 교차표 출력없음(T)

확인 붙여넣기(P) 재설정(R) 취소 도움말

제2장. 두 모비율 차이 검정-2×2 교차표 : 대응표본의 경우(예)

• 출력결과

이전 * 이후 교차표

| 빈도 | | 이후 | | 전체 |
|----|-----|-----|----|-----|
| | | yes | no | |
| 이전 | yes | 63 | 4 | 67 |
| | no | 21 | 12 | 33 |
| 전체 | | 84 | 16 | 100 |

카이제곱 검정

| | 값 | 자유도 | 점근 유의확률 (양측검정) | 정확한 유의확률 (양측검정) | 정확한 유의확률 (1사) |
|-------------------|---------------------|-----|-------------------|-----------------------|---------------------|
| Pearson 카이제곱 | 15.197 ^a | 1 | .000 | | |
| 연속수정 ^b | 13.019 | 1 | .000 | | |
| 우도비 | 14.369 | 1 | .000 | | |
| Fisher의 정확한 검정 | | | | .000 | .000 |
| 선형 대 선형결합 | 15.045 | 1 | .000 | | |
| McNemar 검정 | | | | .001 ^c | |
| 유효 케이스 수 | 100 | | | | |

a. 0 셀 (.0%)은(는) 5보다 작은 기대 빈도를 가지는 셀입니다. 최소 기대빈도는 5.28입니다.

b. 2x2 표에 대해서만 계산됨

c. 미항분포를 사용함.

대칭적 측도

| | 값 | 점근 표준오차 ^a | 근사 T 값 ^b | 근사 유의확률 |
|----------|------|-------------------------|---------------------|---------|
| 일치 측도 카파 | .350 | .097 | 3.898 | .000 |
| 유효 케이스 수 | 100 | | | |

a. 영가설을 가정하지 않음.

b. 영가설을 가정하는 점근 표준오차 사용

제3장. 교차표 분석(2×2에서 확장)

두 개의 범주형 변수 A와 B에 대한 교차표(cross table) 혹은 분할표(contingency table)에서, 변수 A가 r 개의 범주, 변수 B가 c 개의 범주를 가지고 있다면 $r \times c$ 교차표는 다음과 같은 형태를 가진다.



| 변수 A | 변수 B | | | | | 행 합계 |
|----------|----------|-----|----------|-----|----------|----------|
| | 1 | ... | j | ... | c | |
| 1 | n_{11} | ... | n_{1j} | ... | n_{1c} | $n_{1.}$ |
| \vdots | \vdots | ... | \vdots | ... | \vdots | \vdots |
| i | n_{i1} | ... | n_{ij} | ... | n_{ic} | $n_{i.}$ |
| \vdots | \vdots | ... | \vdots | ... | \vdots | \vdots |
| r | n_{r1} | ... | n_{rj} | ... | n_{rc} | $n_{r.}$ |
| 열 합계 | $n_{.1}$ | ... | $n_{.j}$ | ... | $n_{.c}$ | n |

| 등급 | 기계 | | | 행 합계 |
|------|------|------|------|------|
| | 기계 1 | 기계 2 | 기계 3 | |
| 1급 | 78 | 65 | 68 | 211 |
| 2급 | 22 | 8 | 30 | 60 |
| 3급 | 20 | 2 | 7 | 29 |
| 열 합계 | 120 | 75 | 105 | 300 |

제3장. 교차표 분석

교차분석 : 범주에 따른 빈도수와 기대도수를 이용하여 범주의 특성을 분석

교차분석의 예

- 적합도 검정(Goodness of fit test) : 관측값이 특정한 분포를 따르는지를
검정
- 독립성 검정(test of independence) : 두 요인 간에 관계가 있는지를 검정
- 동일성 검정 (test of homogeneity) : 몇 개의 모집단이 분석하고자 하는
문제의 특성에 대하여 동질성을 갖는지
검정

제3장. 교차표 분석

'청량음료' 데이터

| 연령 | 청량음료 | | | | 행 합계 |
|-----|------|-------|-------|--------|------|
| | coke | pepsi | fanta | others | |
| 20대 | 10 | 14 | 4 | 12 | 40 |
| 30대 | 13 | 9 | 10 | 8 | 40 |
| 40대 | 12 | 8 | 10 | 10 | 40 |

동일성 검정

'교육수준과 소득수준' 데이터

| 교육수준 | 소득수준 | | | 행 합계 |
|------|------|-----|-----|------|
| | 상 | 중 | 하 | |
| 대졸 | 255 | 105 | 81 | 441 |
| 고졸 | 110 | 92 | 66 | 268 |
| 중졸 | 90 | 113 | 88 | 291 |
| 열 합계 | 455 | 310 | 235 | 1000 |

독립성 검정

$$H_0 : (p_{11}, p_{12}, \dots, p_{1c}) = \dots = (p_{r1}, p_{r2}, \dots, p_{rc}). \quad H_0 : p_{ij} = p_{i.} p_{.j}, \quad i = 1, \dots, r; \quad j = 1, \dots, c.$$

- 교차분석 : 2개의 범주형 변수에 대해 빈도표 그린 것(이차원 빈도표, 교차표)->일단 교차표 만듦.
- 동일성 : 연령에 따라 선호하는 청량 음료의 분포가 동일한가?
- 독립성 : 교육수준과 소득수준이 관련이 있는가?
- 두 양적 변수의 연관성->상관계수, 두 질적 변수의 연관성->교차표를 이용

제3장. 교차표 분석-관찰도수와 기대도수

관찰도수

n_{ij}

| 연령/상품 | A | B | C | 전체 |
|--------|-----------|------------|-----------|-----|
| 30세 이하 | 20 20% | 20 20% | 60 60% | 100 |
| 30세 이상 | 70 35% | 100 50% | 30 15% | 200 |
| 전체 | 90 30% | 120 40% | 90 30% | 300 |

기대도수

e_{ij}

$$e_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$$

| 연령/상품 | A | B | C | 전체 |
|--------|-----------|------------|-----------|-----|
| 30세 이하 | 30 30% | 40 40% | 30 30% | 100 |
| 30세 이상 | 60 30% | 80 40% | 60 30% | 200 |
| 전체 | 90 30% | 120 40% | 90 30% | 300 |

제3장. 교차표 분석-카이제곱 (Chi-Square) 검정

$$\chi^2 = \sum_i \sum_j \frac{(n_{ij} - e_{ij})^2}{e_{ij}} : \text{Pearson's chi - square}$$

IxJ 분할표에서 카이제곱에 대한 준거분포는 자유도가 (I-1)x(J-1)인 카이제곱 분포 (대표본 이론)

$$\text{예) } (20-30)^2/30 + (20-40)^2/40 + \dots + (30-60)^2/60 = 65$$

$$p\text{-값} = 0.001, \chi^2(2) = 5.99$$

⇒ 연령과 상품이 독립이라는 귀무가설을 기각

※ 카이제곱 검증을 위한 충분조건(Cochran의 기준) :

80% 이상의 칸(cell)들에서 기대빈도 $E_{ij} \geq 5$.

※ 이 조건을 만족시키지 못하는 경우에는 인접(유사)행 또는 열의 병합 후 카이제곱 검정을 적용하거나 Fisher's Exact Test를 수행해야 함.

제3장. 교차표 분석 : 동일성 검정



satisf.sav [데이터집합3] - PASW Statistics Data Editor

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 창(W) 도움말(H)

1 : gender 0

| | gender | agecat | regular | reason1 | reason2 | dept | purchase | payment | followup | distance | store | contact | price | numitems |
|----|--------|--------|---------|---------|---------|------|----------|---------|----------|----------|-------|---------|-------|----------|
| 1 | 0 | 3 | 1 | . | 2 | 6 | 1 | 3 | 1 | 2 | 2 | 1 | 4 | 5 |
| 2 | 0 | 1 | 2 | . | 2 | 7 | 1 | 5 | 2 | 2 | 4 | 1 | 4 | 5 |
| 3 | 1 | 3 | 1 | . | 1 | 1 | 1 | 3 | 5 | 4 | 3 | 0 | 4 | 2 |
| 4 | 0 | 3 | 1 | . | 2 | 3 | 1 | 5 | 3 | 2 | 1 | 0 | 3 | 2 |
| 5 | 1 | 4 | 2 | . | 2 | 3 | 1 | 2 | 5 | 2 | 1 | 1 | 2 | 1 |
| 6 | 0 | 3 | 2 | . | 2 | 7 | 1 | 2 | 3 | 2 | 1 | 1 | 2 | 4 |
| 7 | 1 | 4 | 3 | . | 1 | 3 | 0 | 4 | 1 | 2 | 3 | 1 | 2 | 3 |
| 8 | 0 | 3 | 3 | . | 1 | 1 | 1 | 5 | 1 | 3 | 3 | 1 | 5 | 5 |
| 9 | 1 | 4 | 3 | . | 1 | 7 | 0 | 3 | 5 | 2 | 1 | 1 | 3 | 2 |
| 10 | 0 | 3 | 1 | . | 2 | 6 | 1 | 4 | 5 | 1 | 1 | 1 | 3 | 4 |
| 11 | 0 | 3 | 0 | 2 | . | 1 | 1 | 4 | 5 | 4 | 4 | 1 | 3 | 5 |
| 12 | 0 | 4 | 2 | . | 2 | 6 | 1 | 4 | 1 | 1 | 2 | 1 | 5 | 5 |
| 13 | 1 | 4 | 1 | . | 2 | 7 | 1 | 3 | 5 | 3 | 1 | 1 | 1 | 2 |
| 14 | 0 | 3 | 2 | . | 3 | 7 | 1 | 4 | 1 | 4 | 2 | 1 | 4 | 1 |
| 15 | 1 | 3 | 1 | . | 3 | 4 | 1 | 3 | 5 | 5 | 4 | 1 | 2 | 1 |
| 16 | 0 | 3 | 3 | . | 1 | 3 | 1 | 4 | 5 | 2 | 2 | 1 | 2 | 1 |
| 17 | 1 | 3 | 4 | . | 2 | 7 | 0 | 5 | 2 | 4 | 4 | 0 | 5 | 5 |
| 18 | 1 | 2 | 2 | . | 4 | 4 | 1 | 3 | 5 | 2 | 4 | 1 | 5 | 5 |
| 19 | 1 | 4 | 1 | . | 2 | 4 | 0 | 2 | 5 | 5 | 4 | 0 | 4 | 3 |
| 20 | 1 | 2 | 3 | . | 4 | 7 | 1 | 3 | 4 | 1 | 3 | 1 | 3 | 2 |
| 21 | 0 | 3 | 2 | . | 2 | 4 | 1 | 3 | 1 | 1 | 3 | 1 | 4 | 3 |

