

# glmnetUI : une interface interactive pour le package glmnet (Lasso et Elastic Net)

*Traduction française, préparée avec l'assistance d'une IA et revue par la rédaction ; la version anglaise fait foi.*

William Bert Craytor

Volume 1, Numéro 1 · Été 2026

Publié le 1<sup>er</sup> juillet 2026

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

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

Cet article en libre accès est diffusé selon les termes de la [licence Creative Commons Attribution 4.0](#).

**Résumé.** glmnetUI est une interface graphique pour le package R glmnet, qui ajuste des modèles linéaires généralisés régularisés au moyen des pénalités Lasso, ridge et elastic-net. Elle propose trois modes d'utilisation — modélisation prédictive générale, évaluation immobilière et analyse de zone de marché — et guide l'utilisateur à travers l'importation des données, la configuration du modèle, l'ajustement par validation croisée, les chemins de régularisation des coefficients et les rapports téléchargeables. Cet article documente les exigences de format de données de glmnetUI, le flux de travail de modélisation, les affichages de sortie et la référence complète des fonctionnalités.

**Mots-clés :** Lasso, régression ridge, elastic net, régression régularisée, glmnet, R, interface graphique, évaluation immobilière

**Note de l'éditeur (juillet 2026) :** Peu avant la mise sous presse, glmnetUI a été intégré à earthUI (version 0.11.0 et ultérieures), qui fournit désormais le flux de travail elastic-net décrit ici aux côtés des flux de travail MARS et GAM. glmnetUI demeure disponible dans sa dernière version publiée (0.5.0) à l'adresse <https://github.com/wcraytor/glmnetUI> ; les développements futurs se poursuivent dans earthUI. Le présent article s'applique à glmnetUI tel que publié.

## Table des matières

|          |                                                     |          |
|----------|-----------------------------------------------------|----------|
| <b>1</b> | <b>Introduction</b>                                 | <b>3</b> |
| 1.1      | Historique et contexte                              | 3        |
| 1.1.1    | Le Lasso                                            | 3        |
| 1.1.2    | Régression ridge et elastic net                     | 4        |
| 1.1.3    | Le package glmnet                                   | 4        |
| 1.1.4    | Comparaison avec MARS (earth) et GAM (mgcv)         | 4        |
| 1.1.5    | Flux de travail recommandé                          | 5        |
| 1.2      | Qu'est-ce que glmnetUI ?                            | 6        |
| 1.3      | Trois modes Purpose                                 | 6        |
| 1.4      | Fonctionnalités propres à l'immobilier              | 7        |
| 1.5      | Premiers pas                                        | 7        |
| <b>2</b> | <b>Exigences relatives aux données d'entrée MLS</b> | <b>9</b> |
| 2.1      | Format et structure du fichier                      | 9        |
| 2.2      | Colonnes requises en mode évaluation                | 9        |
| 2.3      | Conventions de nommage des colonnes spéciales       | 10       |
| 2.4      | Qualité et complétude des données                   | 10       |
| 2.5      | Emplacement de la ligne du bien à évaluer           | 11       |

|          |                                                    |           |
|----------|----------------------------------------------------|-----------|
| <b>3</b> | <b>Mode General Purpose</b>                        | <b>12</b> |
| 3.1      | Vue d'ensemble                                     | 12        |
| 3.1.1    | Skip First Row                                     | 12        |
| 3.1.2    | Paramètres propres à chaque finalité               | 12        |
| 3.2      | Le flux de travail de la barre latérale            | 12        |
| 3.3      | Projets                                            | 13        |
| 3.4      | Importer depuis earthUI (facultatif)               | 14        |
| 3.5      | Onglets du panneau principal                       | 14        |
| 3.6      | Persistance des paramètres                         | 15        |
| 3.7      | Mode sombre                                        | 15        |
| 3.8      | Locale et paramètres régionaux                     | 15        |
| 3.8.1    | Enregistrer les valeurs par défaut                 | 15        |
| <b>4</b> | <b>Mode évaluation</b>                             | <b>16</b> |
| 4.1      | Traitement de la ligne du bien à évaluer           | 16        |
| 4.2      | Date d'effet et ancienneté de la vente             | 16        |
| 4.3      | Désignations spéciales de colonnes                 | 16        |
| 4.4      | Aperçu des ajustements RCA                         | 16        |
| <b>5</b> | <b>Mode analyse de zone de marché</b>              | <b>17</b> |
| 5.1      | Différences avec le mode évaluation                | 17        |
| 5.2      | Quand utiliser le mode marché                      | 17        |
| <b>6</b> | <b>Sélection des variables</b>                     | <b>18</b> |
| 6.1      | Variable cible                                     | 18        |
| 6.2      | Le tableau des prédicteurs                         | 18        |
| 6.3      | Détection des types de données et substitutions    | 18        |
| 6.4      | Signes attendus et application                     | 18        |
| 6.5      | Référence des types de colonnes spéciales          | 19        |
| <b>7</b> | <b>Choix des paramètres</b>                        | <b>20</b> |
| 7.1      | Alpha (paramètre de mélange)                       | 20        |
| 7.2      | Choix de lambda                                    | 20        |
| 7.3      | Famille                                            | 20        |
| 7.4      | Standardize                                        | 21        |
| 7.5      | Contraintes de signe                               | 21        |
| 7.6      | Relaxed lasso                                      | 21        |
| 7.7      | Graine aléatoire                                   | 21        |
| 7.8      | Matrice d'interactions                             | 22        |
| 7.9      | Exclusion de l'effet principal (interaction seule) | 22        |
| 7.10     | Paramètres avancés                                 | 22        |
| 7.11     | Valeurs par défaut des paramètres                  | 23        |
| <b>8</b> | <b>Ajustement du modèle</b>                        | <b>24</b> |
| 8.1      | Le bouton d'ajustement                             | 24        |
| 8.2      | Ce qui se passe pendant l'ajustement               | 24        |
| <b>9</b> | <b>Onglets de résultats</b>                        | <b>25</b> |
| 9.1      | Data Preview                                       | 25        |
| 9.2      | Equation                                           | 25        |
| 9.3      | Correlation                                        | 25        |
| 9.4      | Summary                                            | 25        |

|           |                                                          |           |
|-----------|----------------------------------------------------------|-----------|
| 9.5       | Coefficients                                             | 26        |
| 9.6       | Variable Importance                                      | 26        |
| 9.7       | Contributions                                            | 26        |
| 9.8       | ANOVA                                                    | 26        |
| 9.9       | Diagnostics                                              | 26        |
| 9.10      | Glmnet Output                                            | 27        |
| 9.11      | Report                                                   | 27        |
| <b>10</b> | <b>Téléchargement des données</b>                        | <b>28</b> |
| 10.1      | Colonnes de sortie                                       | 28        |
| 10.2      | Scores CQA                                               | 28        |
| <b>11</b> | <b>Calculs RCA et téléchargement</b>                     | <b>29</b> |
| 11.1      | Ouverture de la boîte de dialogue RCA                    | 29        |
| 11.2      | Interpolation du score CQA                               | 29        |
| 11.3      | Colonnes de sortie                                       | 29        |
| <b>12</b> | <b>Sales Comparison Grid</b>                             | <b>30</b> |
| 12.1      | Boîte de dialogue de sélection des comparables           | 30        |
| 12.2      | Structure du classeur                                    | 30        |
| 12.3      | Protection des feuilles et cellules résiduelles          | 30        |
| <b>13</b> | <b>Téléchargement des rapports</b>                       | <b>31</b> |
| 13.1      | Le flux de travail Quarto en deux étapes                 | 31        |
| 13.2      | Formats de rapport                                       | 31        |
| 13.3      | Contenu des rapports                                     | 31        |
| <b>14</b> | <b>Comparaison avec earthUI</b>                          | <b>33</b> |
| 14.1      | Principales différences                                  | 33        |
| 14.2      | Fonctionnalités communes                                 | 33        |
| 14.3      | Lequel utiliser, et quand                                | 33        |
| <b>15</b> | <b>Jeu de données de démonstration : Appraisal_1.csv</b> | <b>35</b> |
| 15.1      | Description                                              | 35        |
| 15.2      | Colonnes                                                 | 35        |
| 15.3      | Prise en main rapide suggérée                            | 36        |
| <b>16</b> | <b>Configuration requise et résolution des problèmes</b> | <b>37</b> |
| 16.1      | Plateformes prises en charge                             | 37        |
| 16.2      | Dépendances facultatives                                 | 37        |
| 16.3      | Remarques par plateforme                                 | 37        |
| 16.4      | Dégradation progressive                                  | 37        |
| 16.5      | Résolution des problèmes                                 | 37        |

## 1 Introduction

### 1.1 Historique et contexte

#### 1.1.1 Le Lasso

Le **Lasso** (**L**east **A**bsolute **S**hrinkage and **S**election **O**perator, soit « opérateur de rétrécissement et de sélection par valeur absolue ») a été introduit par Robert Tibshirani en 1996. Le nom

est un jeu de mots délibéré sur le lasso du cow-boy — une corde qui contraint et qui sélectionne. Tibshirani l’a choisi parce que la méthode « lasso » littéralement les coefficients : elle contraint leur valeur absolue et en ramène certains exactement à zéro, sélectionnant ainsi les variables qui demeurent dans le modèle.

L’intuition de Tibshirani a consisté à ajouter une pénalité  $L_1$  (somme des valeurs absolues des coefficients) au critère des moindres carrés. Contrairement à la pénalité  $L_2$  de la régression ridge (Hoerl et Kennard, 1970), qui rétrécit les coefficients vers zéro sans jamais l’atteindre, la pénalité  $L_1$  produit des solutions *parcimonieuses* dans lesquelles de nombreux coefficients sont exactement nuls. Le Lasso est donc à la fois une méthode de régularisation et une méthode de sélection de variables.

### 1.1.2 Régression ridge et elastic net

La **régression ridge** a été proposée par Arthur Hoerl et Robert Kennard en 1970. Elle ajoute une pénalité  $L_2$  (somme des carrés des coefficients) qui rétrécit tous les coefficients proportionnellement mais conserve toutes les variables dans le modèle. Ridge est efficace lorsque les prédicteurs sont corrélés (multicolinéarité), mais elle ne simplifie pas le modèle.

L’**elastic net**, introduit par Hui Zou et Trevor Hastie en 2005, combine les deux pénalités. Le paramètre de mélange  $\alpha$  contrôle le dosage :  $\alpha = 1$  correspond au Lasso pur,  $\alpha = 0$  à la régression ridge pure, et les valeurs intermédiaires donnent un compromis. L’elastic net surmonte une limite du Lasso : lorsque les prédicteurs sont fortement corrélés, le Lasso tend à en sélectionner un seul et à ignorer les autres, tandis que l’elastic net regroupe les prédicteurs corrélés.

