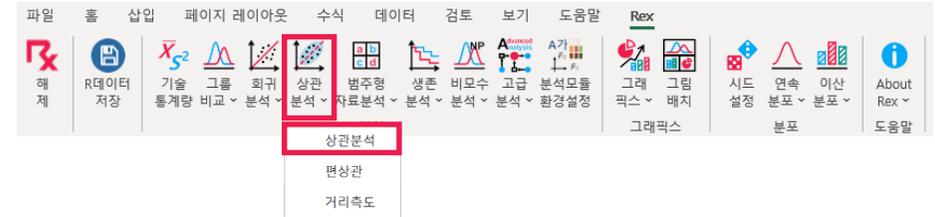


상관분석

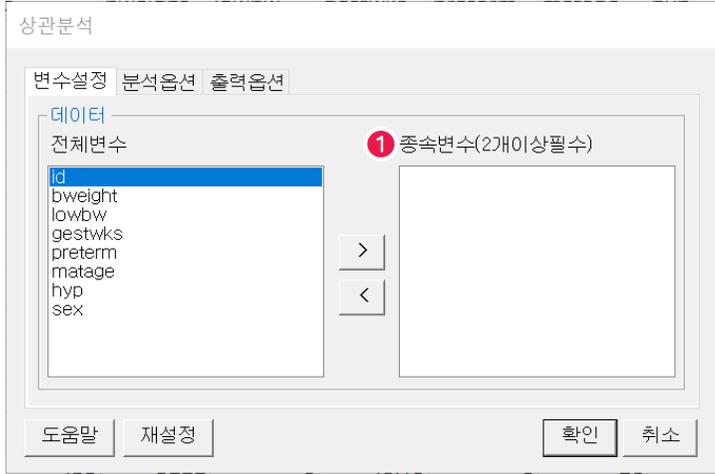
상관분석은 두 변수가 서로 선형적으로 어떤 관계를 갖는지 파악해볼 수 있는 분석입니다. 연속적(continuous) 이며 무작위적인(random) 두 양적 변수 간의 연관성을 파악해 볼 수 있는 분석 방법으로, 상관계수(correlation coefficient)는 두 변수의 연관성의 크기와 방향을 나타내며 -1부터 1까지의 범위를 갖는다. Rex에서는 변수 지정만으로도 모수적, 비모수적인 상관분석이 가능하며, 상관행렬 그림도 손쉽게 그릴 수 있습니다.

메뉴 호출하기

- Rex > 상관분석 > 상관분석



• 변수설정 탭



• 분석옵션 탭



메뉴 요소	설명
① 종속변수	상관관계를 분석하고자 하는 변수들을 전체변수로부터 선택합니다. 반드시 2개 이상의 양적 변수가 선택되어야 합니다.
② 상관계수 추정방법	<p>종속변수로 지정된 변수들의 상관계수를 추정하는 방법으로, 다음 3가지 옵션 중 1개를 선택합니다.</p> <ul style="list-style-type: none"> Pearson : 두 변수 간 선형적인 상관관계의 크기를 모수적으로 추정하는 방식입니다. Kendall : 두 변수가 순서형인 경우 또는 변수가 정규분포에서 크게 벗어나는 경우 자료를 순서형으로 변환하여 비모수적으로 상관관계를 추정하는 방식입니다. 표본수가 적거나 데이터에 동물이 많은 경우 유용하다고 알려져 있습니다. Spearman : 두 변수가 순서형인 경우 또는 변수가 정규분포에서 크게 벗어나는 경우 자료를 순서형으로 변환하여 비모수적으로 상관관계를 추정하는 방식입니다. 데이터 내의 편차에 민감하며, 일반적으로 Kendall의 상관계수보다 높게 추정되는 편입니다.
③ 정확 검정	[상관계수 추정방법]에서 [Kendall] 또는 [Spearman]이 선택된 경우 활성화됩니다. 표본수가 작을 경우, 순위간 정확한 계산을 통해 p-value를 도출하는 방식입니다. 이 옵션이 선택되지 않는 경우 정규근사방법으로 도출된 p-value가 출력됩니다.
④ 연속성 수정	[상관계수 추정방법]에서 [Kendall] 또는 [Spearman]이 선택된 경우 활성화됩니다. 표본수가 작을 경우 연속성 수정된 검정통계량을 통해 계산된 p-value가 출력됩니다.
⑤ 가설 검정	<p>대립가설에 맞는 검정방법을 다음 3가지 중 1개를 선택합니다.</p> <ul style="list-style-type: none"> 양측검정 (Default) : $\rho \neq 0$ 좌측단측검정 : $\rho < 0$ 우측단측검정 : $\rho > 0$

