이변량 분석 개요

o 두 변수 간의 관계분석

X \ Y	범주형	측정형
범주형	교차분석*)	분산분석*)
측정형	로지스틱 회귀분석	상관분석*)

교차분석 Cross Tabulation χ^2-검정

o 두 범주형 변수의 빈도표를 교차함

o 교차표 작성 시 '~에 따른' 해당하는 변수를 행에 넣는다. (*)

■교차표

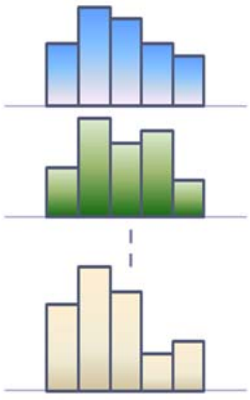
- π_{ij}: (X, Y) 결합밀도함수
- π_{i+}: (X) 주변밀도함수

X \ Y	1	2	...	C	Total
1	π_{11}	π_{12}	...	π_{1c}	π_{1+}
2	π_{21}	π_{22}	...	π_{2c}	Π_{2+}
...	...	...	...	...	...
R	π_{r1}	π_{r2}	...	π_{rc}	Π_{r+}
Total	π_{+1}	π_{+2}	...	π_{+c}	π_{++}

■Homogeneity (동질성)

- 각 행에 대해 열의 분포가 동일한가?

$$H_0 : \pi_{ij} = \pi_{kj} \ for \ j = 1,2,...,c \ \text{and} \ k \neq i$$



■Independence (독립성)

- (X, Y)는 서로 독립인가? $H_0 : \pi_{ij} = \pi_{i+}\pi_{+j}$

■검정통계량 $\dfrac{\sum_{i=1}^{k}(O_i - E_i)^2}{E_i} \sim \chi^2(k-1)$

Professor Kwon, Sehyug | Dept. of Statistics, HANNAM Univ. 010.6365.7622 http://wolfpack.hnu.ac.kr

X \ Y	1	2	…	C	Total
1	n_{11}	n_{12}	…	n_{1c}	n_{1+}
2	n_{21}	n_{22}	…	n_{2c}	n_{2+}
…	…	…	…	…	…
R	n_{r1}	n_{r2}	…	n_{rc}	n_{r+}
Total	n_{+1}	n_{+2}	…	n_{+c}	n_{++}

$$TS = \sum_{i}\sum_{j=1}^{k}\frac{(n_{ij}-E_{ij})^2}{E_{ij}} \sim \chi^2(df=(r-1)*(c-1))$$

기대빈도 $E_{ij} = n_{++}\frac{n_{i+}}{n_{++}}\frac{n_{+j}}{n_{++}}$ (귀무가설⇔ 두 범주형 변수는 독립, 하에서)

교차분석의 검정통계량은 χ^2-분포에 근사한다. (조건은 기대빈도 5 이하 20% 미만 => Cochran 법칙) => 셀을 합치거나 응답자 (n++) 수를 늘린다.

연습문제

전공과 사회 진출분야의 관계?

Major	Oil	Chemical	Electrical	Computer
Business	30	15	15	40
Engineering	30	30	20	20

```
Data one;
   input major $ area $ f;
   datalines;
   B Oil 30
   B Che 15
   B Ele 15
   B Com 40
   E Oil 30
   E Che 30
   E Ele 20
   E Com 20
run;

PROC FREQ DATA=ONE;
   TABLE MAJOR*AREA / CHISQ NOPERCENT NOCOL CROSSLIST STDRES;
   WEIGHT F;
RUN;
```

Professor Kwon, Sehyug | Dept. of Statistics, HANNAM Univ. 010.6365.7622 http://wolfpack.hnu.ac.kr

Frequency Row Pct	Table of major by area				
	area				
major	Che	Com	Ele	Oil	Total
B	15 15.00	40 40.00	15 15.00	30 30.00	100
E	30 30.00	20 20.00	20 20.00	30 30.00	100
Total	45	60	35	60	200

Statistics for Table of major by area

Statistic	DF	Value	Prob
Chi-Square	3	12.3810	0.0062
Likelihood Ratio Chi-Square	3	12.6097	0.0056

유의확률이 0.62%로 귀무가설이 기각되어 전공과 진출분야는 연관이 있다.

행 퍼센트 (ROW PERCENT) 해석 => 비즈니스 전공은 컴퓨터, 공학은 화학, 오일 분야로 주로 진출한다.

Table of major by area				
major	area	Frequency	Std Residual	Row Percent
B	Che	15	-2.5400	15.00
	Com	40	3.0861	40.00
	Ele	15	-0.9305	15.00
	Oil	30	0	30.00
	Total	100		100.00
E	Che	30	2.5400	30.00
	Com	20	-3.0861	20.00
	Ele	20	0.9305	20.00
	Oil	30	0	30.00
	Total	100		100.00

$$R_{ij} = \frac{(n_{ij} - E_{ij})}{\sqrt{E_{ij}(1-p_{i+})(1-p_{+j})}}$$

다른 분야에 비해 컴퓨터 분야에 전공에 따른 차이가 가장 크다. (±3.086)

연습문제 2 🖫EXAMPLE_CROSS

고객의 쇼핑성향과 (1=보수, 2=전통, 3=현대) 직업형태(1=무직, 2=파트타임, 3=정규직)과의 관계분석을 실시하고 해석하시오.

	A	B	C
1	성격	수입	직업
2	1	56	1
3	3	32	3
4	3	31	3
5	3	40	3
6	2	54	1
7	1	52	3

연관 검정, 행 퍼센트 해석, 표준화 잔차 해석