35
36
37

1

데이터 보기(D) 변수 보기(V)

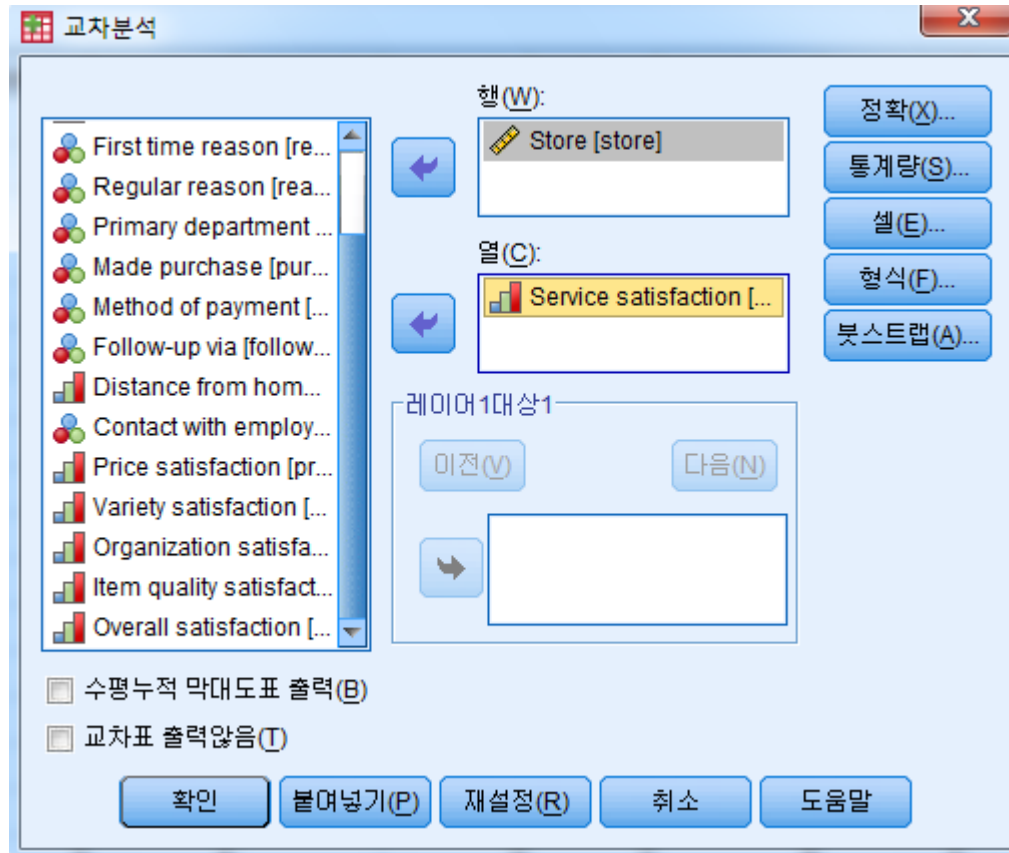
제3장. 교차표 분석 : 동일성 검정

satisf.sav [데이터집합3] - PASW Statistics Data Editor

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 창(W) 도움말(H)

| | 이름 | 유형 | 너비 | 소수점미... | 설명 | 값 | 결측값 | 열 | 맞춤 | 측도 |
|----|----------|----|----|---------|---------------------|------------------|-----|---|-----|-------|
| 1 | gender | 숫자 | 4 | 0 | Gender | {0, Male}... | 없음 | 6 | 오른쪽 | 명목(N) |
| 2 | agecat | 숫자 | 4 | 0 | Age category | {1, 18-24}... | 없음 | 6 | 오른쪽 | 순서(O) |
| 3 | regular | 숫자 | 4 | 0 | Shopping frequ... | {0, First tim... | 없음 | 7 | 오른쪽 | 순서(O) |
| 4 | reason1 | 숫자 | 4 | 0 | First time reason | {1, Disliked ... | 없음 | 7 | 오른쪽 | 명목(N) |
| 5 | reason2 | 숫자 | 4 | 0 | Regular reason | {1, Prices}... | 없음 | 7 | 오른쪽 | 명목(N) |
| 6 | dept | 숫자 | 4 | 0 | Primary depart... | {1, Automoti... | 없음 | 6 | 오른쪽 | 명목(N) |
| 7 | purchase | 숫자 | 4 | 0 | Made purchase | {0, No}... | 없음 | 8 | 오른쪽 | 명목(N) |
| 8 | payment | 숫자 | 4 | 0 | Method of pay... | {1, Cash}... | 없음 | 7 | 오른쪽 | 명목(N) |
| 9 | followup | 숫자 | 4 | 0 | Follow-up via | {1, Mail}... | 없음 | 8 | 오른쪽 | 명목(N) |
| 10 | distance | 숫자 | 4 | 0 | Distance from h... | {1, < 1 mile}... | 없음 | 8 | 오른쪽 | 순서(O) |
| 11 | store | 숫자 | 4 | 0 | Store | {1, Store 1}... | 없음 | 6 | 오른쪽 | 척도(S) |
| 12 | contact | 숫자 | 4 | 0 | Contact with e... | {0, No}... | 없음 | 7 | 오른쪽 | 명목(N) |
| 13 | price | 숫자 | 4 | 0 | Price satisfaction | {1, Strongly ... | 없음 | 6 | 오른쪽 | 순서(O) |
| 14 | numitems | 숫자 | 4 | 0 | Variety satisfac... | {1, Strongly ... | 없음 | 8 | 오른쪽 | 순서(O) |
| 15 | org | 숫자 | 4 | 0 | Organization sa... | {1, Strongly ... | 없음 | 6 | 오른쪽 | 순서(O) |
| 16 | service | 숫자 | 4 | 0 | Service satisfac... | {1, Strongly ... | 없음 | 7 | 오른쪽 | 순서(O) |
| 17 | quality | 숫자 | 4 | 0 | Item quality sat... | {1, Strongly ... | 없음 | 7 | 오른쪽 | 순서(O) |
| 18 | overall | 숫자 | 4 | 0 | Overall satisfac... | {1, Strongly ... | 없음 | 7 | 오른쪽 | 순서(O) |

제3장. 교차표 분석 : 동일성 검정



H0 : 지점에 관계없이 고객들의 서비스 만족도의 분포는 동일하다.

제3장. 교차표 분석 : 동일성 검정

교차분석: 통계량

☒ 카이제곱(H) ☐ 상관관계(R)

명목 데이터

☒ 분할계수(O) 순서

☒ 파이 및 크레이머의 V(P) ☐ 감마(G)

☒ 람다(L) ☐ Somers의 d(S)

☐ 불확실성 계수(U) ☐ Kendall의 타우-b

☐ Kendall의 타우-c

명목 대 등간 척도

☐ 에타(E) ☐ 카파(K)

☐ McNemar

☐ Cochran 및 Mantel-Haenszel 통계량

검정 공통승산비가 동일(T): 1

계속 취소 도움말

교차분석: 셀 출력

빈도(T)

☒ 관측빈도(O) ☐ 기대빈도(E)

퍼센트

☒ 행(R) 잔차

☐ 열(C) ☐ 비표준화(U)

☐ 전체(T) ☐ 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

제3장. 교차표 분석 : 동일성 검정

Store * Service satisfaction 교차표

| | | | Service satisfaction | | | | | 전체 |
|-------|-----------|-----------|----------------------|-------------------|---------|-------------------|-------------------|--------|
| | | | Strongly Negative | Somewhat Negative | Neutral | Somewhat Positive | Strongly Positive | |
| Store | Store 1 | 빈도 | 25 | 20 | 38 | 30 | 33 | 146 |
| | | Store 중 % | 17.1% | 13.7% | 26.0% | 20.5% | 22.6% | 100.0% |
| | Store 2 | 빈도 | 26 | 30 | 34 | 27 | 19 | 136 |
| | | Store 중 % | 19.1% | 22.1% | 25.0% | 19.9% | 14.0% | 100.0% |
| | Store 3 | 빈도 | 15 | 20 | 41 | 33 | 29 | 138 |
| | | Store 중 % | 10.9% | 14.5% | 29.7% | 23.9% | 21.0% | 100.0% |
| | Store 4 | 빈도 | 27 | 35 | 44 | 22 | 34 | 162 |
| | | Store 중 % | 16.7% | 21.6% | 27.2% | 13.6% | 21.0% | 100.0% |
| 전체 | 빈도 | 93 | 105 | 157 | 112 | 115 | 582 | |
| | Store 중 % | 16.0% | 18.0% | 27.0% | 19.2% | 19.8% | 100.0% | |

카이제곱 검정

| | 값 | 자유도 | 점근 유의확률 (양측검정) |
|--------------|---------------------|-----|---|
| Pearson 카이제곱 | 16.293 ^a | 12 |  |
| 우도비 | 17.012 | 12 | .149 |
| 선형 대 선형결합 | .084 | 1 | .772 |
| 유효 케이스 수 | 582 | | |

a. 0 셀 (.0%)은(는) 5보다 작은 기대 빈도를 가지는 셀입니다. 최소 기대빈도는 21.73입니다.