• 출력옵션 탭

상관분석

변수설정 분석옵션 출력옵션

출력

① P-value

② 공분산

③ 샘플수

그래프

④ 산점도 행렬

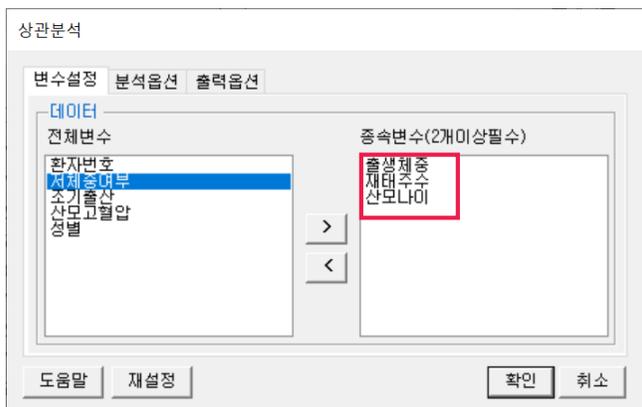
도움말 재설정 확인 취소

메뉴 요소	설명
① 출력 > P-value	상관분석 수행 후 Rex output에 출력되는 교차표(Cross correlation table)에 상관계수의 p-value를 포함합니다.
② 출력 > 공분산	상관분석 수행 후 Rex output에 출력되는 교차표(Cross correlation table)에 종속변수들의 공분산을 포함합니다.
③ 출력 > 샘플수	상관분석 수행 후 Rex output에 출력되는 교차표(Cross correlation table)에 상관분석 시 사용된 샘플수(결측치 제외)를 포함합니다.
④ 그래프 > 산점도행렬	종속변수들의 산점도 행렬을 출력합니다.

• 예제

감마인터페론의 만성 육아종증에 대한 효과를 평가하기 위한 임상시험데이터인 birth 데이터 중, 임신 주수와 출생시 몸무게 데이터이다. 임신주 수(gestwks)와 출생시 몸무게(bweight), 산모나이(matage)의 표본상관계수를 구하라. [데이터 : manual_data.xlsx > birth]

1. [변수설정] > [종속변수] > '출생체중', '재태주수', '산모나이' 지정
2. [분석옵션] > [상관계수 추정방법] > [Pearson] 선택
3. [분석옵션] > [가설 검정] > [양측검정] 선택
4. [출력옵션] > [출력] > [P-value] 체크
5. [출력옵션] > [출력] > [공분산] 체크
6. [출력옵션] > [출력] > [샘플수] 체크
7. [출력옵션] > [그래프] > [산점도 행렬] 체크
8. [확인]



예제 - 결과창

Correlation Analysis

Data Structure

No. of total observations	500
No. of used variable(s)	3

Variable List

Category	Variable	N	N.valid	(%.valid)	N.miss	(%.miss)
Dependent variable	출생체중	500	500	(100.00%)	0	(0.00%)
Dependent variable	재태주수	500	490	(98.00%)	10	(2.00%)
Dependent variable	산모나이	500	500	(100.00%)	0	(0.00%)

Analysis Description

Method	Pearson correlation
Test direction	Two-sided
H1	$r \neq 0$

Analysis Result

출생체중 vs. 재태주수

Result	Value
Pearson correlation	0.7122
T-value	22.4135
Df	488
P-value	0

출생체중 vs. 산모나이

Result	Value
Pearson correlation	0.0149
T-value	0.3329
Df	498
P-value	0.7393

[Data Structure]

- No. of total observations : 전체 개체 수
- No. of used observations : 분석에 사용된 개체 수
- No. of used variable(s) : 분석에 사용된 변수의 수

[Variable List]

- 분석에 사용된 개별 변수들의 유효개수(N.valid), 유효퍼센트(%.valid), 결측수(N.miss), 결측퍼센트(%.miss)가 출력됩니다.

