

가중선형회귀분석

가중선형회귀분석은 종속변수(dependent variable)의 등분산성 가정을 만족하지 못할 때, 가중변수(weighted variable)를 활용하여 선형회귀분석을 수행하는 통계분석 모듈입니다. 선형회귀분석의 중요한 가정 중 하나인 종속변수의 등분산성 가정이 위배될 경우 최적의 모수 추정치를 구할 수 없는데, 이를 보완하기 위하여 가중변수를 사용합니다. 즉, 종속변수와 설명변수(explanatory variable), 가중변수를 적절히 지정하여 선형회귀분석을 실시하게 되는데, 이를 통해 최적의 모수 추정치를 계산할 수 있습니다. 가중선형회귀분석 모듈에서는 변수설정만으로도 최적의 모수 추정치를 손쉽게 계산하고, 변수 선택이 용이하며, 자료분할을 통해 모형의 성능을 평가할 수 있습니다.

메뉴 호출하기

- 회귀분석 > 가중선형회귀분석



- 변수설정 탭

가중선형회귀분석

변수설정 | 자료분할 | 출력옵션 | 변수선택

데이터

전체변수

id
lowbw
preterm
sex

① 종속변수(1개이상필수)

> bweight
<

변수유형

② 질적변수(선택-1개이상가능)

> hyp
<

③ 양적변수(선택-1개이상가능)

> gestwks
matage
<

④ ▼주효과 ⑤ ▼교호작용

⑥ 최종모형

matage
hyp
gestwks

삭제

⑦ 가중변수(필수)

gestwks

⑧ ☐ 최적이중변수 추정

⑨ ☐ 상수항 포함하지 않음

도움말 | 재설정 | 확인 | 취소

메뉴 요소	설명
① 종속변수	모형화하고자 하는 종속변수를 전체변수로부터 선택합니다. 반드시 1개 이상의 양적 변수가 선택되어야 합니다.
② 질적변수	설명변수에 포함된 변수들의 유형을 지정해줍니다. 질적변수로 지정된 변수는 문자로 인식되어 분석에 사용됩니다. 엑셀 시트 상에 가장 먼저 등장하는 수준이 기저범주(reference)로 인식됩니다.
③ 양적변수	설명변수에 포함된 변수들의 유형을 지정해줍니다. 문자형 변수는 선택될 수 없으며, 선택된 경우 분석에서 제외됩니다. 질적변수에 지정된 변수와 중복되어 선택될 수 없습니다.
④ 주효과	[질적변수]와 [양적변수]에서 유형이 지정된 변수를 1개 이상 선택한 상태에서 [주효과] 버튼을 클릭하면, 해당 변수들이 최종모형에 각각 주효과로 포함됩니다.
⑤ 교호작용	[질적변수]와 [양적변수]에서 유형이 지정된 변수를 2개 이상 선택한 상태에서 [교호작용] 버튼을 클릭하면, 해당 변수들의 교호작용이 최종모형에 포함됩니다.
⑥ 최종모형	주효과 또는 교호작용으로 정의된 변수들이 설명변수로 간주되어 모형에 포함됩니다. 포함된 주효과 또는 교호작용 중 삭제하고자 하는 항목이 있는 경우, 해당 항목을 선택한 뒤 [삭제] 버튼을 클릭하면 최종 모형에서 제외됩니다.

- 변수설정 탭

가중선행회귀분석

변수설정 | 자료분할 | 출력옵션 | 변수선택

데이터

전체변수

id
lowbw
preterm
sex

① 종속변수(1개이상필수)

> bweight
<

변수유형

② 질적변수(선택-1개이상가능)

> hyp
<

③ 양적변수(선택-1개이상가능)

> gestwks
matage
<

④ ▼주효과 ⑤ ▼교호작용

⑥ 최종모형

matage
hyp
gestwks

삭제

⑦ 가중변수(필수)

gestwks

⑧ ☐ 최적가중변수 추정

⑨ ☐ 상수항 포함하지 않음

도움말 | 재설정 | 확인 | 취소

메뉴 요소	설명
⑦ 가중변수	[질적변수]와 [양적변수]에서 유형이 지정된 변수들 중 1개를 선택합니다. 질적변수를 지정할 경우 각 수준별로 가중치가 추정되며, 양적변수를 지정할 경우 $1/(\text{양적변수})^2$ 형태로 가중치가 추정됩니다.
⑧ 최적가중변수 추정	[가중변수]에 지정된 변수에 대해 기본적인 방식 외에 최적의 가중치를 새로 추정하고자 하는 경우 선택합니다.
⑨ 상수항 포함하지 않음	모형에 상수항(intercept)을 포함하지 않으려면 이 옵션을 선택합니다. 최종모형에 1개 이상의 설명변수가 포함되지 않은 경우, 이 옵션을 선택할 수 없습니다.

