



I. Excel 상관분석

II. R 상관분석

## 1. 공분산

- 두 변수가 함께 변하는 정도를 측정

$$\text{Cov}(X,Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$$

- 공분산의 부호와 크기로 두 변수의 관계를 판단

- 부호가 양(+) : 두 변수가 같은 방향으로 변화 즉, X변수가 증가(감소)하면 Y변수도 증가(감소)

예를 들면, 통화량 증가율과 물가 상승률

크기가 클수록 두 변수는 강한 양(+)의 관계를 가짐

- 부호가 음(-) : 두 변수가 반대 방향으로 변화 즉, X변수가 증가(감소)하면 Y변수는 감소(증가)

예를 들면, 경제성장률과 실업률

크기가 작을수록 두 변수는 강한 음(-)의 관계를 가짐

- 크기가 0 : 두 변수는 관계가 없음

- 두 변수의 측정단위에 따라 공분산의 부호는 변하지 않지만 공분산의 크기는 달라짐

- 예를 들어 X의 측정단위 변화로 X의 값이 원래보다 10배 증가하면 공분산의 크기도 10배 증가

- X 및 Y의 측정단위 변화로 X 및 Y의 값이 각각 원래보다 10배씩 증가하면 공분산의 크기는 100배 증가

- 공분산의 크기는 측정단위에 영향을 받기 때문에 동일한 집단이라도 측정단위에 따라 달라지는 한계로 많이 이용되지 않음

## 2. 상관계수

- 상관분석은 독립변수와 종속변수의 구분이 없는 두 확률변수 간의 선형성의 정도를 측정한 상관계수를 추정하고 검정하는 것을 말함
- 상관계수는 두 변수 간 선형관계의 밀접도를 측정하는 통계량

$$\text{Corr}(X,Y) = \rho_{X,Y} = \frac{\text{Cov}(X,Y)}{S_X S_Y}$$

단,  $S_X, S_Y$ 는 X 및 Y의 표준편차를 각각 나타냄

- 상관계수의 부호와 크기로 두 변수의 관계를 판단
  - 부호가 양(+) : 두 변수들이 서로 같은 방향으로 변화 즉, X변수가 증가(감소)하면 Y변수도 증가(감소)
  - 부호가 음(-) : 두 변수들이 서로 반대 방향으로 변화 즉, X변수가 증가(감소)하면 Y변수는 감소(증가)
  - 크기가 0 : 두 변수는 관계가 없음
- 상관계수는 -1과 1사이의 값을 가짐
  - 상관계수의 크기가 양(+)이면서 1에 가까워질수록 두 변수는 강한 양(+)의 관계를 가짐
  - 상관계수의 크기가 음(-)이면서 -1에 가까워질수록 두 변수는 강한 음(-)의 관계를 가짐
  - 상관계수의 크기가 0이면 두 변수는 관계가 없음
- 한편, 상관계수는 두 변수의 측정단위 변화에 따라 크기 및 부호가 영향을 받지 않으므로(이를 unit-free라고 함) 많이 이용됨

- (예) 제3장의 기술통계량 계산 시 사용한 자료(describe.xlsx)로 공분산 및 상관계수를 구해보라
- <http://kanggc.ipstime.org/book/data/describe.xlsx>로 파일을 다운로드(왼쪽 그림)
- 데이터-분석-데이터 분석을 실행하면 나타나는 통계 데이터분석 대화상자에서 상관분석(공분산분석)을 실행
- 이때 계열이름이 있는 셀과 자료가 입력되어 있는 셀을 모두 선택한 후 아래 그림과 같은 상관분석 대화상자에서 첫째 행 이름표 사용에 ☒

상관 분석

입력  
입력 범위(I):

데이터 방향:  
☒ 열(C)  
☐ 행(R)

☒ 첫째 행 이름표 사용(L)

출력 옵션  
☐ 출력 범위(O):   
☒ 새로운 워크시트(P):   
☐ 새로운 통합 문서(W)