제3장. 교차표 분석 : 동일성 검정

방향성 측도

| | | | 값 | 점근 표준오차 ^a | 근사 T 값 ^b | 근사 유의확률 |
|-------------|---------------------|-------------------------|------|-------------------------|---------------------|-------------------|
| 명목척도 대 명목척도 | 람다 | 대칭적 | .013 | .009 | 1.486 | .137 |
| | | Store 종속 | .026 | .017 | 1.486 | .137 |
| | | Service satisfaction 종속 | .000 | .000 | . ^c | . ^c |
| | Goodman과 Kruskal 타우 | Store 종속 | .009 | .004 | | .183 ^d |
| | | Service satisfaction 종속 | .007 | .003 | | .226 ^d |

a. 영가설을 가정하지 않음.

b. 영가설을 가정하는 점근 표준오차 사용

c. 점근 표준오차가 0이므로 계산할 수 없습니다.

d. 카이제곱 근사법을 기준으로

대칭적 측도

| | | 값 | 근사 유의확률 |
|-------------|-----------|------|---------|
| 명목척도 대 명목척도 | 파이 | .167 | .178 |
| | Cramer의 V | .097 | .178 |
| | 분할계수 | .165 | .178 |
| 유효 케이스 수 | | 582 | |

제3장. 교차표 분석 : 동일성 검정

*freq.sav [데이터집합4] - PASW Statistics Data Editor

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 창(W) 도움말(H)

1 : store 1

| | store | service | freq | 변수 | 변수 | 변수 | 변수 | 변수 | 변수 | 변수 |
|----|-------|---------|------|----|----|----|----|----|----|----|
| 1 | 1 | 1 | 25 | | | | | | | |
| 2 | 1 | 2 | 20 | | | | | | | |
| 3 | 1 | 3 | 38 | | | | | | | |
| 4 | 1 | 4 | 30 | | | | | | | |
| 5 | 1 | 5 | 33 | | | | | | | |
| 6 | 2 | 1 | 26 | | | | | | | |
| 7 | 2 | 2 | 30 | | | | | | | |
| 8 | 2 | 3 | 34 | | | | | | | |
| 9 | 2 | 4 | 27 | | | | | | | |
| 10 | 2 | 5 | 19 | | | | | | | |
| 11 | 3 | 1 | 15 | | | | | | | |
| 12 | 3 | 2 | 20 | | | | | | | |
| 13 | 3 | 3 | 41 | | | | | | | |
| 14 | 3 | 4 | 33 | | | | | | | |
| 15 | 3 | 5 | 29 | | | | | | | |
| 16 | 4 | 1 | 27 | | | | | | | |
| 17 | 4 | 2 | 35 | | | | | | | |
| 18 | 4 | 3 | 44 | | | | | | | |
| 19 | 4 | 4 | 22 | | | | | | | |
| 20 | 4 | 5 | 34 | | | | | | | |
| 21 | | | | | | | | | | |

가중 케이스

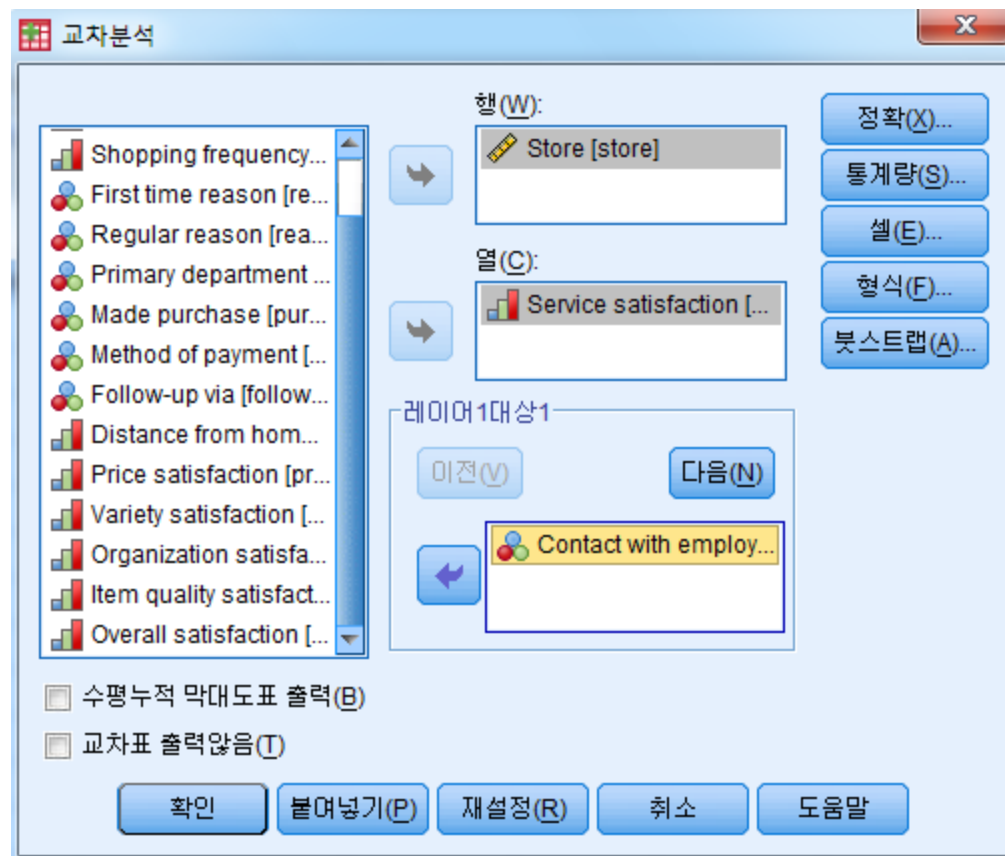
☐ 가중 케이스 사용않음(D)
☒ 가중 케이스 지정(W)

빈도변수(F): freq

현재 상태:가중케이스 지정freq

확인 붙여넣기(P) 재설정(R) 취소 도움말

제3장. 교차표 분석 : 동일성 검정



제3장. 교차표 분석 : 동일성 검정

Store * Service satisfaction * Contact with employee 교차표

| Contact with employee | | | | Service satisfaction | | | | | 전체 |
|-----------------------|---------|----|-----------|----------------------|-------------------|---------|-------------------|-------------------|--------|
| | | | | Strongly Negative | Somewhat Negative | Neutral | Somewhat Positive | Strongly Positive | |
| No | Store 1 | 빈도 | 빈도 | 16 | 9 | 18 | 17 | 19 | 79 |
| | | | Store 중 % | 20.3% | 11.4% | 22.8% | 21.5% | 24.1% | 100.0% |
| | Store 2 | 빈도 | 빈도 | 2 | 15 | 16 | 13 | 12 | 58 |
| | | | Store 중 % | 3.4% | 25.9% | 27.6% | 22.4% | 20.7% | 100.0% |
| | Store 3 | 빈도 | 빈도 | 9 | 14 | 23 | 22 | 14 | 82 |
| | | | Store 중 % | 11.0% | 17.1% | 28.0% | 26.8% | 17.1% | 100.0% |
| | Store 4 | 빈도 | 빈도 | 17 | 14 | 19 | 10 | 10 | 70 |
| | | | Store 중 % | 24.3% | 20.0% | 27.1% | 14.3% | 14.3% | 100.0% |
| | 전체 | 빈도 | 빈도 | 44 | 52 | 76 | 62 | 55 | 289 |
| | | | Store 중 % | 15.2% | 18.0% | 26.3% | 21.5% | 19.0% | 100.0% |
| Yes | Store 1 | 빈도 | 빈도 | 9 | 11 | 20 | 13 | 14 | 67 |
| | | | Store 중 % | 13.4% | 16.4% | 29.9% | 19.4% | 20.9% | 100.0% |
| | Store 2 | 빈도 | 빈도 | 24 | 15 | 18 | 14 | 7 | 78 |
| | | | Store 중 % | 30.8% | 19.2% | 23.1% | 17.9% | 9.0% | 100.0% |
| | Store 3 | 빈도 | 빈도 | 6 | 6 | 18 | 11 | 15 | 56 |
| | | | Store 중 % | 10.7% | 10.7% | 32.1% | 19.6% | 26.8% | 100.0% |
| | Store 4 | 빈도 | 빈도 | 10 | 21 | 25 | 12 | 24 | 92 |
| | | | Store 중 % | 10.9% | 22.8% | 27.2% | 13.0% | 26.1% | 100.0% |
| | 전체 | 빈도 | 빈도 | 49 | 53 | 81 | 50 | 60 | 293 |
| | | | Store 중 % | 16.7% | 18.1% | 27.6% | 17.1% | 20.5% | 100.0% |

카이제곱 검정

| Contact with employee | 값 | 자유도 | 점근유의확률 (양측검정) | |
|-----------------------|--------------|---------------------|------------------|------|
| No | Pearson 카이제곱 | 20.898 ^a | 12 | .052 |
| | 우도비 | 22.937 | 12 | .028 |
| | 선형 대 선형결합 | 3.514 | 1 | .061 |
| | 유효 케이스 수 | 289 | | |
| Yes | Pearson 카이제곱 | 25.726 ^b | 12 | .012 |
| | 우도비 | 25.777 | 12 | .012 |
| | 선형 대 선형결합 | 1.993 | 1 | .158 |
| | 유효 케이스 수 | 293 | | |



제3장. 교차표 분석 : 동일성 검정

방향성 속도

| Contact with employee | | | | 값 | 점근 표준오차 ^a | 근사 T 값 ^b | 근사 유의확률 |
|-----------------------|---------------------|----|-------------------------|------|-------------------------|---------------------|-------------------|
| No | 명목척도 대 명목척도 | 람다 | 대칭적 | .036 | .030 | 1.178 | .239 |
| | | | Store 종속 | .068 | .044 | 1.498 | .134 |
| | | | Service satisfaction 종속 | .005 | .028 | .164 | .869 |
| | Goodman과 Kruskal 타우 | | Store 종속 | .023 | .009 | | .067 ^c |
| | | | Service satisfaction 종속 | .016 | .006 | | .112 ^c |
| | | | | | | | |
| Yes | 명목척도 대 명목척도 | 람다 | 대칭적 | .053 | .029 | 1.806 | .071 |
| | | | Store 종속 | .080 | .037 | 2.081 | .037 |
| | | | Service satisfaction 종속 | .028 | .030 | .927 | .354 |
| | Goodman과 Kruskal 타우 | | Store 종속 | .031 | .012 | | .007 ^c |
| | | | Service satisfaction 종속 | .020 | .008 | | .021 ^c |
| | | | | | | | |

a. 영가설을 가정하지 않음.

b. 영가설을 가정하는 점근 표준오차 사용

c. 카이제곱 근사법을 기준으로

대칭적 속도

| Contact with employee | | | | 값 | 근사 유의확률 |
|-----------------------|-------------|-------------|----|------|---------|
| No | 명목척도 대 명목척도 | 파이 | | .269 | .052 |
| | | Cramer의 V | | .155 | .052 |
| | | 분할계수 | | .260 | .052 |
| | | 유효 케이스 수 | | 289 | |
| | Yes | 명목척도 대 명목척도 | 파이 | | .296 |
| | | Cramer의 V | | .171 | .012 |
| | | 분할계수 | | .284 | .012 |
| | | 유효 케이스 수 | | 293 | |

제3장. 교차표 분석-행 간 평균 차이(열이 순서형)

- SPSS사례분석-데이터(CAT06.sav)

한 정화시설 처리장에서 세가지 처리(행:무처리(placebo), 표준처리(standard), 새로운 처리(test))에 있어서 물이 깨끗해지는 정도(열: 물이 깨끗해지는 시간 0, 1, 2, 3, 4시간, 즉 몇 시간이 지나야 물이 깨끗해지는가?)를 77개의 물의 샘플을 대상으로 조사하였다.
세 처리집단의 분포가 같다고 할 수 있는가?

| 처리 완화시간 \ | 무처리 | 표준처리 | 새로운처리 |
|--------------|-----|------|-------|
| 0 | 6 | 1 | 2 |
| 1 | 9 | 4 | 5 |
| 2 | 6 | 6 | 6 |
| 3 | 3 | 6 | 8 |
| 4 | 1 | 8 | 6 |

귀무가설: 세 처리 집단의 분포가 같다.