• 자료분할 탭

가중선형회귀분석

변수설정 | **자료분할** | 출력옵션 | 변수선택

변수목록

id
lowbw
preterm
sex

① 훈련 및 검증(필수)

② 분할검증

☒ 모든 데이터를 훈련에 이용
☐ 비율에 따라 임의로 분할
 훈련(train) 자료 %
 시험(test) 자료 %
☐ 변수로 분할
 분할변수(1-훈련, 2-시험)

③ 교차검증

☐ Leave-one-out 교차검증
☒ K-fold 교차검증 K

④ 예측(선택)

분할변수(1-예측, 2-훈련 및 검증)

도움말 | 재설정 | **확인** | 취소

메뉴 요소	설명
① 훈련 및 검증	<p>회귀모형 적합에 사용될 데이터를 훈련자료(training data)와 시험자료(test data)로 분할하는 방식으로 다음 2가지 옵션 중 1개를 선택할 수 있습니다.</p> <ul style="list-style-type: none"> 분할검증 (Default) : 훈련자료와 시험자료로 분할된 자료로 모형을 1회 검증하는 방법입니다. 교차검증 : 훈련자료와 시험자료를 변경해가며 여러 차례 반복 검증하는 방법입니다.
② 분할검증	<p>[분할검증]을 선택하는 경우 다음의 3가지 옵션이 활성화되어 이 중 1개를 선택할 수 있습니다.</p> <ul style="list-style-type: none"> 모든 데이터를 훈련에 이용 (Default) : 시험자료 없이 모든 개체를 회귀모형 적합에 사용합니다. 비율에 따라 임의로 분할 : 훈련자료와 시험자료의 비율을 설정하여 임의로 분할하는 방식입니다. Default 값은 훈련자료 70%, 시험자료가 30% 입니다. 사용자는 훈련자료에 0~100을 입력할 수 있으며, 시험자료에는 100에서 입력한 값을 뺀 수치가 자동으로 입력됩니다. 임의로 분할된 개체들 중 훈련자료와 시험자료의 인덱스를 저장하려면 [출력옵션]-[저장]-[자료분할지표]를 선택합니다. 변수로 분할 : 훈련자료와 시험자료로 사용될 개체가 결정되어 있는 경우 이 옵션을 선택합니다. 이 때, 훈련자료에 해당하는 개체는 1, 시험자료에 해당하는 개체는 2의 값을 갖는 인덱스 변수를 분할변수로 지정해주어야 합니다.

• 자료분할 탭

가중선형회귀분석

변수설정 | **자료분할** | 출력옵션 | 변수선택

변수목록

id
lowbw
preterm
sex

① 훈련 및 검증(필수)

② 분할검증

☒ 모든 데이터를 훈련에 이용
☐ 비율에 따라 임의로 분할
 훈련(train) 자료 %
 시험(test) 자료 %
☐ 변수로 분할
 분할변수(1-훈련, 2-시험)

③ 교차검증

☐ Leave-one-out 교차검증
☒ K-fold 교차검증 K

④ 예측(선택)

분할변수(1-예측, 2-훈련 및 검증)