### 1.1.3 Le package glmnet

Le package R `glmnet` a été développé par Jerome Friedman, Trevor Hastie et Robert Tibshirani à l’université Stanford. Il met en œuvre la régularisation elastic net pour les modèles linéaires généralisés au moyen d’un algorithme de descente par coordonnées extrêmement efficace (FRIEDMAN, HASTIE et TIBSHIRANI 2010), étendu depuis aux modèles à risques proportionnels de Cox (SIMON et al. 2011) et à toutes les familles de modèles linéaires généralisés (TAY, NARASIMHAN et HASTIE 2023). Le package prend en charge les familles gaussian (linéaire), binomial (logistique), poisson et d’autres familles GLM.

Parmi les fonctionnalités clés de `glmnet` :

- **Chemin de régularisation** : ajuste le modèle sur toute une grille de valeurs de  $\lambda$  en une seule passe, ce qui est bien plus rapide qu’un ajustement séparé pour chaque  $\lambda$ .
- **Validation croisée** : `cv.glmnet()` sélectionne automatiquement l’intensité optimale de régularisation par validation croisée à  $k$  blocs.
- **Bornes sur les coefficients** : prise en charge de limites supérieures et inférieures sur les coefficients individuels (contraintes de signe).
- **Relaxed lasso** (lasso relâché) : réestime le modèle sur les variables sélectionnées avec une pénalisation réduite ou nulle, ce qui diminue le biais.

### 1.1.4 Comparaison avec MARS (earth) et GAM (mgcv)

L’application `earth` et ses deux Post Processing Modules (modules de post-traitement) `glmnet` et `mgcv` reposent sur des moteurs de modélisation différents. Le tableau suivant résume les principales différences :

| Caractéristique           | glmnet (elastic net)                                                     | earth (MARS)                                                                               | mgcv (GAM)                                                                      |
|---------------------------|--------------------------------------------------------------------------|--------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|
| Type de modèle            | Linéaire (avec régularisation)                                           | Linéaire par morceaux (splines adaptatives)                                                | Non linéaire lisse (splines pénalisées)                                         |
| Non-linéarité             | Uniquement par interactions ou expansion de base                         | Automatique, par fonctions charnières à nœuds déduits des données                          | Automatique, par fonctions de lissage (thin plate, cubique, etc.)               |
| Sélection de variables    | Intégrée, via la pénalité $L_1$ (Lasso)                                  | Intégrée, par sélection pas à pas avant/arrière avec élagage GCV                           | Facultative, par double pénalité de rétrécissement ( <code>select=TRUE</code> ) |
| Interactions              | Produits croisés deux à deux (difficiles à interpréter)                  | Produits de charnières (interprétables, jusqu'au degré 3)                                  | Lissages par produit tensoriel ( <code>te()</code> , <code>ti()</code> )        |
| Interprétabilité          | Les coefficients sont des poids linéaires, faciles à expliquer un par un | Les g-fonctions montrent des effets partiels linéaires par morceaux, variable par variable | Courbes d'effets partiels lisses ; très intuitives                              |
| Domaine de prédilection   | Données de grande dimension, sélection de variables, petits échantillons | Détection automatique des non-linéarités et des interactions                               | Relations non linéaires lisses, modélisation souple                             |
| Ajustements RCA           | Ajustements linéaires par variable                                       | Ajustements linéaires par morceaux via les g-fonctions                                     | Ajustements lisses via les effets partiels                                      |
| Contrôle du surajustement | Lambda par validation croisée                                            | Élagage fondé sur le GCV                                                                   | Estimation du paramètre de lissage par REML/GCV                                 |

**MARS** (Multivariate Adaptive Regression Splines) a été introduit par Jerome Friedman en 1991. La méthode construit des modèles linéaires par morceaux en sélectionnant de manière adaptative des fonctions charnières  $\max(0, x - k)$  et la position de leurs nœuds à partir des données. L'implémentation en R est le package **earth** de Stephen Milborrow. MARS excelle à découvrir automatiquement les relations non linéaires et les interactions sans que l'utilisateur ait à les spécifier. Ses g-fonctions (termes de base regroupés par variable) fournissent des courbes d'effets partiels hautement interprétables.

Les **GAM** (Generalized Additive Models, modèles additifs généralisés) ont été formalisés par Hastie et Tibshirani (1990). L'implémentation en R, **mgcv** de Simon Wood, propose des splines de régression pénalisées avec sélection automatique du degré de lissage par REML ou GCV. Les GAM produisent des courbes d'effets partiels lisses, intuitives à visualiser et à expliquer. Lorsque **earthUI** exporte les positions de nœuds vers **mgcvUI**, les nœuds issus d'**earth** servent de points de départ aux termes de lissage du GAM, combinant le placement adaptatif des nœuds propre à MARS et l'estimation lisse propre aux GAM.

### 1.1.5 Flux de travail recommandé

Pour l'évaluation immobilière et les applications analogues qui exigent des modèles interprétables et défendables :

1. **Commencer par earthUI** pour découvrir les variables importantes, les relations non linéaires et les interactions présentes dans les données. La sélection automatique de base

d'earth et ses g-fonctions fournissent d'excellents éclairages initiaux.

2. **Affiner avec mgcvUI** si des effets partiels plus lisses sont souhaités. L'importation des nœuds d'earthUI dans mgcvUI assure une transition sans rupture des modèles linéaires par morceaux vers les modèles lisses.
3. **Utiliser glmnetUI** lorsqu'une sélection de variables agressive (Lasso) est nécessaire, lorsque le nombre de prédicteurs approche ou dépasse le nombre d'observations, ou lorsqu'on souhaite un modèle purement linéaire avec régularisation, pour des raisons de défendabilité.

## 1.2 Qu'est-ce que glmnetUI ?

glmnetUI est une interface graphique pour le package R **glmnet**. Elle s'exécute sous la forme d'une application Shiny locale — sans identifiant, sans serveur et sans compte. On la lance depuis R, on importe un jeu de données (CSV ou Excel), on configure le modèle et on l'ajuste de manière interactive.

Depuis la version actuelle, glmnetUI n'est plus un package installé séparément : il est livré à l'intérieur du package earthUI, dont il constitue le Post Processing Module elastic net, lancé par `earthUI::launch_glmnet()`. L'application elle-même — et tout ce qui est décrit dans le présent article — demeure inchangée ; seules l'installation et la commande de lancement diffèrent. Les trois routines (earth, glmnet, mgcv) partagent une même arborescence de projets, une même base de données de paramètres et une même copie de l'infrastructure de support.

L'application offre un flux de travail complet : importation des données, configuration des variables, ajustement du modèle, graphiques de diagnostic, importance des variables, équations du modèle et rapports téléchargeables aux formats Word, PDF ou HTML.

La **régression elastic net** combine deux techniques de régularisation :

- **Régression ridge** ( $\alpha = 0$ ) : rétrécit proportionnellement tous les coefficients vers zéro. Adaptée lorsqu'on estime que toutes les variables sont pertinentes.
- **Régression Lasso** ( $\alpha = 1$ ) : peut ramener des coefficients exactement à zéro, réalisant une sélection automatique de variables. Produit des modèles plus simples et plus interprétables.
- **Elastic net** ( $0 < \alpha < 1$ ) : un mélange des deux approches. Utile lorsque les prédicteurs sont corrélés.

## 1.3 Trois modes Purpose

Chaque projet glmnetUI possède un **Purpose** (finalité), choisi à sa création (section 1, *Project*). La finalité du projet actif détermine les outils et les éléments d'interface qui apparaissent ; pour travailler dans un autre mode, il faut créer ou ouvrir un projet doté de cette finalité. Il existe trois modes :

- **General** (général) — régression elastic net pour tout type de population ou de jeu de données. Il s'agit du mode par défaut. Il offre le flux de travail complet de régression régularisée, sans aucun ajout propre à un domaine particulier.
- **For Appraisal** (pour l'évaluation immobilière) — régression elastic net adaptée à l'évaluation immobilière. Ce mode ajoute des fonctionnalités propres à l'évaluation d'un bien individuel, notamment la prise en charge du bien à évaluer, les désignations spéciales de colonnes et la Residual Constraint Approach (RCA) (fr. : approche par contraintes résiduelles).
- **Market Area Analysis** (analyse de zone de marché) — régression elastic net adaptée aux études de zone de marché. Ce mode ajoute des fonctionnalités permettant d'analyser des groupes de biens dans un marché défini, avec exclusion facultative de la ligne du bien à évaluer.

Dans les trois modes, le moteur de modélisation central est identique — on ajuste toujours un modèle elastic net (glmnet). Le réglage Purpose détermine uniquement les outils et les éléments d'interface supplémentaires disponibles.

## 1.4 Fonctionnalités propres à l'immobilier

Lorsque **For Appraisal** ou **Market Area Analysis** est sélectionné, glmnetUI active plusieurs fonctionnalités conçues pour l'analyse immobilière :

- **Désignations spéciales de colonnes** — chaque prédicteur peut être doté d'un rôle spécial tel que `contract_date`, `dom`, `concessions`, `latitude`, `longitude`, `living_area`, `lot_size`, `actual_age`, `effective_age`, `area`, `site_dimensions` ou `display_only`. Ces désignations déterminent le traitement de la colonne lors de l'ajustement et dans les sorties.
- **Arrondi de la latitude et de la longitude** — les colonnes désignées comme latitude ou longitude sont automatiquement arrondies à 3 décimales afin d'éviter le surajustement.
- **Colonne Sale Age** (ancienneté de la vente) — lorsqu'une colonne est désignée comme `contract_date` et qu'une Effective Date (date d'effet) est fournie, glmnetUI calcule une colonne `sale_age` (nombre de jours entre la date de vente et la date d'effet) et la substitue comme prédicteur.
- **Calculs RCA (évaluation uniquement)** — en mode évaluation, après l'ajustement du modèle, glmnetUI peut produire la sortie Residual Constraint Approach (RCA), qui fournit des ajustements par comparable, des synthèses d'ajustements nets et bruts, ainsi qu'un prix de vente ajusté pour le bien à évaluer.

***Conseil :** le mode General convient à tout jeu de données — financier, scientifique, technique ou immobilier. Les modes évaluation et zone de marché ne font qu'ajouter des fonctionnalités de commodité destinées aux professionnels de l'immobilier.*

## 1.5 Premiers pas

Pour utiliser glmnetUI :

1. **Installer earthUI** dans R : `install.packages("earthUI")` — la routine glmnet est incluse. (Voir « Installing earthUI and vProlog » dans l'article consacré à earthUI dans ce numéro pour la version de développement et les composants facultatifs.)
2. **Lancer l'application** : exécuter `earthUI::launch_glmnet()` dans R. L'application s'ouvre dans le navigateur web sur le port 7879. Il est également possible d'y accéder directement en se rendant à l'adresse <http://localhost:7879>. L'application mémorise le dernier mode Purpose utilisé et le restaure automatiquement.
3. **Créer ou ouvrir un projet** dans la section 1 de la barre latérale. Un projet fixe la finalité et le lieu de travail (pays, État, comté, ville) ainsi que ses dossiers d'entrée et de sortie. Les projets sont stockés sous une racine `regProj` partagée et sont communs aux applications sœurs earthUI et mgcvUI. C'est en ouvrant un projet doté d'une autre finalité que l'on change de mode.
4. **Importer les données** dans la section 2 en sélectionnant un fichier CSV ou Excel dans le dossier d'entrée du projet actif. Ajouter les fichiers à ce dossier, puis cliquer sur **Refresh** (actualiser) pour les faire apparaître dans la liste.
5. **Importer depuis earthUI (facultatif)** — la section 3 permet d'importer un fichier de résultat earthUI `.rds` afin d'utiliser les fonctions de base charnières d'earth avec la régularisation de glmnet. Cette étape peut être ignorée en l'absence de résultat earthUI.
6. **Configurer les variables** — choisir la variable cible et les prédicteurs, définir les types de données, les signes attendus et attribuer les éventuels rôles spéciaux de colonnes.

