

통계기반 데이터 분석 (Ch 5)

데이터사이언스 & AI 전공

정화민 교수



상관분석

상관 분석(相關 分析, Correlation analysis)

- '상관관계' 또는 '상관'은 **확률론**과 **통계학**에서 두 변수간에 어떤 선형적 또는 비선형적 관계를 갖고 있는 지를 분석하는 방법이다.
- 두변수는 서로 독립적인 관계이거나 상관된 관계일 수 있으며 이때 두 변수간의 관계의 강도를 상관관계(Correlation, Correlation coefficient)라 한다. 상관분석에서는 상관관계의 정도를 나타내는 단위로 모상관계수로 ρ 를 사용하며 표본 상관 계수로 r 을 사용한다.

Source: https://ko.wikipedia.org/wiki/%EC%83%81%EA%B4%80_%EB%B6%84%EC%84%9D



Source http://star.ohmynews.com/NWS_Web/OhmyStar/

통계기법을 이용한 상관관계 분석실습 및 코딩방법

덕선이 남편 찾기



엑셀 코딩 예)

설문지번호	덕선	택이	선우	동릉	정환
1	7	7	6	4	7
2	7	7	6	5	5
3	6	5	7	5	5
4	5	4	4	7	4
5	5	5	6	6	7

분석결과 : 직접 데이터 수집, 저장
분석을 해 보시다.

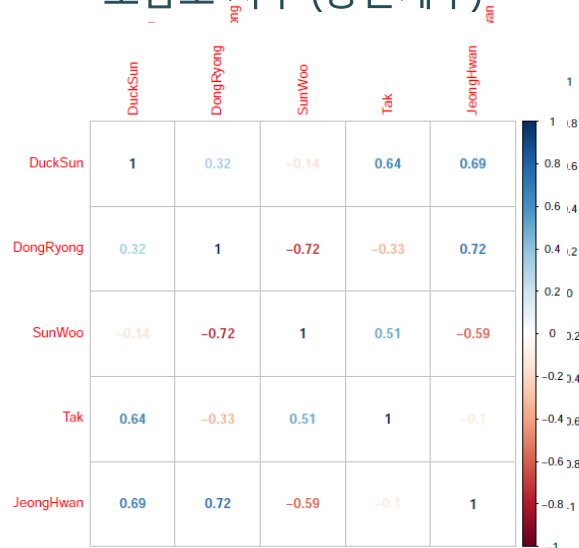
R 프로그래밍 예)

```
install.packages("xlsx")
install.packages("corrplot")
library(xlsx)
setwd("c:/RStudio/데이터분석과R프로그래밍")
xlsx_ds=read.xlsx("dsdata.xlsx",1)
xlsx_ds

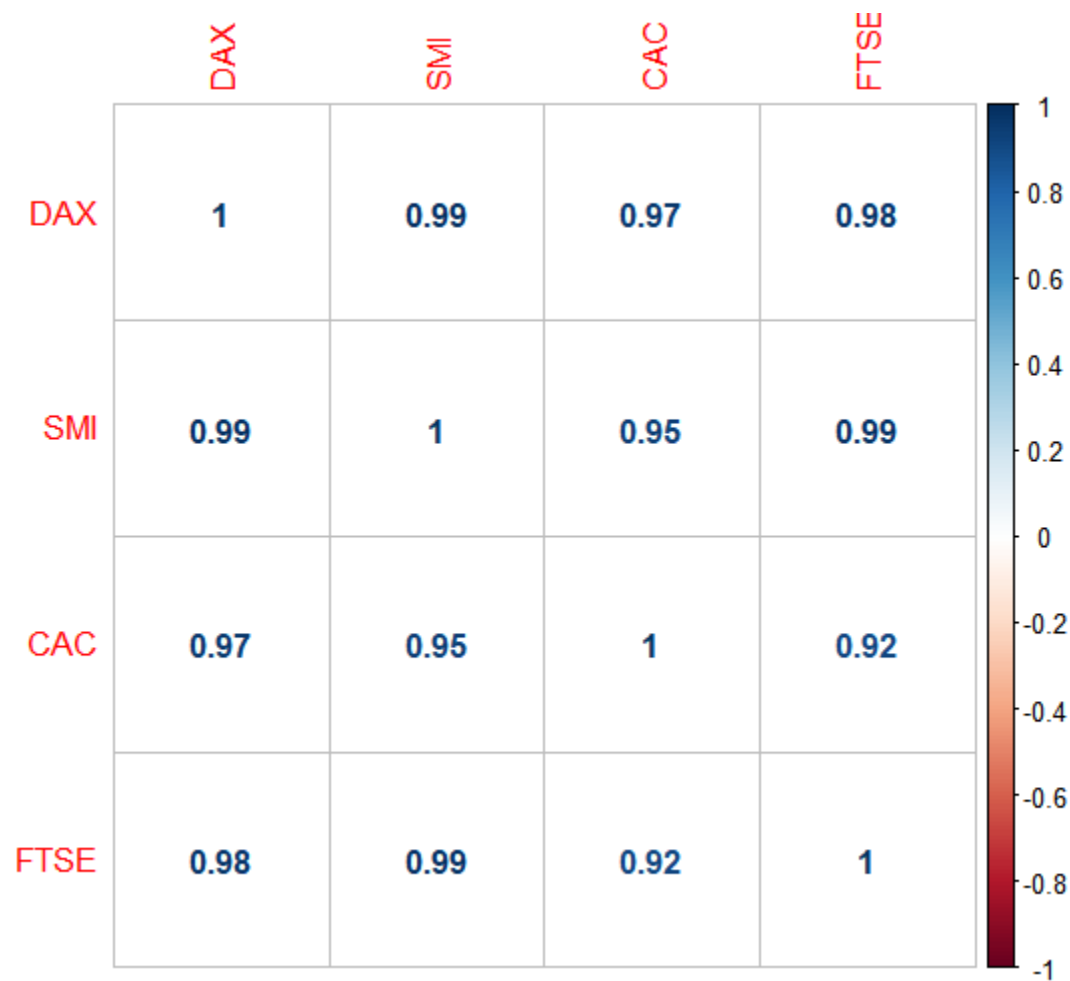
dscorr <- cor(xlsx_ds)
library(corrplot)

corrplot(dscorr, method = "number")
corrplot(dscorr, method = "circle")
corrplot(dscorr, method = "pie")
corrplot(dscorr, method = "ellipse")
```

