

glmnetUI: glmnet(Lasso 및 Elastic Net) 패키지를 위한 대화형 인터페이스 (한국어판)

한국어 번역본. 인공지능의 도움을 받아 번역하고 편집부가 검토했습니다. 영문판이正本입니다.

William Bert Craytor

제1권 제1호 · 2026년 여름호

발행일: 2026년 7월 1일

DOI: [10.67330/ja8s2k39](https://doi.org/10.67330/ja8s2k39)

© 2026 William Bert Craytor. Published by Pacific Vista Net.

이 논문은 오픈 액세스 논문으로 [크리에이티브 커먼즈 저작자표시 4.0 국제 라이선스\(CC BY 4.0\)](#)의 조건에 따라 배포됩니다.

초록. glmnetUI는 R의 glmnet 패키지를 위한 그래픽 사용자 인터페이스(graphical user interface)이며, glmnet은 Lasso, 릿지(ridge), 엘라스틱 넷(elastic net) 벌점(penalty)을 통해 정규화(regularization)된 일반화 선형모형(generalized linear model)을 적합한다. glmnetUI는 세 가지 용도 모드—범용 예측 모형화, 부동산 감정평가, 시장지역 분석(market-area analysis)—를 제공하며, 데이터 가져오기, 모형 구성, 교차검증(cross-validation) 적합, 계수 경로(coefficient path) 확인, 보고서 내려받기까지 사용자를 안내한다. 본 논문은 glmnetUI의 데이터 형식 요건, 모형화 작업 흐름, 출력 화면, 그리고 전체 기능 참조 사항을 기술한다.

키워드: Lasso, 릿지 회귀, 엘라스틱 넷, 정규화 회귀, glmnet, R, 그래픽 사용자 인터페이스, 부동산 감정평가

발행인 주(2026년 7월): 인쇄 직전에 glmnetUI는 earthUI(0.11.0 이후 버전)에 통합되었으며, 본 논문에서 설명하는 엘라스틱 넷 작업 흐름은 이제 MARS 및 GAM 작업 흐름과 함께 earthUI가 제공한다. glmnetUI는 최종 배포판(0.5.0) 형태로 <https://github.com/wcraytor/glmnetUI>에 계속 공개되어 있으며, 향후 개발은 earthUI에서 이어진다. 본 논문은 배포된 상태의 glmnetUI에 적용된다.

목차

1	서론	3
1.1	역사와 배경	3
1.1.1	Lasso	3
1.1.2	릿지 회귀와 엘라스틱 넷	4
1.1.3	glmnet 패키지	4
1.1.4	MARS(earth) 및 GAM(mgcv)과의 비교	4
1.1.5	권장 작업 흐름	5
1.2	glmnetUI란 무엇인가?	5
1.3	세 가지 용도 모드	6
1.4	부동산 전용 기능	6
1.5	시작하기	7
2	MLS 입력 데이터 요건	8
2.1	파일 형식과 구조	8
2.2	감정평가 모드에 필요한 열	8
2.3	특수 열 명명 규칙	9
2.4	데이터 품질과 완전성	9
2.5	대상 행의 위치	9

3	General Purpose 모드	10
3.1	개요	10
3.1.1	Skip First Row	10
3.1.2	용도별 설정	10
3.2	사이드바 작업 흐름	10
3.3	프로젝트	11
3.4	earthUI에서 가져오기 (선택 사항)	11
3.5	주 패널 탭	11
3.6	설정의 보존	12
3.7	다크 모드	12
3.8	로케일과 지역 설정	12
3.8.1	기본값 저장	13
4	Appraisal 모드	14
4.1	대상 행의 처리	14
4.2	기준일과 매매 경과일	14
4.3	특수 열 지정	14
4.4	RCA 보정 개요	14
5	Market Area Analysis 모드	15
5.1	Appraisal 모드와의 차이	15
5.2	시장 모드를 쓸 때	15
6	변수 선택	16
6.1	목표변수	16
6.2	설명변수 표	16
6.3	자료형 감지와 재지정	16
6.4	기대 부호와 적용	16
6.5	특수 열 유형 참조	17
7	모수 선택	18
7.1	Alpha(혼합 모수)	18
7.2	Lambda 선택	18
7.3	Family	18
7.4	Standardize	19
7.5	부호 제약	19
7.6	완화 Lasso	19
7.7	난수 시드	19
7.8	교호작용 행렬	19
7.9	주효과에서 차단(교호작용 전용)	20
7.10	고급 모수	20
7.11	설정 기본값	20
8	모형 적합	22
8.1	Fit 버튼	22
8.2	적합 중에 일어나는 일	22
9	결과 탭	23
9.1	Data Preview	23
9.2	Equation	23
9.3	Correlation	23
9.4	Summary	23

9.5	Coefficients	23
9.6	Variable Importance	24
9.7	Contributions	24
9.8	ANOVA	24
9.9	Diagnostics	24
9.10	Glmnet Output	25
9.11	Report	25
10	데이터 내려받기	26
10.1	출력 열	26
10.2	CQA 점수	26
11	RCA 계산과 내려받기	27
11.1	RCA 대화상자 열기	27
11.2	CQA 점수 보간	27
11.3	출력 열	27
12	거래사례비교표	28
12.1	비교 사례 선택 대화상자	28
12.2	통합문서 구성	28
12.3	시트 보호와 잔여 칸	28
13	보고서 내려받기	29
13.1	두 단계 Quarto 작업 흐름	29
13.2	보고서 형식	29
13.3	보고서 내용	29
14	earthUI와의 비교	31
14.1	주요 차이	31
14.2	공통 기능	31
14.3	어느 것을 언제 쓸 것인가	31
15	예제 데이터셋: Appraisal_1.csv	32
15.1	설명	32
15.2	열	32
15.3	권장 빠른 시작	32
16	시스템 요구사항과 문제 해결	34
16.1	지원 플랫폼	34
16.2	선택적 의존 요소	34
16.3	플랫폼별 참고사항	34
16.4	점진적 기능 축소	34
16.5	문제 해결	34

1 서론

1.1 역사와 배경

1.1.1 Lasso

Lasso(Least Absolute Shrinkage and Selection Operator)는 1996년 Robert Tibshirani가 제안하였다. 그 이름은 카우보이의 올가미(lasso)—구속하고 골라내는 밧줄—를 의도적으로 빗댄 말장

난이다. Tibshirani가 이 이름을 택한 것은, 이 방법이 문자 그대로 계수를 “올가미로 낚아채어” 절댓값을 제약하고 일부를 정확히 0으로 끌어당김으로써 어떤 변수가 모형에 남을지를 선택하기 때문이다.

Tibshirani의 착안은 최소제곱 목적함수에 L_1 벌점(penalty, 계수 절댓값의 합)을 더하는 것이었다. 계수를 0 쪽으로 축소하되 결코 0에 도달하지는 않는 릿지 회귀(ridge regression; Hoerl & Kennard, 1970)의 L_2 벌점과 달리, L_1 벌점은 많은 계수가 정확히 0이 되는 희소(sparse) 해를 만들어 낸다. 이 덕분에 Lasso는 정규화(regularization) 방법이면서 동시에 변수 선택 방법이기도 하다.

1.1.2 릿지 회귀와 엘라스틱 넷

릿지 회귀(ridge regression)는 1970년 Arthur Hoerl과 Robert Kennard가 제안하였다. 이 방법은 L_2 벌점(계수 제곱의 합)을 더하여 모든 계수를 비례적으로 축소하되 모든 변수를 모형에 남겨 둔다. 릿지는 설명변수들이 상관되어 있을 때(다중공선성) 효과적이지만 모형을 단순화하지는 않는다.

엘라스틱 넷(elastic net)은 2005년 Hui Zou와 Trevor Hastie가 제안한 것으로 두 벌점을 결합한다. 혼합 모수 α 가 그 배합을 조절하는데, $\alpha = 1$ 은 순수한 Lasso, $\alpha = 0$ 은 순수한 릿지이며, 그 사이의 값은 절충안을 준다. 엘라스틱 넷은 Lasso의 한계를 극복한다. 설명변수들이 서로 강하게 상관되어 있을 때 Lasso는 하나만 선택하고 나머지를 무시하는 경향이 있는 반면, 엘라스틱 넷은 상관된 설명변수들을 함께 묶어 준다.

1.1.3 glmnet 패키지

R 패키지 **glmnet**은 스탠퍼드 대학교의 Jerome Friedman, Trevor Hastie, Robert Tibshirani가 개발하였다. 이 패키지는 매우 효율적인 좌표하강(coordinate descent) 알고리즘을 사용하여 일반화 선형모형 (generalized linear model)에 대한 엘라스틱 넷 정규화를 구현하며 (Friedman, Hastie, and Tibshirani 2010), 이후 Cox 비례위험모형 (Simon et al. 2011)과 모든 일반화 선형모형 계열 (Tay, Narasimhan, and Hastie 2023)로 확장되었다. 이 패키지는 가우시안(선형), 이항(로지스틱), 포아송을 비롯한 여러 GLM 계열을 지원한다.

glmnet의 주요 기능은 다음과 같다.

- **정규화 경로(regularization path)**: 람다(lambda) 값의 격자 전체에 대해 한 번의 계산으로 모형을 적합하므로, 각 람다마다 따로 적합하는 것보다 훨씬 빠르다.
- **교차검증(cross-validation)**: `cv.glmnet()`이 k -겹 교차검증을 통해 최적의 정규화 강도를 자동으로 선택한다.
- **계수 한계(coefficient bounds)**: 개별 계수에 대한 상한과 하한 (부호 제약)을 지원한다.
- **완화 Lasso(relaxed lasso)**: 선택된 변수만으로 벌점을 줄이거나 없앤 상태에서 다시 적합하여 편향을 줄인다.

1.1.4 MARS(earth) 및 GAM(mgcv)과의 비교

earth 애플리케이션과 그 두 개의 후처리 모듈(Post Processing Module, **glmnet**과 **mgcv**)은 서로 다른 모형화 엔진을 사용한다. 다음 표는 주요 차이를 요약한 것이다.