7. **Régler les paramètres de glmnet** — alpha, lambda, famille, contraintes de signe, interactions et autres options.
8. **Ajuster le modèle** — cliquer sur « Fit Glmnet Model » (ajuster le modèle glmnet) et examiner les résultats dans le panneau principal.
9. **Exporter** — télécharger les prédictions au format Excel, générer et convertir un rapport Quarto ou, en mode évaluation, calculer les ajustements RCA et la Sales Comparison Grid (grille de comparaison).

Les paramètres sont automatiquement conservés dans le stockage local du navigateur et restaurés au rechargement du même fichier d'entrée.

## 2 Exigences relatives aux données d'entrée MLS

Pour les flux de travail d'évaluation immobilière et d'analyse de marché, les données d'entrée proviennent en général d'une exportation de Multiple Listing Service (MLS). Ce chapitre décrit la structure de fichier attendue et les colonnes que glmnetUI sait exploiter.

### 2.1 Format et structure du fichier

glmnetUI accepte les fichiers **CSV** et **Excel** (.xlsx, .xls). À l'import, les noms de colonnes sont automatiquement convertis en **snake\_case** — par exemple, « Living SqFt » devient **living\_sqft**, « Contract Date » devient **contract\_date** et « Sale Price » devient **sale\_price**. Cette normalisation garantit des références de colonnes cohérentes tout au long du flux de travail. Le séparateur CSV et le séparateur décimal utilisés à l'import sont déterminés par les paramètres de locale (voir le chapitre 3, « Locale & Regional Settings »).

Le fichier de données doit être un tableau plat comportant une ligne par bien et une colonne par attribut. La première ligne du fichier doit contenir les en-têtes de colonnes.

### 2.2 Colonnes requises en mode évaluation

Si glmnetUI fonctionne avec n'importe quel ensemble de colonnes, le flux de travail complet d'évaluation tire profit des colonnes suivantes :

| Colonne          | Type spécial           | Rôle                                                                              |
|------------------|------------------------|-----------------------------------------------------------------------------------|
| Sale Price       | (cible)                | Variable réponse du modèle                                                        |
| Contract Date    | <b>contract_date</b>   | Sert à calculer <b>sale_age</b> (jours depuis la date d'effet)                    |
| Listing Date     | <b>listing_date</b>    | Utilisée avec la date de contrat pour calculer le DOM en l'absence de colonne DOM |
| Days on Market   | <b>dom</b>             | Jours de mise en marché ; affiché dans les exportations                           |
| Concessions      | <b>concessions</b>     | Concessions de vente ; prix net = prix de vente – concessions                     |
| Living Area (SF) | <b>living_area</b>     | Active les résidus au pied carré ( <b>residual_sf</b> , <b>cqa_sf</b> )           |
| Lot Size         | <b>lot_size</b>        | Colonne de superficie du terrain                                                  |
| Site Dimensions  | <b>site_dimensions</b> | Regroupée avec la superficie du terrain                                           |
| Latitude         | <b>latitude</b>        | Arrondie à 3 décimales                                                            |
| Longitude        | <b>longitude</b>       | Arrondie à 3 décimales                                                            |
| Area ID          | <b>area</b>            | Identifiant de zone de marché ou de quartier                                      |
| Actual Age       | <b>actual_age</b>      | Âge du bien                                                                       |
| Effective Age    | <b>effective_age</b>   | Âge effectif du bien                                                              |
| Address          | <b>display_only</b>    | Affichée dans les exportations ; exclue du modèle                                 |

Les noms de colonnes du tableur peuvent être rédigés dans une langue étrangère — les noms « spéciaux » sont en anglais afin que le programme R puisse leur réserver un traitement particulier. Pour le reste, les noms de colonnes fournis apparaissent dans les modèles de régression, les graphiques et, s'il s'agit d'évaluations, dans les rapports produits.

Toutes les colonnes ne sont pas obligatoires. glmnetUI s'adapte : si une colonne manque, la fonctionnalité correspondante est simplement omise. Cela dit, pour les modèles de prix immobiliers, certaines colonnes sont vivement recommandées afin d'obtenir un ajustement acceptable :

1. **Sale Age** (ancienneté de la vente) — le nombre de jours entre la date de contrat de vente et la date d'effet de l'évaluation ou de l'analyse. Lorsqu'on utilise un historique de ventes pluriannuel, en particulier sur des périodes de plus de 5 ans, **sale\_age** joue souvent un rôle central dans l'estimation du prix de vente.
2. **Living Area** (surface habitable) — également désignée par « Living Sqft », « GLA » (gross living area) et autres. Il s'agit d'un autre déterminant majeur du prix de vente.
3. **Total Bath Count** (nombre total de salles de bains) — le nombre total de salles de bains complètes, aux trois quarts, à moitié et au quart. Par exemple, deux salles de bains complètes et une demi-salle de bains donnent la valeur 2,5.
4. **Garage Bays** ou **Garage Area** — le nombre de places de garage ou la surface du garage.
5. **Lot Size** (superficie du terrain) — la surface foncière du bien, généralement exprimée en pieds carrés ou en acres.
6. **Longitude**, **Latitude** et, si disponible, **Area ID**. Ces variables de localisation aident le modèle à rendre compte de la variation géographique des prix.

## 2.3 Conventions de nommage des colonnes spéciales

glmnetUI identifie les colonnes par leur **désignation de type spécial**, et non par leur nom. Les colonnes peuvent porter n'importe quel nom dans l'exportation MLS — seule compte l'attribution du bon type spécial dans le tableau Variable Configuration (chapitre 6).

Par exemple, votre MLS peut exporter la surface habitable sous le nom « GLA », « Living SqFt », « liv\_area » ou « gross\_living\_area ». Après l'import (au cours duquel le nom devient `snake_case`), il suffit de la désigner comme `living_area` dans la liste déroulante Special. glmnetUI l'utilisera alors pour les calculs de résidus au pied carré, quel que soit son nom d'origine.

## 2.4 Qualité et complétude des données

- **Valeurs manquantes (NA)** : les lignes comportant des NA dans une colonne prédictrice ou cible sont automatiquement supprimées avant l'ajustement.
- **Colonnes de dates** : les colonnes de type caractère correspondant aux formats de date courants sont automatiquement analysées à l'import. Les formats d'année à 2 chiffres sont privilégiés lorsqu'aucune année à 4 chiffres n'est détectée.
- **Colonnes numériques** : le prix de vente, la surface habitable, la superficie du terrain, les concessions et les champs analogues doivent être numériques. Si votre MLS exporte les prix avec des symboles monétaires ou des séparateurs de milliers (par exemple « \$350,000 » ou « 350.000,00 »), il peut être nécessaire de les nettoyer avant l'import.
- **Colonnes facteurs** : les colonnes de texte, TRUE/FALSE et les facteurs R sont automatiquement traités comme catégoriels — une régression ne peut les traiter que comme des facteurs. Une colonne **numérique** (par exemple un code de zone ou de style numérique) ne devient un facteur que si vous cochez sa case **Factor** ; glmnetUI ne déduit jamais le caractère catégoriel de l'étendue des valeurs d'une variable numérique, de sorte que des prédicteurs numériques discrets comme `bath_count` restent continus. Le rôle Special est indépendant de la décision relative au facteur.

***Conseil** : après l'import, examinez la colonne NA du tableau des prédicteurs. Les colonnes comportant de nombreuses valeurs manquantes peuvent entraîner la suppression de lignes. Il convient d'envisager d'exclure les colonnes à fort taux de NA ou de nettoyer les données avant l'import.*

## 2.5 Emplacement de la ligne du bien à évaluer

En mode **Appraisal**, la ligne 1 doit être le bien à évaluer. Toutes les lignes suivantes sont des ventes comparables. La ligne du bien à évaluer est exclue de l'ajustement du modèle. Après l'ajustement, le modèle produit néanmoins des prédictions pour cette ligne.

En mode **Market Area Analysis**, le placement du bien à évaluer en ligne 1 est facultatif.

En mode **General**, il n'existe aucun traitement particulier des lignes — toutes sont traitées de la même façon.

### 3 Mode General Purpose

#### 3.1 Vue d'ensemble

Le mode General Purpose est le mode par défaut au lancement de glmnetUI. Il offre le flux de travail complet de régression elastic net pour tout jeu de données — et pas seulement en immobilier. glmnetUI peut ainsi servir à des données scientifiques, à des analyses financières, à des études techniques ou à tout problème de régression.

En mode General, l'interface omet les fonctionnalités propres à l'immobilier (colonnes spéciales, ancienneté de la vente, arrondi des coordonnées, RCA). La barre latérale est allégée et se concentre sur la sélection des variables, la configuration des paramètres, l'ajustement du modèle et l'exportation.

##### 3.1.1 Skip First Row

Une case à cocher **Skip first row** (ignorer la première ligne) apparaît en haut de la barre latérale dès qu'un projet General ou Market est ouvert. Lorsqu'elle est cochée, la ligne 1 est exclue de l'ajustement du modèle. C'est utile lorsque la ligne 1 contient une observation cible ou de référence qui ne doit pas influencer le modèle. En mode Appraisal, la ligne 1 (le bien à évaluer) est toujours exclue automatiquement.

##### 3.1.2 Paramètres propres à chaque finalité

Les paramètres (sélection des prédicteurs, paramètres du modèle, interactions) sont enregistrés séparément pour chaque combinaison de fichier d'entrée et de mode Purpose. Le passage d'un mode General, Appraisal ou Market à un autre préserve indépendamment les paramètres de chaque mode.

#### 3.2 Le flux de travail de la barre latérale

La barre latérale est organisée en sections numérotées et repliables qui guident l'utilisateur depuis la sélection du projet jusqu'à l'exportation du rapport. La plupart des sections n'apparaissent qu'une fois un projet actif ouvert, et plusieurs sont propres à la finalité du projet. Le dossier de sortie est déduit du projet actif — il n'existe pas de champ distinct pour le dossier de sortie.

**1. Project** (projet) — créer ou ouvrir un projet. Un projet fixe la finalité et le lieu de travail (pays, État, comté, ville) ainsi que ses dossiers d'entrée (`<os>_in/`) et de sortie. Les projets sont stockés sous une racine `regProj` partagée (par défaut `~/regProj`, modifiable via la variable d'environnement `REGPROJ_ROOT`) et sont communs aux applications sœurs earthUI et mgcvUI, de sorte que les modèles des trois routines coexistent. Cliquer sur **Close Project** (fermer le projet) pour en changer.

