

Chapter 9. 상관분석

상관계수의 개념

❖ 모공분산과 모상관계수

■ 공분산

- 확률변수 X 의 증가에 따른 확률변수 Y 의 증가 또는 감소의 경향을 나타내는 척도
- 공분산이 큰 양의 값을 가진다면 X 가 증가할 때 Y 도 증가하는 경향이 크다는 것을 나타내고,
- 공분산이 큰 음의 값을 가진다면 X 가 증가할 때 Y 는 감소하는 경향이 크다는 것을 나타냄
- 공분산이 0에 가깝다면 두 확률변수의 증감은 서로 체계적인 경향이 없다는 것을 의미

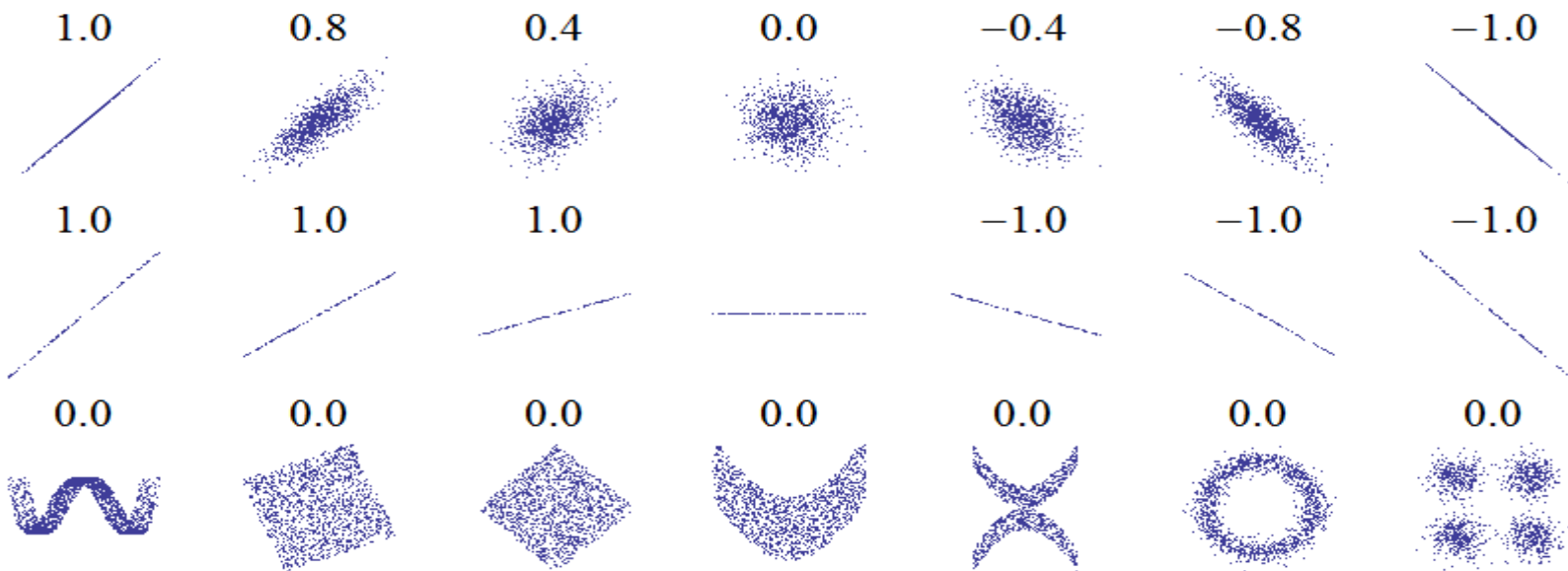
■ 상관계수

- 확률변수 X 와 Y 의 선형성의 정도를 나타내는 척도
- 측정단위에 무관
- -1과 1 사이에 있음

상관계수의 개념

❖ 상관계수의 부호와 크기

- +1 혹은 -1에 가까울수록 직선에 가까운 선형상관관계를 나타냄
- 0에 가까울수록 선형상관관계가 없음



상관계수의 추정과 검정

❖ 피어슨의 상관계수

- 대표본
- 각 변수의 모집단 분포가 정규분포인 경우 사용

$$r_{xy} = \frac{S_{xy}}{S_x S_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

- 가설검정
 - 귀무가설: 두 변수 사이에는 선형관계가 없다
 - 대립가설: 두 변수 사이에는 선형관계가 있다
 - 검정통계량

$$T = \sqrt{n-2} \frac{r}{\sqrt{1-r^2}} \sim t(n-2)$$

상관계수의 추정과 검정

❖ 스피어만의 순위상관계수

- 개념
 - Pearson 상관계수
 - 두 변수를 간격척도 혹은 비율척도로 측정한 경우
 - Spearman 서열상관계분석
 - 변수가 서열척도로 이루어진 경우
- 자료
 - 두 변수 모두 서열척도로 측정된 자료
 - 둘 이상의 대상에 대한 서열이 같은 경우에 가운데 서열을 부여
- 가정
 - 변수의 정규분포 가정이 필요하지 않음(비모수적 방법)