항목	glmnet (엘라스틱 넷)	earth (MARS)	mgcv (GAM)
모형 유형	선형(정규화 포함)	조각별 선형(적응적 스플라인)	매끄러운 비선형(별점 스플라인)
비선형성	교호작용이나 기저 확장을 통해서만 가능	데이터 기반 매듭을 갖는 힌지 함수로 자동 처리	매끄러운 함수(박판, 3차 등)로 자동 처리
변수 선택	L_1 별점(Lasso)으로 내장	GCV 가지치기를 동반한 전진/후진 단계적 선택으로 내장	이중 별점 축소로 선택적 지원(select=TRUE)
교호작용	쌍별 교차곱(해석하기 어려움)	힌지 기반 곱(해석 가능, 3차까지)	텐서곱 평활(te(), ti())
해석 가능성	계수가 선형 가중치이므로 개별 설명이 쉬움	g-함수가 변수별 조각별 선형 부분효과를 보여 줌	매끄러운 부분효과 곡선으로 매우 직관적
적합한 상황	고차원 데이터, 변수 선택, 소표본	비선형성과 교호작용의 자동 탐지	매끄러운 비선형 관계, 유연한 모형화
RCA 보정	변수별 선형 보정	g-함수를 통한 조각별 선형 보정	부분효과를 통한 매끄러운 보정
과적합 제어	교차검증된 람다	GCV 기반 가지치기	REML/GCV 평활모수 추정

MARS(Multivariate Adaptive Regression Splines, 다변량 적응 회귀 스플라인)는 1991년 Jerome Friedman이 제안하였다. MARS는 힌지 함수 $\max(0, x - k)$ 와 그 매듭 위치를 데이터로부터 적응적으로 선택하여 조각별 선형모형을 구성한다. R 구현은 Stephen Milborrow의 **earth** 패키지이다. MARS는 사용자가 따로 지정하지 않아도 비선형 관계와 교호작용을 자동으로 찾아내는 데 탁월하다. 그 g-함수(변수별로 묶인 기저 항)는 매우 해석하기 쉬운 부분효과 곡선을 제공한다.

GAM(Generalized Additive Model, 일반화 가법모형)은 Hastie와 Tibshirani(1990)가 정식화하였다. R 구현인 Simon Wood의 **mgcv**는 REML 또는 GCV를 통한 자동 평활도 선택 기능을 갖춘 별점 회귀 스플라인을 제공한다. GAM은 시각화하고 설명하기에 직관적인 매끄러운 부분효과 곡선을 산출한다. **earthUI**가 매듭 위치를 **mgcvUI**로 내보내면, **earth**에서 도출된 매듭이 GAM 평활항의 출발점 역할을 하여 MARS의 적응적 매듭 배치와 GAM의 매끄러운 추정을 결합하게 된다.

1.1.5 권장 작업 흐름

해석 가능하고 방어 가능한 모형이 필요한 부동산 감정평가 및 이와 유사한 용도에서는 다음을 권장한다.

1. **earthUI**로 시작하여 데이터에서 중요한 변수, 비선형 관계, 교호작용을 발견한다. **earth**의 자동 기저 선택과 g-함수는 훌륭한 초기 통찰을 제공한다.
2. 더 매끄러운 부분효과를 원한다면 **mgcvUI**로 정교화한다. **earthUI**의 매듭을 **mgcvUI**로 가져오면 조각별 선형모형에서 매끄러운 모형으로 매끄럽게 넘어갈 수 있다.
3. 공격적인 변수 선택(Lasso)이 필요할 때, 설명변수 개수가 관측치 개수에 근접하거나 이를 넘어설 때, 또는 방어 가능성을 위해 정규화된 순수 선형모형을 원할 때 **glmnetUI**를 사용한다.

1.2 glmnetUI란 무엇인가?

glmnetUI는 R **glmnet** 패키지를 위한 그래픽 사용자 인터페이스이다. 로컬 Shiny 애플리케이션으로 실행되며—로그인도, 서버도, 계정도 없다. R에서 실행한 뒤 데이터셋(CSV 또는 Excel)을 가져오고, 모형을 구성하고, 대화형으로 적합한다.

현재 배포판 기준으로 glmnetUI는 더 이상 별도로 설치하는 패키지가 아니다. 이제 earthUI 패키지 내부에 엘라스틱 넷 후처리 모듈(Post Processing Module)로 포함되어 `earthUI::launch_glmnet()`으로 시작한다. 애플리케이션 자체—그리고 본 논문에서 설명하는 모든 내용—는 변함이 없으며, 설치 방법과 실행 명령만 달라졌다. 세 루틴(earth, glmnet, mgcv)은 하나의 프로젝트 트리, 하나의 설정 데이터베이스, 하나의 지원 인프라를 공유한다.

이 애플리케이션은 데이터 가져오기, 변수 구성, 모형 적합, 진단 그림, 변수 중요도, 모형 방정식, 그리고 Word, PDF, HTML 형식으로 내려받을 수 있는 보고서에 이르는 완결된 작업 흐름을 제공한다.

엘라스틱 넷 회귀는 두 가지 정규화 기법을 결합한다.

- **릿지 회귀**($\alpha = 0$): 모든 계수를 0 쪽으로 비례적으로 축소한다. 모든 변수가 유의미하다고 볼 때 적합하다.
- **Lasso 회귀**($\alpha = 1$): 계수를 정확히 0으로 축소할 수 있어 자동 변수 선택을 수행한다. 더 단순하고 해석하기 쉬운 모형을 만든다.
- **엘라스틱 넷**($0 < \alpha < 1$): 두 방식의 혼합이다. 설명변수들이 상관되어 있을 때 유용하다.

1.3 세 가지 용도 모드

모든 glmnetUI 프로젝트는 생성 시 선택하는 **용도(purpose)**를 갖는다 (제1절, *Project*). 활성 프로젝트의 용도에 따라 어떤 도구와 인터페이스 요소가 나타날지가 결정되므로, 다른 모드로 작업하려면 해당 용도를 가진 프로젝트를 새로 만들거나 열어야 한다. 세 가지 모드는 다음과 같다.

- **General**(범용) — 모든 유형의 모집단이나 데이터셋에 대한 엘라스틱 넷 회귀이다. 기본 모드이며, 도메인 특화 기능 없이 정규화 회귀의 전 과정을 제공한다.
- **For Appraisal**(감정평가용) — 부동산 감정평가에 맞춘 엘라스틱 넷 회귀이다. 대상 부동산 처리, 특수 열 지정, 잔차 제약 접근법 (Residual Constraint Approach, RCA) 등 단일 부동산 평가에 특화된 기능을 추가한다.
- **Market Area Analysis**(시장지역 분석) — 시장지역 연구에 맞춘 엘라스틱 넷 회귀이다. 정의된 시장 내 부동산 집단을 분석하는 기능을 추가하며, 대상 행(subject row) 제외를 선택적으로 지원한다.

세 모드 모두에서 핵심 모형화 엔진은 동일하다—언제나 엘라스틱 넷 (glmnet) 모형을 적합하는 것이다. 용도 설정은 어떤 추가 도구와 인터페이스 요소를 사용할 수 있는지를 제어할 뿐이다.

1.4 부동산 전용 기능

For Appraisal 또는 Market Area Analysis를 선택하면 glmnetUI는 부동산 분석을 위해 설계된 여러 기능을 활성화한다.

- **특수 열 지정(special column designations)** — 각 설명변수에 `contract_date`, `dom`, `concessions`, `latitude`, `longitude`, `living_area`, `lot_size`, `actual_age`, `effective_age`, `area`, `site_dimensions`, `display_only`와 같은 특수 역할을 지정할 수 있다. 이 지정에 따라 적합과 출력 과정에서 해당 열이 어떻게 처리될지가 결정된다.
- **위도와 경도의 반올림** — 위도 또는 경도로 지정된 열은 과적합을 막기 위해 자동으로 소수점 셋째 자리까지 반올림된다.
- **Sale Age 열** — 어떤 열이 `contract_date`로 지정되고 기준일(Effective Date)이 주어지면 glmnetUI는 `sale_age` 열(매매일과 기준일 사이의 일수)을 계산하여 설명변수로 대체 투입한다.
- **RCA 계산(감정평가 모드 전용)** — 감정평가 모드에서는 모형 적합 후 glmnetUI가 잔차 제약 접근법(Residual Constraint Approach, RCA) 출력을 계산할 수 있으며, 이는 사례별 보정치, 순/총 보정 요약, 그리고 대상 부동산의 보정된 매매가격을 산출한다.

요령: *General* 용도 모드는 금융, 과학, 공학, 부동산 등 어떤 데이터셋에도 적용된다. 감정평가 모드와 시장 모드는 부동산 실무자를 위한 편의 기능을 더할 뿐이다.

1.5 시작하기

glmnetUI를 사용하려면 다음과 같이 한다.

1. R에서 **earthUI**를 설치한다: `install.packages("earthUI")` — glmnet 루틴이 함께 포함되어 있다. (개발 버전과 선택 구성요소에 대해서는 이번 호 earthUI 논문의 “Installing earthUI and vProlog” 절을 참조하라.)
2. 애플리케이션을 실행한다: R에서 `earthUI::launch_glmnet()`을 실행한다. 앱은 웹 브라우저의 7879 포트에서 열린다. 브라우저에서 `http://localhost:7879`로 직접 접속할 수도 있다. 앱은 마지막에 사용한 용도 모드를 기억하여 자동으로 복원한다.
3. 사이드바 제1절에서 프로젝트를 생성하거나 연다. 프로젝트는 용도와 작업 위치(국가, 주, 카운티, 시)를 설정하며 입력/출력 폴더를 갖는다. 프로젝트는 공유되는 `regProj` 루트 아래에 저장되며 형제 앱인 earthUI 및 mgcvUI와 공유된다. 다른 용도를 가진 프로젝트를 여는 것이 모드를 전환하는 방법이다.
4. 제2절에서 활성 프로젝트의 입력 폴더에 있는 CSV 또는 Excel 파일을 선택하여 데이터를 가져온다. 해당 폴더에 파일을 넣은 뒤 **Refresh**를 눌러 목록을 갱신한다.
5. earthUI에서 가져오기(선택 사항) — 제3절에서는 earthUI 결과 `.rds` 파일을 가져와 earth의 `hinz` 기저 함수를 glmnet의 정규화와 함께 사용할 수 있다. earthUI 결과가 없다면 이 단계는 건너뛴다.
6. 변수를 구성한다 — 목표변수와 설명변수를 고르고, 자료형과 기대 부호를 설정하며, 필요한 특수 열 역할을 지정한다.
7. glmnet 모수를 설정한다 — `alpha`, `lambda`, `family`, 부호 제약, 교호작용 및 기타 옵션.
8. 모형을 적합한다 — “Fit Glmnet Model”을 클릭하고 주 패널에서 결과를 검토한다.
9. 내보내기 — 예측값을 Excel로 내려받거나, Quarto 보고서를 생성·변환하거나, (감정평가 모드에서는) RCA 보정치와 거래사례비교표 (Sales Comparison Grid)를 계산한다.

설정은 브라우저의 로컬 저장소에 자동으로 보존되며 동일한 입력 파일을 다시 불러올 때 복원된다.

2 MLS 입력 데이터 요건

부동산 감정평가와 시장분석 작업 흐름에서 입력 데이터는 일반적으로 다중 매물정보 서비스 (Multiple Listing Service, MLS) 내보내기 파일에서 온다. 이 장에서는 기대되는 파일 구조와 glmnetUI가 사용할 수 있는 열을 설명한다.

