

2. 상관관계 분석

상관분석

<http://contents.kocw.or.kr/KOCW/document/2015/dongguk/sonchangkyoon/5.pdf>

<http://contents.kocw.or.kr/KOCW/document/2015/pusan/kimsukwoo/7.pdf>

상관관계분석(Correlation Analysis): 변수들 간의 관련성을 분석하기 위해 사용하는 분석 방법
하나의 변수가 다른 변수와 관련성이 있는지, 있다면 어느 정도의 관련성이 있는지를 개관할 수 있는 분석기법

상관관계 분석 수행 시,

회귀분석에서 변수 간의 인과관계를 분석하기 전에 변수 간의 관련성을 분석하는 선행자료
(가설검정 전 수행) 로 이용

변수 간의 관련성을 위해 상관계수인 피어슨(Pearson) r 계수를 이용하여 관련성 유무와 정도를 파악

상관관계 분석의 척도인 피어슨 상관계수(Pearson correlation coefficient: r)와 상관관계의 정도는
[표 14.3]

[표 14.3] 피어슨 상관계수와 상관관계 정도

피어슨 상관계수 R	상관관계 정도
± 0.9 이상	매우 높은 상관관계
$\pm 0.9 \sim \pm 0.7$	높은 상관관계
$\pm 0.7 \sim \pm 0.4$	다소 높은 상관관계
$\pm 0.4 \sim \pm 0.2$	낮은 상관관계
± 0.2 미만	상관관계 없음

상관계수 r 은 $-1 \sim +1$ 까지의 값을 가진다.

가장 높은 완전 상관관계의 상관계수는 1. 두 변수 간에 전혀 상관관계가 없으면 상관계수는 0

2.1 상관계수 r 과 상관관계 정도

[그림 14.3] 상관계수 r 과 상관관계 정도

선, 원, U자

2.2 상관관계 분석 수행

제품의 친밀도, 적절성, 만족도 변수를 대상으로 변수 간의 상관계수를 통해서 상관관계 분석을 수행

실습 (기술통계량 구하기)

1단계: 데이터 가져오기

```
product <- read.csv("C:/Rwork/product.csv", header = TRUE)
head(product)
```

2단계: 기술통계량

```
summary(product)
sd(product$제품_친밀도); sd(product$제품_적절성); sd(product$제품_만족도)
```

실습 (상관계수 보기)

변수간의 상관계수는 stats패키지에서 제공하는 cor()함수 사용

형식: cor(x, y=NULL, use="everything", method=c("pearson", "kendall", "spearman"))
where
method를 생략하면 pearson이 사용된다(default)

"pearson", "kendall", "spearman"의 비교

<https://bskyvision.com/116>

1단계: 변수 간의 상관계수 보기

```
cor(product$제품_친밀도, product$제품_적절성)
cor(product$제품_친밀도, product$제품_만족도)
```

2단계: 제품_적절성과 제품_만족도의 상관계수 보기

```
cor(product$제품_적절성, product$제품_만족도)
```

3단계: (제품_적절성+제품_친밀도)와 제품_만족도의 상관계수 보기

```
cor(product$제품_적절성 + product$제품_친밀도, product$제품_만족도)
```

실습 (전체 변수 간의 상관계수 보기)

```
cor(product, method = "pearson")
```

대각선은 자기 상관계수를 의미

실습 (방향성 있는 색상으로 표현)

```
install.packages("corrgram")  
library(corrgram)  
corrgram(product)  
corrgram(product, upper.panel = panel.conf)  
corrgram(product, lower.panel = panel.conf)
```

corrgram 패키지 설치

실습 (차트에 밀도곡선, 상관성, 유의확률(별표) 추가)

1단계: 패키지 설치

```
install.packages("PerformanceAnalytics")  
library(PerformanceAnalytics)
```

PerformanceAnalytics 패키지 설치

2단계: 상관성, p값(*), 정규분포(모수 검정 조건) 시각화

```
chart.Correlation(product, histogram = , pch = "+")
```

실습 (서열척도 대상 상관계수)

```
cor(product, method = "spearman")
```

서열척도로 구성된 변수에 대해서 상관계수를 구하기 위해서 "method=spearman"속성 적용

피어슨(Pearson)상관계수: 대상변수가 등간척도 또는 비율척도일 때

스피어만(Spearman) 상관계수: 대상변수가 서열척도일 때

2.3 상관관계 분석 결과 제시

일반적으로 상관관계 분석 결과를 논문이나 보고서에 제시하는 경우, 해당 변수들의 기본적인 기술통계량(평균과 표준편차)와 피어슨 상관계수를 함께 제시

상관관계 분석의 유형

단순 상관관계: X와 Y간의 상관관계

다중(Multiple)상관관계: 둘 이상의 변수들이 다른 한 변수와 관계를 갖는 경우

편(Partial)상관관계: 두 변수 관계의 정도를 파악하고자 할 때 제 3의 변수가 두 변수 모두에 영향을 미치고 있는 경우 이를 통제한 다음 분석

부분(Semi_partial)상관관계: 제3의 변수가 어느 한 변수에만 영향을 미치는 경우 이를 통제한 후 분석