대립가설: 세 처리 집단의 분포가 같지 않다.

제3장. 교차표 분석-행 간 평균 차이(열이 순서형)

- SPSS사례분석-데이터(CAT06.sav)

*cat06.sav [데이터집합6] - PASW Statistics Data Editor

| 1 : treat | treat | relief | cou |
|-----------|-------|--------|-----|
| 1 | 1.00 | 1.00 | |
| 2 | 1.00 | 2.00 | |
| 3 | 1.00 | 3.00 | |
| 4 | 1.00 | 4.00 | |
| 5 | 1.00 | 5.00 | |
| 6 | 2.00 | 1.00 | |
| 7 | 2.00 | 2.00 | |
| 8 | 2.00 | 3.00 | |
| 9 | 2.00 | 4.00 | |
| 10 | 2.00 | 5.00 | |
| 11 | 3.00 | 1.00 | |
| 12 | 3.00 | 2.00 | |
| 13 | 3.00 | 3.00 | |
| 14 | 3.00 | 4.00 | |
| 15 | 3.00 | 5.00 | |
| 16 | | | |
| 17 | | | |
| 18 | | | |
| 19 | | | |
| 20 | | | |
| 21 | | | |

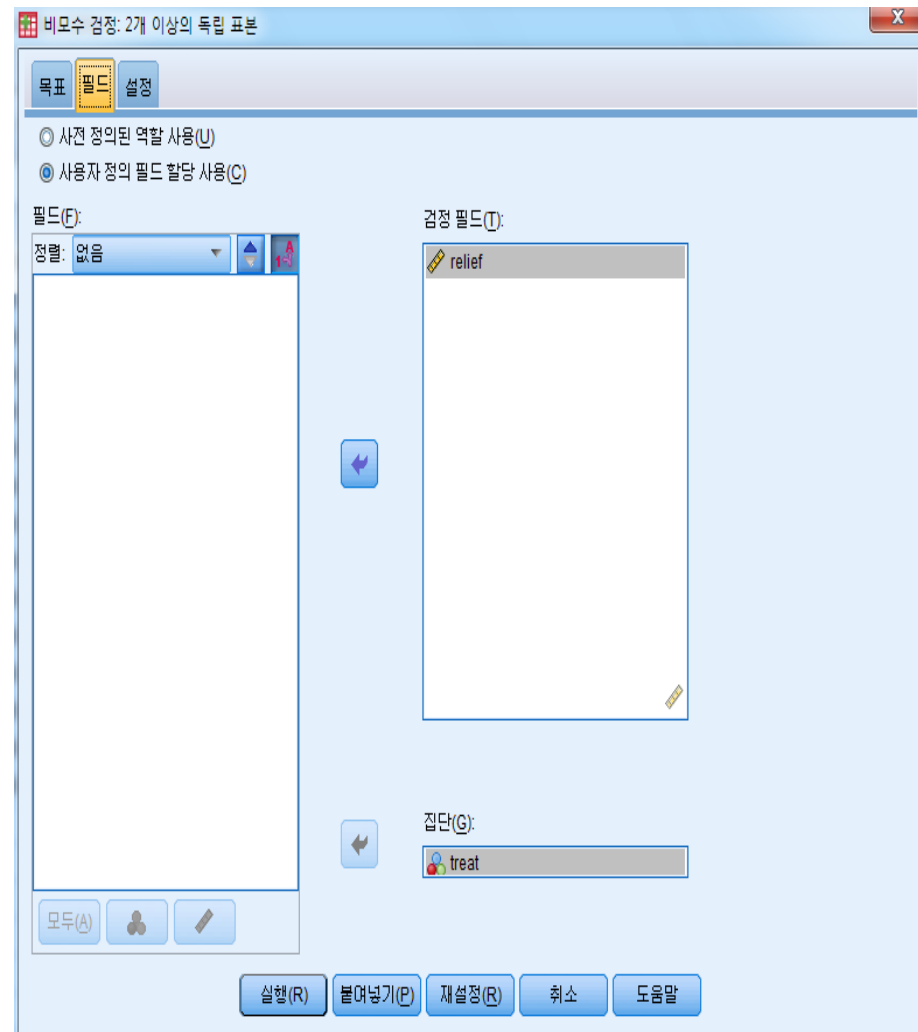
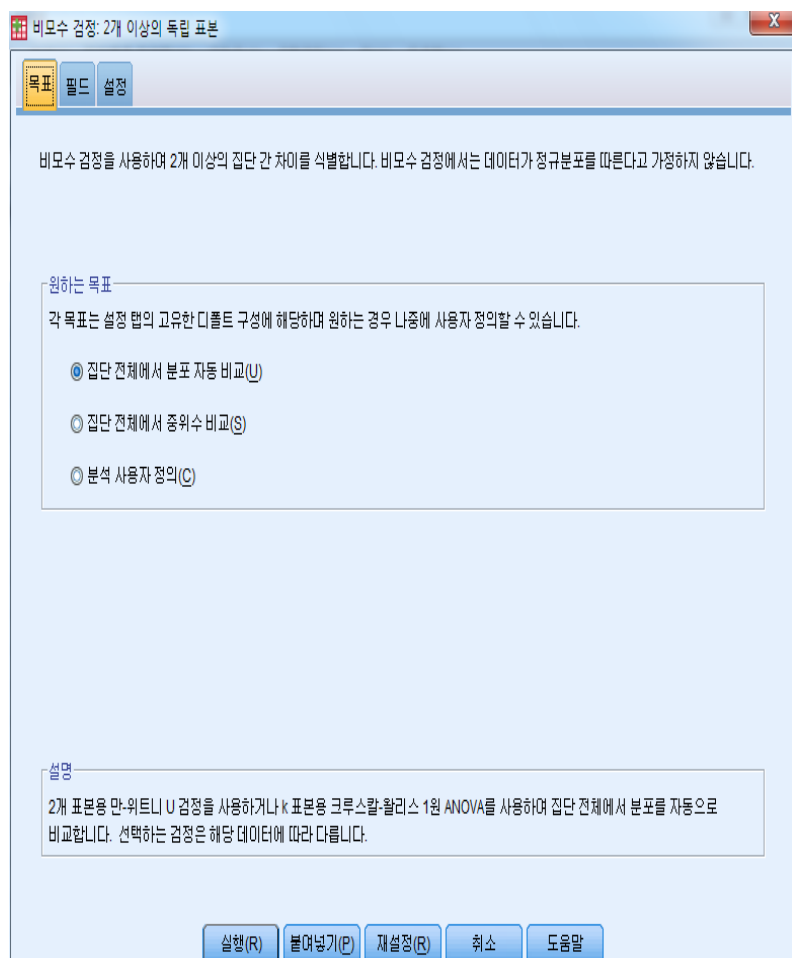
분석(A) 메뉴:

- 보고서(P)
- 기술통계량(E)
- 표
- 평균 비교(M)
- 일반선형모형(G)
- 일반화 선형 모형(Z)
- 혼합 모형(X)
- 상관분석(C)
- 회귀분석(R)
- 로그선형분석(O)
- 신경망(W)
- 분류분석(Y)
- 차원 감소(D)
- 척도(A)
- 비모수 검정(N)**
 - 일표본(O)...
 - 독립 표본(I)...**
 - 대응 표본(R)...
 - 레거시 대화 상자(L) ▶
- 예측(T)
- 생존확률(S)
- 다중응답(U)
- 결측값 분석(V)...
- 다중 대입(T)
- 복합 표본(L)
- 품질 관리(Q)
- ROC 곡선(V)...

| | 이름 | 유형 | 너비 | 소수점이... | 설명 | 값 | 결측값 | 열 | 맞춤 | 측도 |
|---|--------|----|----|---------|----|----|-----|---|-------|-------|
| 1 | treat | 숫자 | 8 | 2 | | 없음 | 없음 | 8 | 클 오른쪽 | 도(S) |
| 2 | relief | 숫자 | 8 | 2 | | 없음 | 없음 | 8 | 클 오른쪽 | 척도(S) |
| 3 | count | 숫자 | 8 | 2 | | 없음 | 없음 | 8 | 클 오른쪽 | 순서(O) |
| 4 | | | | | | | | | | 명목(N) |
| 5 | | | | | | | | | | |

제3장. 교차표 분석-행 간 평균 차이(열이 순서형)

- SPSS사례분석-데이터(CAT06.sav)



제3장. 교차표 분석-행 간 평균 차이(열이 순서형)

• 출력결과



| treat * relief 교차표 | | | | | | |
|--------------------|------|--------|------|------|------|------|
| 빈도 | | relief | | | | |
| | | 1.00 | 2.00 | 3.00 | 4.00 | 5.00 |
| treat | 1.00 | 6 | 9 | 6 | 3 | 1 |
| | 2.00 | 1 | 4 | 6 | 6 | 8 |
| | 3.00 | 2 | 5 | 6 | 8 | 6 |
| 전체 | | 9 | 18 | 18 | 17 | 15 |
| | 전체 | 25 | 25 | 27 | 27 | 77 |

| 카이제곱 검정 | | | |
|--------------|---------------------|-----|----------------|
| | 값 | 자유도 | 점근 유의확률 (양측검정) |
| Pearson 카이제곱 | 14.593 ^a | 8 | .068 |
| 우도비 | 15.504 | 8 | .050 |
| 선형 대 선형결합 | 8.067 | 1 | .005 |
| 유효 케이스 수 | 77 | | |

a. 5 셀 (33.3%)은 (는) 5보다 작은 기대 빈도를 가지는 셀입니다. 최소 기대빈도는 2.92입니다.