**2. Import Data** (importer les données) — sélectionner un fichier CSV ou Excel dans le dossier d'entrée du projet actif. Ajouter les fichiers au dossier via le Finder ou l'Explorateur, puis cliquer sur **Refresh** pour les lister. Pour les fichiers Excel comportant plusieurs feuilles, un sélecteur de feuille apparaît. Les noms de colonnes sont automatiquement convertis en `snake_case`.

**3. Import from earthUI (facultatif)** — importer un fichier de résultat earthUI `.rds` afin d'utiliser les fonctions de base charnières d'earth avec la régularisation elastic net de glmnet. Le bouton de navigation reprend le style du sélecteur de fichiers de la section 2. Cette étape est facultative — il suffit de l'ignorer en l'absence de résultat earthUI à importer.

**4. Variable Configuration** (configuration des variables) — sélecteur de variable cible, tableau des prédicteurs avec cases à cocher Include, Factor et Force, ainsi que des listes déroulantes

Sign. La colonne Special (pour désigner les dates de contrat, les coordonnées, etc.) n'apparaît qu'en modes Appraisal et Market. Voir le chapitre 6 pour le détail complet.

**5. glmnet Call Parameters** (paramètres d'appel de glmnet) — toute la configuration du modèle : alpha, lambda, famille, standardisation, contraintes de signe, relaxed lasso, graine aléatoire, matrice d'interactions et paramètres avancés (lambda.min.ratio, nlambdas, métrique de perte en validation croisée, seuil de convergence, nombre maximal d'itérations, ordonnée à l'origine). Voir le chapitre 7 pour la référence complète des paramètres.

**6. Fit Glmnet Model** (ajuster le modèle glmnet) — le bouton qui exécute le modèle. Les résultats apparaissent dans les onglets de droite.

**7. Download Output** (télécharger la sortie) — exporte les prédictions, les résidus, les scores CQA et les contributions par variable dans un fichier Excel placé dans le dossier de sortie du projet. Disponible dans tous les modes Purpose une fois le modèle ajusté.

**8. Calculate RCA Adjustments & Download** (calculer et télécharger les ajustements RCA) — calcule les ajustements RCA par variable pour chaque comparable et interpole le résidu du bien à évaluer via le CQA. N'apparaît qu'en modes Appraisal et Market, après l'ajustement. Voir le chapitre 11.

**9. Generate Sales Grid & Download** (générer et télécharger la grille de ventes) — génère la Sales Comparison Grid intermédiaire, en classant les comparables par pourcentage d'ajustement brut. N'apparaît qu'en mode Appraisal, après le calcul des ajustements RCA. Voir le chapitre 12.

**10. Generate Quarto Report** (générer le rapport Quarto) — écrit un ensemble de rapport Quarto (.qmd) autonome — avec tous les graphiques et tableaux — dans le dossier de sortie du projet. Disponible une fois le modèle ajusté.

**11. Convert Quarto Report** (convertir le rapport Quarto) — rend l'ensemble .qmd généré au format HTML, Word ou PDF. Le PDF requiert une installation LaTeX ; si aucune n'est détectée, l'option PDF est masquée. Voir le chapitre 13.

En modes Market et General, les étapes propres à l'évaluation sont masquées et les deux sections de rapport sont renumérotées en conséquence.

### 3.3 Projets

glmnetUI organise le travail en **projets** placés sous un dossier racine **regProj** partagé (la même arborescence que celle qu'utilisent earthUI et mgcvUI). Chaque projet possède un dossier d'entrée (<os>\_in) et des dossiers de sortie propres à chaque moteur (par exemple <os>\_out\_glmnet).

Pour créer un projet, cliquer sur + **New...** dans la section 1 et fournir :

- **Purpose** — General, Appraisal ou Market Area.
- **Location** (lieu) — la hiérarchie politique et administrative du projet. Les niveaux à renseigner dépendent du pays choisi, de sorte que les conventions de nommage étrangères sont prises en charge automatiquement — un projet américain utilise État/comté/ville, un projet français région/département/commune, un projet japonais préfecture/ville-ou-arrondissement/district, et ainsi de suite.
- **Project Name** (nom du projet) — une étiquette courte, limitée à 8 caractères (lettres, chiffres, \_ et -).

L'application n'utilise pas le nom saisi tel quel : elle construit un nom de dossier unique et horodaté en préfixant les codes de localisation et la date-heure de création, sous la forme <pays>\_<codes de niveaux administratifs>\_<AAAAAMJJ-HHMMSS>\_<votre nom> — par exemple, un projet américain nommé burl devient us\_ca\_081\_burlin\_20260604-131500\_burl. La limite de 8 caractères ne s'applique qu'à la partie saisie. Seuls les projets nouvellement créés reçoivent le préfixe horodaté ; les projets déjà présents sur le disque conservent leur nom d'origine et restent pleinement accessibles.

### 3.4 Importer depuis earthUI (facultatif)

La section 3 de la barre latérale — **Import from earthUI** — permet d'importer un fichier de résultat earthUI `.rds`. glmnetUI peut ainsi utiliser les fonctions de base charnières d'earth (termes linéaires par morceaux) avec la régularisation elastic net de glmnet, combinant la détection automatique des non-linéarités propre à earth et la sélection de variables et les contraintes de coefficients propres à glmnet.

Lorsqu'un résultat earthUI est importé, ce sont les prédicteurs du modèle earth — et non les cases Include — qui pilotent l'ajustement glmnet, car glmnet s'ajuste sur l'expansion de base d'earth. Le tableau Variable Configuration l'indique explicitement : **Include** et **Type** sont verrouillés, les lignes correspondant aux prédicteurs non retenus par le modèle earth sont désactivées, et une note précise que l'ensemble des prédicteurs provient du modèle earth. **Force** et **Sign** restent modifiables pour les prédicteurs d'earth, puisqu'ils s'appliquent encore à la base earth au moment de l'ajustement.

Le bouton de navigation reprend le style du sélecteur de fichiers de la section 2. Cette étape est entièrement facultative — en l'absence de fichier de résultat earthUI, il suffit d'ignorer la section 3 et de passer à la section 4.

### 3.5 Onglets du panneau principal

Après l'ajustement, le panneau principal propose les onglets suivants :

| Onglet              | Contenu                                                                                                                                                                                                                                                    |
|---------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data Preview        | Aperçu des données importées. En mode évaluation, séparé en deux tableaux : Subject Property et Comparable Sales.                                                                                                                                          |
| Equation            | L'équation du modèle ajusté, affichée en LaTeX/MathJax, montrant chaque coefficient non nul et les termes d'interaction.                                                                                                                                   |
| Correlation         | Carte de chaleur des corrélations entre prédicteurs numériques (disponible avant l'ajustement).                                                                                                                                                            |
| Summary             | Indicateurs clés ( $R^2$ , $R^2$ ajusté, $GR^2$ , $R^2$ de validation croisée, RMSE, MAE) et indicateurs d'évaluation (COD, PRD, ratio médian).                                                                                                            |
| Coefficients        | Tableau des coefficients non nuls, avec avertissements en cas de violation de signe.                                                                                                                                                                       |
| Variable Importance | Diagramme en barres et tableau classé de l'importance des prédicteurs ( $ \beta  \times sd(x)$ ). Survol interactif plotly lorsque disponible.                                                                                                             |
| Contributions       | Graphiques d'effets partiels par variable, avec étiquettes de pente. Nuage de points et droite d'ajustement pour les prédicteurs numériques, boîtes à moustaches pour les facteurs, surface 3D, nuage de points ou carte de chaleur pour les interactions. |
| ANOVA               | Tableau de décomposition de la variance : SS par variable, % du SS du modèle, coefficient.                                                                                                                                                                 |
| Diagnostics         | Quatre graphiques : chemin de régularisation, erreur de validation croisée, valeurs observées vs prédites, résidus vs valeurs ajustées.                                                                                                                    |
| Glmnet Output       | Sortie brute du modèle : graine aléatoire utilisée, impression du modèle, lambda sélectionné, lambda.min/1se, gamma (si relaxed) et vecteur complet des coefficients.                                                                                      |
| Report              | Configuration du rapport et champs d'exportation.                                                                                                                                                                                                          |
| RCA Adjustments     | Résultats de l'analyse RCA (mode évaluation, après calcul).                                                                                                                                                                                                |

### 3.6 Persistance des paramètres

glmnetUI enregistre automatiquement la configuration dans le stockage local du navigateur, indexée par le nom du fichier d'entrée. Au rechargement du même fichier, tous les paramètres sont restaurés : sélection de la cible, cases des prédicteurs, types de données, signes attendus, paramètres de glmnet et matrice d'interactions. Le dernier mode Purpose utilisé est lui aussi conservé globalement et restauré au redémarrage de l'application. Trois options sont proposées :

- **Use last settings for input file** — restaurer les paramètres enregistrés pour ce fichier précis
- **Use default settings** — appliquer les valeurs par défaut globales enregistrées
- **glmnet defaults** — revenir aux valeurs d'usine ( $\alpha = 1$ ,  $\lambda_{1se}$  par validation croisée, 10 blocs, gaussian, standardisation activée)

Un bouton **Save current as default** (enregistrer comme valeur par défaut) enregistre tous les paramètres courants comme valeur par défaut globale.

### 3.7 Mode sombre

Cliquer sur l'icône lune/soleil dans le coin supérieur droit pour basculer entre le thème clair et le thème sombre. La préférence de thème est enregistrée dans le stockage local et persiste d'une session à l'autre. Tous les éléments de l'interface (tableaux, graphiques, cartes, boutons) s'adaptent au thème sélectionné via des variables CSS.

### 3.8 Locale et paramètres régionaux

glmnetUI prend en charge les conventions internationales de format numérique et CSV grâce à un système de locales fondé sur le pays. Le menu déroulant **Settings** de la barre de titre propose des sélecteurs **Country** (pays) et **Paper** (papier) pour 31 pays pris en charge. Chaque préreglage configure :

- **Séparateur CSV** — virgule (,) pour les États-Unis, le Royaume-Uni et le Japon, ou point-virgule (;) pour la majeure partie de l'Europe, où la virgule sert de séparateur décimal.
- **Séparateur décimal** — point (.) ou virgule (,).
- **Séparateur de milliers** — virgule (États-Unis/Royaume-Uni/Japon), point (Allemagne/Italie/Espagne), espace (Finlande/France/Pologne/pays baltes/Ukraine/Russie) ou apostrophe (Suisse).
- **Format de papier** — Letter (États-Unis/Canada/Mexique) ou A4 (partout ailleurs).

#### 3.8.1 Enregistrer les valeurs par défaut

Cliquer sur **Save as my default** pour enregistrer globalement vos préférences de locale. Ces valeurs par défaut s'appliquent à toutes les sessions futures, quel que soit le fichier de données chargé.

