

전산통계 과제#5

컴퓨터소프트웨어 학부

2018008559

신상윤

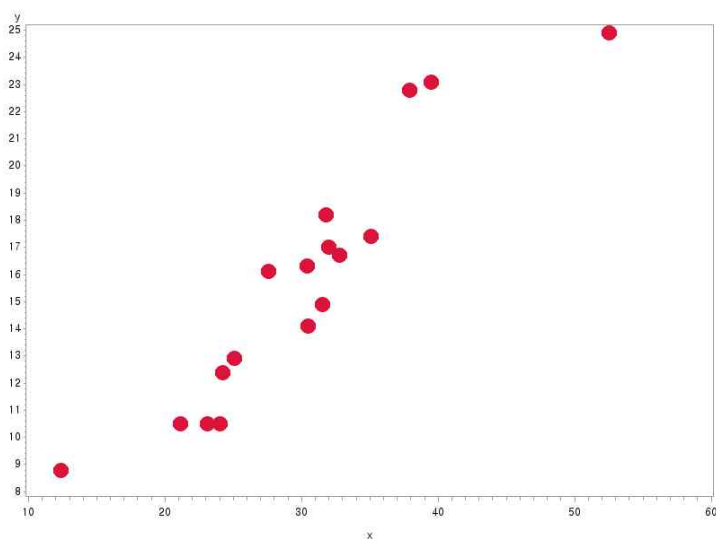
7-1

코드

```
data exe7_1;  
input x y @@;  
cards;  
23.1 10.5 32.8 16.7 31.8 18.2 32.0 17.0 30.4 16.3 24.0 10.5  
37.9 22.8 30.5 14.1 25.1 12.9 12.4 8.8 35.1 17.4 31.5 14.9  
39.5 23.1 24.2 12.4 52.5 24.9 21.1 10.5 27.6 16.1  
;  
run;  
symbol value=dot height=3 color=crimson;  
proc gplot data = exe7_1;  
    plot y*x;  
run;  
quit;  
  
proc corr data = exe7_1 pearson spearman kendall;  
    var x y;  
run;
```

결과

(가)



선형적인 연관관계가 있는 것 같다.

(나)

피어슨 상관 계수, N = 17 H0: Rho=0 가정하에서 Prob > r		
	x	y
x	1.00000	0.93320 <.0001
y	0.93320 <.0001	1.00000

스피어만 상관 계수, N = 17 H0: Rho=0 가정하에서 Prob > r		
	x	y
x	1.00000	0.96069 <.0001
y	0.96069 <.0001	1.00000

켄달 타우 b 상관 계수, N = 17 H0: Tau=0 가정하에서 Prob > tau		
	x	y
x	1.00000	0.86994 <.0001
y	0.86994 <.0001	1.00000

$$r_{\text{피어슨}} = \frac{\frac{1}{16} \times 616.8823529}{8.7981449 \times 4.6958602} = 0.933202932$$

$$r_{\text{스피어만}} = 1 - \frac{6 \times 32}{17(17^2 - 1)} = 0.9607843$$

동률이면 SAS proc rank 기본 설정인 mean을 따르는 것 같다.

$$r_{\text{켄달}} = \frac{C-D}{17C_2} = \frac{125-8}{136} = 0.860294$$

<계산에 사용한 값>

MEANS 프로시저

변수	합계	평균	표준편차
x	511.5000000	30.0882353	8.7981449
y	267.1000000	15.7117647	4.6958602
z	8653.45	509.0264706	297.2500720

$$z = x*y$$

```
23.1 10.5 32.8 16.7 31.8 18.2 32.0 17.0 30.4 16.3 24.0 10.5
37.9 22.8 30.5 14.1 25.1 12.9 12.4 8.8 35.1 17.4 31.5 14.9
39.5 23.1 24.2 12.4 52.5 24.9 21.1 10.5 27.6 16.1
0.860294
Process returned 0 (0x0)   execution time : 6.933 s
Press any key to continue.
```

OBS	x	y	z
1	3	3	0
2	13	11	4
3	11	14	9
4	12	12	0
5	8	10	4
6	4	3	1
7	15	15	0
8	9	7	4
9	6	6	0
10	1	1	0
11	14	13	1
12	10	8	4
13	16	16	0
14	5	5	0
15	17	17	0
16	2	3	1
17	7	9	4

$$z = (x-y)*(x-y)$$

피어슨 상관계수, 순위상관계수는 비슷하게 0.93, 0.96 정도로 높게 나왔다. 그러나 켄달의 타우는 다른 상관계수에 비해 조금 낮은 0.86 정도로 나왔다. 그래도 모두 높은 상관계수 값이 나타났다. 이는 x, y 사이에 어떤 선형의 관계가 있을 가능성이 높음을 의미한다.