Kruskal-Wallis 검정

순위

| | treat | N | 평균순위 |
|--------|-------|----|-------|
| relief | 1.00 | 25 | 25.90 |
| | 2.00 | 25 | 47.28 |
| | 3.00 | 27 | 43.46 |
| | 합계 | 77 | |

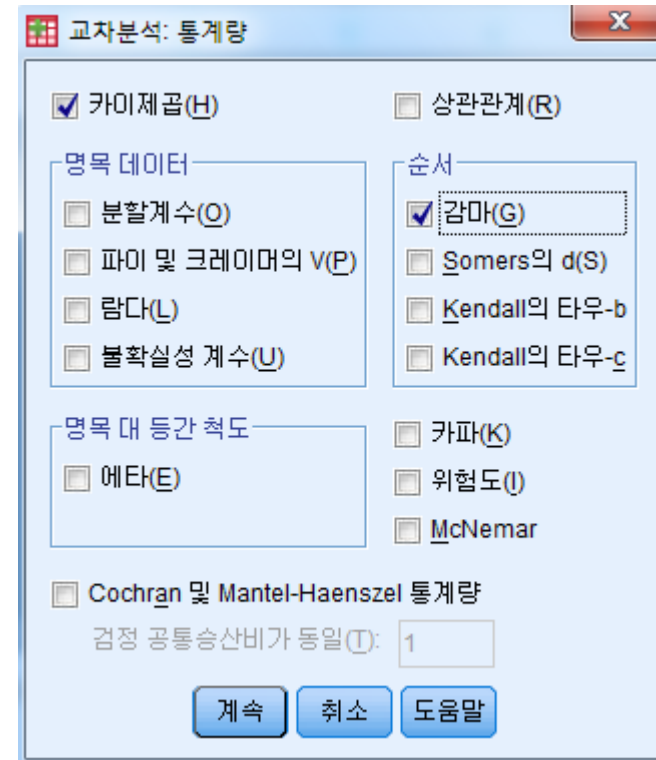
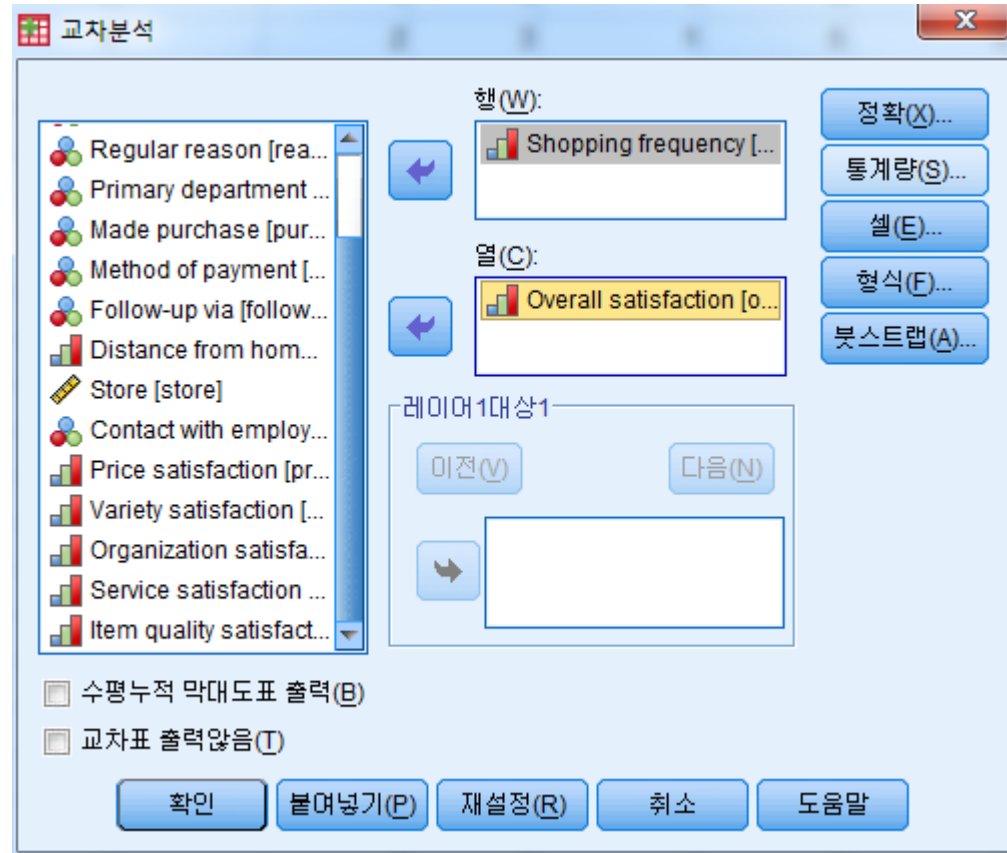
검정 통계량^{a,b}

| | relief |
|---------|--------|
| 카이제곱 | 13.689 |
| 자유도 | 2 |
| 근사 유의확률 | .001 |

a. Kruskal Wallis 검정

b. 집단변수: treat

제3장. 교차표 분석-독립성 검정



제3장. 교차표 분석-독립성 검정

Shopping frequency * Overall satisfaction 교차표

빈도

| | | Overall satisfaction | | | | | 전체 |
|--------------------|------------|----------------------|-------------------|---------|-------------------|-------------------|-----|
| | | Strongly Negative | Somewhat Negative | Neutral | Somewhat Positive | Strongly Positive | |
| Shopping frequency | First time | 5 | 13 | 15 | 14 | 5 | 52 |
| | < 1/month | 26 | 38 | 39 | 34 | 16 | 153 |
| | 1/month | 27 | 43 | 46 | 55 | 30 | 201 |
| | 1/week | 7 | 36 | 33 | 40 | 26 | 142 |
| | > 1/week | 1 | 8 | 10 | 5 | 10 | 34 |
| 전체 | | 66 | 138 | 143 | 148 | 87 | 582 |

카이제곱 검정

| | 값 | 자유도 | 점근 유의확률 (양측검정) |
|--------------|---------------------|-----|-------------------|
| Pearson 카이제곱 | 26.121 ^a | 16 | .052 |
| 우도비 | 27.320 | 16 | .038 |
| 선형 대 선형결합 | 9.873 | 1 | .002 |
| 유효 케이스 수 | 582 | | |

a. 1 셀 (4.0%)은(는) 5보다 작은 기대 빈도를 가지는
셀입니다. 최소 기대빈도는 3.86입니다.

대칭적 측도

| | 값 | 점근 표준오차 ^a | 근사 T 값 ^b | 근사 유의확률 |
|-------------------|------|-------------------------|---------------------|---------|
| 순서척도 대 순서척도 감마 | .140 | .043 | 3.267 | .001 |
| 유효 케이스 수 | 582 | | | |

제3장. 교차표 분석-상관계수(행과 열 모두 순서형)

- SPSS 사례분석- 데이터(CAT07.sav)

The screenshot shows the SPSS Statistics Data Editor window with the file name *cat07.sav. The 'Analyze' menu is open, and the path 'Analyze > Descriptive Statistics > Crosstabs > Chi-Square' is highlighted. The data table has columns 't', 'wash', and 'cou'. The 't' column has values 1.00, 2.00, and 3.00. The 'wash' column has values 1.00, 2.00, and 3.00. The 'cou' column has values 1.00, 2.00, and 3.00.

메뉴 경로: 분석(A) > 기술통계량(E) > 표 > 평균 비교(M) > 일반선행모형(G) > 일반화 선행 모형(Z) > 혼합 모형(X) > 상관분석(C) > 회귀분석(R) > 로그선행분석(O) > 신경망(W) > 분류분석(Y) > 차원 감소(D) > 척도(A) > 비모수 검정(N) > 예측(T) > 생존확률(S) > 다중응답(U) > 결측값 분석(V)... > 다중 대입(T) > 복합 표본(L) > 품질 관리(Q) > ROC 곡선(V)...

하위 메뉴: 레거시 대화 상자(L) > 카이제곱검정(C)...

데이터 테이블:

| | t | wash | cou |
|----|------|------|------|
| 1 | 1.00 | 1.00 | 1.00 |
| 2 | 1.00 | 2.00 | 1.00 |
| 3 | 1.00 | 3.00 | 1.00 |
| 4 | 2.00 | 1.00 | 1.00 |
| 5 | 2.00 | 2.00 | 1.00 |
| 6 | 2.00 | 3.00 | 1.00 |
| 7 | 3.00 | 1.00 | 1.00 |
| 8 | 3.00 | 2.00 | 1.00 |
| 9 | 3.00 | 3.00 | 1.00 |
| 10 | | | |
| 11 | | | |
| 12 | | | |
| 13 | | | |
| 14 | | | |
| 15 | | | |
| 16 | | | |
| 17 | | | |
| 18 | | | |
| 19 | | | |
| 20 | | | |
| 21 | | | |
| 22 | | | |
| 23 | | | |
| 24 | | | |
| 25 | | | |

제3장. 교차표 분석-상관계수(행과 열 모두 순서형)

- SPSS 사례분석- 데이터(CAT07.sav)

*cat07.sav [데이터집합3] - PASW Statistics Data Editor

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 창(W) 도움말(H)

| | trt | wash | count | 변수 | 변수 | 변수 | 변수 | 변수 | 변수 |
|----|------|------|-------|----|----|----|----|----|----|
| 1 | 1.00 | 1.00 | 27.00 | | | | | | |
| 2 | 1.00 | 2.00 | 14.00 | | | | | | |
| 3 | 1.00 | 3.00 | 5.00 | | | | | | |
| 4 | 2.00 | 1.00 | 10.00 | | | | | | |
| 5 | 2.00 | 2.00 | 17.00 | | | | | | |
| 6 | 2.00 | 3.00 | 26.00 | | | | | | |
| 7 | 3.00 | 1.00 | 5.00 | | | | | | |
| 8 | 3.00 | 2.00 | 12.00 | | | | | | |
| 9 | 3.00 | 3.00 | 50.00 | | | | | | |
| 10 | | | | | | | | | |
| 11 | | | | | | | | | |
| 12 | | | | | | | | | |
| 13 | | | | | | | | | |
| 14 | | | | | | | | | |
| 15 | | | | | | | | | |
| 16 | | | | | | | | | |
| 17 | | | | | | | | | |
| 18 | | | | | | | | | |
| 19 | | | | | | | | | |
| 20 | | | | | | | | | |

독립 K-표본 비모수검정

검정변수(T): wash

집단변수(G): trt(1 3)

범위 지정(D)...