***Astuce :** le formatage des nombres sur les axes des graphiques et sur les étiquettes de pente s'adapte automatiquement à votre locale. La locale allemande utilise des points pour les milliers (200.000), la locale finlandaise des espaces (200 000), la locale suisse des apostrophes (200'000). Aucun symbole monétaire n'est affiché — glmnetUI est indépendant de la devise.*

## 4 Mode évaluation

Lorsque **For Appraisal** est sélectionné comme Purpose, glmnetUI se configure pour l'évaluation d'un bien individuel. Toutes les fonctionnalités décrites au chapitre 3 restent disponibles ; le présent chapitre ne traite que des ajouts propres à l'évaluation.

### 4.1 Traitement de la ligne du bien à évaluer

En mode évaluation, **la ligne 1 du jeu de données est le bien à évaluer** et toutes les lignes suivantes sont des ventes comparables. Le fichier d'entrée doit être organisé en conséquence (voir le chapitre 2). Le prix de vente du bien à évaluer peut être laissé vide ou renseigné avec n'importe quelle valeur — glmnetUI le traite automatiquement comme NA lors de l'ajustement.

Après l'import, l'onglet Data Preview se scinde en deux sections : **Subject Property** (ligne 1) et **Comparable Sales** (lignes 2 et suivantes). La ligne 1 est toujours exclue de l'ajustement du modèle. Après l'ajustement, le modèle produit néanmoins des prédictions pour cette ligne.

### 4.2 Date d'effet et ancienneté de la vente

En modes évaluation et marché, un champ **Effective Date** (date d'effet) apparaît dans la section Variable Configuration (initialisé à la date du jour). Si une colonne est désignée comme **contract\_date** dans la liste déroulante Special, glmnetUI calcule une colonne **sale\_age** — le nombre entier de jours entre la date de contrat de chaque vente et la date d'effet. Cette colonne est ajoutée comme prédicteur.

Lorsque la date d'effet change, **sale\_age** est automatiquement recalculée.

### 4.3 Désignations spéciales de colonnes

En modes évaluation et marché, une liste déroulante **Special** apparaît pour chaque prédicteur du tableau Variable Configuration. Voir le chapitre 6 pour la référence complète des types spéciaux et de leurs effets.

### 4.4 Aperçu des ajustements RCA

Le bouton **Calculate RCA Adjustments & Download** (section 8 de la barre latérale, visible en modes Appraisal et Market après l'ajustement) calcule des ajustements déduits du marché pour chaque comparable, relativement au bien à évaluer. Le flux de travail RCA complet est décrit au chapitre 11.

***Conseil :** le score CQA (état, qualité, attrait) que vous attribuez au bien à évaluer détermine la part de la distribution des résidus attribuée à celui-ci. Un score de 5.00 place le bien à évaluer à la médiane des biens comparables.*

## 5 Mode analyse de zone de marché

Lorsque **Market Area Analysis** est sélectionné comme Purpose, glmnetUI offre les mêmes fonctionnalités propres à l'immobilier que le mode évaluation (colonnes spéciales, ancienneté de la vente, arrondi des coordonnées), mais il est orienté vers l'analyse d'un groupe de biens plutôt que vers l'évaluation d'un bien unique.

### 5.1 Différences avec le mode évaluation

- **Le traitement de la ligne du bien à évaluer est souple** — en mode marché, la ligne 1 n'est pas automatiquement considérée comme un bien à évaluer.
- **Pas de Sales Comparison Grid** — l'étape « Calculate RCA Adjustments & Download » (section 8) est disponible en mode marché, mais l'étape Sales Grid (section 9) est réservée au mode évaluation.

### 5.2 Quand utiliser le mode marché

Le mode Market Area Analysis convient lorsqu'il s'agit de :

- construire un modèle de régression pour un quartier ou une zone de marché afin de comprendre les déterminants de la valeur ;
- analyser la façon dont des variables telles que la surface, l'âge, la superficie du terrain et la localisation influent sur les prix de vente au sein d'un groupe de biens ;
- préparer les éléments justificatifs d'une analyse des conditions de marché ou d'une délimitation de quartier ;
- travailler sur un jeu de données pour lequel les fonctionnalités de colonnes spéciales (ancienneté de la vente, arrondi des coordonnées) sont souhaitées sans que le flux de travail d'ajustements RCA soit nécessaire.

***Conseil :** le mode marché est également utile pour la régression immobilière générale lorsque vous souhaitez bénéficier des fonctionnalités de colonnes spéciales sans avoir besoin du flux de travail d'ajustements RCA.*

## 6 Sélection des variables

La section 4 de la barre latérale — **Variable Configuration** — est l'endroit où l'on choisit les colonnes qui participent au modèle et la façon dont elles sont traitées.

### 6.1 Variable cible

La liste déroulante **Target (response) variable** (variable cible ou réponse), en haut de la section 4, énumère toutes les colonnes du jeu de données. Sélectionner une colonne comme variable réponse (par exemple le prix de vente). La colonne cible est automatiquement exclue de la liste des prédicteurs.

### 6.2 Le tableau des prédicteurs

Sous le sélecteur de cible, un tableau énumère toutes les colonnes restantes avec les champs suivants :

| Colonne         | Description                                                                                                                |
|-----------------|----------------------------------------------------------------------------------------------------------------------------|
| <b>Variable</b> | Nom de la colonne                                                                                                          |
| <b>Type</b>     | Liste déroulante de type de données : <b>numeric</b> , <b>integer</b> , <b>character</b> , <b>Date</b> , <b>POSIXct</b>    |
| <b>Include</b>  | Case à cocher — inclure cette colonne comme prédicteur dans le modèle                                                      |
| <b>Factor</b>   | Case à cocher — traiter la variable comme un facteur catégoriel (crée des colonnes indicatrices dans la matrice du modèle) |
| <b>Force</b>    | Case à cocher — imposer la variable dans le modèle (facteur de pénalité = 0, jamais éliminée par le lasso)                 |
| <b>Special</b>  | Liste déroulante (évaluation et marché uniquement) — voir la référence des types de colonnes spéciales ci-dessous          |
| <b>Sign</b>     | Signe attendu du coefficient : positif, négatif ou indifférent                                                             |
| <b>NAs</b>      | Nombre de valeurs manquantes                                                                                               |

### 6.3 Détection des types de données et substitutions

glmnetUI détecte automatiquement les types de données à l'import. Les colonnes numériques, entières et de dates sont reconnues. Les colonnes de type caractère qui ressemblent à des dates (motifs de format de date courants) sont classées comme **Date**.

Toute détection peut être remplacée en modifiant la liste déroulante **Type**. Le changement de type modifie l'encodage de la colonne dans la matrice du modèle.

Les colonnes de texte, TRUE/FALSE et les facteurs R sont automatiquement traités comme catégoriels. Une colonne numérique ne devient un facteur que si vous cochez sa case **Factor** — glmnetUI ne déduit jamais le caractère catégoriel de l'étendue des valeurs d'une variable numérique, de sorte que des prédicteurs numériques discrets comme un nombre de salles de bains restent continus sauf indication contraire. Le réglage Factor est conservé via localStorage.

### 6.4 Signes attendus et application

Pour chaque prédicteur, la liste déroulante **Sign** précise si son coefficient est attendu positif, négatif ou dans l'une ou l'autre direction. Après l'ajustement :

- les coefficients dont le signe est inattendu sont **signalés dans le tableau Coefficients** ;

- des avertissements de signe apparaissent dans l'onglet Summary.

Lorsque la case **Enforce Sign Constraints** (imposer les contraintes de signe) est activée dans les paramètres de `glmnet` (chapitre 7), le modèle est contraint de produire des coefficients conformes aux signes attendus. Cela est mis en œuvre au moyen de `upper.limits` et `lower.limits` dans `glmnet`.

## 6.5 Référence des types de colonnes spéciales

En modes évaluation et marché, la liste déroulante **Special** propose les options suivantes :  
**Pondération :**

- **weight** — colonne de pondération des observations (disponible dans tous les modes ; une seule autorisée ; les lignes de poids = 0 sont exclues de l'ajustement)

**Types de date et d'heure :**

- **contract\_date** — déclenche le calcul automatique de **sale\_age** à partir de la date d'effet
- **listing\_date** — utilisée comme solution de repli pour calculer les jours de mise en marché
- **dom** — identifie la colonne des jours de mise en marché
- **sale\_age** — désigne une colonne d'ancienneté de vente existante, de sorte que n'importe quelle colonne puisse tenir ce rôle au lieu que `glmnetUI` la calcule à partir de **contract\_date**

**Types monétaires :**

- **concessions** — identifie les concessions de vente (crédits du vendeur, incitations à l'acheteur, etc.)

**Types de transaction :**

- **sale\_type** — identifie la colonne de type de vente ou de transaction

**Types de taille et de localisation :**

- **latitude** — valeurs automatiquement arrondies à 3 décimales
- **longitude** — même traitement d'arrondi que la latitude
- **area** — identifiant de zone de marché ou de quartier (par exemple `area_id`, `neighborhood`, `market_area`)
- **living\_area** — active les calculs de résidus au pied carré (**residual\_sf** et **cqa\_sf**)
- **lot\_size** — colonne de superficie du terrain
- **site\_dimensions** — regroupée avec la superficie du terrain (par exemple « 75x120 »)

**Types d'âge :**

- **actual\_age** — colonne d'âge du bien
- **effective\_age** — âge effectif du bien

**Types d'affichage :**

- **display\_only** — la colonne est incluse dans les exportations Excel mais entièrement exclue de l'ajustement du modèle. À utiliser pour les champs d'adresse, les numéros MLS, les identifiants de parcelle ou d'autres données de référence. Plusieurs colonnes peuvent porter cette désignation.

***Conseil :** les colonnes désignées comme **display\_only** demeurent dans le jeu de données et figurent dans l'exportation Excel, mais sont entièrement exclues de la table des prédicteurs. À utiliser pour les colonnes d'identifiants, les adresses ou d'autres champs de référence.*

## 7 Choix des paramètres

La section 5 de la barre latérale — **glmnet Call Parameters** — donne accès à toutes les options de configuration du modèle elastic net. Chaque paramètre est accompagné d'une icône d'aide bleue (?) dont l'infobulle fournit une explication.

### 7.1 Alpha (paramètre de mélange)

Le paramètre alpha détermine le type de régularisation :

| Valeur d'alpha | Type        | Comportement                                                                  |
|----------------|-------------|-------------------------------------------------------------------------------|
| 0              | Ridge       | Rétrécit proportionnellement tous les coefficients ; n'en annule jamais aucun |
| 1              | Lasso       | Peut ramener des coefficients exactement à zéro (sélection de variables)      |
| Entre 0 et 1   | Elastic net | Mélange de ridge et de lasso                                                  |

Deux méthodes de choix d'alpha sont proposées :

- **Fixed** (fixe) — définir une valeur unique d'alpha à l'aide du curseur (par défaut : 1,0)
- **Grid Search** (recherche sur grille) — tester plusieurs valeurs d'alpha et retenir celle dont l'erreur de validation croisée est la plus faible. L'étendue et le nombre de valeurs sont configurables.