도움말 | 재설정 | **확인** | 취소

메뉴 요소	설명
③ 교차검증	<p>[교차검증]을 선택하는 경우 다음의 2가지 옵션이 활성화되어 이 중 1개를 선택할 수 있습니다.</p> <ul style="list-style-type: none"> Leave-one-out 교차검증 : 한 개체를 시험자료로 사용하고 나머지 개체를 모두 훈련자료로 하여 모델을 적합하는 방식으로 모든 개체에 대해 이 과정을 반복한 뒤, 전체 개체 수만큼의 모형으로부터 얻은 정확도의 평균을 모형의 최종 정확도로 계산합니다. K-fold 교차검증 : 전체 개체를 K개의 그룹으로 임의로 분할하여, 하나의 그룹을 시험자료로 사용하고 나머지 그룹을 모두 훈련자료로 하여 모델을 적합하는 방식으로 K개의 그룹에 대해 이 과정을 반복한 뒤, 그룹 수만큼의 모형으로부터 얻은 정확도의 평균을 모형의 최종 정확도로 계산합니다. <p>- K : [교차검증]-[K-fold 교차검증]을 선택할 경우 활성화됩니다. K-fold 교차검증에 사용할 K의 값을 입력합니다. 2 이상의 정수만 입력 가능하며, 전체 개체 수보다 더 큰 정수가 입력되는 경우 자동으로 Leave-one-out 교차검증을 실시합니다. Default는 10입니다.</p>
④ 예측 > 분할변수	<p>회귀모형 적합에 사용될 훈련 및 검증 데이터와 해당 모형으로부터 예측값을 얻을 예측 데이터가 분할되어 있는 경우 사용됩니다. 훈련 및 검증에 사용되는 개체는 2, 예측에 사용되는 개체는 1의 값을 갖는 인덱스 변수를 분할변수로 지정해주어야 합니다. 예측분할변수를 지정하지 않아도 분석이 가능합니다. 예측분할변수가 지정된 경우, 예측에 해당하는 개체에 해당하는 예측값이 엑셀 시트에 "Predicted_pred_WLM"라는 변수명으로 저장됩니다.</p>

• 출력옵션 탭

가중선형회귀분석

변수설정 | 자료분할 | **출력옵션** | 변수선택

출력

① 회귀계수

☒ 신뢰구간 ☐ 분산팽창지수(VIF)
신뢰수준

② ☒ 분산분석표

제공합유형

☐ Type I ☐ Type II ☒ Type III

③ ☐ 적합도검정 ④ ☐ 잔차진단그래프

⑤ 요약표

☐ 단순가중선형회귀 추가

저장

⑥ 훈련자료

☐ 적합값 ☐ 비표준화 잔차 ☐ 쿼의 거리
☐ 신뢰구간 ☐ 표준화 잔차 ☐ 헤트 행렬의 대각원소
☐ 예측구간 ☐ 스튜던트화 잔차 ☐ (최종)가중변수
신뢰수준

⑦ 시험자료

☐ 예측값 ⑧ 예측자료

☐ 신뢰구간 ☐ 예측값의 신뢰구간
☐ 예측구간 ☐ 예측값의 예측구간
신뢰수준

⑨ ☐ 자료분할지표

도움말 | 재설정 | **확인** | 취소

메뉴 요소	설명
① 회귀계수	<ul style="list-style-type: none"> 신뢰구간 : 회귀계수 추정값의 신뢰구간을 출력합니다. 신뢰수준 : [신뢰구간] 옵션이 선택되는 경우 활성화됩니다. 신뢰수준을 입력합니다. 0에서 1 사이의 값을 입력할 수 있으며, Default는 0.95입니다. 분산팽창지수 (VIF) : 설명변수들의 다중공선성(multicollinearity)을 진단하는 지표인 분산팽창지수(variance inflation factor, VIF)를 출력합니다.
② 분산분석표	<p>회귀모형식에 대한 분산분석표(ANOVA table)이 출력됩니다. 종속변수의 변동량에 대한 회귀모형의 설명력을 판단하기 위한 제공합 계산 방식 3가지 중 하나를 선택할 수 있습니다.</p> <ul style="list-style-type: none"> Type I : 설명변수를 순차적으로 하나씩 추가하면서 제공합의 증가량을 계산하는 방식입니다. Type II : 전체 회귀모형에서 주효과를 하나씩 제거하면서 제공합의 감소량을 계산하는 방식입니다. 최종모형이 주효과만으로 구성된 경우에 적합합니다. Type III (Default) : 전체 회귀모형에서 주효과와 교호작용을 포함한 모든 효과를 하나씩 제거하면서 제공합의 감소량을 계산하는 방식입니다. 최종모형에 교호작용도 포함된 경우 적합합니다.

