



AI기초통계



상관분석과 회귀분석



학습 목표

1. 상관분석
2. 회귀분석



강의 계획



기술통계	<ul style="list-style-type: none">• 숫자로 자료 정리<ul style="list-style-type: none">• 대표값• 산포도• 그림으로 자료 정리
추측통계	<ul style="list-style-type: none">• 모집단분포, 표본분포, 표집분포• 중심극한정리• 가설설정과 유의수준
집단간 차이분석	<ul style="list-style-type: none">• t검정• 일원분산분석• x^2검정
관계분석	<ul style="list-style-type: none">• 상관분석• 회귀분석

상관분석



상관분석

- 두 변인 간의 선형적인 관계를 설명하기 위한 통계적 방법
- 두 변인이 함께 변하는 정도를 의미
- Pearson 적률상관계수
 - ▣ 두 변수 모두 연속적인 양적변수일때 두 변수 간의 관계를 나타내는 상관계수
- Spearman 등위상관계수 – 등간척도
 - ▣ 순서척도에 의한 두 서열변수 사이의 관계를 나타내는 상관계수
- 상관계수 ≠ 인과관계

상관분석

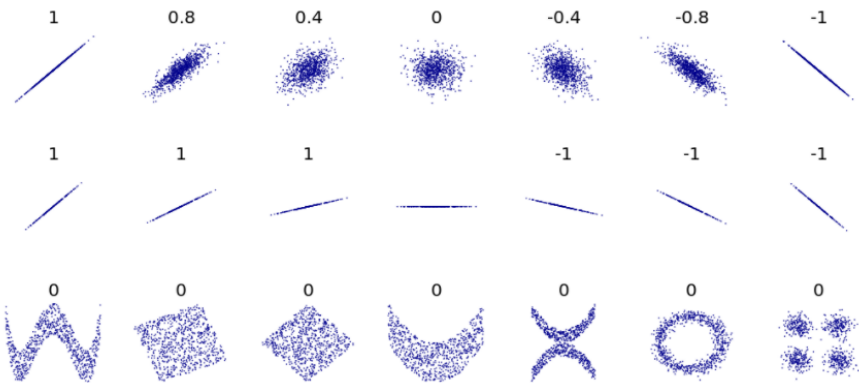
- 1 변수가 변하는 정도 - 분산(variance)
- 2 변수가 변하는 정도 - 공분산(covariance)

상관분석 해석

p137

r 피어슨 상관계수 (correlation coefficient)

- $-1 < r < 1$ 사이 값을 가지며,
- 1이나 -1에 가까워질수록 두 변수가 양 혹은 음의 상관관계가 크다고 말할 수 있고,
- 0에서는 상관관계가 없다