### 7.2 Choix de lambda

Lambda contrôle l'intensité de la régularisation :

| Méthode                   | Description                                                                          |
|---------------------------|--------------------------------------------------------------------------------------|
| <b>Validation croisée</b> | Recommandée. Teste de nombreuses valeurs de lambda par validation croisée à k blocs. |
| <b>Manuelle</b>           | Saisir une valeur précise de lambda si l'on a une raison de le faire.                |

En validation croisée, deux choix de lambda sont possibles :

- **lambda.1se** (par défaut) — le plus grand lambda situé à moins d'une erreur type du minimum. Produit des modèles plus simples et plus régularisés. Recommandé pour les travaux d'évaluation.
- **lambda.min** — le lambda qui minimise l'erreur de validation croisée. Produit le meilleur ajustement, mais un modèle potentiellement plus complexe.

Le paramètre **Number of CV Folds** (nombre de blocs de validation croisée) détermine la façon dont les données sont découpées pour la validation croisée (par défaut : 10 blocs).

### 7.3 Famille

Choisir la famille de distribution de la variable réponse :

| Famille         | Cas d'usage                                                          |
|-----------------|----------------------------------------------------------------------|
| <b>gaussian</b> | Réponses continues (par exemple le prix de vente). La plus courante. |
| <b>binomial</b> | Résultats binaires (par exemple vendu / non vendu).                  |
| <b>poisson</b>  | Données de comptage (par exemple le nombre de ventes).               |

## 7.4 Standardize

Lorsque la case est cochée (par défaut), tous les prédicteurs sont ramenés à une moyenne de 0 et un écart type de 1 avant l'ajustement. La pénalité traite ainsi tous les prédicteurs de la même façon, quelle que soit leur échelle d'origine. Les coefficients sont restitués à l'échelle initiale. Il est en général préférable de laisser cette option activée.

## 7.5 Contraintes de signe

Lorsque **Enforce Sign Constraints** est activé, les coefficients sont contraints de respecter les signes attendus définis dans le tableau des variables (chapitre 6) :

- un signe « positive » impose un coefficient  $\geq 0$  ;
- un signe « negative » impose un coefficient  $\leq 0$  ;
- les variables réglées sur « either » ne sont pas contraintes.

Cela est mis en œuvre au moyen des paramètres `upper.limits` et `lower.limits` de `glmnet()`.

## 7.6 Relaxed lasso

Le lasso ordinaire utilise la même pénalité pour la sélection des variables ET pour l'estimation des coefficients, ce qui peut rétrécir excessivement vers zéro des coefficients importants. Le **relaxed lasso** (lasso relâché) sépare ces deux étapes :

1. d'abord, sélectionner les variables par lasso (la pénalité détermine quels coefficients survivent) ;
2. ensuite, réestimer les variables retenues avec une pénalité réduite ou nulle, afin d'obtenir des estimations moins biaisées.

Le curseur **Gamma** règle le degré de relâchement :

| Gamma         | Effet                                                         |
|---------------|---------------------------------------------------------------|
| 0             | Entièrement relâché (réestimation MCO, aucun rétrécissement)  |
| 1             | Aucun relâchement (identique à <code>glmnet</code> ordinaire) |
| Intermédiaire | Mélange des ajustements relâché et pénalisé                   |

En validation croisée, gamma est choisi automatiquement pour une performance optimale.

## 7.7 Graine aléatoire

Un champ texte **Random seed** (graine aléatoire), prérempli avec un entier aléatoire, rend reproductible l'affectation des blocs de validation croisée — `set.seed()` est appelée avant l'ajustement, et la graine utilisée est indiquée dans l'onglet `Glmnet Output` ainsi que dans le message d'état de l'ajustement. Après chaque ajustement, une nouvelle graine est générée automatiquement (à partir de l'heure système, et non du générateur de nombres aléatoires de R). Les 5 dernières graines utilisées pour le fichier courant sont affichées sous forme de liens cliquables permettant de les rappeler aisément ; l'historique des graines est réinitialisé au chargement d'un nouveau fichier.

## 7.8 Matrice d'interactions

Une matrice triangulaire supérieure interactive permet de choisir quelles paires de variables sont autorisées à interagir. Chaque cellule contient une case à cocher — cochée, elle signifie que le terme d'interaction  $x_1:x_2$  est inclus dans la matrice du modèle.

- **Allow All** (tout autoriser) — cocher toutes les paires d'interaction
- **Clear All** (tout effacer) — décocher toutes les paires (par défaut, tout est décoché pour un nouveau fichier)
- **Cliquer sur le nom d'une variable** — basculer toutes les interactions de cette variable

Les sélections d'interactions sont enregistrées dans le localStorage du navigateur, par fichier d'entrée, et restaurées au rechargement du même fichier.

Dès qu'une interaction est activée, le rapport exporté comprend une matrice **Enabled Interactions** (interactions activées) — les prédicteurs sur les deux axes, une coche là où le terme produit  $x_1:x_2$  a été inclus — rendue en format paysage et calquée sur la matrice Allowed Interactions d'earthUI. Il convient de noter que le lasso peut néanmoins avoir ramené à zéro le coefficient d'un terme activé.

***Conseil :** avec de nombreux prédicteurs, le lasso annulera automatiquement les termes d'interaction inutiles. Partir de « Allow All » et laisser le modèle sélectionner constitue une approche raisonnable.*

### Avertissement

L'activation des interactions dans glmnet produit des ajustements qui peuvent être difficiles à interpréter et à expliquer devant un tribunal ou lors d'un audit. Les interactions doivent être employées avec prudence et il convient de vérifier que les ajustements obtenus sont raisonnables.

## 7.9 Exclusion de l'effet principal (interaction seule)

Un **clic droit** sur le nom d'une variable dans la matrice d'interactions l'empêche d'entrer dans le modèle comme effet principal. Un indicateur « 1 » en gras apparaît après le nom de la variable. Lorsqu'une variable est ainsi exclue :

- elle est **retirée** des termes principaux de la formule ;
- elle **peut encore** figurer dans des termes d'interaction ( $x_1:x_2$ ) ;
- un nouveau clic droit lève l'exclusion ;
- l'état est conservé via le localStorage, par fichier d'entrée.

Cas d'usage typique : lorsqu'on modélise des ajustements temporels qui doivent varier avec la taille du bien, on exclut `sale_age` de l'effet principal et on ne l'autorise qu'en interaction avec `living_area`. On obtient ainsi un terme d'interaction `sale_age:living_area` au lieu d'un coefficient `sale_age` uniforme — le taux d'appréciation dans le temps varie alors selon la taille de la maison.

## 7.10 Paramètres avancés

Une section repliable « Advanced » (avancé), au bas de glmnet Call Parameters, expose des réglages supplémentaires. Ceux-ci ont des valeurs par défaut judicieuses, mais restent visibles afin que tous les réglages du modèle puissent être documentés pour un tribunal ou un audit.

- **Lambda Min Ratio** — rapport entre le plus petit et le plus grand lambda du chemin de régularisation. Par défaut : 0,00001. Pour les petits jeux de données (< 50 observations), une valeur plus élevée (par exemple 0,01) peut améliorer les performances.

- **Number of Lambda Values** — nombre de valeurs de lambda du chemin. Par défaut : 100.
- **CV Loss Metric** — fonction de perte pour la validation croisée : MSE (par défaut pour gaussian), MAE (robuste aux valeurs aberrantes) ou déviance.
- **Convergence Threshold** — seuil de convergence de la descente par coordonnées. Par défaut : 1e-07.
- **Max Iterations** — nombre maximal d'itérations de la descente par coordonnées. Par défaut : 100 000. À augmenter si des avertissements de convergence apparaissent.
- **Fit Intercept** — inclure un terme constant (ordonnée à l'origine). Par défaut : activé. L'ordonnée à l'origine constitue la « valeur de base » dans la RCA en évaluation.

### 7.11 Valeurs par défaut des paramètres

Trois options déterminent l'initialisation des paramètres au chargement d'un fichier :

- **Use last settings for input file** — restaurer exactement les paramètres utilisés la dernière fois avec ce fichier
- **Use default settings** — appliquer les valeurs par défaut globales enregistrées
- **glmnet defaults** — revenir aux réglages d'usine ( $\alpha = 1$ ,  $\lambda_{1se}$  par validation croisée, 10 blocs, gaussian, standardisation activée)

Cliquer sur **Save current as default** pour enregistrer tous les paramètres courants comme valeur par défaut globale pour les fichiers futurs.

## 8 Ajustement du modèle

### 8.1 Le bouton d'ajustement

La section 6 de la barre latérale contient le bouton **Fit Glmnet Model**. Un clic exécute le modèle selon la configuration courante.

### 8.2 Ce qui se passe pendant l'ajustement

Lorsqu'on clique sur Fit, glmnetUI :

1. **Prépare les données** — supprime les lignes comportant des NA, applique les pondérations (si une colonne est désignée), exclut la ligne du bien à évaluer (en mode évaluation) et encode les variables facteurs sous forme de colonnes indicatrices.
2. **Construit la matrice du modèle** — crée la matrice de plan d'expérience, incluant les éventuels termes d'interaction autorisés par la matrice d'interactions.
3. **Applique les contraintes** — si l'application des signes est activée, définit `upper.limits` et `lower.limits` sur les coefficients.
4. **Exécute la validation croisée** — si la validation croisée est retenue, appelle `cv.glmnet()` pour trouver le lambda optimal. Si le relaxed lasso est activé, appelle `cv.glmnet(..., relax = TRUE)`.
5. **Ajuste le modèle final** — enregistre le modèle ajusté, les coefficients, les prédictions et les résidus.
6. **Met à jour tous les onglets de résultats** — Summary, Coefficients, Equation, Variable Importance, Contributions, ANOVA et Diagnostics sont renseignés.

Un message d'état, sous le bouton Fit, indique le nombre d'observations, les lignes exclues et le lambda retenu.

## 9 Onglets de résultats

### 9.1 Data Preview

Affiche les données importées sous forme de DataTable interactif. En modes évaluation et marché, l'aperçu est scindé en deux tableaux : **Subject Property** (ligne 1) et **Comparable Sales** (lignes 2 et suivantes).

Chaque cellule tient sur une seule ligne et est tronquée à la largeur de la colonne, de sorte que les champs de texte libre étendus (tels que les remarques sur le bien) n'étirent pas une ligne sur toute la hauteur de la page. **Un double-clic sur une cellule** ouvre une fenêtre contextuelle affichant son contenu complet, non tronqué.

### 9.2 Equation

Affiche l'équation du modèle ajusté, rendue en LaTeX via MathJax. Tous les coefficients non nuls sont présentés avec une mise en forme mathématique correcte :

- coefficients positifs précédés du signe +
- coefficients négatifs précédés du signe −
- termes d'interaction affichés sous la forme  $x_1 \times x_2$
- nombres formatés avec séparateurs de milliers et un nombre approprié de décimales

### 9.3 Correlation

Une carte de chaleur matricielle présentant les corrélations de Pearson entre tous les prédicteurs numériques et la variable réponse. Disponible immédiatement après l'import des données — aucun ajustement de modèle n'est requis.