2.1 파일 형식과 구조

glmnetUI는 CSV와 Excel(.xlsx, .xls) 파일을 받는다. 가져오기 시 열 이름은 자동으로 snake_case로 변환된다—예를 들어 “Living SqFt”는 living_sqft로, “Contract Date”는 contract_date로, “Sale Price”는 sale_price로 바뀐다. 이 정규화 덕분에 작업 흐름 전반에서 열 참조가 일관되게 유지된다. 가져오기에 사용되는 CSV 구분자와 소수점 기호는 로케일 설정에 따라 결정된다(제3장 “Locale & Regional Settings” 참조).

데이터 파일은 부동산 한 건당 한 행, 속성 하나당 한 열로 구성된 평평한 표여야 한다. 파일의 첫 행에는 반드시 열 머리글이 있어야 한다.

2.2 감정평가 모드에 필요한 열

glmnetUI는 어떤 열 구성에서도 동작하지만, 완전한 감정평가 작업 흐름을 위해서는 다음 열들이 있는 편이 좋다.

열	특수 유형	용도
Sale Price	(목표변수)	모형의 반응변수
Contract Date	contract_date	sale_age(기준일로부터의 일수) 계산에 사용
Listing Date	listing_date	DOM 열이 없을 때 계약일과 함께 DOM 계산에 사용
Days on Market	dom	시장 노출 일수; 내보내기 파일에 표시
Concessions	concessions	매매 양허; 순매매가 = 매매가격 - 양허
Living Area (SF)	living_area	제곱피트당 잔차(residual_sf, cqa_sf) 사용 가능
Lot Size	lot_size	대지 규모 열
Site Dimensions	site_dimensions	대지 규모와 함께 묶임
Latitude	latitude	소수점 셋째 자리 반올림
Longitude	longitude	소수점 셋째 자리 반올림
Area ID	area	시장지역/근린 식별자
Actual Age	actual_age	부동산 실제 연령
Effective Age	effective_age	부동산 유효 연령
Address	display_only	내보내기에는 표시하되 모형에서는 제외

스프레드시트의 열 이름은 외국어여도 무방하다—R 프로그램이 특별하게 처리할 수 있도록 “특수” 이름만 영어로 되어 있을 뿐이다. 그 외에는 주어진 열 이름이 회귀모형, 그래프, 그리고 감정평가를 수행하는 경우 출력 보고서에 그대로 나타난다.

모든 열이 필수는 아니다. glmnetUI는 상황에 맞게 적응한다—어떤 열이 없으면 그에 대응하는 기능이 단순히 생략된다. 다만 부동산 가격 모형에서 수용 가능한 적합도를 얻으려면 다음 열들이 강력히 권장된다.

1. **Sale Age**(매매 경과일) — 계약 매매일과 감정평가 또는 분석의 기준일 사이의 일수이다. 여러 해에 걸친 매매 이력을 사용하는 경우, 특히 5년이 넘는 기간이라면 sale_age가 매매 가격 추정에서 중심적인 역할을 하는 경우가 많다.
2. **Living Area**(주거면적) — “Living Sqft”, “GLA”(총 주거면적) 등으로도 불린다. 이 역시 매매가격의 주된 결정요인이다.
3. **Total Bath Count**(총 욕실 수) — 전용, 1/4, 1/2, 3/4 욕실을 모두 합한 수이다. 예를 들어 전용 욕실 두 개와 반욕실 하나는 2.5가 된다.

4. **Garage Bays**(차고 칸 수) 또는 **Garage Area**(차고 면적) — 차고 주차 공간의 수 또는 차고 면적(제곱피트).
5. **Lot Size**(대지 규모) — 부동산의 토지 면적으로, 보통 제곱피트나 에이커 단위이다.
6. **Longitude, Latitude**, 그리고 가능하다면 **Area ID**. 이 위치 변수들은 모형이 지리적 가격 변동을 반영하도록 돕는다.

2.3 특수 열 명명 규칙

glmnetUI는 열 이름이 아니라 특수 유형 지정(**special type designation**)으로 열을 식별한다. MLS 내보내기 파일에서 열 이름은 원하는 대로 붙여도 되며, 중요한 것은 변수 구성표(제6장)에서 올바른 특수 유형을 지정하는 것이다.

예를 들어 사용자의 MLS는 주거면적을 “GLA”, “Living SqFt”, “liv_area”, “gross_living_area” 등으로 내보낼 수 있다. 가져오기(이때 snake_case로 바뀐다) 후에 Special 드롭다운에서 living_area로 지정하기만 하면 된다. 그러면 glmnetUI는 원래 이름과 무관하게 그 열을 제곱피트당 잔차 계산에 사용한다.

2.4 데이터 품질과 완전성

- **결측값(NA)**: 설명변수나 목표변수 열에 NA가 있는 행은 적합 전에 자동으로 제거된다.
- **날짜 열**: 흔한 날짜 형식과 일치하는 문자열 열은 가져오기 시 자동으로 해석된다. 네 자리 연도가 감지되지 않으면 두 자리 연도 형식이 우선 적용된다.
- **수치형 열**: 매매가격, 주거면적, 대지 규모, 양허 등과 같은 항목은 반드시 수치형이어야 한다. MLS가 통화 기호나 천 단위 구분기호가 포함된 가격을 내보낸다면(예: “\$350,000” 또는 “350.000,00”) 가져오기 전에 정리해야 할 수 있다.
- **요인형 열**: 텍스트, TRUE/FALSE, R 요인(factor) 열은 자동으로 범주형으로 취급된다—회귀분석은 이들을 요인으로만 다룰 수 있기 때문이다. 수치형 열(예: 수치 코드로 된 지역이나 양식 코드)은 **Factor** 체크박스를 선택할 때에만 요인이 된다. glmnetUI는 수치의 값 범위만 보고 범주형이라고 추론하지 않으므로 bath_count 같은 이산 수치형 설명변수는 연속형으로 남는다. Special 역할은 요인 여부와는 무관하다.

요령: 가져오기 후 설명변수 표의 NA 열을 검토하라. 결측값이 많은 열은 행이 통째로 버려지는 원인이 될 수 있다. NA가 많은 열은 제외하거나 가져오기 전에 데이터를 정리하는 것을 고려하라.

2.5 대상 행의 위치

Appraisal(감정평가) 모드에서는 1행이 반드시 대상 부동산이어야 한다. 나머지 행은 모두 비교 거래사례이다. 대상 행은 모형 적합에서 제외된다. 적합 후에도 모형은 대상 행에 대한 예측값을 생성한다.

Market Area Analysis(시장지역 분석) 모드에서는 대상을 1행에 두는 것이 선택 사항이다. **General**(범용) 모드에는 특별한 행 처리가 없다—모든 행이 동등하게 취급된다.

3 General Purpose 모드

3.1 개요

General Purpose(범용) 모드는 glmnetUI를 실행했을 때의 기본값이다. 이 모드는 부동산에 국한되지 않고 어떤 데이터셋에 대해서도 완전한 엘라스틱 넷 회귀 작업 흐름을 제공한다. 과학 데이터, 재무 분석, 공학 연구, 또는 어떤 회귀 문제에도 glmnetUI를 사용할 수 있다.

General 모드에서는 부동산 전용 기능(특수 열, 매매 경과일, 좌표 반올림, RCA)이 인터페이스에서 생략된다. 사이드바는 변수 선택, 모수 구성, 모형 적합, 내보내기에 집중하도록 간소화된다.

3.1.1 Skip First Row

General 또는 Market 프로젝트가 열려 있으면 사이드바 맨 위에 **Skip first row**(첫 행 건너뛰기) 체크박스가 나타난다. 이를 선택하면 1행이 모형 적합에서 제외된다. 1행에 모형에 영향을 주어서는 안 되는 대상 관측치나 기준 관측치가 들어 있을 때 유용하다. Appraisal 모드에서는 1행(대상 부동산)이 항상 자동으로 제외된다.

3.1.2 용도별 설정

설정(설명변수 선택, 모형 모수, 교호작용)은 입력 파일과 용도 모드의 조합마다 따로 저장된다. General, Appraisal, Market 모드를 오가더라도 각 모드의 설정은 독립적으로 유지된다.

3.2 사이드바 작업 흐름

사이드바는 번호가 매겨진 접이식 절들로 구성되어, 프로젝트 선택부터 보고서 내보내기까지 사용자를 안내한다. 대부분의 절은 활성 프로젝트가 열려 있어야 나타나며, 몇몇 절은 프로젝트의 용도에 따라서만 나타난다. 출력 폴더는 활성 프로젝트로부터 도출되므로 별도의 출력 폴더 입력란은 없다.

1. Project — 프로젝트를 생성하거나 연다. 프로젝트는 용도와 작업 위치(국가, 주, 카운티, 시), 그리고 입력 (<os>_in/) 및 출력 폴더를 설정한다. 프로젝트는 공유되는 regProj 루트(기본값 ~/regProj, REGPROJ_ROOT 환경변수로 변경 가능) 아래에 저장되며 형제 앱인 earthUI 및 mgcvUI와 공유되므로 세 앱의 모형이 나란히 놓인다. 전환하려면 **Close Project**를 클릭한다.

2. Import Data — 활성 프로젝트의 입력 폴더에서 CSV 또는 Excel 파일을 선택한다. Finder/탐색기로 폴더에 파일을 넣은 뒤 **Refresh**를 눌러 목록을 갱신한다. 시트가 여러 개인 Excel 파일은 시트 선택기가 나타난다. 열 이름은 자동으로 snake_case로 변환된다.

3. Import from earthUI (optional) — earthUI 결과 .rds 파일을 가져와 earth의 힌지 기저 함수를 glmnet의 엘라스틱 넷 정규화와 함께 사용한다. 찾아보기 버튼은 제2절의 파일 선택기와 같은 모양으로 꾸며져 있다. 이 단계는 선택 사항이며—가져올 earthUI 결과가 없다면 건너뛴다.

4. Variable Configuration — 목표변수 선택기, Include, Factor, Force 체크박스과 Sign 드롭다운이 있는 설명변수 표. 계약일, 좌표 등을 지정하는 Special 열은 Appraisal 및 Market 모드에서만 나타난다. 자세한 내용은 제6장을 참조하라.

5. glmnet Call Parameters — 모형 구성 전체: alpha, lambda, family, standardize, 부호 제약, 완화 lasso, 난수 시드, 교호작용 행렬, 그리고 고급 모수(lambda.min.ratio, nlambda, CV 손실 지표, 수렴 임계값, 최대 반복 횟수, 절편). 전체 모수 참조는 제7장에 있다.