검정 유형

☐ Kruskal-Wallis의 H(K) ☐ 중위수(M)

☒ Jonckheere-Terpstra

확인 불여넣기(P) 재설정(R) 취소 도움말

제3장. 교차표 분석-상관계수(행과 열 모두 순서형)

• 출력결과

| trt * wash 교차표 | | | | | |
|----------------|---|------|----|----|-----|
| 빈도 | | | | | |
| | | wash | | | 전체 |
| | | 1 | 2 | 3 | |
| trt | 1 | 27 | 14 | 5 | 46 |
| | 2 | 10 | 17 | 26 | 53 |
| | 3 | 5 | 12 | 50 | 67 |
| 전체 | | 42 | 43 | 81 | 166 |

| 카이제곱 검정 | | | |
|--------------|---------------------|-----|-------------------|
| | 값 | 자유도 | 점근 유의확률 (양측검정) |
| Pearson 카이제곱 | 55.088 ^a | 4 | .000 |
| 우도비 | 58.037 | 4 | .000 |
| 선형 대 선형결합 | 50.602 | 1 | .000 |
| 유효 케이스 수 | 166 | | |

a. 0 셀 (.0%)은(는) 5보다 작은 기대 빈도를 가지는 셀입니다. 최소 기대빈도는 11.64입니다.

| 비모수 검정 | |
|-------------------------------------|----------|
| Jonckheere-Terpstra 검정 ^a | |
| | wash |
| trt의 수준의 수 | 3 |
| N | 166 |
| 관측된 J-T 통계량 | 6741.500 |
| 평균 J-T 통계량 | 4535.500 |
| J-T 통계량의 표준편차 | 309.369 |
| 표준화 J-T 통계량 | 7.131 |
| 근사 유의확률(양측) | .000 |

a. 집단변수: trt

제3장. 교차표 분석-정확 검정 (exact test)

□ SPSS 사례분석 - 데이터(CAT03.sav)

*CAT03.sav [데이터집합4] - PASW Statistics Data Editor

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 창(W) 도움말(H)

1 : treat placebo

| | treat | favorun | count | 변수 | 변수 | 변수 | 변수 | 변수 | 변수 | 변수 | 변수 |
|---|---------|---------|-------|----|----|----|----|----|----|----|----|
| 1 | placebo | favor | 10 | | | | | | | | |
| 2 | placebo | unfavor | 2 | | | | | | | | |
| 3 | test | favor | 2 | | | | | | | | |
| 4 | test | unfavor | 4 | | | | | | | | |
| 5 | | | | | | | | | | | |
| 6 | | | | | | | | | | | |

교차분석: 통계량

☒ 카이제곱(H) ☐ 상관관계(R)

명목 데이터

- ☐ 분할계수(O)
- ☐ 파이 및 크레이머의 V(P)
- ☐ 람다(L)
- ☐ 불확실성 계수(U)

순서

- ☐ 감마(G)
- ☐ Somers의 d(S)
- ☐ Kendall의 타우-b
- ☐ Kendall의 타우-c

명목 대 등간 척도

- ☐ 에타(E)
- ☐ 카파(K)
- ☐ 위험도(I)
- ☐ McNemar

☐ Cochran 및 Mantel-Haenszel 통계량

검정 공통승산비가 동일(T): 1

계속 취소 도움말

교차분석

행(W): treat

열(C): favorun

레이어1대상1

이전(V) 다음(N)

수평누적 막대도표 출력(B)

교차표 출력없음(T)

확인 붙여넣기(P) 재설정(R) 취소 도움말

제3장. 교차표 분석-정확 검정 (exact test)

• 출력결과

케이스 처리 요약

| | 케이스 | | | | | |
|-----------------|-----|--------|----|-----|----|--------|
| | 유효 | | 결측 | | 전체 | |
| | N | 퍼센트 | N | 퍼센트 | N | 퍼센트 |
| treat * favorun | 18 | 100.0% | 0 | .0% | 18 | 100.0% |

treat * favorun 교차표

빈도

→

| | | favorun | | 전체 |
|-------|---------|---------|---------|----|
| | | favor | unfavor | |
| treat | placebo | 10 | 2 | 12 |
| | test | 2 | 4 | 6 |
| 전체 | | 12 | 6 | 18 |

카이제곱 검정

| | 값 | 자유도 | 점근 유의 확률 (양측검정) | 정확한 유의 확률 (양측검정) | 정확한 유의 확률 (단측검정) |
|-------------------|--------------------|-----|--------------------|------------------------|------------------------|
| Pearson 카이제곱 | 4.500 ^a | 1 | .034 | | |
| 연속수정 ^b | 2.531 | 1 | .112 | | |
| 우도비 | 4.463 | 1 | .035 | | |
| Fisher의 정확한 검정 | | | | .107 | .057 |
| 유효 케이스 수 | 18 | | | | |

a. 3 셀 (75.0%)은(는) 5보다 작은 기대 빈도를 가지는 셀입니다. 최소 기대빈도는 2.00입니다.

b. 2x2 표에 대해서만 계산됨



제3장. 교차표 분석-연관성 측도

- 파이 계수 (Phi Coefficient)

$$\phi = \frac{n_{11} n_{22} - n_{12} n_{21}}{\sqrt{n_{1.} n_{2.} n_{.1} n_{.2}}} \quad \text{for } 2 \times 2 \text{ tables} \quad -1 \leq \phi \leq 1$$

$$\phi = \sqrt{Q_P/n} \quad \text{otherwise} \quad 0 \leq \phi \leq \min(\sqrt{R-1}, \sqrt{C-1})$$

- 크래머 V (Cramer's V)

$$V = \phi \quad \text{for } 2 \times 2 \text{ tables} \quad -1 \leq V \leq 1$$

$$V = \sqrt{\frac{Q_P/n}{\min(R-1, C-1)}} \quad \text{otherwise} \quad 0 \leq V \leq 1$$

- 분할계수 (Contingency Coefficient)

$$P = \sqrt{\frac{Q_P}{Q_P + n}} \quad 0 \leq P \leq \sqrt{(m-1)/m}, \text{ where } m = \min(R, C)$$

제3장. 교차표 분석-연관성 측도 : 람다 (lambda)

✓ 2차원 분할표에서 행과 열의 결합정도를 수량화

- $\lambda(C|R)$: 명목형 행과 열에 적용.
- 행 수준의 유무에 따라 열에 대한 "예측오류감소"계산.

(예) 1987년 대통령 선거

| 5대 도시 | 노태우 | 김영삼 | 김대중 | 김종필 | 합계 |
|-------|------|------|------|-----|-------|
| 서울 | 1683 | 1637 | 1833 | 461 | 5614 |
| 부산 | 641 | 1117 | 182 | 52 | 1992 |
| 대구 | 800 | 275 | 30 | 23 | 1128 |
| 인천 | 326 | 249 | 177 | 76 | 828 |
| 광주 | 23 | 2 | 450 | 1 | 476 |
| 합계 | 3473 | 3280 | 2672 | 613 | 10038 |

제3장. 교차표 분석-연관성 측도 : 람다 (lambda)

(1) 행에 대한 정보를 없애면

| 5대 도시 | 노태우 | 김영삼 | 김대중 | 김종필 |
|-------|------|------|------|-----|
| 합계 | 3473 | 3280 | 2672 | 613 |

$$\text{최소 예측오류수} = 0 + 3280 + 2672 + 613 = 6565$$

(2) 행에 대한 정보를 활용하면 최소 예측오류수 = 5512

| 5대 도시 | 노태우 | 김영삼 | 김대중 | 김종필 |
|-------|------|------|------|-----|
| 서울 | 1683 | 1637 | 1833 | 461 |
| 부산 | 641 | 1117 | 182 | 52 |
| 대구 | 800 | 275 | 30 | 23 |
| 인천 | 326 | 249 | 177 | 76 |
| 광주 | 23 | 2 | 450 | 1 |

따라서 오류감소 비

$$= 1 - 5512/6565 = 0.16 = \lambda(C|R) : \text{정의, } 0 \leq \lambda(C|R) \leq 1$$

제3장. 교차표 분석-연관성 측도 : 람다 (lambda)

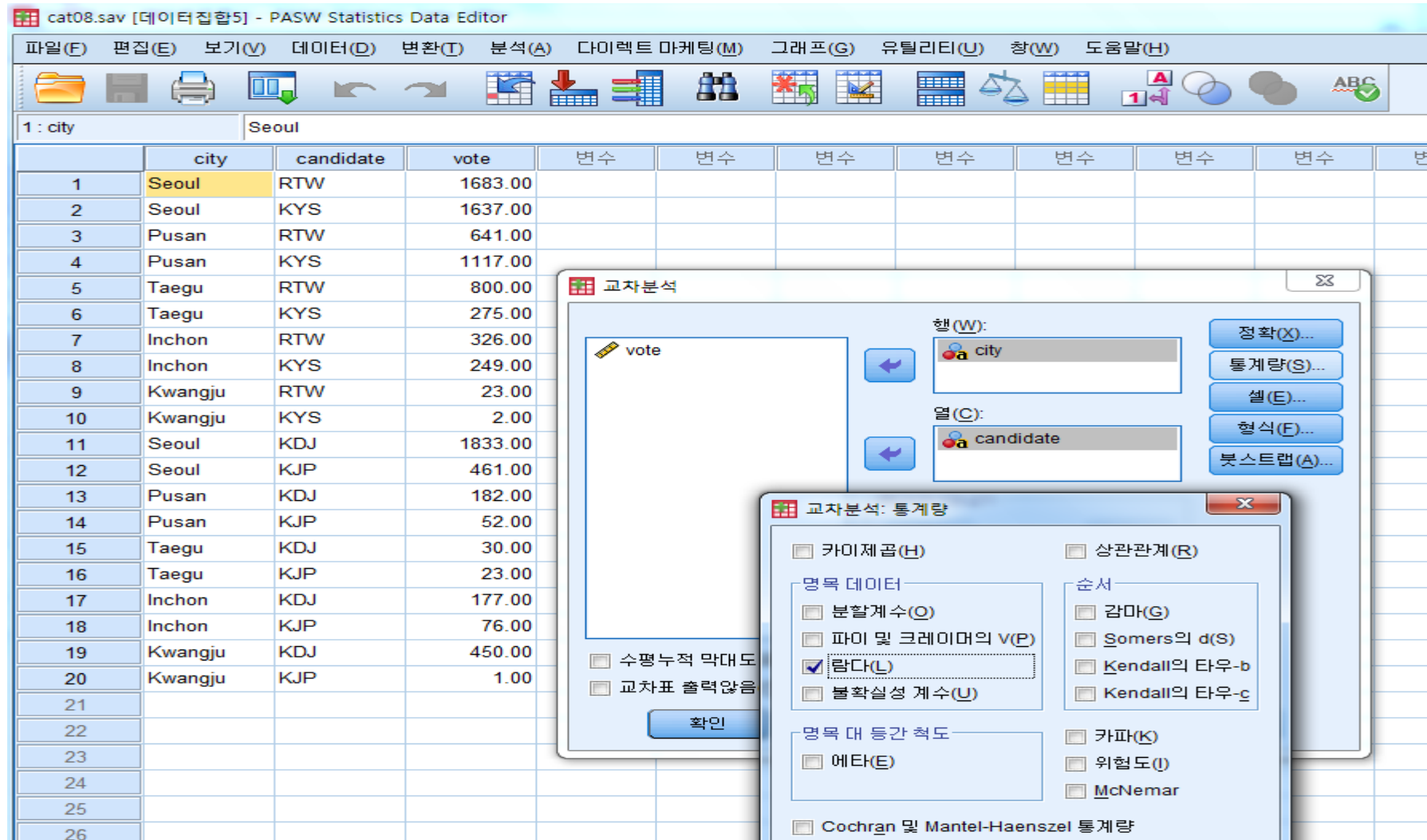
- 분할표에서 두 명목형 변수 사이의 연관성을 재는 측도. ($0 \leq \lambda \leq 1$)