- Dégradé de couleurs : bleu (−1) → blanc (0) → rouge (+1)
- Valeurs de corrélation imprimées dans chaque cellule
- Taille du texte et des axes adaptée au nombre de variables
- Utile pour repérer la multicollinéarité avant l'ajustement

### 9.4 Summary

Statistiques d'ajustement du modèle, affichées sous forme de cartes :

- **$R^2$**  — proportion de variance expliquée
- **$R^2$  ajusté** — ajusté selon le nombre de prédicteurs
- **$GR^2$**  —  $R^2$  généralisé, issu du rapport de déviance de glmnet au lambda retenu
- **$R^2$  de validation croisée** —  $R^2$  de validation croisée (lorsque la validation croisée est utilisée)
- **RMSE** — racine de l'erreur quadratique moyenne
- **MAE** — erreur absolue moyenne

En modes évaluation et marché, des cartes supplémentaires affichent :

- **Median Ratio** — médiane du rapport (prédit / observé)
- **COD** — coefficient de dispersion (plus il est faible, mieux c'est)
- **PRD** — différentiel lié au prix (Price-Related Differential)

Un **avertissement de surajustement** apparaît lorsque le  $R^2$  d'apprentissage dépasse de plus de 0,1 le  $R^2$  de validation croisée.

Sous les cartes, un tableau de coefficients avec avertissements de signe reprend, par commodité, le contenu de l'onglet Coefficients.

## 9.5 Coefficients

Un tableau de tous les coefficients non nuls indiquant :

- le nom de la variable
- la valeur du coefficient
- le signe attendu (issu de la configuration des variables)
- un indicateur d'avertissement si le sens du coefficient ne correspond pas aux attentes

## 9.6 Variable Importance

Amplitude standardisée du coefficient :  $|\beta| \times \text{sd}(x)$  pour chaque prédicteur, agrégée sur les colonnes indicatrices dans le cas des variables facteurs. Deux modes d'affichage :

- **plotly interactif** (lorsque le package `plotly` est installé) : diagramme en barres horizontales avec infobulles au survol indiquant l'importance exacte, le coefficient et le pourcentage relatif.
- **ggplot2 statique** (solution de repli) : diagramme en barres horizontales avec étiquettes d'axes formatées avec séparateurs de milliers.

Sous le graphique, un `DataTable` affiche Variable, Importance, Coefficient et % relatif.

## 9.7 Contributions

Graphiques d'effets partiels par variable, montrant la contribution de chaque prédicteur aux prédictions du modèle. La variable se choisit dans une liste déroulante.

**Pour les prédicteurs numériques** : un nuage de points des valeurs  $x$  de la variable en regard de sa contribution ( $\beta \times x$ ), avec un segment de droite rouge et une étiquette de pente. L'étiquette de pente indique l'effet marginal par unité (par exemple **+1,234.56/unit**), avec des unités adaptatives pour les variables de faible étendue comme la latitude et la longitude (par exemple **+0.12/0.001**). Le sous-titre indique la valeur de l'ordonnée à l'origine (valeur de base).

**Pour les variables facteurs** : une boîte à moustaches représentant la distribution des contributions par niveau du facteur.

**Pour les interactions à deux variables** : trois visualisations superposées — une surface 3D (`plotly` interactif lorsque disponible, `persp()` statique sinon) de la contribution conjointe, un nuage de points dont les points sont colorés selon la contribution, et une carte de chaleur de la contribution moyenne sur l'étendue des deux variables.

## 9.8 ANOVA

Tableau de décomposition de la variance indiquant, pour chaque prédicteur :

- **SS** — somme des carrés :  $\sum(\text{contribution}_j - \overline{\text{contribution}})^2$
- **% du SS du modèle** — pourcentage de la somme des carrés totale du modèle
- **Coefficient** — le coefficient dominant pour ce prédicteur

Le tableau comprend une ligne d'ordonnée à l'origine et une ligne TOTAL MODEL. Les termes d'interaction (contenant « : ») sont présentés comme des groupes distincts.

## 9.9 Diagnostics

Quatre graphiques de diagnostic à grands caractères (taille de base 16 pt), 15 graduations d'axe et étiquettes formatées avec séparateurs de milliers (sans notation scientifique) :

1. **Coefficient Path** (chemin de régularisation) — montre l'évolution de tous les coefficients à mesure que  $\log(\lambda)$  augmente. Les courbes sont colorées par variable. Aide à visualiser le chemin de régularisation et la sélection de variables.

2. **CV Error** (erreur de validation croisée) — erreur de validation croisée en fonction de  $\log(\lambda)$ . Les points représentent l'erreur moyenne de validation croisée ; les barres d'erreur indiquent  $\pm 1$  erreur type. Des lignes verticales pointillées marquent  $\lambda_{\min}$  (bleu) et  $\lambda_{1se}$  (vert).
3. **Actual vs Predicted** (observé vs prédit) — nuage de points des valeurs observées en regard des valeurs prédites. Les points devraient se regrouper autour de la droite de référence pointillée à 45 degrés. Une dispersion plus large signale une erreur de prédiction plus importante.
4. **Residuals vs Fitted** (résidus vs valeurs ajustées) — nuage de points des résidus en regard des valeurs ajustées. La dispersion devrait être aléatoire autour de la ligne zéro (pointillée). Toute structure suggère une mauvaise spécification du modèle (par exemple non-linéarité, hétéroscédasticité).

Chaque graphique dispose d'un bouton **Download PNG** permettant de l'enregistrer à 150 DPI.

### 9.10 Glmnet Output

Sortie brute du modèle indiquant : la graine aléatoire utilisée, l'impression du modèle, le  $\lambda$  retenu,  $\lambda_{\min}/\lambda_{1se}$ ,  $\gamma$  (si le relaxed lasso est activé) et le vecteur complet des coefficients.

### 9.11 Report

Permet de renseigner les champs du rapport (nom de l'évaluateur, adresse du bien, date du rapport, numéro de dossier) et d'exporter directement depuis l'onglet au format Word (.docx) ou PDF. Pour le flux de travail complet de rapport fondé sur Quarto (HTML, Word ou PDF), voir le chapitre 13.

## 10 Téléchargement des données

Après l'ajustement, le bouton **Download Output (Excel)** (section 7 de la barre latérale, disponible dans tous les modes Purpose) exporte un fichier Excel contenant les prédictions et les diagnostics dans le dossier de sortie du projet. L'intitulé de la section est « Download Estimated Sale Prices & Residuals » en modes évaluation et marché, et « Download Estimated Target Variable(s) & Residuals » en mode General.

### 10.1 Colonnes de sortie

| Colonne                               | Description                                                           |
|---------------------------------------|-----------------------------------------------------------------------|
| <code>est_&lt;target&gt;</code>       | Prédiction du modèle (par exemple <code>est_sale_price</code> )       |
| <code>residual</code>                 | Observé – prédit                                                      |
| <code>cqa</code>                      | Score Condition-Quality-Appeal (état, qualité, attrait), échelle 0–10 |
| <code>residual_sf</code>              | Résidu / surface habitable (si <code>living_area</code> est désignée) |
| <code>cqa_sf</code>                   | CQA calculé à partir du classement selon <code>residual_sf</code>     |
| <code>&lt;var&gt;_contribution</code> | Contribution de chaque variable à la prédiction                       |
| <code>basis</code>                    | Valeur de l'ordonnée à l'origine (identique pour toutes les lignes)   |

Une coche blanche apparaît sur le bouton de téléchargement une fois l'opération réussie.

### 10.2 Scores CQA

Le CQA classe le résidu de chaque ligne par rapport à tous les autres, sur une échelle de 0 à 10 :

- **CQA élevé (~9–10)** : vendu bien plus cher que prédit — probablement un état, une qualité ou un attrait supérieurs
- **CQA faible (~0–1)** : vendu bien moins cher que prédit — probablement un état inférieur ou une vente en difficulté
- **CQA ~5** : proche de la médiane

En modes évaluation et marché, les lignes sont triées par `residual_sf` décroissant lorsqu'une colonne `living_area` est désignée.

## 11 Calculs RCA et téléchargement

Le flux de travail RCA (Residual Constraint Approach, approche par contraintes résiduelles) est disponible en modes **Appraisal** et **Market Area Analysis**, après l'ajustement du modèle. La Sales Comparison Grid qui en découle (chapitre 12) est réservée au mode évaluation.

### 11.1 Ouverture de la boîte de dialogue RCA

Cliquer sur le bouton **Calculate RCA Adjustments & Download** dans la section 8 de la barre latérale. Choisir :

- le type de score, **CQA** ou **CQA per SF** (CQA au pied carré). Si « CQA per SF » est retenu, le résidu du bien à évaluer est mis à l'échelle de la surface habitable.
- Saisir le score CQA du bien à évaluer (0,00–9,99, par défaut 5,00)

### 11.2 Interpolation du score CQA

1. Les scores CQA et les résidus des comparables sont triés
2. Une interpolation linéaire fait correspondre la valeur CQA choisie à un résidu
3. Valeur du bien à évaluer = prédiction du modèle + résidu interpolé

### 11.3 Colonnes de sortie

| Colonne             | Description                                                 |
|---------------------|-------------------------------------------------------------|
| subject_value       | Prédiction du modèle + résidu interpolé                     |
| <var>_adjustment    | Contribution du bien à évaluer – contribution du comparable |
| residual_adjustment | Résidu du bien à évaluer – résidu du comparable             |
| net_adjustments     | Somme de tous les ajustements                               |
| gross_adjustments   | Somme des valeurs absolues des ajustements                  |
| adjusted_sale_price | Prix de vente du comparable + ajustements nets              |

Une coche blanche apparaît sur le bouton une fois le calcul réussi.

## 12 Sales Comparison Grid

La Sales Comparison Grid (grille de comparaison des ventes) n'est disponible qu'en mode **Appraisal**, après le calcul des ajustements RCA (étape 8). Elle génère un classeur Excel mis en forme, prêt à être intégré aux rapports d'évaluation.

### 12.1 Boîte de dialogue de sélection des comparables

Un clic sur le bouton **Generate Sales Grid & Download** de la section 9 de la barre latérale ouvre une fenêtre modale énumérant toutes les ventes comparables issues de la sortie RCA. Les comparables sont répartis en deux groupes :

- **Recommended comps** (comparables recommandés, cochés d'office) : ajustement brut < 25 % du prix de vente, triés par ancienneté de la vente (les plus récents d'abord).
- **Other comps** (autres comparables, décochés par défaut) : les comparables restants, triés par ancienneté de la vente.

Sélectionner jusqu'à 30 comparables, puis cliquer sur **Generate Sales Grid** pour créer le classeur.