6. Fit Glmnet Model — 모형을 실행하는 버튼이다. 결과는 오른쪽 탭에 나타난다.

7. Download Output — 예측값, 잔차, CQA 점수, 변수별 기여도를 Excel 파일로 프로젝트의 출력 폴더에 내보낸다. 모형이 적합되고 나면 모든 용도 모드에서 사용할 수 있다.

8. Calculate RCA Adjustments & Download — 각 비교 사례에 대한 변수별 RCA 보정치를 계산하고 CQA를 통해 대상의 잔차를 보간한다. Appraisal 및 Market 모드에서 적합 이후에만 나타난다. 제11장을 참조하라.

9. Generate Sales Grid & Download — 총 보정 비율로 비교 사례를 순위화한 중간 단계의 거래사례비교표(Sales Comparison Grid)를 생성한다. Appraisal 모드에서 RCA 보정치가 계산된 뒤에만 나타난다. 제12장을 참조하라.

10. Generate Quarto Report — 모든 그림과 표를 포함한 자기완결적 Quarto(.qmd) 보고서 묶음을 프로젝트의 출력 폴더에 기록한다. 모형이 적합되고 나면 사용할 수 있다.

11. Convert Quarto Report — 생성된 .qmd 묶음을 HTML, Word, PDF로 렌더링한다. PDF에는 LaTeX 설치가 필요하며, 설치가 감지되지 않으면 PDF 옵션은 숨겨진다. 제13장을 참조하라.

Market 및 General 모드에서는 감정평가 전용 단계들이 숨겨져 건너뛰어지고 두 개의 보고서 절 번호가 그에 맞게 다시 매겨진다.

3.3 프로젝트

glmnetUI는 작업을 공유되는 regProj 루트 폴더(earthUI 및 mgcvUI가 사용하는 것과 동일한 트리) 아래의 프로젝트로 조직한다. 각 프로젝트는 입력 폴더(<os>_in)와 엔진별 출력 폴더(예: <os>_out_glmnet)를 갖는다.

프로젝트를 만들려면 제1절에서 + **New...**를 클릭하고 다음을 입력한다.

- **Purpose**(용도) — General, Appraisal, Market Area 중 하나.
- **Location**(위치) — 프로젝트의 행정 구역 체계. 선택한 국가에 따라 채워야 할 단계가 달라지므로 외국의 명명 관행이 자동으로 지원된다—미국 프로젝트는 주/카운티/시를, 프랑스 프로젝트는 지역권/도(département)/코뮌을, 일본 프로젝트는 도도부현/시구/지구를 사용하는 식이다.
- **Project Name**(프로젝트 이름) — 8자 이내의 짧은 이름 (영문자, 숫자, _, -).

앱은 입력한 이름을 그대로 쓰지 않는다. 위치 코드와 생성 일시를 앞에 붙여 고유한 시각 표시 폴더 이름을 <country>_<admin level codes>_<YYYYMMDD-HHMMSS>_<your name> 형식으로 만든다—예를 들어 burl이라는 이름의 미국 프로젝트는 us_ca_081_burlin_20260604-131500_burl이 된다. 8자 제한은 사용자가 입력하는 부분에만 적용된다. 시각 접두어는 새로 생성한 프로젝트에만 붙으며, 이미 디스크에 있는 프로젝트는 원래 이름을 유지한 채 그대로 완전히 사용할 수 있다.

3.4 earthUI에서 가져오기 (선택 사항)

사이드바 제3절 — **Import from earthUI** — 에서는 earthUI 결과 .rds 파일을 가져올 수 있다. 이를 통해 glmnetUI는 earth의 힌지 기저 함수(조각별 선형 항)를 glmnet의 엘라스틱 넷 정규화와 함께 사용할 수 있으며, earth의 자동 비선형성 탐지와 glmnet의 변수 선택 및 계수 제약이 결합된다.

earthUI 결과를 가져오면 Include 체크박스가 아니라 earth 모형의 설명변수가 glmnet 적합을 주도한다. glmnet이 earth의 기저 확장 위에서 적합하기 때문이다. 변수 구성표는 이를 명시적으로 보여 준다. **Include**와 **Type**은 잠기고, earth 모형이 사용하지 않은 설명변수의 행은 비활성화되며, 설명변수 집합이 earth 모형에서 왔음을 알리는 안내문이 표시된다. **Force**와 **Sign**은 earth의 설명변수에 대해 계속 편집할 수 있는데, 적합 시점에 earth 기저에 여전히 적용되기 때문이다.

찾아보기 버튼은 제2절의 파일 선택기와 같은 모양으로 꾸며져 있다. 이 단계는 전적으로 선택 사항이다—가져올 earthUI 결과 파일이 없다면 제3절을 건너뛰고 제4절로 진행하면 된다.

3.5 주 패널 탭

적합이 끝나면 주 패널에 다음 탭들이 나타난다.

탭	내용
Data Preview	가져온 데이터의 미리보기. 감정평가 모드에서는 대상 부동산 표와 비교 거래사례 표로 나뉜다.
Equation	적합된 모형 방정식을 LaTeX/MathJax로 표시하며, 0이 아닌 각 계수와 교호작용 항을 보여 준다.
Correlation	수치형 설명변수 상관관계수의 히트맵(적합 전에도 사용 가능).
Summary	주요 지표(R^2 , 조정 R^2 , GR^2 , $CV R^2$, RMSE, MAE)와 감정평가 지표(COD, PRD, 중위 비율).
Coefficients	부호 위반 경고가 표시된, 0이 아닌 계수 표.
Variable Importance	설명변수 중요도($ b \times sd(x)$)의 막대그래프와 순위 표. 가능한 경우 plotly 대화형 호버 지원.
Contributions	기울기 표시가 있는 변수별 부분효과 그림. 수치형 설명변수는 산점도와 적합선, 요인형은 상자그림, 교호작용은 3차원 곡면/산점도/히트맵으로 표시.
ANOVA	분산 분해 표: 변수별 제곱합, 모형 제곱합 대비%, 계수.
Diagnostics	네 가지 그림: 계수 경로, CV 오차, 실제값 대 예측값, 잔차 대 적합값.
Glmnet Output	원시 모형 출력: 사용된 난수 시드, 모형 출력, 선택된 람다, $\lambda_{min}/1se$, 감마(완화 적합인 경우), 전체 계수 벡터.
Report	보고서 구성 및 내보내기 입력란.
RCA Adjustments	RCA 분석 결과(감정평가 모드에서 계산 후).

3.6 설정의 보존

glmnetUI는 사용자의 구성을 입력 파일 이름을 키로 하여 브라우저의 로컬 저장소에 자동으로 저장한다. 같은 파일을 다시 불러오면 목표변수 선택, 설명변수 체크박스, 자료형, 기대 부호, glmnet 모수, 교호작용 행렬 등 모든 설정이 복원된다. 마지막에 사용한 용도 모드도 전역적으로 보존되어 앱을 다시 실행할 때 복원된다. 세 가지 옵션이 있다.

- **Use last settings for input file** — 이 특정 파일에 대해 저장된 설정을 복원
- **Use default settings** — 저장해 둔 전역 기본값을 적용
- **glmnet defaults** — 공장 기본값으로 초기화($\alpha = 1$, CV $\lambda_{min}/1se$, 10겹, gaussian, standardize 켜)

Save current as default 버튼은 현재의 모든 설정을 전역 기본값으로 저장한다.

3.7 다크 모드

오른쪽 위 모서리의 달/해 아이콘을 클릭하면 밝은 테마와 어두운 테마를 전환할 수 있다. 테마 설정은 로컬 저장소에 저장되어 세션이 바뀌어도 유지된다. 모든 UI 요소(표, 그림, 카드, 버튼)는 CSS 변수로 선택된 테마에 맞추어 조정된다.

3.8 로케일과 지역 설정

glmnetUI는 국가 기반 로케일 체계를 통해 국제적인 숫자 및 CSV 형식 관행을 지원한다. 제목 표 시줄의 **Settings** 드롭다운은 지원되는 31개 국가에 대해 **Country**와 **Paper** 선택기를 제공한다. 각 프리셋은 다음을 설정한다.

- **CSV 구분자** — 미국/영국/일본은 쉼표(,), 쉼표를 소수점 기호로 쓰는 유럽 대부분은 세미콜론(;).
- **소수점 기호** — 마침표(.) 또는 쉼표(,).
- **천 단위 구분기호** — 쉼표(미국/영국/일본), 마침표(독일/이탈리아/스페인), 공백(핀란드/프랑스/폴란드/발트 3국/우크라이나/러시아), 아포스트로피(스위스).
- **용지 크기** — Letter(미국/캐나다/멕시코) 또는 A4(그 밖의 모든 지역).

3.8.1 기본값 저장

Save as my default를 클릭하면 로케일 설정이 전역으로 저장된다. 이 기본값은 어떤 데이터 파일을 불러오든 이후의 모든 세션에 적용된다.

요령: 그림 축과 기울기 표시의 숫자 형식은 로케일에 맞추어 자동으로 조정된다. 독일 로케일은 천 단위에 마침표(200.000)를, 핀란드는 공백(200 000)을, 스위스는 아포스트로피(200'000)를 사용한다. 통화 기호는 표시되지 않는다—*glmnetUI*는 통화에 중립적이다.

4 Appraisal 모드

용도로 **For Appraisal**을 선택하면 glmnetUI는 단일 부동산 평가에 맞추어 스스로를 구성한다. 제3장에서 설명한 모든 기능은 그대로 사용할 수 있으며, 이 장에서는 감정평가에 특화된 추가 사항만 다룬다.

4.1 대상 행의 처리

감정평가 모드에서는 데이터셋의 **1행이 대상 부동산**이고 나머지 행은 모두 비교 거래사례이다. 입력 파일은 그에 맞게 구성되어야 한다 (제2장 참조). 대상의 매매가격은 비워 두거나 아무 값이나 넣어도 된다—glmnetUI가 적합 시 자동으로 NA로 취급한다.

가져오기 후 Data Preview 탭은 두 부분으로 나뉜다. **Subject Property**(1행)와 **Comparable Sales**(2행 이하)이다. 1행은 언제나 모형 적합에서 제외된다. 적합 후에도 모형은 대상 행에 대한 예측값을 생성한다.

4.2 기준일과 매매 경과일

감정평가 모드와 시장 모드에서는 변수 구성 절에 **Effective Date**(기준일) 입력란이 나타난다 (기본값은 오늘 날짜). Special 열 드롭다운에서 어떤 열을 **contract_date**로 지정하면 glmnetUI는 **sale_age** 열—각 매매의 계약일과 기준일 사이의 정수 일수—을 계산한다. 이 열은 설명변수로 추가된다.

기준일이 바뀌면 **sale_age**는 자동으로 다시 계산된다.

4.3 특수 열 지정

감정평가 모드와 시장 모드에서는 변수 구성표의 각 설명변수마다 **Special** 드롭다운이 나타난다. 특수 유형과 그 효과의 전체 참조는 제6장에 있다.