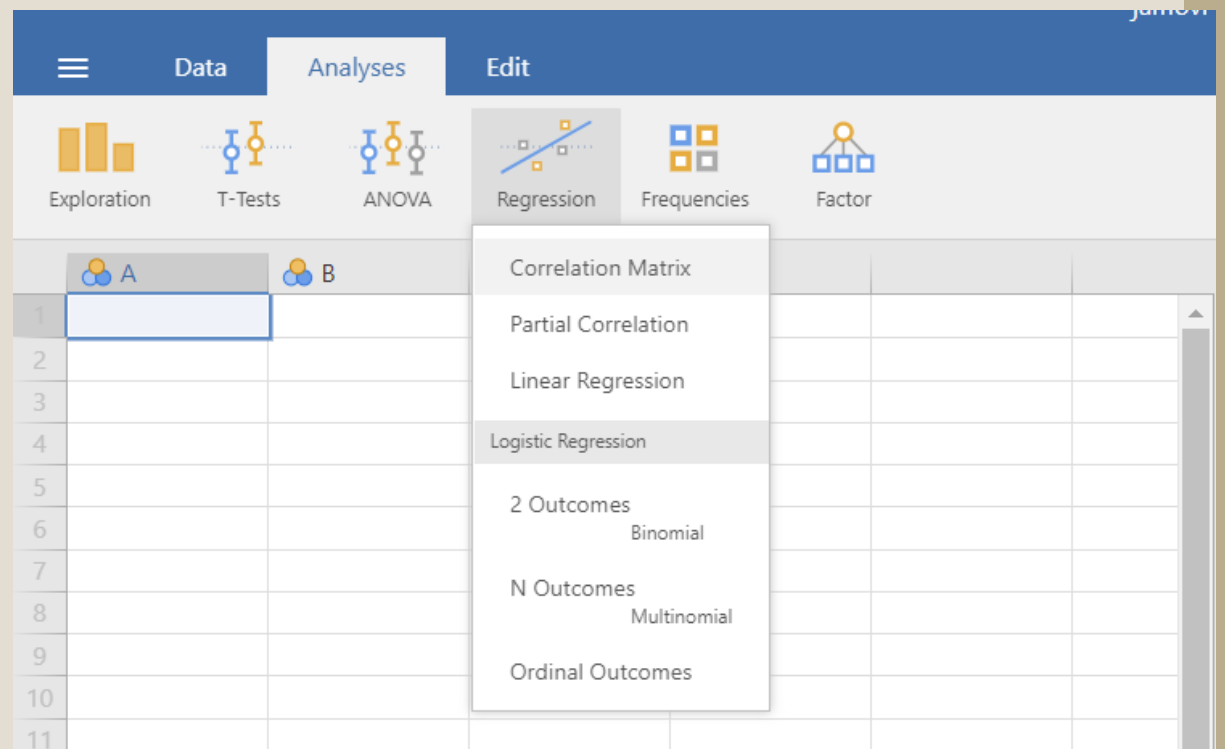


R² 결정계수 (Coefficient of Determination)

- 피어슨 상관계수 r은 제곱하여 R²이라 함.
- $0 < R^2 < 1$ 사이의 값을 가지며 실제값, 예측값, 평균값을 비교하는 분산분석을 통해서도 구할 수 있습니다

상관분석_자모비 실습

- ❑ 실습파일 : big5.omv
- ❑ Analyses – Correlation Matrix 클릭



회귀분석



회귀분석

- 독립변수와 종속변수 사이의 선형식을 구하여 독립변수의 값이 주어졌을 때 종속변수의 값을 예측하고, 종속변수에 대한 독립변수의 예측력(영향력)을 분석하는 방법
- 모형설정, 모형추정, 모형진단의 단계를 거쳐 자료에 가장 적합한 모형을 선택
- 독립변수가 1개 - 단순회귀분석
 - ▣ 종속변수는 양적변수이고, 독립변수는 양적 또는 질적변수임.
 - ▣ 종속변수는 정규분포 가정을 충족하여야 함
- 독립변수가 2개이상 - 중다회귀분석

단순회귀분석

- 종속변수는 양적변수이고, 독립변수는 양적 또는 질적변수임.
- 종속변수는 정규분포 가정을 충족하여야 함
- 1개의 독립변수와 종속변수 간의 선형적 관계를 가정함
- $Y = a + bX$