• 출력옵션 탭

가중선행회귀분석

변수설정 | 자료분할 | **출력옵션** | 변수선택

출력

① 회귀계수
☒ 신뢰구간 ☐ 분산팽창지수(VIF)
신뢰수준

② 분산분석표
제공할유형
☐ Type I ☐ Type II ☒ Type III

③ 적합도검정 ④ 잔차진단그래프

⑤ 요약표
☐ 단순가중선행회귀 추가

저장

⑥ 훈련자료
☐ 적합값 ☐ 비표준화 잔차 ☐ 쿡의 거리
☐ 신뢰구간 ☐ 표준화 잔차 ☐ 헤트 행렬의 대각원소
☐ 예측구간 ☐ 스튜던트화 잔차 ☐ (최종)가중변수
신뢰수준

⑦ 시험자료 ⑧ 예측자료

☐ 예측값
☐ 신뢰구간 ☐ 예측값의 신뢰구간
☐ 예측구간 ☐ 예측값의 예측구간
신뢰수준

⑨ 자료분할지표

도움말 | 재설정 | **확인** | 취소

메뉴 요소	설명
③ 적합도검정	주어진 데이터에 대해 최종모형이 얼마나 적합한지 확인하기 위한 적합도 통계량을 계산하고, 적합도 검정을 수행합니다. <ul style="list-style-type: none"> 적합도 통계량 : Deviance, Pearson's chi-square, -2(log-likelihood), AIC, BIC (2개 이상의 모형에 대해 통계량이 작을수록 적합도가 더 좋다고 판단합니다) 적합도 검정 : 설명변수가 없는 NULL 모형 대비 현재의 최종모형이 통계적으로 유의한 차이를 보이는지 검정하는 Likelihood Ratio test (LRT)를 수행합니다.
④ 잔차진단그래프	회귀모형에 대한 회귀진단을 위한 잔차진단 그래프 6가지가 출력됩니다. <ul style="list-style-type: none"> Residuals vs Fitted : 적합값 대비 잔차의 그래프 Normal Q-Q : 잔차의 QQ그림 Scale-Location : 적합값 대비 표준화잔차의 그래프 Cook's distance : 개체별 쿡의 거리 그래프 Constant Leverage : 요인별 표준화잔차의 그래프 Cook's distance vs Leverage : 쿡의 거리와 지렛값의 산점도
⑤ 단순가중선행회귀 추가	[변수설정] 탭에서 2개 이상의 설명변수가 최종모형에 포함된 경우 이 옵션이 활성화됩니다. 이 옵션을 선택할 경우, 최종모형에 포함된 2개 이상의 설명변수에 대해 단순회귀모형을 적합한 회귀계수표가 출력됩니다.

• 출력옵션 탭

가중선행회귀분석

변수설정 | 자료분할 | **출력옵션** | 변수선택

출력

① 회귀계수
☒ 신뢰구간 ☐ 분산팽창지수(VIF)
신뢰수준

② ☒ 분산분석표
제공할유형
☐ Type I ☐ Type II ☒ Type III

③ ☐ 적합도검정 ④ ☐ 잔차진단그래프

⑤ 요약표
☐ 단순가중선행회귀 추가

저장

⑥ **훈련자료**
☐ 적합값 ☐ 비표준화 잔차 ☐ 쿡의 거리
☐ 신뢰구간 ☐ 표준화 잔차 ☐ 해트 행렬의 대각원소
☐ 예측구간 ☐ 스튜던트화 잔차 ☐ (최종)가중변수
신뢰수준