4.4 RCA 보정 개요

Calculate RCA Adjustments & Download 버튼(사이드바 제8절, 적합 후 Appraisal 및 Market 모드에서 표시)은 대상 대비 각 비교 사례의 시장 도출 보정치를 계산한다. RCA 작업 흐름 전체는 제11장에서 설명한다.

요령: 대상에 부여하는 **CQA** 점수는 잔차 분포 가운데 얼마만큼을 대상에 귀속시킬지를 결정한다. 5.00점은 대상을 비교 사례들의 중앙값에 놓는다.

5 Market Area Analysis 모드

용도로 **Market Area Analysis**를 선택하면 glmnetUI는 감정평가 모드와 동일한 부동산 전용 기능(특수 열, 매매 경과일, 좌표 반올림)을 제공하되, 단일 대상의 가치를 평가하기보다 부동산 집단을 분석하는 데 초점을 맞춘다.

5.1 Appraisal 모드와의 차이

- 대상 행 처리가 유연하다 — 시장 모드에서는 1행이 자동으로 대상 부동산으로 취급되지 않는다.
- 거래사례비교표가 없다 — “Calculate RCA Adjustments & Download” 단계(제8절)는 시장 모드에서도 사용할 수 있지만, Sales Grid 단계(제9절)는 감정평가 모드 전용이다.

5.2 시장 모드를 쓸 때

Market Area Analysis 모드는 다음과 같은 경우에 적합하다.

- 가치 결정요인을 이해하기 위해 근린이나 시장지역에 대한 회귀모형을 구축할 때
- 면적, 연령, 대지 규모, 위치 같은 변수가 부동산 집단 전반의 매매가격에 어떻게 영향을 미치는지 분석할 때
- 시장 상황 분석이나 근린 확정에 대한 근거를 준비할 때
- 특수 열 기능(매매 경과일, 좌표 반올림)은 원하지만 RCA 보정 작업 흐름은 필요하지 않은 데이터셋을 다룰 때

요령: 시장 모드는 특수 열 기능은 원하지만 RCA 보정 작업 흐름은 필요하지 않은 일반적인 부동산 회귀분석에도 유용하다.

6 변수 선택

사이드바 제4절 — **Variable Configuration** — 은 어떤 열을 모형에 참여시키고 그것을 어떻게 다룰지 정하는 곳이다.

6.1 목표변수

제4절 맨 위의 **Target (response) variable** 드롭다운에는 데이터셋의 모든 열이 나열된다. 한 열을 반응변수(예: 매매가격)로 선택한다. 목표변수 열은 자동으로 설명변수 목록에서 제외된다.

6.2 설명변수 표

목표변수 선택기 아래에는 남은 모든 열이 다음 항목과 함께 표로 나열된다.

열	설명
Variable	열 이름
Type	자료형 드롭다운: <code>numeric</code> , <code>integer</code> , <code>character</code> , <code>Date</code> , <code>POSIXct</code>
Include	체크박스 — 이 열을 모형의 설명변수로 포함
Factor	체크박스 — 변수를 범주형 요인으로 취급(모형 행렬에 가변수 열 생성)
Force	체크박스 — 모형에 강제 투입(벌점 계수 = 0, lasso가 걸코 제거하지 않음)
Special	드롭다운(감정평가/시장 모드 전용) — 아래의 특수 열 유형 참조를 보라
Sign	기대되는 계수 부호: 양, 음, 또는 무관
NAs	결측값의 개수

6.3 자료형 감지와 재지정

glmnetUI는 가져오기 시 자료형을 자동으로 감지한다. 수치형, 정수형, 날짜형 열이 인식된다. 날짜처럼 보이는 문자형 열(흔한 날짜 형식 패턴)은 `Date`로 분류된다.

Type 드롭다운을 바꾸어 어떤 감지 결과든 재지정할 수 있다. 자료형을 바꾸면 해당 열이 모형 행렬에 어떻게 부호화되는지가 달라진다.

텍스트, TRUE/FALSE, R 요인 열은 자동으로 범주형으로 취급된다. 수치형 열은 **Factor** 체크박스를 선택할 때에만 요인이 된다—glmnetUI는 수치의 값 범위만 보고 범주형이라고 추론하지 않으므로, 욕실 수 같은 이산 수치형 설명변수는 사용자가 달리 지정하지 않는 한 연속형으로 남는다. Factor 설정은 localStorage를 통해 유지된다.

6.4 기대 부호와 적용

각 설명변수에 대해 **Sign** 드롭다운은 그 계수가 양이 될 것으로 기대하는지, 음이 될 것으로 기대하는지, 아니면 어느 쪽이든 무방한지를 지정한다. 적합 후에는 다음과 같이 처리된다.

- 기대와 다른 부호를 가진 계수는 **Coefficients** 표에 표시된다
- 부호 경고가 Summary 탭에 나타난다

glmnet 모수(제7장)에서 **Enforce Sign Constraints** 체크박스를 활성화하면 모형이 기대 부호에 부합하는 계수를 산출하도록 강제된다. 이는 glmnet의 `upper.limits`와 `lower.limits`로 구현된다.

6.5 특수 열 유형 참조

감정평가 모드와 시장 모드에서 **Special** 드롭다운은 다음 옵션을 제공한다.

가중치:

- **weight** — 관측치 가중치 열(모든 모드에서 사용 가능; 하나만 허용; 가중치가 0인 행은 적합에서 제외)

날짜 및 시간 유형:

- **contract_date** — 기준일로부터 **sale_age**를 자동 계산하도록 함
- **listing_date** — 시장 노출 일수 계산의 대체 근거로 사용
- **dom** — 시장 노출 일수 열을 식별
- **sale_age** — 기존의 매매 경과일 열을 지정하므로, glmnetUI가 **contract_date**로부터 계산하는 대신 임의의 열을 매매 경과일로 쓸 수 있다

금액 유형:

- **concessions** — 매매 양허(매도인 신용, 매수인 유인책 등)를 식별

거래 유형:

- **sale_type** — 매매/거래 유형 열을 식별

규모 및 위치 유형:

- **latitude** — 값이 자동으로 소수점 셋째 자리까지 반올림됨
- **longitude** — 위도와 동일한 반올림 처리
- **area** — 시장지역 또는 근린 식별자(예: **area_id**, **neighborhood**, **market_area**)
- **living_area** — 제곱피트당 잔차 계산 (**residual_sf**, **cqa_sf**)을 가능하게 함
- **lot_size** — 대지 규모 열
- **site_dimensions** — 대지 규모와 함께 묶임(예: “75x120”)

연령 유형:

- **actual_age** — 부동산 실제 연령 열
- **effective_age** — 부동산 유효 연령

표시 유형:

- **display_only** — 해당 열은 Excel 내보내기에는 포함되지만 모형 적합에서는 완전히 제외된다. 주소 항목, MLS 번호, 필지 ID, 기타 참조 자료에 사용한다. 여러 열에 이 지정을 줄 수 있다.

요령: *display_only*로 지정된 열은 데이터셋에 남아 Excel 내보내기에 나타나지만 설명변수 표에서는 완전히 제외된다. ID 열, 주소, 기타 참조 항목에 사용하라.

7 모수 선택

사이드바 제5절 — **glmnet Call Parameters** — 은 엘라스틱 넷 모형의 모든 구성 옵션에 접근할 수 있게 한다. 각 모수에는 설명 툴팁이 달린 파란색 도움말 아이콘(?)이 있다.

7.1 Alpha(혼합 모수)

alpha 모수는 정규화의 유형을 조절한다.

Alpha 값	유형	동작
0	릿지	모든 계수를 비례적으로 축소하 되 어느 것도 0으로 만들지 않음
1	Lasso	계수를 정확히 0으로 만들 수 있 음(변수 선택)
0에서 1 사이	엘라스틱 넷	릿지와 lasso의 혼합

alpha를 선택하는 방법은 두 가지이다.

- **Fixed** — 슬라이더로 단일 alpha 값을 설정(기본값: 1.0)
- **Grid Search** — 여러 alpha 값을 시험하여 CV 오차가 가장 낮은 값을 선택한다. 범위와 값의 개수를 설정할 수 있다.

7.2 Lambda 선택

lambda는 정규화의 강도를 조절한다.

방법	설명
Cross-validation	권장. k겹 CV를 사용하여 많은 lambda 값을 시험한다.
Manual	타당한 이유가 있다면 특정 lambda 값을 직접 입력한다.

교차검증을 사용할 때는 두 가지 lambda 선택지가 있다.

- **lambda.1se**(기본값) — 최솟값으로부터 1 표준오차 이내에 있는 가장 큰 lambda이다. 더 단순하고 더 강하게 정규화된 모형을 만든다. 감정평가 업무에 권장된다.
- **lambda.min** — 교차검증 오차를 최소화하는 lambda이다. 적합도는 가장 좋지만 모형이 더 복잡해질 수 있다.

Number of CV Folds 모수는 교차검증을 위해 데이터를 어떻게 나눌지를 조절한다(기본값: 10겹).

7.3 Family

반응변수의 분포 계열을 선택한다.

Family	사용 상황
gaussian	연속형 반응(예: 매매가격). 가장 흔하다.
binomial	이항 결과(예: 매도됨/매도되지 않음).
poisson	계수(count) 자료(예: 매매 건수).

7.4 Standardize

선택하면(기본값) 적합 전에 모든 설명변수가 평균 0, 표준편차 1이 되도록 척도화된다. 이로써 원래 척도와 무관하게 별점이 모든 설명변수를 동등하게 다루게 된다. 계수는 원래 척도로 되돌려 반환된다. 대개 커 둔 채로 두는 것이 좋다.

7.5 부호 제약

Enforce Sign Constraints가 활성화되면 계수는 변수 표(제6장)에서 설정한 기대 부호에 부합하도록 제약된다.

- “positive” 부호는 계수를 ≥ 0 이 되도록 강제한다
- “negative” 부호는 계수를 ≤ 0 이 되도록 강제한다
- “either”로 설정된 변수는 제약을 받지 않는다

이는 `glmnet()`의 `upper.limits` 및 `lower.limits` 모수로 구현된다.

7.6 완화 Lasso

일반적인 lasso는 변수 선택과 계수 추정에 동일한 벌점을 쓰기 때문에 중요한 계수를 0 쪽으로 과도하게 축소할 수 있다. **완화 lasso**(relaxed lasso)는 이 두 단계를 분리한다.

1. 먼저 lasso로 변수를 선택한다(벌점이 어떤 계수가 살아남을지 결정한다)
2. 그다음 살아남은 변수들을 더 작은 벌점 또는 무벌점으로 다시 적합하여 편향이 덜한 추정치를 얻는다

Gamma 슬라이더는 완화의 정도를 조절한다.

Gamma	효과
0	완전 완화(OLS 재적합, 축소 전혀 없음)
1	완화 없음(일반 glmnet과 동일)
그 사이	완화 적합과 벌점 적합을 혼합

교차검증을 쓰면 최적 성능을 위해 감마가 자동으로 선택된다.