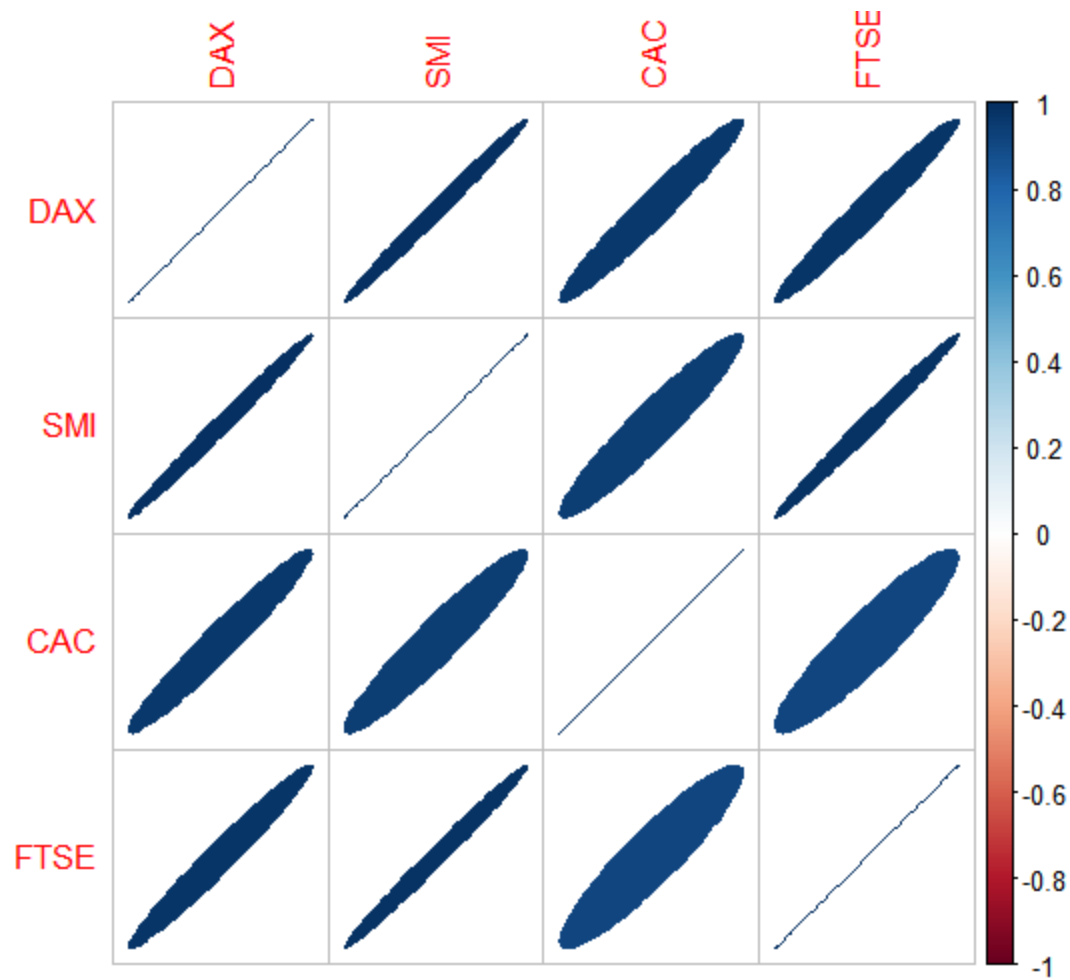
호감도 지수 (상관계수)