⑦ 시험자료 ⑧ 예측자료
☐ 예측값 ☐ 예측값의 신뢰구간
☐ 신뢰구간 ☐ 예측값의 예측구간
☐ 예측구간 신뢰수준

⑨ ☐ 자료분할지표

도움말 | 재설정 | **확인** | 취소

메뉴 요소	설명
⑥ 훈련자료	<p>최종모형 적합에 사용된 훈련자료에 대하여 다음 중 선택되는 통계량을 엑셀시트에 저장합니다. 괄호 안에 표기된 변수명으로 저장됩니다.</p> <ul style="list-style-type: none"> 적합값 : 최종모형으로 예측된 적합값 (Fitted_train_WLM) 신뢰구간 : [적합값]이 선택된 경우 활성화. 적합값의 신뢰구간 (Fitted_95CI_Lower_train_WLM / Fitted_95CI_Upper_train_WLM) 예측구간 : [적합값]이 선택된 경우 활성화. 적합값의 예측구간 (Fitted_95PI_Lower_train_WLM / Fitted_95PI_Upper_train_WLM) 신뢰수준 : [신뢰구간] 혹은 [예측구간]을 선택한 경우 활성화됩니다. 신뢰구간과 예측구간 계산 시 사용되는 신뢰수준입니다. 0과 1 사이 값을 입력할 수 있으며, Default는 0.95입니다. 비표준화 잔차 : 적합값과 실제값의 차이 (unstdResid_train_WLM) 표준화 잔차 : 잔차를 표준편차로 나눈 값 (stdResid_train_WLM) 스튜던트화 잔차 : 해당 개체를 제외한 상태에서 계산된 표준편차로 잔차를 나눈 값 (studResid_train_WLM) 쿡의 거리 : 개별 개체들이 모형에 미치는 영향력을 평가하기 위해, 잔차와 지렛값을 동시에 고려한 척도 (CookDist_train_WLM) 해트 행렬의 대각원소 : 해당 개체와 나머지 개체의 평균의 차이인 지렛값 (HatValue_train_WLM) (최종)가중변수 : 최종적으로 사용된 가중치 (weightVar_WLM)

• 출력옵션 탭

가중선형회귀분석

변수설정 | 자료분할 | **출력옵션** | 변수선택

출력

① 회귀계수
☒ 신뢰구간 ☐ 분산팽창지수(VIF)
신뢰수준

② 분산분석표
제공합유형
☐ Type I ☐ Type II ☒ Type III

③ 적합도검정 ④ 잔차진단그래프

⑤ 요약표
☐ 단순가중선형회귀 추가

저장

⑥ 훈련자료
☐ 적합값 ☐ 비표준화 잔차 ☐ 쿡의 거리
☐ 신뢰구간 ☐ 표준화 잔차 ☐ 헤트 행렬의 대각원소
☐ 예측구간 ☐ 스튜던트화 잔차 ☐ (최종)가중변수
신뢰수준

⑦ 시험자료 ⑧ 예측자료

☐ 예측값 ☐ 예측값의 신뢰구간
☐ 신뢰구간 ☐ 예측값의 예측구간
☐ 예측구간 신뢰수준

⑨ 자료분할지표

도움말 | 재설정 | **확인** | 취소

메뉴 요소	설명
⑦ 시험자료	<p>[자료분할] 탭에서 '분할검증'이 선택되고, [비율에 따라 임의로 분할] 또는 [변수로 분할]이 선택된 경우 활성화됩니다. 시험자료에 대하여 다음 중 선택되는 통계량을 엑셀시트에 저장합니다. 괄호 안에 표기된 변수명으로 저장됩니다.</p> <ul style="list-style-type: none"> 예측값 : 최종모형으로 예측된 값 (Predicted_testing_WLM) 신뢰구간 : [예측값]이 선택된 경우 활성화. 예측값의 신뢰구간 (Predicted_95CI_Lower_testing_WLM / Predicted_95CI_Upper_testing_WLM) 예측구간 : [예측값]이 선택된 경우 활성화. 예측값의 예측구간 (Predicted_95PI_Lower_testing_WLM / Predicted_95PI_Upper_testing_WLM) 신뢰수준 : [신뢰구간] 혹은 [예측구간]을 선택한 경우 활성화됩니다. 신뢰구간과 예측구간 계산 시 사용되는 신뢰수준입니다. 0과 1 사이 값을 입력할 수 있으며, Default는 0.95입니다.