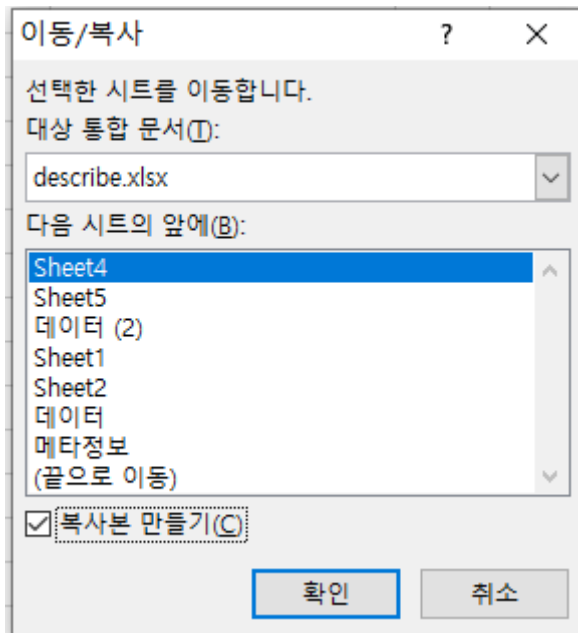
- 모든 공분산이 상관계수와 마찬가지로 양의 부호를 가지고 있고, 대각행렬에 있는 값은 각 변수의 분산

|   | A           | B           | C           | D           | E           |
|---|-------------|-------------|-------------|-------------|-------------|
| 1 |             | 소비자물가상승률    | 주택담보대출금리    | GDP성장률      | GDP디플레이터상승률 |
| 2 | 소비자물가상승률    | 1.215822222 |             |             |             |
| 3 | 주택담보대출금리    | 1.122755556 | 1.551715556 |             |             |
| 4 | GDP성장률      | 0.279866667 | 0.781333333 | 2.869066667 |             |
| 5 | GDP디플레이터상승률 | 0.568488889 | 0.615755556 | 0.1352      | 1.261155556 |
| 6 |             |             |             |             |             |

- 모든 상관계수가 양의 부호를 가지고 있고, 소비자물가상승률과 주택담보대출금리 간 상관계수가 0.817로 가장 크고, GDP성장률과 GDP디플레이터상승률 간 상관계수가 0.071로 가장 작음

|   | A           | B           | C           | D           | E           |
|---|-------------|-------------|-------------|-------------|-------------|
| 1 |             | 소비자물가상승률    | 주택담보대출금리    | GDP성장률      | GDP디플레이터상승률 |
| 2 | 소비자물가상승률    | 1           |             |             |             |
| 3 | 주택담보대출금리    | 0.81741785  | 1           |             |             |
| 4 | GDP성장률      | 0.149846229 | 0.370305341 | 1           |             |
| 5 | GDP디플레이터상승률 | 0.459095054 | 0.44016748  | 0.071075847 | 1           |

- 상승률, 이자율, 성장률의 측정 단위를 현재 %에서 100으로 나누어 준 후 공분산 및 상관계수를 계산
- 데이터 시트를 마우스 오른쪽쪽을 클릭하여 이동/복사를 선택하면 나타나는 이동/복사 대화상자에서 복사본 만들기에 ☒하고 확인
- F2셀에 식 =B2/100을 입력하여 %를 100으로 나누어 주고, F2셀을 선택하고 I2셀까지 마우스를 끌어서 클릭하여 복사
- F2셀부터 I2셀까지 블록으로 선택하고 I16셀까지 마우스를 끌어서 클릭하여 복사
- F2셀부터 I16셀까지 블록으로 선택하고 복사를 하고 B2셀에서 마우스 오른쪽을 클릭하여 선택하여 붙여넣기-값을 선택하고 확인을 클릭
- 공분산과 상관계수를 계산해 보면 공분산의 크기는 원래 크기의 10000분의 1로 축소되고, 상관계수의 크기는 변화 없음