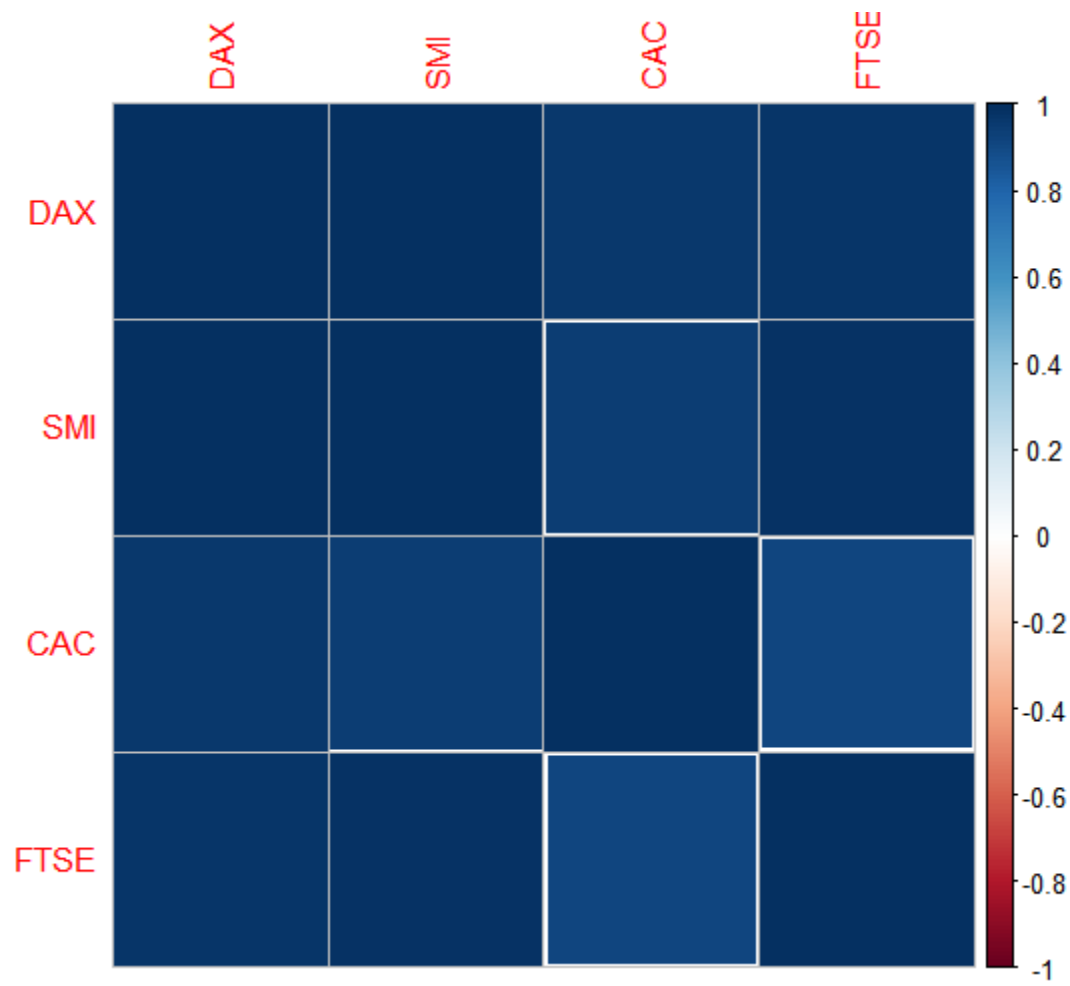
number



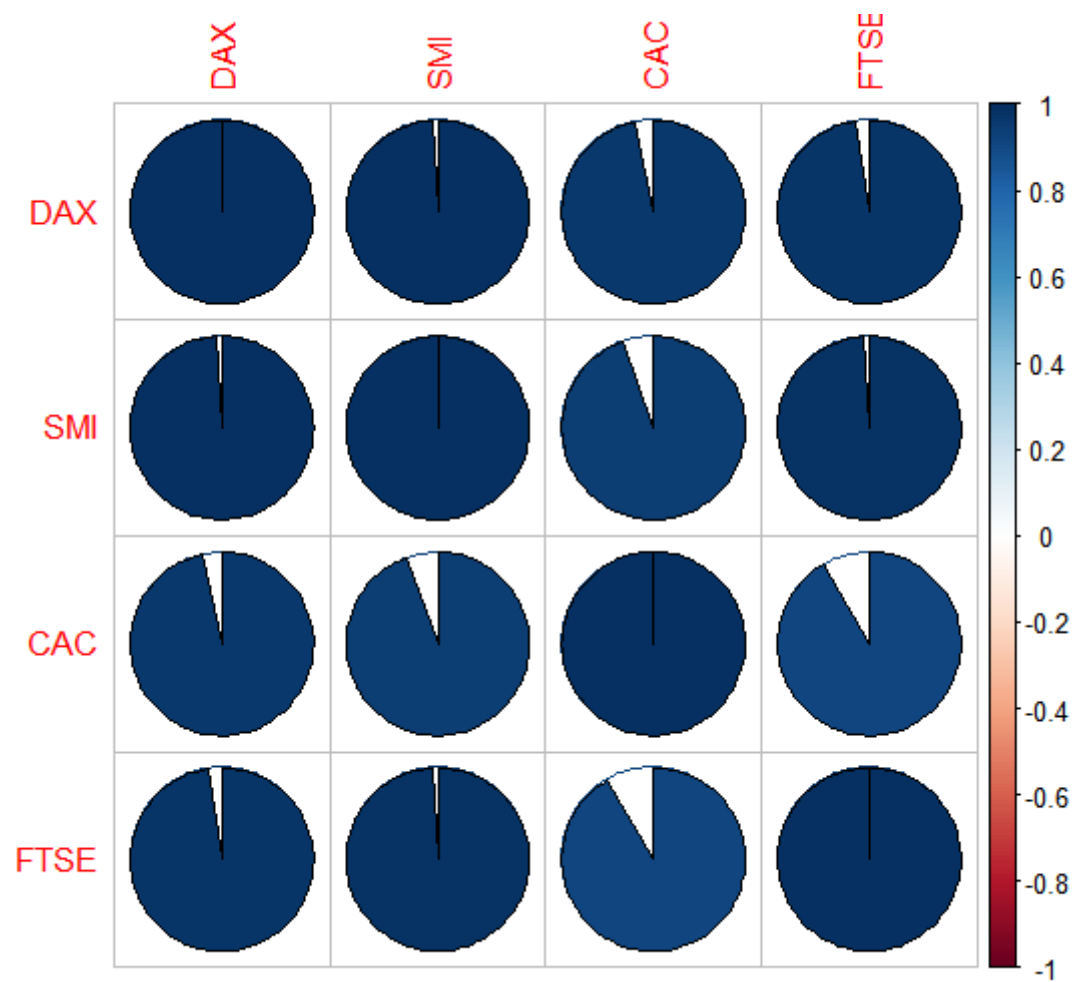
ellips



square



pie



상관분석 예)

<표 5> 표본의 일반적 특성 (N, %)

항목	기업형태				전체	
	개인사업자		법인사업자			
1 제조	24	8.30%	5	8.77%	29	8.38%
2 건설	4	1.38%	7	12.28%	11	3.18%
3 도·소매	107	37.02%	27	47.37%	134	38.73%
4 음식	80	27.68%	2	3.51%	82	23.70%
5 숙박	2	0.69%	0	0.00%	2	0.58%
6 출판·영상·방송통신 등	6	2.08%	2	3.51%	8	2.31%
7 IT	6	2.08%	6	10.53%	12	3.47%
8 개인서비스	49	16.96%	5	8.77%	54	15.61%
9 기타	11	3.81%	3	5.26%	14	4.05%
합계	289	100%	57	100%	346	100%

4.4 상관관계 분석

요인분석과 신뢰도 분석을 통하여 얻어진 결과를 가지고 통계프로그램 SPSS 21.0상에서 상관관계 분석을 실시하였다. 상관관계란 변수들 간의 관계를 말하는 것으로 한 변수가 변화함에 따라 다른 변수가 어떻게 변화하는지와 같은 변화의 강도와 방향을 상관관계라 한다.