기본가정

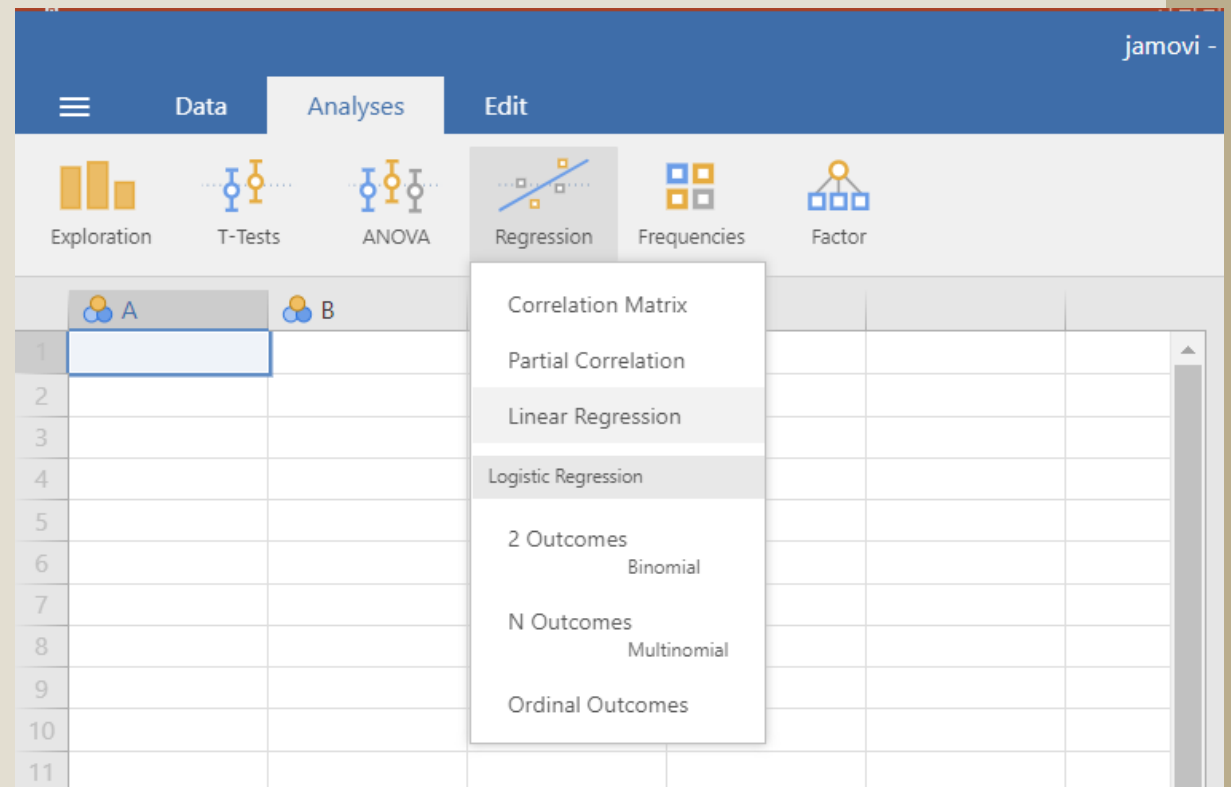
- 선형성
- 동분산성
- 잔차의 정규분포성

회귀분석_자모비 실습

- ❑ 실습파일 : mReg.omv
- ❑ TV : tv시청시간
- ❑ living : 거주(명명변수, 1:부부가 함께 거주, 0:주말부부)
- ❑ age
- ❑ education
- ❑ area