7.7 난수 시드

무작위 정수가 미리 채워진 **Random seed** 텍스트 입력란은 교차검증 겹 배정을 재현 가능하게 만든다—적합 전에 `set.seed()`가 호출되며, 사용된 시드는 Glmnet Output 탭과 적합 상태 메시지에 표시된다. 매 적합 후에는 (R의 난수 생성기가 아니라 시스템 시각을 사용하여) 새 시드가 자동 생성된다. 현재 파일에 대해 마지막으로 사용한 시드 5개는 클릭 가능한 링크로 표시되어 쉽게 다시 불러올 수 있으며, 새 파일을 불러오면 시드 이력은 초기화된다.

7.8 교호작용 행렬

대화형 상삼각 행렬을 통해 어떤 변수 쌍의 교호작용을 허용할지 조절할 수 있다. 각 칸에는 체크 박스가 있으며, 선택되어 있으면 교호작용 항 $x1:x2$ 가 모형 행렬에 포함된다.

- **Allow All** — 모든 교호작용 쌍을 선택
- **Clear All** — 모든 쌍의 선택을 해제(새 파일의 기본값은 전부 해제 상태)
- **변수 이름을 클릭** — 해당 변수의 모든 교호작용을 토글

교호작용 선택은 입력 파일별로 브라우저 localStorage에 저장되어 같은 파일을 다시 불러올 때 복원된다.

교호작용이 하나라도 활성화되면 내보낸 보고서에 **Enabled Interactions** 행렬(양 축에 설명변수, $x1:x2$ 곱 항이 포함된 자리에 체크 표시)이 가로 방향으로 실려 earthUI의 Allowed

Interactions 행렬과 짝을 이룬다. 다만 활성화된 항이라도 lasso가 그 계수를 0으로 축소했을 수 있다는 점에 유의하라.

요령: 설명변수가 많을 때 lasso는 불필요한 교호작용 항을 자동으로 0으로 만든다. "Allow All"로 시작해 모형이 선택하도록 두는 것도 합리적인 접근이다.

경고

glmnet 교호작용을 활성화하면 범정이나 감사 상황에서 해석하고 설명하기 어려운 보정치가 산출될 수 있다. 교호작용은 신중하게 사용하고, 그 결과로 나온 보정치가 합리적인지 검증하라.

7.9 주효과에서 차단(교호작용 전용)

교호작용 행렬에서 변수 이름을 오른쪽 클릭하면 그 변수가 주효과로 모형에 들어가지 못하도록 차단할 수 있다. 변수 이름 뒤에 굵은 "1" 표시가 나타난다. 차단되면 다음과 같다.

- 해당 변수는 주효과 항에서 제거된다
- 해당 변수는 교호작용 항($x_1:x_2$)에는 여전히 나타날 수 있다
- 다시 오른쪽 클릭하면 차단이 해제된다
- 이 상태는 입력 파일별로 localStorage를 통해 유지된다

전형적인 사용 사례는 다음과 같다. 부동산 규모에 비례해야 하는 시점 보정을 모형화할 때 sale_age를 주효과에서 차단하고 living_area와의 교호작용에서만 허용한다. 그러면 평평한 sale_age 계수 대신 sale_age:living_area 교호작용 항이 만들어져—시간에 따른 가격 상승률이 주택 규모에 따라 달라진다.

7.10 고급 모수

glmnet Call Parameters 맨 아래의 접이식 "Advanced" 절에는 추가 설정이 있다. 이들은 합리적인 기본값을 사용하지만, 범정이나 감사를 위해 모든 모형 설정을 문서화할 수 있도록 노출되어 있다.

- **Lambda Min Ratio** — 정규화 경로에서 가장 작은 lambda와 가장 큰 lambda의 비율. 기본값: 0.00001. 소규모 데이터셋 (관측치 50개 미만)에서는 더 큰 값(예: 0.01)이 성능을 개선할 수 있다.
- **Number of Lambda Values** — 경로상의 lambda 값 개수. 기본값: 100.
- **CV Loss Metric** — 교차검증의 손실함수: MSE (gaussian의 기본값), MAE(이상치에 강건), Deviance.
- **Convergence Threshold** — 좌표하강의 수렴 임계값. 기본값: $1e-07$.
- **Max Iterations** — 좌표하강의 최대 반복 횟수. 기본값: 100,000. 수렴 경고가 나타나면 늘린다.
- **Fit Intercept** — 절편(상수) 항 포함 여부. 기본값: 켜. 절편은 감정평가 RCA에서 "기준 값(basis value)"에 해당한다.

7.11 설정 기본값

파일을 불러올 때 모수를 어떻게 초기화할지를 세 가지 옵션이 제어한다.

- **Use last settings for input file** — 이 파일에 대해 지난번에 사용한 설정을 그대로 복원
- **Use default settings** — 저장해 둔 전역 기본값을 적용
- **glmnet defaults** — 공장 설정으로 초기화($\alpha = 1$, CV lambda.1se, 10겹, gaussian, standardize 켜)

Save current as default를 클릭하면 현재의 모든 설정이 이후 파일들에 적용될 전역 기본값으로 저장된다.

8 모형 적합

8.1 Fit 버튼

사이드바 제6절에는 **Fit Glmnet Model** 버튼이 있다. 이를 클릭하면 현재 구성으로 모형이 실행된다.

8.2 적합 중에 일어나는 일

Fit을 클릭하면 glmnetUI는 다음을 수행한다.

1. 데이터를 준비한다 — NA가 있는 행을 제거하고, (지정되어 있다면) 가중치를 적용하며, (감정평가 모드에서는) 대상 행을 제외하고, 요인 변수를 가변수 열로 부호화한다.
2. 모형 행렬을 구성한다 — 교호작용 행렬에서 허용된 교호작용 항을 포함하여 설계행렬을 만든다.
3. 제약을 적용한다 — 부호 적용이 활성화되어 있으면 계수에 `upper.limits`와 `lower.limits`를 설정한다.
4. 교차검증을 수행한다 — CV가 선택되어 있으면 `cv.glmnet()`을 호출하여 최적 `lambda`를 찾는다. 완화 lasso가 활성화되어 있으면 `cv.glmnet(..., relax = TRUE)`를 호출한다.
5. 최종 모형을 적합한다 — 적합된 모형, 계수, 예측값, 잔차를 저장한다.
6. 모든 결과 탭을 갱신한다 — Summary, Coefficients, Equation, Variable Importance, Contributions, ANOVA, Diagnostics가 채워진다.

Fit 버튼 아래의 상태 메시지는 관측치 수, 제외된 행 수, 선택된 `lambda`를 보여 준다.

9 결과 탭

9.1 Data Preview

가져온 데이터를 대화형 DataTable로 보여 준다. 감정평가/시장 모드에서는 미리보기가 두 표로 나뉜다. **Subject Property**(1행)와 **Comparable Sales**(2행 이하)이다.

각 칸은 한 줄로 유지되고 열 너비에 맞추어 잘리므로, 폭이 넓은 자유 서술 항목(예: 매물 설명)이 행을 아래로 늘어뜨리지 않는다. 아무 칸이나 더블클릭하면 잘리지 않은 전체 내용을 보여 주는 팝업이 열린다.

9.2 Equation

MathJax를 통해 LaTeX로 렌더링된 적합 모형 방정식을 표시한다. 0이 아닌 모든 계수를 적절한 수학 서식으로 보여 준다.

- 양의 계수는 + 부호로 표시
- 음의 계수는 - 부호로 표시
- 교호작용 항은 $x_1 \times x_2$ 로 표시
- 숫자는 쉼표와 적절한 소수 자릿수로 서식 지정

9.3 Correlation

모든 수치형 설명변수와 반응변수 사이의 피어슨 상관계수를 보여 주는 히트맵 행렬이다. 데이터를 가져온 직후 바로 사용할 수 있다—모형 적합이 필요하지 않다.

- 색상 그레데이션: 파랑(-1) → 흰색(0) → 빨강(+1)
- 각 칸에 상관계수 값이 인쇄됨
- 변수 개수에 따라 글자와 축 크기가 조정됨
- 적합 전에 다중공선성을 확인하는 데 유용함

9.4 Summary

모형 적합 통계량이 카드 형태로 표시된다.

- R^2 — 설명된 분산의 비율
- 조정 R^2 — 설명변수 개수를 반영해 조정한 값
- GR^2 — 선택된 lambda에서 glmnet의 이탈도 비율로부터 얻은 일반화 R^2
- $CV R^2$ — 교차검증 R^2 (CV를 사용한 경우)
- **RMSE** — 평균제곱근오차
- **MAE** — 평균절대오차

감정평가/시장 모드에서는 다음 카드가 추가로 표시된다.

- **Median Ratio**(중위 비율) — (예측값 / 실제값)의 중앙값
- **COD** — 산포계수(Coefficient of Dispersion, 낮을수록 좋음)
- **PRD** — 가격관련편차(Price-Related Differential)

훈련 R^2 가 $CV R^2$ 를 0.1 넘게 초과하면 **과적합 경고**가 나타난다.

카드 아래에는 편의를 위해 Coefficients 탭과 동일한, 부호 경고가 표시된 계수 표가 반복해서 실린다.

9.5 Coefficients

0이 아닌 모든 계수를 다음과 함께 보여 주는 표이다.

- 변수 이름
- 계수 값

- 기대 부호(변수 구성에서 지정한 값)
- 계수의 방향이 기대와 맞지 않을 경우의 부호 경고 표시

9.6 Variable Importance

표준화된 계수 크기, 즉 각 설명변수에 대한 $|\beta| \times \text{sd}(x)$ 이며, 요인 변수의 경우 가변수 열들에 걸쳐 합산된다. 두 가지 표시 방식이 있다.

- 대화형 **plotly**(plotly 패키지가 설치된 경우): 정확한 중요도, 계수, 상대 백분율을 보여주는 호버 툴팁이 달린 가로 막대그래프.
- 정적 **ggplot2**(대체 방식): 심표 서식의 축 이름표가 달린 가로 막대그래프.

그래프 아래의 DataTable은 Variable, Importance, Coefficient, Relative %를 보여 준다.

9.7 Contributions

각 설명변수가 모형의 예측에 기여하는 정도를 보여 주는 변수별 부분효과 그림이다. 드롭다운에서 변수를 선택한다.

수치형 설명변수의 경우: 변수의 x값 대 그 기여도 ($\beta \times x$)의 산점도이며, 빨간 선분과 기울기 표시가 함께 나타난다. 기울기 표시는 단위당 한계효과(예: +1,234.56/unit)를 보여 주며, 위도/경도처럼 값의 범위가 좁은 변수에는 적응적 단위(예: +0.12/0.001)를 사용한다. 부제에는 절편(기준) 값이 표시된다.