본 연구에서는 가설검증을 하고자 하는 변수들 간의 관련성의 정도와 방향을 파악하기 위하여 상관관계 분석을 실시하였다. 상관관계 분석결과 전체 표본은 346부로 나타나 있다. 그리고 <표 11>과 같이 모든 변수들의 상관계수는 유의수준 0.01에서 유의한 것으로 나타났다. 컨설팅과 서비스만족의 상관계수는 .532로 비교적 높은 상관관계를 나타내고 있으며, 자금지원과 서비스만족의 상관관계는 .578로 높은 상관관계를 보이고 있다. 교육과 서비스만족의 상관관계도 .551, 서비스만족과 지각된 유용성과 지각된 용이성은 각각 .535, .490 나타났다. 상관관계 분석결과 각각의 변수들은 유의확률 0.01 유의수준 하에서 정(+) 유의한 상관관계가 것으로 나타나 상관관계가 있는 것으로 나타났으며, 그 관계가 0.8 이상인 것이 없고, 공선성 검증에서도 공차한계값이 0.1 이하 VIF도 1.0 이상으로 각 변수별 다중공선성의 문제는 없다고 할 수 있다.

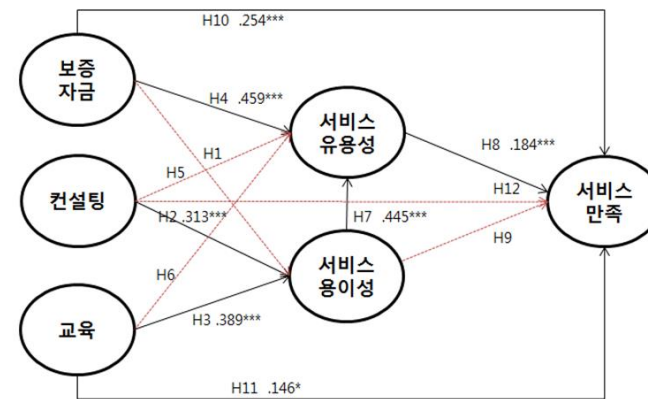
<표 11> 상관관계 분석 결과 N=346

항목	컨설팅	보증_자금	교육	유용성	용이성	서비스만족
컨설팅	1					
자금지원	.596**	1				
교육	.678**	.568**	1			
지각된유용성	.521**	.597**	.514**	1		
지각된용이성	.593**	.462**	.620**	.546**	1	
서비스만족	.532**	.578**	.551**	.535**	.490**	1

**상관계수는 0.01수준(양쪽)에서 유의

<표 12> 확인적 요인분석

개념	측정변수	표준화 계수	비표준화된 요인적재치	S.E.	C.R.	CCR	AVE
컨설팅	컨설팅1	0.863	1			0.965	0.903
	컨설팅2	0.948	1.142	0.047	24.337		
	컨설팅3	0.883	1.071	0.048	22.275		
자금지원	자금지원1	0.848	1			0.959	0.886
	자금지원2	0.927	1.04	0.049	21.424		
	자금지원3	0.859	0.973	0.049	19.938		
교육	교육1	0.904	1			0.971	0.916
	교육2	0.922	1.081	0.04	27.223		
	교육3	0.925	1.114	0.041	27.353		
서비스 유용성	유용성1	0.842	1			0.945	0.852
	유용성2	0.89	1.093	0.056	19.554		
	유용성3	0.863	0.957	0.05	19.066		
서비스 용이성	용이성1	0.805	1			0.908	0.767
	용이성2	0.828	1.195	0.084	14.189		
	용이성3	0.764	1.131	0.082	13.733		
서비스 만족	서비스만족1	0.831	1			0.964	0.901
	서비스만족2	0.952	1.064	0.048	22.051		
	서비스만족3	0.88	0.983	0.048	20.595		



* $P < 0.05$, ** $P < 0.01$, *** $P < 0.0001$

<그림 2> 가설검증을 위한 구조방정식 분석