|   | A           | B           | C           | D           | E           |
|---|-------------|-------------|-------------|-------------|-------------|
| 1 |             | 소비자물가상승률    | 주택담보대출금리    | GDP성장률      | GDP디플레이터상승률 |
| 2 | 소비자물가상승률    | 0.000121582 |             |             |             |
| 3 | 주택담보대출금리    | 0.000112276 | 0.000155172 |             |             |
| 4 | GDP성장률      | 2.79867E-05 | 7.81333E-05 | 0.000286907 |             |
| 5 | GDP디플레이터상승률 | 5.68489E-05 | 6.15756E-05 | 0.00001352  | 0.000126116 |

|   | A           | B           | C           | D           | E           |
|---|-------------|-------------|-------------|-------------|-------------|
| 1 |             | 소비자물가상승률    | 주택담보대출금리    | GDP성장률      | GDP디플레이터상승률 |
| 2 | 소비자물가상승률    | 1           |             |             |             |
| 3 | 주택담보대출금리    | 0.81741785  | 1           |             |             |
| 4 | GDP성장률      | 0.149846229 | 0.370305341 | 1           |             |
| 5 | GDP디플레이터상승률 | 0.459095054 | 0.44016748  | 0.071075847 | 1           |

b3-ch5-1-rev.R

```
library(openxlsx)
```

```
df<-read.xlsx("http://kanggc.iptime.org/book/data/describe-e.xlsx")
```

```
df_dat<-data.matrix(df)
```

```
cpi<-df_dat[,2]
```

```
interest<-df_dat[,3]
```

```
gdp<-df_dat[,4]
```

```
deflator<-df_dat[,5]
```

```
df_new<-cbind(cpi, interest, gdp, deflator)
```

```
df_new
```

```
(var<-var(df_new))
```

```
df_var<-data.matrix(var)
```

```
(sd<-sqrt(diag(df_var)))
```

```
corr<-cor(df_new)
```

```
corr
```

```
df_new_1<-cbind(cpi/100, interest/100, gdp/100, deflator/100)
```

```
(var_new<-var(df_new_1))
```

```
(corr<-cor(df_new_1))
```

```
> (var<-var(df_new))
```

|          | cpi       | interest  | gdp       | deflator  |
|----------|-----------|-----------|-----------|-----------|
| cpi      | 1.3026667 | 1.2029524 | 0.2998571 | 0.6090952 |
| interest | 1.2029524 | 1.6625524 | 0.8371429 | 0.6597381 |
| gdp      | 0.2998571 | 0.8371429 | 3.0740000 | 0.1448571 |
| deflator | 0.6090952 | 0.6597381 | 0.1448571 | 1.3512381 |

```
> corr
```

|          | cpi       | interest  | gdp        | deflator   |
|----------|-----------|-----------|------------|------------|
| cpi      | 1.0000000 | 0.8174179 | 0.14984623 | 0.45909505 |
| interest | 0.8174179 | 1.0000000 | 0.37030534 | 0.44016748 |
| gdp      | 0.1498462 | 0.3703053 | 1.00000000 | 0.07107585 |
| deflator | 0.4590951 | 0.4401675 | 0.07107585 | 1.00000000 |

```
> (var_new<-var(df_new_1))
```

|      | [,1]         | [,2]         | [,3]         | [,4]         |
|------|--------------|--------------|--------------|--------------|
| [1,] | 1.302667e-04 | 1.202952e-04 | 2.998571e-05 | 6.090952e-05 |
| [2,] | 1.202952e-04 | 1.662552e-04 | 8.371429e-05 | 6.597381e-05 |
| [3,] | 2.998571e-05 | 8.371429e-05 | 3.074000e-04 | 1.448571e-05 |
| [4,] | 6.090952e-05 | 6.597381e-05 | 1.448571e-05 | 1.351238e-04 |

```
> (corr<-cor(df_new_1))
```

|      | [,1]      | [,2]      | [,3]       | [,4]       |
|------|-----------|-----------|------------|------------|
| [1,] | 1.0000000 | 0.8174179 | 0.14984623 | 0.45909505 |
| [2,] | 0.8174179 | 1.0000000 | 0.37030534 | 0.44016748 |
| [3,] | 0.1498462 | 0.3703053 | 1.00000000 | 0.07107585 |
| [4,] | 0.4590951 | 0.4401675 | 0.07107585 | 1.00000000 |