상관계수의 추정과 검정

❖ CORR 프로시저의 이용

- 예제: 나이, 월평균 실질소득, 월평균 지출액의 상관분석

나이	소득	지출액	나이	소득	지출액
25	170	67	28	177	62
20	165	53	16	150	48
19	160	58	21	160	59
22	173	60	16	169	57
20	169	70	19	170	71
20	179	63	26	180	75
23	174	82	16	179	60
25	189	82	17	169	74
30	180	77			

상관계수의 추정과 검정

상관분석

```
DATA student;  
INPUT age income expense @@;  
CARDS;  
25 170 67 28 177 62 20 165 53 16 150 48 19 160 58 21 160 59  
22 173 60 16 169 57 20 169 70 19 170 71 20 179 63 26 180 75  
23 174 82 16 179 60 25 189 82 17 169 74 30 180 77  
;RUN;  
  
PROC CORR DATA=student PEARSON SPEARMAN NOSIMPLE;  
VAR age income expense;  
RUN;
```

편상관계수

❖ 개요

- 개념
 - 어떤 두 변수가 다른 제 3의 변수와의 상관관계가 높으면 두 변수의 상관관계는 순수한 상관관계보다 높게 나타날 수 있음
 - 편상관관계(Partial correlations) 분석
 - 제 3의 변수를 통제된 상태에서 관심을 갖는 두 변수의 상관관계를 분석하는 것
 - 예)
 - 여름 휴가시즌의 평일날 어느 해변가 낚은 콘도에 출몰한 개미의 수와 그날 콘도 앞을 지난 승용차의 수와의 상관관계 조사
 - 개미의 수와 승용차의 수의 순수한 상관관계를 조사하려면 온도를 통제변수로 설정하고 편상관계수를 계산
- 자료와 가정
 - Pearson 상관관계분석의 경우와 동일

편상관계수

❖ 예제

- 특정 제품에 대해서 20명의 고객으로부터 연령, 기능만족도, 디자인만족도를 조사하여 얻은 제품만족도 데이터에 대해서 변수 연령을 통제하였을 때 기능만족도와 디자인만족도의 편상관계수를 구하고, 그 의미를 해석

편상관분석

```
DATA satis;  
INPUT age age_level satis1 satis2 @@;  
CARDS;  
28 2 0 70 23 2 0 55 26 2 5 65 27 2 5 65  
25 2 10 60 26 2 20 65 29 2 25 70 31 3 25 75  
32 3 25 80 34 3 40 85 31 3 40 75 33 3 50 80  
39 3 55 95 36 3 60 90 30 3 65 75 36 3 65 90  
32 3 80 80 39 3 85 95 31 3 90 75 32 3 95 80  
;RUN;
```

```
PROC CORR DATA=satis;  
VAR age satis1 satis2;RUN;  
PROC CORR DATA=satis;  
VAR satis1 satis2;  
PARTIAL age;RUN;
```

편상관계수

연령대별로 계산된 단순상관계수(계속)

```
PROC SORT DATA=satis;  
BY age_level;  
RUN;  
PROC CORR DATA=satis NOSIMPLE;  
VAR satis1 satis2;  
BY age_level;  
RUN;
```

연습문제1

- ❖ 다음은 눈 속에 포함된 수분 함유량(X)과 용수량(Y)을 조사하여 얻은 것이다.
- 선형적인 연관관계가 있는지를 산점도를 통해서 확인해 보아라.
 - 피어슨 상관계수, 순위상관계수를 구하고 그 결과를 비교해 보아라.

X	Y	X	Y
37.9	22.8	32.8	16.7
30.5	14.1	31.8	18.2
25.1	12.9	32.0	17.0
12.4	8.8	30.4	16.3
35.1	17.4	24.0	10.5
31.5	14.9	39.5	23.1
24.2	12.4	52.5	24.9
21.1	10.5	27.6	16.1
23.1	10.5		



연습문제 2

- ❖ 다음은 판촉비, 광고비, 매출을 조사하여 얻은 자료이다.
 - 판촉비가 일정한 경우(통제된 경우) 광고비와 매출 간에 상관관계가 있는가?

광고비	판촉비	매출
1.2	1.0	101
0.8	0.7	92
1.0	0.8	110
1.3	0.9	120
0.7	0.7	90
0.8	0.7	82
1.0	0.8	93
0.6	0.5	75
0.9	0.6	91
1.1	0.9	105