요인 변수의 경우: 요인 수준별 기여도 값의 분포를 보여 주는 상자그림이다.

두 변수 교호작용의 경우: 세 가지 시각화가 위아래로 배치된다 — 결합 기여도의 3차원 곡면 그림(가능하면 대화형 plotly, 그렇지 않으면 정적 **persp()**), 기여도에 따라 점의 색을 달리한 산점도, 그리고 두 변수의 범위 전반에 걸친 평균 기여도의 히트맵이다.

9.8 ANOVA

각 설명변수에 대해 다음을 보여 주는 분산 분해 표이다.

- **SS** — 제곱합: $\sum (\text{contribution}_j - \overline{\text{contribution}_j})^2$
- **% of Model SS** — 모형 총제곱합에서 차지하는 비율
- **Coefficient** — 해당 설명변수의 주된 계수

절편 행과 TOTAL MODEL 행이 포함된다. 교호작용 항(“:”을 포함하는 항)은 별도의 그룹으로 표시된다.

9.9 Diagnostics

큰 글자(기본 크기 16pt), 15개의 축 눈금, 심표 서식의 이름표(지수 표기 없음)를 갖춘 네 가지 진단 그림이다.

1. **계수 경로(Coefficient Path)** — $\log(\lambda)$ 가 커짐에 따라 모든 계수가 어떻게 변하는지 보여 준다. 선은 변수별로 색이 구분된다. 정규화 경로와 변수 선택을 시각적으로 파악하는 데 도움이 된다.
2. **CV 오차(CV Error)** — $\log(\lambda)$ 의 함수로 나타낸 교차검증 오차이다. 점은 평균 CV 오차를, 오차막대는 ± 1 표준오차를 나타낸다. 파선 수직선은 $\lambda_{\min}$ (파랑)과 λ_{1se} (초록)를 표시한다.
3. **실제값 대 예측값(Actual vs Predicted)** — 관측값 대 예측값의 산점도이다. 점들은 45도 파선 기준선 주변에 모여야 한다. 산포가 넓을수록 예측 오차가 크다는 뜻이다.
4. **잔차 대 적합값(Residuals vs Fitted)** — 잔차 대 적합값의 산점도이다. 0선(파선) 주위에 무작위로 흩어져야 한다. 어떤 패턴이 보인다면 모형 설정 오류(예: 비선형성, 이분산성)를 시사한다.

각 그림에는 150 DPI로 저장할 수 있는 **Download PNG** 버튼이 있다.

9.10 Glmnet Output

원시 모형 출력으로, 사용된 난수 시드, 모형 출력, 선택된 λ , $\lambda_{\min}/\lambda_{1se}$, 감마 (완화 lasso가 활성화된 경우), 전체 계수 벡터를 보여 준다.

9.11 Report

보고서 항목(감정평가사 이름, 부동산 주소, 보고서 일자, 문서 번호)을 구성하고 이 탭에서 곧바로 Word(.docx) 또는 PDF로 내보낸다. Quarto 기반 보고서 작업 흐름 전체(HTML, Word, PDF)는 제13장을 참조하라.

10 데이터 내려받기

적합 후 **Download Output (Excel)** 버튼(사이드바 제7절, 모든 용도 모드에서 사용 가능)은 예측값과 진단 정보가 담긴 Excel 파일을 프로젝트의 출력 폴더로 내보낸다. 이 절의 제목은 감정평가 모드와 시장 모드에서는 “Download Estimated Sale Prices & Residuals”로, General 모드에서는 “Download Estimated Target Variable(s) & Residuals”로 표시된다.

10.1 출력 열

열	설명
est_<target>	모형 예측값(예: est_sale_price)
residual	실제값 - 예측값
cqa	상태-품질-선호도(Condition-Quality-Appeal) 점수(0-10 척도)
residual_sf	잔차 / 주거면적(living_area가 지정된 경우)
cqa_sf	residual_sf 순위로부터 계산한 CQA
<var>_contribution	예측값에 대한 변수별 기여도
basis	절편 값(모든 행에서 동일)

성공적으로 완료되면 내려받기 버튼에 흰색 체크 표시가 나타난다.

10.2 CQA 점수

CQA는 각 행의 잔차를 다른 모든 행과 견주어 0-10 척도로 순위화한다.

- 높은 CQA(~9-10): 예측보다 훨씬 비싸게 팔림 — 상태/품질/선호도가 우수할 가능성이 높다
- 낮은 CQA(~0-1): 예측보다 훨씬 싸게 팔림 — 상태가 열등하거나 급매일 가능성이 높다
- CQA ~5: 중앙값 부근

감정평가/시장 모드에서는 living_area 열이 지정되어 있으면 행이 residual_sf 내림차순으로 정렬된다.

11 RCA 계산과 내려받기

RCA(잔차 제약 접근법, Residual Constraint Approach) 작업 흐름은 모형 적합 후 **Appraisal** 및 **Market Area Analysis** 모드에서 사용할 수 있다. 후속 단계인 거래사례비교표(제12장)는 감정 평가 모드 전용이다.

11.1 RCA 대화상자 열기

사이드바 제8절의 **Calculate RCA Adjustments & Download** 버튼을 클릭한다. 그다음 선택한다.

- **CQA** 또는 **CQA per SF** 점수 유형. “CQA per SF”를 선택하면 대상의 잔차가 주거면적으로 척도화된다.
- 대상의 CQA 점수를 입력한다(0.00–9.99, 기본값 5.00)

11.2 CQA 점수 보간

1. 비교 사례들의 CQA 점수와 잔차가 정렬된다
2. 선형 보간으로 입력한 CQA 값을 잔차로 대응시킨다
3. 대상 가치 = 모형 예측값 + 보간된 잔차

11.3 출력 열

열	설명
subject_value	모형 예측값 + 보간된 잔차
<var>_adjustment	대상 기여도 – 비교 사례 기여도
residual_adjustment	대상 잔차 – 비교 사례 잔차
net_adjustments	모든 보정치의 합
gross_adjustments	보정치 절댓값의 합
adjusted_sale_price	비교 사례 매매가격 + 순보정치

계산이 성공적으로 끝나면 버튼에 흰색 체크 표시가 나타난다.

12 거래사례비교표

거래사례비교표(Sales Comparison Grid)는 RCA 보정치를 계산한 뒤(제8단계) **Appraisal** 모드에서만 사용할 수 있다. 이 기능은 감정평가서에 넣기에 적합한 서식화된 Excel 통합문서를 생성한다.

12.1 비교 사례 선택 대화상자

사이드바 제9절의 **Generate Sales Grid & Download** 버튼을 클릭하면 RCA 출력에 있는 모든 비교 거래사례를 나열하는 모달 대화상자가 열린다. 비교 사례는 두 그룹으로 나뉜다.

- **권장 사례**(미리 선택됨): 총 보정치가 매매가격의 25% 미만인 사례로, 매매 경과일 순(가장 최근 순)으로 정렬된다.
- **기타 사례**(기본적으로 선택 해제): 나머지 사례로, 매매 경과일 순으로 정렬된다.

최대 30개의 사례를 선택한 뒤 **Generate Sales Grid**를 클릭하면 통합문서가 만들어진다.

12.2 통합문서 구성

생성된 통합문서는 최대 10개의 시트를 가지며, 각 시트에는 비교 사례 3건이 나란히 배치된다.

- **머리글 행**: 대상 및 비교 사례의 주소, 매매가격, 매매일, 시장 노출 일수
- **묶음 행**: 위치(위도, 경도, 지역), 대지 규모 (대지 면적, 대지 치수), 연령(실제 연령, 유효 연령)을 그룹 머리글과 하위 항목으로 구성
- **모형 변수 행**: 0이 아닌 모형 계수마다 한 행씩, 대상 기여도, 비교 사례 기여도, 보정 비율을 표시
- **잔여 특성 행**: 모형이 포착하지 못한 특성(예: 상태, 품질, 조망)에 대해 감정평가사가 입력할 수 있는 빈 잠금 해제 칸
- **요약 행**: 순보정%, 총보정%, 그리고 실시간 Excel 수식이 들어간 보정 매매가격

12.3 시트 보호와 잔여 칸

각 시트는 실수로 편집되는 것을 막기 위해 암호로 보호되지만, 잔여 특성 값 칸은 명시적으로 잠금이 해제되어 있다. 이로써 감정평가사는 수식 기반 보정 계산을 그대로 유지하면서 모형에 없는 특성의 값을 입력할 수 있다.

13 보고서 내려받기

13.1 두 단계 Quarto 작업 흐름

사이드바에서의 보고서 내보내기는 두 단계의 Quarto 작업 흐름이다.

1. **Generate Quarto Report**(사이드바 제10절) — 자기완결적 Quarto 묶음(.qmd 원본과 그림, report_data.rds, reference.docx)을 프로젝트의 glmnet 출력 폴더에 기록한다.
2. **Convert Quarto Report**(사이드바 제11절) — .qmd 파일을 선택한 출력 형식으로 렌더링한다. 경로는 가장 최근에 생성된 묶음이 기본값이지만, 어떤 .qmd든 찾아볼 수 있다.

13.2 보고서 형식

변환 단계에서는 세 가지 형식을 쓸 수 있다.

- **HTML** — 자기완결적 HTML 문서. pandoc 외에 추가 소프트웨어 없이 모든 플랫폼에서 작동한다. 권장 기본값이다.
- **Word (.docx)** — 서식화된 문서. 편집과 배포에 적합하며 모든 플랫폼에서 작동한다.
- **PDF** — LaTeX을 사용하는 Quarto로 생성된다. 용지 크기는 로케일 설정(Letter 또는 A4)을 따른다. **LaTeX** 설치가 필요하다(시스템 요구사항 부록 참조). LaTeX이 감지되지 않으면 형식 체크박스에서 PDF 옵션이 자동으로 숨겨진다.

보고서는 가능한 경우 Quarto로 생성되며, 그렇지 않으면 rmarkdown으로 대체된다. Report 탭은 이와 별도로 Word 내보내기와 PDF 내보내기 버튼을 직접 제공한다.

13.3 보고서 내용

보고서에는 모든 탭의 내용이 포함된다.

- 데이터셋 설명(파일 이름, 관측치 수, 반응변수, 설명변수, 용도)
- 모형 사양(alpha, lambda, family, standardize, 완화 여부, 0이 아닌 계수)
- 결과 요약(R^2 , 조정 R^2 , 일반화 R^2 , CV R^2 , RMSE, MAE)
- 모형 방정식(LaTeX으로 렌더링)
- 계수 표
- 변수 중요도(막대그래프와 표)
- Contributions 탭과 동일한 변수별 기여도 그림:
 - 수치형 설명변수: 선형 적합선이 있는 산점도
 - 요인 변수: 수준별 상자그림
 - 교호작용: 기여도로 색을 구분한 산점도, 평균 기여도의 히트맵, 정적 3차원 곡면 스냅샷(persp)
- Enabled Interactions 행렬(교호작용이 하나라도 활성화된 경우) — 양 축에 설명변수를 두고 x1:x2 곱 항이 포함된 자리에 체크 표시를 하며, 가로 방향으로 실린다
- 상관 행렬 히트맵
- ANOVA 분해 표
- 진단 그림:
 - 계수 경로
 - 교차검증 오차
 - 실제값 대 예측값
 - 잔차 대 적합값
 - 정규 Q-Q 그림
- 모형 출력

모든 축 이름표와 색상 범례는 심표 서식의 숫자를 사용한다(지수 표기 없음).