$$\text{람다} = 1 - \frac{\text{행정보를 이용할 때 열정보 예측에서의 오류수}}{\text{행정보를 무시할 때 열정보 예측에서의 오류수}}$$

- $\lambda=1$
 - 행의 정보를 이용할 때 열의 정보를 예측함에 전혀 오류가 없다.
 - 행변수는 열 변수를 예측하는데 중요한 요인이다.
 - 두 변수간 큰 연관관계가 있다.
- $\lambda=0$
 - 열에 대한 정보를 예측하는데 행의 정보가 아무런 역할이 없다.
 - 두 변수간에 연관관계가 없다.

제3장. 교차표 분석-연관성 측도 : 람다 (lambda)

- SPSS 사례분석 - 데이터(CAT08.sav)



cat08.sav [데이터집합5] - PASW Statistics Data Editor

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 창(W) 도움말(H)

1 : city Seoul

| | city | candidate | vote | 변수 | 변수 | 변수 | 변수 | 변수 | 변수 | 변수 | 변수 |
|----|---------|-----------|---------|----|----|----|----|----|----|----|----|
| 1 | Seoul | RTW | 1683.00 | | | | | | | | |
| 2 | Seoul | KYS | 1637.00 | | | | | | | | |
| 3 | Pusan | RTW | 641.00 | | | | | | | | |
| 4 | Pusan | KYS | 1117.00 | | | | | | | | |
| 5 | Taegu | RTW | 800.00 | | | | | | | | |
| 6 | Taegu | KYS | 275.00 | | | | | | | | |
| 7 | Inchon | RTW | 326.00 | | | | | | | | |
| 8 | Inchon | KYS | 249.00 | | | | | | | | |
| 9 | Kwangju | RTW | 23.00 | | | | | | | | |
| 10 | Kwangju | KYS | 2.00 | | | | | | | | |
| 11 | Seoul | KDJ | 1833.00 | | | | | | | | |
| 12 | Seoul | KJP | 461.00 | | | | | | | | |
| 13 | Pusan | KDJ | 182.00 | | | | | | | | |
| 14 | Pusan | KJP | 52.00 | | | | | | | | |
| 15 | Taegu | KDJ | 30.00 | | | | | | | | |
| 16 | Taegu | KJP | 23.00 | | | | | | | | |
| 17 | Inchon | KDJ | 177.00 | | | | | | | | |
| 18 | Inchon | KJP | 76.00 | | | | | | | | |
| 19 | Kwangju | KDJ | 450.00 | | | | | | | | |
| 20 | Kwangju | KJP | 1.00 | | | | | | | | |
| 21 | | | | | | | | | | | |
| 22 | | | | | | | | | | | |
| 23 | | | | | | | | | | | |
| 24 | | | | | | | | | | | |
| 25 | | | | | | | | | | | |
| 26 | | | | | | | | | | | |

교차분석

행(W): city

열(C): candidate

정확(X)... 통계량(S)... 셀(E)... 형식(F)... 북스트랩(A)...

교차분석: 통계량

☐ 카이제곱(H) ☐ 상관관계(R)

명목 데이터

☐ 분할계수(O) ☐ 순서

☐ 파이 및 크레이머의 V(P) ☐ 감마(G)

☒ 람다(L) ☐ Somers의 d(S)

☐ 불확실성 계수(U) ☐ Kendall의 타우-b

명목 대 등간 척도

☐ 에타(E) ☐ Kendall의 타우-c

☐ 카파(K) ☐ 위험도(I) ☐ McNemar

☐ Cochran 및 Mantel-Haenszel 통계량

확인

제3장. 교차표 분석-연관성 측도 : 람다 (lambda)

• 출력결과

city * candidate 교차표

| | | candidate | | | | 전체 |
|------|---------|-----------|-----|------|------|-------|
| | | KDJ | KJP | KYS | RTW | |
| city | Inchon | 177 | 76 | 249 | 326 | 828 |
| | Kwangju | 450 | 1 | 2 | 23 | 476 |
| | Pusan | 182 | 52 | 1117 | 641 | 1992 |
| | Seoul | 1833 | 461 | 1637 | 1683 | 5614 |
| | Taegu | 30 | 23 | 275 | 800 | 1128 |
| 전체 | | 2672 | 613 | 3280 | 3473 | 10038 |

카이제곱 검정

| | 값 | 자유도 | 점근 유의확률 (양측검정) |
|--------------|-----------------------|-----|-------------------|
| Pearson 카이제곱 | 2660.777 ^a | 12 | .000 |
| 우도비 | 2654.271 | 12 | .000 |
| 유효 케이스 수 | 10038 | | |

a. 0 셀 (.0%)은 (는) 5보다 작은 기대 빈도를 가지는
셀입니다. 최소 기대빈도는 29.07입니다.

방향성 측도

| | | 값 | 점근 표준오차 ^a | 근사 T 값 ^b | 근사 유의확률 |
|---------------------|--------------|------|-------------------------|---------------------|-------------------|
| 명목척도 대 명목척도 | 람다 | .096 | .006 | 14.026 | .000 |
| | 대칭적 | | | | |
| | city 종속 | .000 | .000 | . ^c | . ^c |
| Goodman과 Kruskal 타우 | candidate 종속 | .160 | .011 | 14.026 | .000 |
| | city 종속 | .056 | .003 | | .000 ^d |
| | candidate 종속 | .108 | .003 | | .000 ^d |

제3장. 교차표 분석-연관성 측도 : 감마 (gamma)

γ : 순서형 행과 열에 적용.

행과 열 결합의 일치성·비일치성의 경우 수를 비교.

(예) 지위에 따른 업무 만족도

| 지위 | 만족도 | | 합계 |
|----|-----|----|----|
| | 낮음 | 높음 | |
| 하 | 20 | 10 | 30 |
| 상 | 5 | 25 | 30 |
| 합계 | 25 | 35 | 60 |

지위와 만족도 사이의

일치쌍(concordant pair)의 수 = $20 \times 25 = 500$

비일치쌍(discordant pair)의 수 = $5 \times 10 = 50$

정의 : 감마 =
$$\frac{\text{일치쌍의 수} - \text{비일치쌍의 수}}{\text{일치쌍의 수} + \text{비일치쌍의 수}}$$

= $450/550 = 0.82$

제3장. 교차표 분석-연관성 측도 : 감마 (gamma)

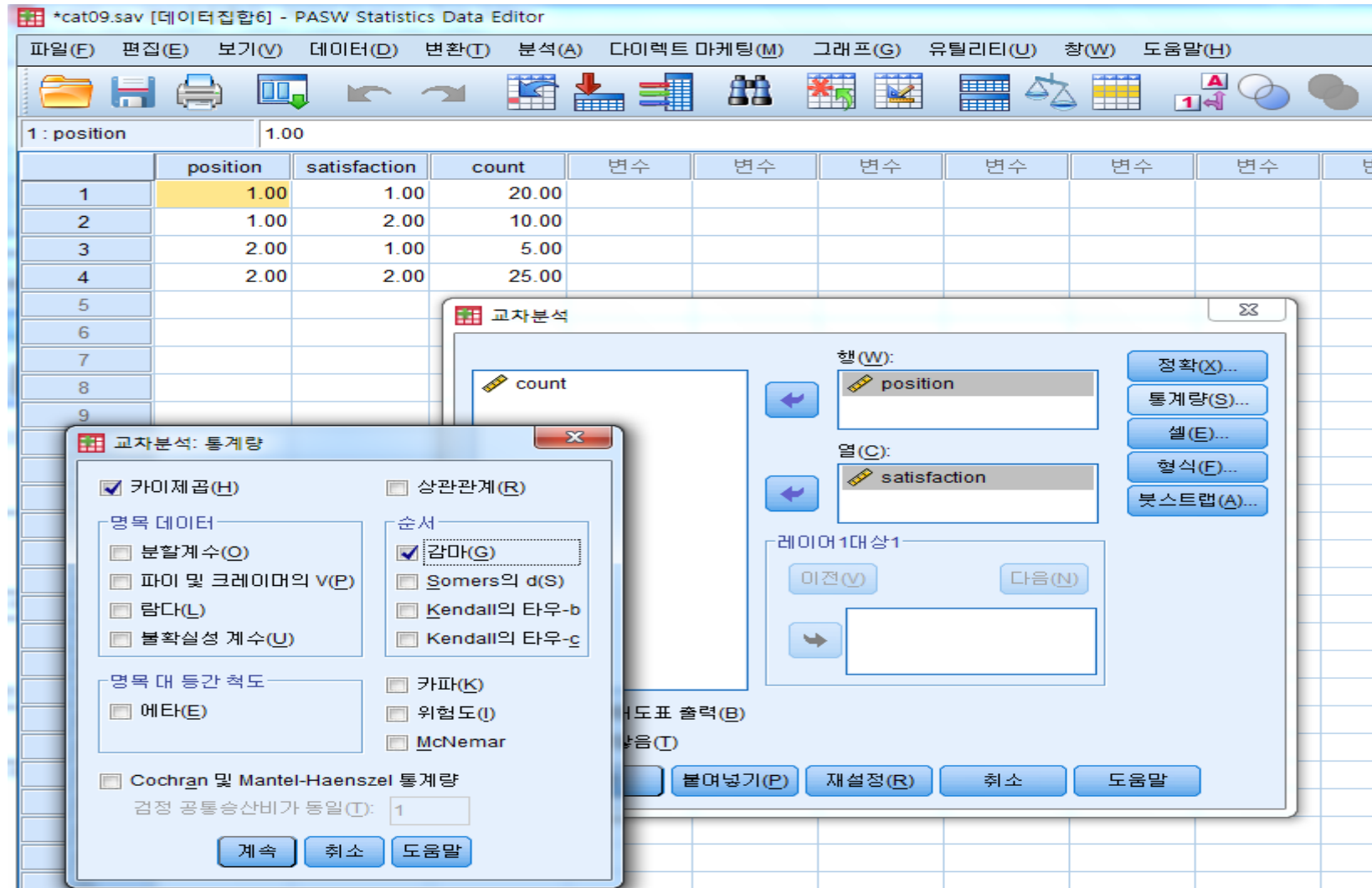
- 분할표에서 두 순서형 변수 사이의 연관도를 재는 측도. $(-1 \leq \gamma \leq 1)$
- 일치쌍 (concordant pair) : 각 변수에 대한 관측값이 크기순서에서 같은 방향에 있는 한 쌍의 관측개체. (C)
- 불일치쌍 (discordant pair) : 각 변수에 대한 관측값이 크기순서에서 반대방향에 있는 한 쌍의 관측개체. (D)