[Analysis Description]

- Method : [분석옵션] > [상관계수 추정방법] 지정 항목 (Pearson's correlation)
- Test direction : [분석옵션] > [가설 검정] 지정 항목 (양측검정)
- H1 : [분석옵션] > [가설 검정]에 의한 대립 가설 (상관계수가 0과 같지 않다)

[Analysis Result]

- 출생체중 vs. 재태주수 : 출생체중과 재태주수의 상관계수는 $r=0.7122$ 로 강한 양의 상관관계를 보입니다. $p<0.0001$ 로 이 상관계수는 통계적으로 0과 유의하게 다르다고 판단됩니다.
- 출생체중 vs. 산모나이 : 출생체중과 산모나이의 상관계수는 $r=0.0149$ 로 매우 약한 양의 상관관계를 보입니다. $p=0.7393$ 으로 이 상관계수는 통계적으로 0과 유의하게 다르다고 할 수 없습니다.

예제 - 결과창

재태주수 vs. 산모나이

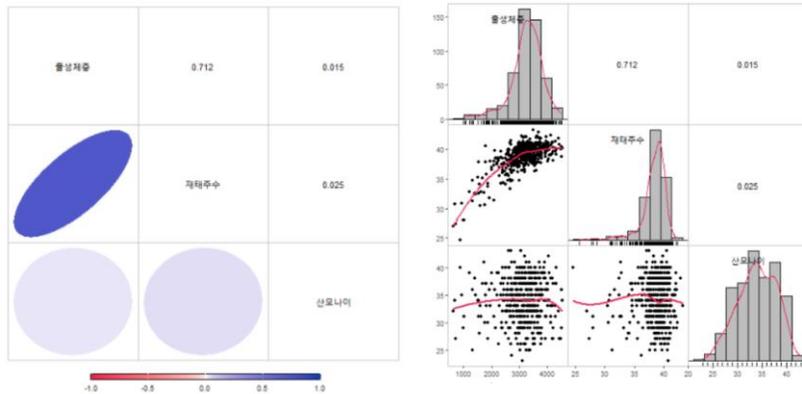
Result	Value
Pearson correlation	0.0249
T-value	0.5492
Df	488
P-value	0.5832

Cross Correlation Table

- Correlation Coefficients
- p-value
- Covariance
- No. of Observations

	출생체중	재태주수	산모나이
출생체중	0.7122 0 1054.8614 490	0.0149 0.7393 37.0674 500	0.0249 0.5832 0.2243 490
재태주수	0.7122 0 1054.8614 490	0.0249 0.5832 0.2243 490	
산모나이	0.0149 0.7393 37.0674 500	0.0249 0.5832 0.2243 490	

Scatter Matrix



[Analysis Result]

- 재태주수 vs 산모나이 : 재태주수와 산모나이의 상관계수는 $r=0.0249$ 로 매우 약한 양의 상관관계를 보입니다. $p=0.5832$ 로 이 상관계수는 통계적으로 0과 유의하게 다르지 않다고 판단됩니다.

[Cross Correlation Table]

- 상관계수행렬로 출력된 결과입니다.
- [출력옵션] > [출력]에서 P-value, 공분산행렬, 샘플수를 체크한 경우 해당 순서대로 출력됩니다.
- 해석은 [Analysis Result]에서와 동일합니다.

[Scatter Matrix]

- [출력옵션] > [그래프] > [산점도행렬]을 선택한 경우 출력됩니다.
- 변수간의 산점도, 상관계수와 각 변수의 히스토그램을 통한 데이터의 분포를 그래프를 통해 살펴볼 수 있습니다.

[Used R Packages]

- 본 분석에 사용된 R 패키지들이 나열되어 있습니다.

Used R Packages

- Main results : 'cor.test' of R package 'stats'
- Analysis of Covariance : 'cov' of R package 'stats'
- All results other than those mentioned above were written with basic functions of R.