보고서가 성공적으로 생성되면 보고서 버튼에 흰색 체크 표시가 나타나며, 보고서를 렌더링하는 동안에는 경과 시간이 표시된다.

14 earthUI와의 비교

earth 애플리케이션과 그 glmnet 후처리 모듈은 회귀 모형화를 위한 하나의 도구 모음이다. 두 도구는 동일한 데이터 형식, 특수 열 유형, RCA 작업 흐름, 예제 데이터셋을 공유하지만 모형화 엔진은 서로 다르다.

14.1 주요 차이

항목	glmnetUI	earthUI
방법	엘라스틱 넷(glmnet)	MARS/Earth(earth)
변수 선택	lasso 별점을 통한 자동 선택	전진/후진 가지치기를 통한 자동 선택
계수 관계	선형 효과만	조각별 선형(힌지 함수)
교호작용	교호작용 행렬을 통한 쌍별 교호작용	3차까지
계수 부호	제약을 통해 강제 가능	직접 강제 불가
완화 적합	완화 lasso 사용 가능	해당 없음
Alpha 조정	격자 탐색 사용 가능	해당 없음(혼합 모수 없음)
거래사례비교표	사용 가능(제9단계)	사용 가능(제8단계)
g-함수 그림	해당 없음	묶음 향 시각화
보고서 형식	HTML, Word, PDF	HTML, Word, PDF

14.2 공통 기능

두 도구 모두 다음을 제공한다.

- snake_case 열 변환을 동반한 CSV 및 Excel 가져오기
- CSV 구분자, 소수점 기호, 용지 크기를 포함한 31개국 로케일 지원
- 특수 열 지정(contract_date, living_area, area, latitude, longitude 등)
- 기준일로부터의 sale_age 자동 계산
- 다크/라이트 모드 전환
- localStorage를 통한 설정 보존
- 상관 히트맵, 변수 중요도, 기여도, ANOVA, 진단
- CQA 점수 산정과 RCA 보정 작업 흐름
- 비교 사례 선택과 서식화된 Excel 출력을 갖춘 거래사례비교표
- 동일한 예제 데이터셋(Appraisal_1.csv)과 호환

14.3 어느 것을 언제 쓸 것인가

- 자동 변수 선택을 갖춘 단순하고 해석하기 쉬운 선형모형을 원한다면 **glmnetUI**를 사용하라. 엘라스틱 넷은 설명변수가 많고 모형이 가장 중요한 것들을 골라 주기를 바랄 때 특히 적합하다.
- 비선형 관계를 예상하거나(예: 주거면적이 가격에 미치는 효과가 규모에 따라 달라지는 경우) 임계효과를 포착할 힌지 함수가 필요하다면 **earthUI**를 사용하라. earth 모형은 더 높은 차수의 교호작용(2차, 3차)도 포착할 수 있다. earthUI는 또한 glmnetUI에는 대응물이 없는 실험적인 Prolog/DCG 기반 매물 설명 추출 기능을 제공한다.
- 같은 데이터에 대해 선형모형과 비선형모형을 비교하려면 둘 다 사용하라. 두 모형의 R^2 값이 비슷하다면 해석 가능성 측면에서 더 단순한 glmnet 모형이 선호될 수 있다.

15 예제 데이터셋: Appraisal_1.csv

15.1 설명

예제 데이터셋은 glmnetUI에 함께 들어 있으며 earthUI와도 공유된다. 다음으로 위치를 확인할 수 있다.

```
demo_file <- system.file("extdata", "Appraisal_1.csv", package = "glmnetUI")
```

이 파일에는 모의 MLS 내보내기 자료로 만든 주거용 매매 1,502건(그리고 1행의 대상 부동산 1건)이 들어 있다. 데이터는 규모, 연령, 위치가 다양한 여러 지역 시장의 단독주택 매매를 나타낸다.

이는 실제 데이터가 아니지만 북부 캘리포니아의 현실적인 근린을 바탕으로 한 것이다. 모든 식별 정보는 변경되거나 삭제되었다.

15.2 열

열	자료형	특수 유형	설명
weight	numeric	weight	관측치 가중치(0 = 적합에서 제외)
id	numeric	display_only	내부 레코드 ID
property_id	numeric	display_only	MLS 부동산 식별자
listing_id	character	display_only	MLS 매물 번호
parcel_number	character	display_only	카운티 과세평가관 필지번호(APN)
street_address	character	display_only	부동산 주소
city_name	character	display_only	시
postal_code	character	display_only	우편번호
county_name	character	display_only	카운티
contract_date	Date	contract_date	매매 계약일(sale_age를 계산)
sale_age	numeric	—	계약일부터 기준일까지의 일수
sale_price	numeric	(목표변수)	매매가격 — 반응변수
living_sqft	numeric	living_area	총 주거면적(제곱피트)
beds_total	integer	—	침실 수
baths_total	numeric	—	총 욕실 수(예: 2.5 = 전용 2 + 반욕실 1)
lot_size	numeric	lot_size	대지 면적(제곱피트)
area_id	integer	area	MLS 지역 식별자
age	numeric	actual_age	부동산 연령(년)
latitude	numeric	latitude	위도(모형에서는 소수점 셋째 자리로 반올림)
longitude	numeric	longitude	경도(모형에서는 소수점 셋째 자리로 반올림)
garage_spaces	integer	—	차고 칸 수
days_on_market	integer	dom	시장 노출 일수
listing_date	Date	listing_date	매물 등록일
sale_concessions	numeric	concessions	매도인 양허

15.3 권장 빠른 시작

1. glmnet 루틴을 실행한다: `earthUI::launch_glmnet()`
2. 용도를 **For Appraisal**로 하여 새 프로젝트를 만든다(제1절)
3. Appraisal_1.csv를 프로젝트의 입력 폴더에 넣고 제2절에서 선택한다(필요하면 **Refresh**를 클릭)
4. sale_price를 목표변수로 선택한다
5. 위 표에 나온 대로 특수 유형을 지정한다

6. 설명변수를 포함시킨다: `sale_age`, `living_sqft`, `baths_total`, `lot_size`, `area_id`(요인으로 지정), `age`, `latitude`, `longitude`, `garage_spaces`
7. `alpha = 1`(Lasso), `lambda.1se`를 쓰는 CV를 그대로 둔다
8. **Fit Glmnet Model**을 클릭한다
9. Summary, Coefficients, Diagnostics 탭을 검토한다
10. 출력을 내려받고(제7단계) CQA 순위를 검토한다
11. CQA 점수를 ~5.00으로 하여 RCA 보정치를 계산한다 (제8단계)
12. 권장 비교 사례로 거래사례비교표를 생성한다(제9단계)

16 시스템 요구사항과 문제 해결

16.1 지원 플랫폼

glmnetUI는 **macOS**, **Windows**, **Linux**에서 실행된다. RStudio Desktop(2023.06 이상)을 적극 권장하는데—보고서에 필요한 **pandoc**이 함께 들어 있기 때문이다.

16.2 선택적 의존 요소

- **LaTeX**(TinyTeX, MiKTeX 또는 MacTeX): PDF 보고서 생성에만 필요하다. 감지되지 않으면 PDF 옵션이 자동으로 숨겨진다. 설치 방법: `tinytex::install_tinytex()`
- **sysfonts + showtext**: 그림용 Roboto Condensed 글꼴. 사용할 수 없거나 오프라인이면 시스템 산세리프로 자동 대체된다.
- **earth**($\geq 5.3.0$): earthUI 가져오기 파이프라인에만 필요하다.

16.3 플랫폼별 참고사항

macOS: 문제가 가장 적다. PDF를 위해 TinyTeX을 설치하라: `tinytex::install_tinytex()`

Windows: RStudio와 잘 작동한다. 기업용/보안이 강화된 컴퓨터는 임시 디렉터리 제한이 있을 수 있는데—앱은 콘솔에 경고를 표시하되 설정 보존 없이 계속 실행된다.

Linux: 시스템 라이브러리가 필요할 수 있다: `sudo apt install libcurl4-openssl-dev libssl-dev libxml2-dev libsqlite3-dev libfontconfig1-dev`

16.4 점진적 기능 축소

없는 구성요소	동작
LaTeX 없음	PDF 옵션이 숨겨짐. HTML과 Word는 계속 사용 가능.
인터넷 없음 / 글꼴 실패	시스템 산세리프 사용. 콘솔에 메시지 기록.
SQLite 없음 / 읽기 전용 파일시스템	설정이 보존되지 않음. 앱은 정상 실행.
임시 디렉터리에 쓸 수 없음	보고서 생성이 명확한 오류와 함께 실패. TMPDIR을 쓰기 가능한 위치로 설정하라.

16.5 문제 해결

“**PDF option not available**”: R에서 `tinytex::install_tinytex()`을 실행한 뒤 앱을 다시 시작하라.

“**Settings will not persist**”: 사용자 데이터 디렉터리의 권한을 확인하라(macOS: `~/Library/Application Support/R/glmnetUI/`, Linux: `~/.local/share/R/glmnetUI/`, Windows: `%APPDATA%/R/data/glmnetUI/`).

7879 포트가 이미 사용 중임: `lsof -ti:7879 | xargs kill`(macOS/Linux) 또는 `Stop-Process -Id (Get-NetTCPConnection -LocalPort 7879).OwningProcess` (Windows PowerShell)를 실행하라.

참고문헌

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani (2010). “Regularization Paths for Generalized Linear Models via Coordinate Descent”. In: *Journal of Statistical Software* 33.1, pp. 1–22. DOI: [10.18637/jss.v033.i01](https://doi.org/10.18637/jss.v033.i01).

- Simon, Noah et al. (2011). “Regularization Paths for Cox’s Proportional Hazards Model via Coordinate Descent”. In: *Journal of Statistical Software* 39.5, pp. 1–13. DOI: [10.18637/jss.v039.i05](https://doi.org/10.18637/jss.v039.i05).
- Tay, J. Kenneth, Balasubramanian Narasimhan, and Trevor Hastie (2023). “Elastic Net Regularization Paths for All Generalized Linear Models”. In: *Journal of Statistical Software* 106.1, pp. 1–31. DOI: [10.18637/jss.v106.i01](https://doi.org/10.18637/jss.v106.i01).