출력옵션 탭

가중선형회귀분석

변수설정

자료분할

출력옵션

변수선택

출력

회귀계수

☒ 신뢰구간
신뢰수준 0.95

☐ 분산팽창지수(VIF)

분산분석표

제공할유형
☐ Type I ☐ Type II ☒ Type III

☐ 적합도검정

☐ 잔차진단그래프

요약표

☐ 단순가중선형회귀 추가

저장

훈련자료

☐ 적합값

☐ 신뢰구간

☐ 예측구간

신뢰수준

☐ 비표준화 잔차

☐ 표준화 잔차

☐ 스튜던트화 잔차

☐ 쿼의 거리

☐ 헤트 행렬의 대각원소

☐ (최종)가중변수

시험자료

☐ 예측값

☐ 신뢰구간

☐ 예측구간

신뢰수준

☐ 예측값의 신뢰구간

☐ 예측값의 예측구간

신뢰수준

예측자료

☐ 예측값

☐ 신뢰구간

☐ 예측구간

신뢰수준

☐ 예측값의 신뢰구간

☐ 예측값의 예측구간

신뢰수준

자료분할지표

도움말

재설정

확인

취소

메뉴 요소	설명
⑧ 예측자료	<p>[자료분할] 탭에서 '예측 분할변수'가 선택된 경우 활성화됩니다. 예측자료에 대하여 다음 중 선택되는 통계량을 엑셀시트에 저장합니다. 괄호 안에 표기된 변수명으로 저장합니다.</p> <ul style="list-style-type: none">예측값의 신뢰구간 : 예측값의 신뢰구간 (Predicted_95CI_Lower_pred_WLM / Predicted_95CI_Upper_pred_WLM)예측구간 : [예측값]이 선택된 경우 활성화. 예측값의 예측구간 (Predicted_95PI_Lower_pred_WLM / Predicted_95PI_Upper_pred_WLM)신뢰수준 : [예측값의 신뢰구간] 혹은 [예측값의 예측구간]을 선택한 경우 활성화됩니다. 신뢰수준을 입력합니다. 0과 1 사이 값을 입력할 수 있으며, Default는 0.95입니다.
⑨ 자료분할지표	<p>[자료분할] 탭에서 '분할검증'이 선택된 경우, 분할된 개체들 중 훈련자료와 시험자료의 인덱스를 엑셀 시트에 저장합니다. 저장된 변수명은 "Partition_idx_WLM"으로 훈련자료의 경우 'training', 시험자료의 경우 'testing'의 인덱스를 갖습니다.</p>

• 변수선택 탭

가중선형회귀분석

변수설정 | 자료분할 | 출력옵션 | **변수선택**

① ☒ 변수선택

② 변수선택방법

☒ 전진선택법 ☐ 후진제거법 ☐ 단계별선택법

③ 고정변수

설명변수

matage
hyp

gestwks

>

<

도움말 | 재설정 | **확인** | 취소

메뉴 요소	설명
① 변수선택	[변수설정] 탭에서 2개 이상의 설명변수가 최종모형에 포함된 경우 이 옵션이 활성화됩니다. 이 옵션을 선택할 경우, AIC(Akaike Information Criterion)을 기반으로 최적의 모형에 해당하는 변수를 자동으로 선택합니다. 최적의 모형은 Rex output의 Variable Selection에서 확인할 수 있습니다.
② 변수선택방법	[변수선택]을 선택한 경우 활성화됩니다. 다음 3가지 옵션 중 1개를 선택할 수 있습니다. <ul style="list-style-type: none"> • 전진선택법 (Default) : 상수항만 있는 모형에서 시작하여 최종모형에 선택된 설명변수들 중 종속변수에 가장 큰 영향력을 갖는 변수부터 하나씩 차례대로 추가하면서 최적의 모형을 선택하는 방식입니다. • 후진제거법 : 최종모형에 선택된 설명변수들을 모두 포함된 모형에서 시작하여 종속변수에 가장 작은 영향력을 갖는 변수부터 하나씩 차례대로 제거하면서 최적의 모형을 선택하는 방식입니다. • 단계별선택법 : 상수항만 있는 모형에서 시작하여 최종모형에 선택된 설명변수들 중 하나씩 모형에 추가하면서 매번 모든 설명변수를 평가하는 방식으로 전진선택법과 후진선택법을 반복적으로 수행하는 방식입니다.
③ 고정변수	[변수설정] 탭에서 선택된 2개 이상의 변수가 나열된 설명변수 창에서 변수선택과정과 상관없이 최종 모형에 반드시 포함되어야 하는 변수들을 고정변수로 지정합니다.