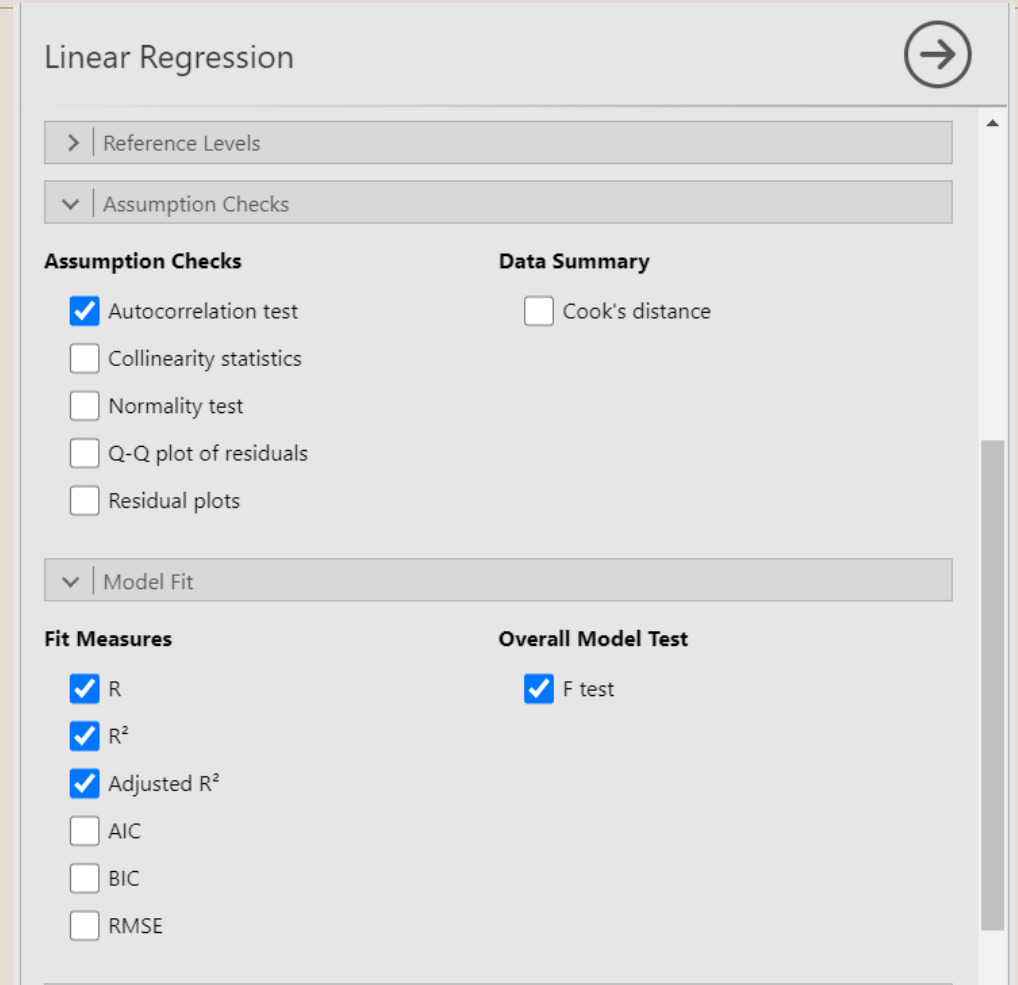
자모비 실습

- Analyses – Linear Regression 클릭



자모비 실습

- Assumption Checks
 - ▣ 잔차의 독립성 검정을 위해 'Autocorrelation test'에 체크
- Model Fit
 - ▣ 모형의 설명력을 확인 : $R, R^2, \text{Adjusted } R^2$
 - ▣ 모형의 적합도 확인 : F test



Linear Regression

> Reference Levels

▼ Assumption Checks

Assumption Checks

- Autocorrelation test
- Collinearity statistics
- Normality test
- Q-Q plot of residuals
- Residual plots

Data Summary

- Cook's distance

▼ Model Fit

Fit Measures

- R
- R^2
- Adjusted R^2
- AIC
- BIC
- RMSE

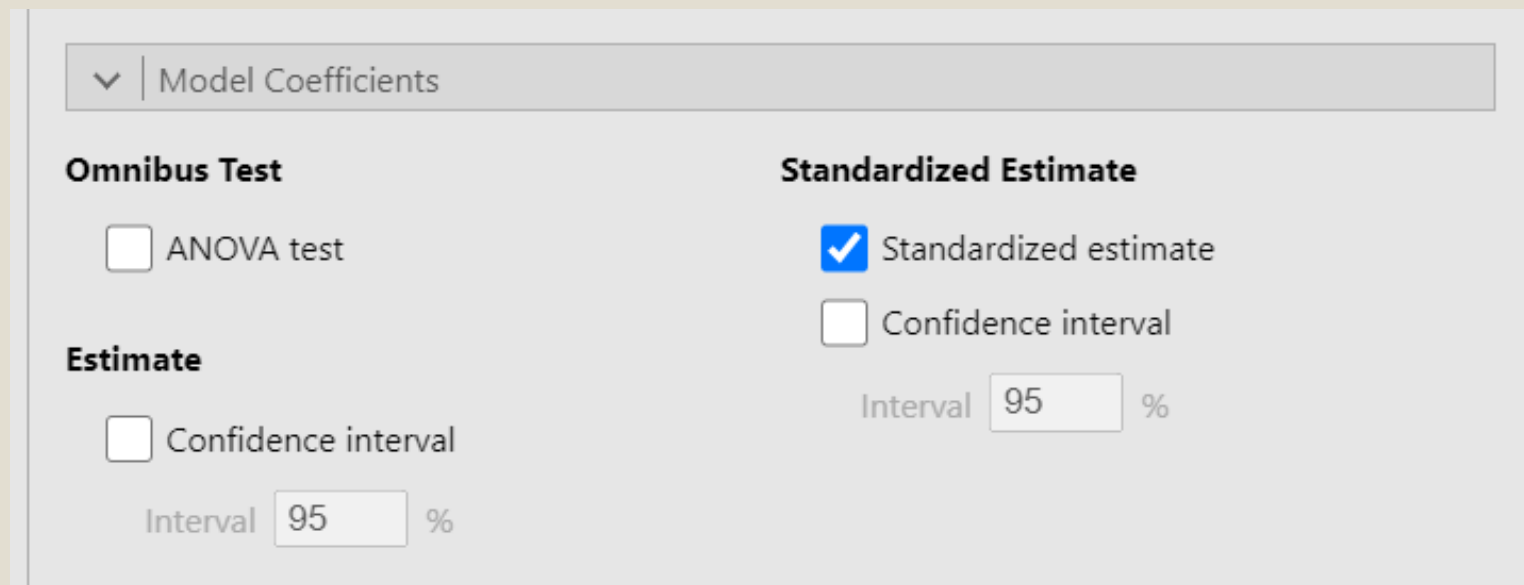
Overall Model Test

- F test

자모비 실습

□ Model Coefficients

- ▣ 표준화 회귀계수를 확인하기 위해 'standardized estimate'에 체크



Model Coefficients

Omnibus Test

ANOVA test

Estimate

Confidence interval
Interval 95 %

Standardized Estimate

Standardized estimate

Confidence interval
Interval 95 %