(다)

$$H_0 : \rho = 0 \quad H_1 : \rho \neq 0$$

$r_{\text{피어슨}}$: 검정통계량 $T = 10.05783025$

p-value = 0

유의수준 5%에서 H_0 를 기각한다. 따라서 선형의 관계가 있다고 할 수 있다.

$r_{\text{스피어만}}$: 검정통계량 $Z = 3.8431372$

p-value = 0.00006

유의수준 5%에서 H_0 를 기각한다. 따라서 선형의 관계가 있다고 할 수 있다.

$r_{\text{켄달}}$: 검정통계량 $Z = 4.819536785$

p-value = 0

유의수준 5%에서 H_0 를 기각한다. 따라서 선형의 관계가 있다고 할 수 있다.

따라서 유의수준 5% 하에서 모든 상관계수에 대하여 선형의 관계가 있다고 할 수 있다.

7-2

결과

(가) 분자가 0이므로 피어슨 상관계수 $r = 0$ 이다. 이는 x 와 y 간의 선형 관계가 없다는 것을 의미한다.

(나) $H_0 : \rho = 0 \quad H_1 : \rho \neq 0$

검정통계량 $T = 0$, p-value = 0.5 기각역에 포함되지 않으므로 H_0 를 채택

따라서 유의수준 5% 하에서 x, y 는 선형관계가 없다고 할 수 있다.

(다) 이변량 정규분포의 경우 피어슨의 상관계수가 0인 것과 서로 독립인 것이 동치관계에 있다. 위의 경우 상관계수는 0이지만, 두 데이터 간의 연관이 있다고 봐야하기 때문에 이변량 정규분포로부터 얻은 표본이 아니다.

(라) $x = (x-3)*(x-3)$ 변환 후 x 와 y 의 상관분석을 진행한다.

7-4

코드

```
data exe7_4;
input x y @@;
cards;
45 30 37 55 22 28
43 52 34 25 58 44
46 45 30 30 60 61
49 38 31 40 52 58
50 62 27 45 26 17
;
run;
data exe7_4_1;
set exe7_4;
if 27 <= x and x <= 37 then delete;
run;
symbol value=dot height=3 color=crimson;
proc gplot data = exe7_4;
    plot y*x;
run;
quit;
symbol value=dot height=3 color=crimson;
proc gplot data = exe7_4_1;
    plot y*x;
run;
quit;
```

```
proc corr data = exe7_4 pearson spearman kendall;
```

```
var x y;
```

```
run;
```

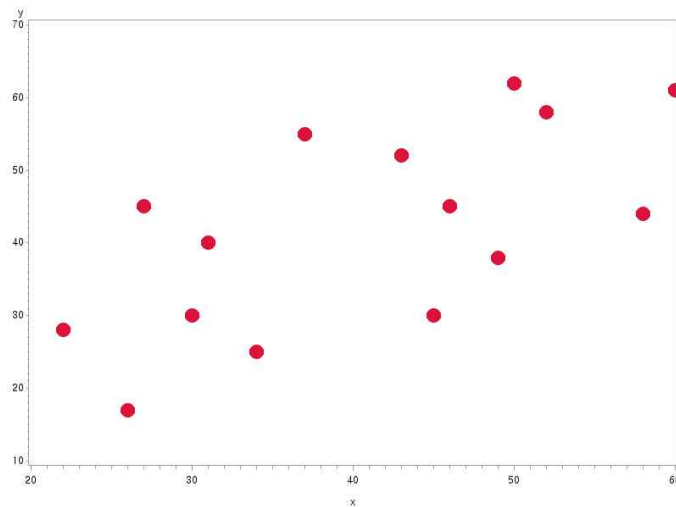
```
proc corr data = exe7_4_1 pearson spearman kendall;
```

```
var x y;
```

```
run;
```

결과

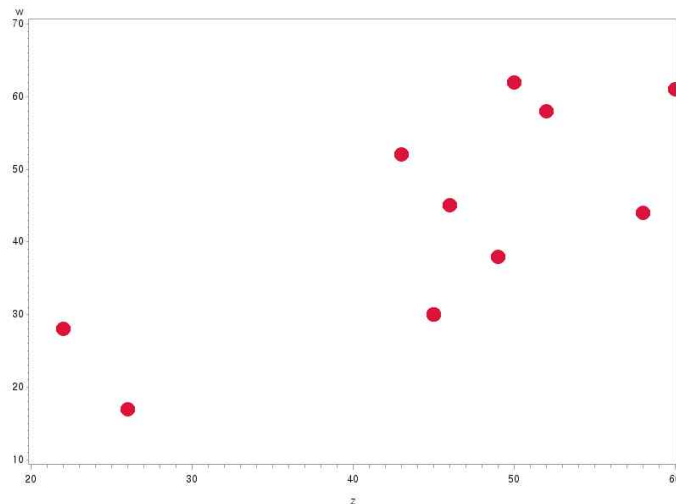
(가)