$$\gamma = \frac{(C - D)}{(C + D)}$$

- 감마가 0에 가까울수록 두 범주간 연관관계가 없고, 1에 가까울수록 양(+)의 연관관계, -1에 가까울 수록 음(-)의 연관관계를 가진다.

제3장. 교차표 분석-연관성 측도 : 감마 (gamma)

- SPSS 사례분석-데이터(CAT09.sav)



The image shows the SPSS Statistics Data Editor window with a dataset named 'cat09.sav'. The data is organized into a table with columns for 'position', 'satisfaction', and 'count'. The 'position' variable has two categories: 1.00 and 2.00. The 'satisfaction' variable has two categories: 1.00 and 2.00. The 'count' column shows the frequency of each combination.

| | position | satisfaction | count |
|---|----------|--------------|-------|
| 1 | 1.00 | 1.00 | 20.00 |
| 2 | 1.00 | 2.00 | 10.00 |
| 3 | 2.00 | 1.00 | 5.00 |
| 4 | 2.00 | 2.00 | 25.00 |

The '교차분석' (Crosstabs) dialog box is open, showing the 'count' variable in the '행(W):' (Row) field and the 'satisfaction' variable in the '열(C):' (Column) field. The '레이어1대상1' (Layer 1 Target 1) field is empty. The '교차분석: 통계량' (Crosstabs: Statistics) sub-dialog box is also open, showing the '카이제곱(H)' (Chi-square) option selected under '명목 데이터' (Nominal data). The '감마(G)' (Gamma) option is selected under '순서' (Order). The 'Cochran 및 Mantel-Haenszel 통계량' (Cochran and Mantel-Haenszel statistics) option is also selected. The '검정 공통승산비가 동일(T):' (Test of homogeneity of proportions) is set to 1.

교차분석: 통계량

☒ 카이제곱(H) ☐ 상관관계(R)

명목 데이터

☐ 분할계수(O) ☒ 감마(G)

☐ 파이 및 크레이머의 V(P) ☐ Somers의 d(S)

☐ 람다(L) ☐ Kendall의 타우-b

☐ 불확실성 계수(U) ☐ Kendall의 타우-c

명목 대 등간 척도

☐ 에타(E) ☐ 카파(K)

☐ 위험도(I) ☐ McNemar

☒ Cochran 및 Mantel-Haenszel 통계량

검정 공통승산비가 동일(T): 1

계속 취소 도움말

제3장. 교차표 분석-연관성 측도 : 감마 (gamma)

• 출력결과

position * satisfaction 교차표

| 빈도 | | satisfaction | | 전체 |
|----------|------|--------------|------|----|
| | | 1.00 | 2.00 | |
| position | 1.00 | 20 | 10 | 30 |
| | 2.00 | 5 | 25 | 30 |
| 전체 | | 25 | 35 | 60 |

카이제곱 검정

| | 값 | 자유도 | 점근 유의확률 (양측검정) | 정확한 유의확률 (양측검정) | 정확한 유의확률 (단측검정) |
|-------------------|---------------------|-----|-------------------|-----------------------|-----------------------|
| Pearson 카이제곱 | 15.429 ^a | 1 | .000 | .000 | .000 |
| 연속수정 ^b | 13.440 | 1 | .000 | | |
| 우도비 | 16.279 | 1 | .000 | | |
| Fisher의 정확한 검정 | | | | | |
| 선형 대 선형결합 | 15.171 | 1 | .000 | | |
| 유효 케이스 수 | 60 | | | | |

a. 0 셀 (.0%)은(는) 5보다 작은 기대 빈도를 가지는 셀입니다. 최소 기대빈도는 12.50입니다.

b. 2x2 표에 대해서만 계산됨

대칭적 측도

| | 값 | 점근 표준오차 ^a | 근사 T 값 ^b | 근사 유의확률 |
|----------------|------|-------------------------|---------------------|---------|
| 순서척도 대 순서척도 감마 | .818 | .103 | 4.557 | .000 |
| 유효 케이스 수 | 60 | | | |

a. 영가설을 가정하지 않음.

b. 영가설을 가정하는 점근 표준오차 사용

제3장. 교차표 분석-심슨의 파라독스 (Simpson's Paradox)

✓ 분할표 분석에 있어 전체분석결과와 세부분석의 결과가 모순되는 현상

(예) Berkeley Admission Data, 1973

| | 합격 | 불합격 | 지원자 계 |
|----|------------|------------|-------------|
| 남자 | 1400 (52%) | 1291 (48%) | 2691 (100%) |
| 여자 | 772 (42%) | 1063 (58%) | 1835 (100%) |
| 전체 | 2172 (48%) | 2354 (52%) | 4526 (100%) |

성차별 주장. 그러나....

| | 남 | | 여 | |
|----|-----|-----|-----|-----|
| 분야 | 지원자 | 합격률 | 지원자 | 합격률 |
| A | 825 | 62% | 108 | 82% |
| B | 560 | 63% | 25 | 68% |
| C | 325 | 37% | 593 | 34% |
| D | 417 | 33% | 375 | 35% |
| E | 191 | 28% | 393 | 24% |
| F | 373 | 60% | 341 | 70% |

제3장. 교차표 분석-심슨의 파라독스 (Simpson's Paradox)

| 전체 | 성공 | 실패 | 합계 |
|-----|----------|----------|-----------|
| 실험군 | 20 (50%) | 20 (50%) | 40 (100%) |
| 대조군 | 16 (40%) | 24 (60%) | 40 (100%) |

실험군이 대조군보다 치료율에 있어 10% 포인트 우세한 것으로 보임.

그러나 전체 사례를 증상의 정도에 따라 분류하여 보면,

결론은 그 정반대로 나온다.

| 경증 | 성공 | 실패 | 합계 | 중증 | 성공 | 실패 | 합계 |
|-----|----------|----------|-----------|-----|---------|----------|-----------|
| 실험군 | 18 (60%) | 12 (40%) | 30 (100%) | 실험군 | 2 (20%) | 8 (80%) | 10 (100%) |
| 대조군 | 7 (70%) | 3 (30%) | 10 (100%) | 대조군 | 9 (30%) | 21 (70%) | 30 (100%) |

제3장. 교차표 분석-코호트 연구(Chort Study)

- 코호트 연구 (Chort Study) : 원인을 알고 결과를 추적해 가는 것
- 전향적 연구 (prospective study) : 흡연유무에 따른 폐암 발생율 차이
- 추적조사를 해야 하므로 시간이 많이 걸리고 방대한 비용이 들며 연구관리가 상당히 어렵다.

| | 비폐암 (Y=0) | 폐암 (Y=1) |
|-----------------|-------------------------------|------------------------------|
| 흡연그룹 (그룹 0) | 99.30% $p_{00}=n_{00}/n_0$ | 0.70% $p_{01}=n_{01}/n_0$ |
| 비흡연그룹 (그룹 1) | 99.95% $p_{10}=n_{10}/n_1$ | 0.05% $p_{11}=n_{11}/n_1$ |

- 비율의 차 (difference in rate)

$$\Delta = p_{11} - p_{01} = 0.70\% - 0.05\% = 0.65\%$$

- 상대비 (relative rate)

$$\Delta = p_{11}/p_{01} = 0.70\%/0.05\% = 14.0$$

- 오즈비 (odds ratio)

$$\begin{aligned}\Psi &= (p_{11}/p_{10})/(p_{01}/p_{00}) \\ &= (0.70/99.30)/(0.05/99.95) = 14.09\end{aligned}$$

- 반응(Y=1)의 출현비율이 아주 작은 경우 오즈비는 상대비율과 거의 동일한 값을 갖는다.

제3장. 교차표 분석-사례-대조 연구 (Case-Control Study)

- 후향적 연구 (retrospective study)
- 시간과 비용 측면에서 경제적이다.
- 관측 편의 개입소지 있다.

| | 비폐암 대조군 | 폐암 사례군 |
|-----------------|-----------------|-----------------|
| 흡연그룹 (그룹 0) | 200 n_{00} | 780 n_{01} |
| 비흡연그룹 (그룹 1) | 800 n_{10} | 220 n_{11} |
| 합계 | 1000 | 1000 |

- 오즈비 (odds ratio)

$$\begin{aligned}\Psi &= (n_{01} \times n_{10}) / (n_{00} \times n_{11}) \\ &= (780 \times 800) / (200 \times 220) = 14.2\end{aligned}$$

$$= \frac{\text{흡연그룹의 폐암 대 비폐암의 비율}}{\text{비흡연그룹의 폐암 대 비폐암의 비율}}$$

- 사례-대조 연구에서는 p_{ij} 를 알 수 없으므로 비율의 차나 상대비율을 계산할 수 없다.
- 이 경우 오즈비가 상대비율의 추정치로 사용될 수 있다.