피어슨 상관 계수, N = 15 H0: Rho=0 가정하에서 Prob > r		
	x	y
x	1.00000	0.64828 0.0090
y	0.64828 0.0090	1.00000

표본상관계수 = 0.64828

(나)



피어슨 상관 계수, N = 10 H0: Rho=0 가정하에서 Prob > r		
	x	y
x	1.00000	0.75751 0.0112
y	0.75751 0.0112	1.00000

표본상관계수 = 0.75751

(다) (가)에서 구한 15명에 대한 표본상관계수보다 (나)에서 구한 10명에 대한 표본상관계수가 더 크다. (1에 가깝다) 이는 15명일 때 보다 10명일때에 선형관계를 강하게 가진다고 할 수 있고, 27세에서 37세 사이의 사람을 빼고 측정하는 것이 유리하다고 할 수 있다.

7-5

코드

```
data exe7_5;
input q1 q2 q3 q4 q5 q6 q7 q8 q9 q10;
cards;
2      1      5      2      4      2      3      5      5      2
1      1      2      4      2      1      2      2      4      4
2      2      3      3      3      1      1      2      4      2
4      4      4      1      5      5      5      4      1      1
1      2      5      1      5      2      1      5      5      1
5      5      1      5      1      4      5      2      2      5
5      5      5      1      4      4      5      5      1      1
1      1      3      4      3      1      2      3      4      3
3      3      4      2      5      3      3      5      4      2
5      4      1      4      2      5      5      1      1      5
4      5      4      1      5      4      4      5      2      1
2      2      5      1      4      4      4      5      3      1
3      4      5      2      5      3      3      4      3      2
2      2      2      5      2      1      1      2      5      5
4      3      4      2      4      4      4      5      4      1
5      4      1      4      1      5      5      1      1      5
5      5      3      2      2      4      4      3      1      3
4      5      2      5      1      5      5      1      2      5
3      3      2      5      2      3      3      2      3      4
;
run;
data exe7_5;
set exe7_5;
```

```

q4 = 6-q4;
q10 = 6-q10;
q9 = 6-q9;
run;
proc corr data=exe7_5 alpha;
    var q3 q4 q5 q8 q10;
run;
proc corr data=exe7_5 alpha;
    var q1 q2 q6 q7 q9;
run;

```

결과

크론바흐의 α 계수	
변수	α 계수
원데이터	0.976731
표준화	0.977073

변수를 제외했을 때의 크론바흐의 α 계수				
제외한 변수	데이터 변수		표준화된 변수	
	합계에 대한 상관 계수	α 계수	합계에 대한 상관 계수	α 계수
q3	0.937476	0.970546	0.937324	0.970766
q4	0.916337	0.973407	0.915447	0.974076
q5	0.919702	0.972961	0.919620	0.973447
q8	0.933411	0.970837	0.933941	0.971280
q10	0.956331	0.967471	0.956229	0.967884

크론바흐의 α 계수	
변수	α 계수
원데이터	0.964115
표준화	0.964129

변수를 제외했을 때의 크론바흐의 α 계수				
제외한 변수	데이터 변수		표준화된 변수	
	합계에 대한 상관 계수	α 계수	합계에 대한 상관 계수	α 계수
q1	0.923907	0.951245	0.923906	0.951268
q2	0.865073	0.960781	0.864893	0.960834
q6	0.906930	0.954019	0.907163	0.954004
q7	0.903270	0.954626	0.903264	0.954639
q9	0.889204	0.956914	0.889220	0.956917

‘직장에 대한 만족도’ 관련 질문의 신뢰도 분석을 진행할 때, 4번과 10번 문항은 부정적 즉, 반대방향 질문을 하고 있으므로 값을 뒤집어 주었다. $\alpha=0.976$ 으로 매우 높은 신뢰도를 가지고 있으며, 어떤 변수를 제외해도 신뢰도는 0.97로 유지되므로 모든 항목을 가져간다.

‘직업에 대한 적극적 태도’ 관련 질문의 신뢰도 분석을 진행할 때, 9번 문항은 부정적 즉, 반대방향 질문을 하고 있으므로 값을 뒤집어 주었다. $\alpha=0.964$ 로 매우 높은 신뢰도를 가지고 있으며, 어떤 변수를 제외해도 신뢰도는 0.95로 유지되므로 모든 항목을 가져간다.