### 12.2 Structure du classeur

Le classeur produit comporte jusqu'à 10 feuilles, chacune contenant 3 comparables côte à côte :

- **Lignes d'en-tête** : adresses du bien à évaluer et des comparables, prix de vente, dates de vente et jours de mise en marché
- **Lignes groupées** : localisation (latitude, longitude, zone), taille du terrain (superficie, dimensions du site) et âge (âge réel, âge effectif), avec en-têtes de groupe et sous-éléments
- **Lignes de variables du modèle** : une ligne par coefficient non nul du modèle, indiquant la contribution du bien à évaluer, les contributions des comparables et les pourcentages d'ajustement
- **Lignes de caractéristiques résiduelles** : cellules vides et déverrouillées, destinées à la saisie par l'évaluateur des caractéristiques non captées par le modèle (par exemple état, qualité, vue)
- **Lignes de synthèse** : % d'ajustement net, % d'ajustement brut et prix de vente ajusté, calculés par des formules Excel actives

### 12.3 Protection des feuilles et cellules résiduelles

Chaque feuille est protégée par un mot de passe afin d'éviter les modifications accidentelles, mais les cellules de valeur des caractéristiques résiduelles sont explicitement déverrouillées. Les évaluateurs peuvent ainsi saisir des valeurs pour des caractéristiques absentes du modèle tout en préservant les calculs d'ajustement pilotés par les formules.

## 13 Téléchargement des rapports

### 13.1 Le flux de travail Quarto en deux étapes

L'exportation d'un rapport depuis la barre latérale suit un flux de travail Quarto en deux étapes :

1. **Generate Quarto Report** (section 10 de la barre latérale) — écrit un ensemble Quarto autonome (la source `.qmd` ainsi que les graphiques, `report_data.rds` et `reference.docx`) dans le dossier de sortie glmnet du projet.
2. **Convert Quarto Report** (section 11 de la barre latérale) — rend un fichier `.qmd` dans les formats de sortie choisis. Le chemin proposé par défaut est celui du dernier ensemble généré, mais il est possible de naviguer vers n'importe quel `.qmd`.

### 13.2 Formats de rapport

Trois formats sont proposés à l'étape de conversion :

- **HTML** — document HTML autonome. Fonctionne sur toutes les plateformes sans logiciel supplémentaire au-delà de pandoc. Format par défaut recommandé.
- **Word (.docx)** — document mis en forme. Adapté à l'édition et à la diffusion. Fonctionne sur toutes les plateformes.
- **PDF** — généré via Quarto avec LaTeX. Le format de papier suit le réglage de locale (Letter ou A4). **Requiert une installation LaTeX** (voir l'annexe consacrée à la configuration requise). Si aucune installation LaTeX n'est détectée, l'option PDF est automatiquement masquée dans les cases de format.

Les rapports sont générés au moyen de Quarto lorsqu'il est disponible, avec repli sur rmarkdown. L'onglet Report propose en outre des boutons d'export direct vers Word et vers PDF.

### 13.3 Contenu des rapports

Les rapports reprennent le contenu de tous les onglets :

- Description du jeu de données (nom du fichier, observations, réponse, prédicteurs, finalité)
- Spécification du modèle (alpha, lambda, famille, standardisation, relaxed, coefficients non nuls)
- Synthèse des résultats ( $R^2$ ,  $R^2$  ajusté,  $R^2$  généralisé,  $R^2$  de validation croisée, RMSE, MAE)
- Équation du modèle (rendue en LaTeX)
- Tableau des coefficients
- Importance des variables (diagramme en barres et tableau)
- Graphiques de contribution par variable, correspondant à l'onglet Contributions :
  - **Prédicteurs numériques** : nuage de points avec droite d'ajustement linéaire
  - **Variables facteurs** : boîte à moustaches par niveau
  - **Interactions** : nuage de points coloré selon la contribution, carte de chaleur de la contribution moyenne et instantané statique de la surface 3D (persp)
- Matrice Enabled Interactions (dès qu'une interaction est activée) — les prédicteurs sur les deux axes, avec une coche là où le terme produit  $x_1:x_2$  a été inclus, rendue en format paysage
- Carte de chaleur de la matrice de corrélation
- Tableau de décomposition ANOVA
- Graphiques de diagnostic :

- Chemin de régularisation
- Erreur de validation croisée
- Observé vs prédit
- Résidus vs valeurs ajustées
- Diagramme quantile-quantile normal
- Sortie du modèle

Toutes les étiquettes d'axes et les légendes de couleurs emploient des nombres formatés avec séparateurs de milliers (sans notation scientifique).

Une coche blanche apparaît sur le bouton de rapport une fois la génération réussie, et un chronomètre s'affiche pendant le rendu du rapport.

## 14 Comparaison avec earthUI

L'application earth et son Post Processing Module glmnet forment un ensemble d'outils unique pour la modélisation par régression. Ils partagent le même format de données, les mêmes types de colonnes spéciales, le même flux de travail RCA et les mêmes jeux de données de démonstration, mais reposent sur des moteurs de modélisation différents.

### 14.1 Principales différences

| Caractéristique              | glmnetUI                                   | earthUI                                       |
|------------------------------|--------------------------------------------|-----------------------------------------------|
| Méthode                      | Elastic net (glmnet)                       | MARS/Earth (earth)                            |
| Sélection de variables       | Automatique via la pénalité lasso          | Automatique via élagage avant/arrière         |
| Relations entre coefficients | Effets linéaires uniquement                | Linéaires par morceaux (fonctions charnières) |
| Interactions                 | Deux à deux, via la matrice d'interactions | Jusqu'au degré 3                              |
| Signes des coefficients      | Imposables par contraintes                 | Non imposables directement                    |
| Ajustement relâché           | Relaxed lasso disponible                   | Sans objet                                    |
| Réglage d'alpha              | Recherche sur grille disponible            | Sans objet (pas de paramètre de mélange)      |
| Sales Comparison Grid        | Disponible (étape 9)                       | Disponible (étape 8)                          |
| Graphiques de g-fonctions    | Sans objet                                 | Visualisation des termes groupés              |
| Formats de rapport           | HTML, Word, PDF                            | HTML, Word, PDF                               |

### 14.2 Fonctionnalités communes

Les deux outils offrent :

- l'import CSV et Excel avec conversion des noms de colonnes en snake\_case ;
- la prise en charge des locales de 31 pays, avec séparateur CSV, séparateur décimal et format de papier ;
- les désignations spéciales de colonnes (contract\_date, living\_area, area, latitude, longitude, etc.) ;
- le calcul automatique de sale\_age à partir de la date d'effet ;
- le basculement entre mode sombre et mode clair ;
- la persistance des paramètres via localStorage ;
- la carte de chaleur des corrélations, l'importance des variables, les contributions, l'ANOVA et les diagnostics ;
- le score CQA et le flux de travail d'ajustements RCA ;
- la Sales Comparison Grid avec sélection des comparables et sortie Excel mise en forme ;
- la compatibilité avec les mêmes jeux de données de démonstration (Appraisal\_1.csv).

### 14.3 Lequel utiliser, et quand

- **Utiliser glmnetUI** lorsqu'on souhaite un modèle linéaire simple et interprétable, doté d'une sélection automatique de variables. L'elastic net est particulièrement adapté lors-

qu'on dispose de nombreux prédicteurs et que l'on souhaite laisser le modèle retenir les plus importants.

- **Utiliser earthUI** lorsqu'on s'attend à des relations non linéaires (par exemple lorsque l'effet de la surface habitable sur le prix varie selon la taille) ou lorsque des fonctions charnières sont nécessaires pour capter des effets de seuil. Les modèles earth peuvent en outre capter des interactions d'ordre supérieur (degrés 2 et 3). earthUI propose par ailleurs une fonctionnalité expérimentale d'extraction des remarques fondée sur Prolog/DCG, sans équivalent dans glmnetUI.
- **Utiliser les deux** pour comparer modèles linéaires et non linéaires sur les mêmes données. Si les deux produisent des  $R^2$  voisins, le modèle glmnet, plus simple, peut être préféré pour son interprétabilité.

## 15 Jeu de données de démonstration : Appraisal\_1.csv

### 15.1 Description

Le jeu de données de démonstration est fourni avec glmnetUI (et partagé avec earthUI). Il se localise ainsi :

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

Le fichier contient 1 502 ventes résidentielles (plus 1 bien à évaluer en ligne 1) issues d'une exportation MLS simulée. Les données représentent des ventes de maisons individuelles dans un marché comportant plusieurs zones, avec une gamme variée de tailles, d'âges et de localisations.

Il ne s'agit pas de données réelles, mais elles s'inspirent d'un quartier réaliste du nord de la Californie. Toutes les informations d'identification ont été modifiées ou supprimées.

### 15.2 Colonnes

| Colonne          | Type      | Type spécial  | Description                                                              |
|------------------|-----------|---------------|--------------------------------------------------------------------------|
| weight           | numeric   | weight        | Pondération de l'observation (0 = exclue de l'ajustement)                |
| id               | numeric   | display_only  | Identifiant interne d'enregistrement                                     |
| property_id      | numeric   | display_only  | Identifiant MLS du bien                                                  |
| listing_id       | character | display_only  | Numéro d'annonce MLS                                                     |
| parcel_number    | character | display_only  | Numéro de parcelle du service du cadastre (APN)                          |
| street_address   | character | display_only  | Adresse du bien                                                          |
| city_name        | character | display_only  | Ville                                                                    |
| postal_code      | character | display_only  | Code postal                                                              |
| county_name      | character | display_only  | Comté                                                                    |
| contract_date    | Date      | contract_date | Date du contrat de vente (permet de calculer sale_age)                   |
| sale_age         | numeric   | —             | Nombre de jours entre la date de contrat et la date d'effet              |
| sale_price       | numeric   | (cible)       | Prix de vente — variable réponse                                         |
| living_sqft      | numeric   | living_area   | Surface habitable brute en pieds carrés                                  |
| beds_total       | integer   | —             | Nombre de chambres                                                       |
| baths_total      | numeric   | —             | Nombre total de salles de bains (par exemple 2,5 = 2 complètes + 1 demi) |
| lot_size         | numeric   | lot_size      | Superficie du terrain en pieds carrés                                    |
| area_id          | integer   | area          | Identifiant de zone MLS                                                  |
| age              | numeric   | actual_age    | Âge du bien en années                                                    |
| latitude         | numeric   | latitude      | Latitude (arrondie à 3 décimales pour le modèle)                         |
| longitude        | numeric   | longitude     | Longitude (arrondie à 3 décimales pour le modèle)                        |
| garage_spaces    | integer   | —             | Nombre de places de garage                                               |
| days_on_market   | integer   | dom           | Jours de mise en marché                                                  |
| listing_date     | Date      | listing_date  | Date de mise en annonce                                                  |
| sale_concessions | numeric   | concessions   | Concessions du vendeur                                                   |

### 15.3 Prise en main rapide suggérée

1. Lancer la routine glmnet : `earthUI::launch_glmnet()`
2. Créer un nouveau projet (section 1) avec la finalité **For Appraisal**
3. Placer **Appraisal\_1.csv** dans le dossier d'entrée du projet et le sélectionner dans la section 2 (cliquer sur **Refresh** si nécessaire)
4. Sélectionner **sale\_price** comme variable cible
5. Attribuer les types spéciaux comme indiqué dans le tableau ci-dessus
6. Inclure les prédicteurs : **sale\_age**, **living\_sqft**, **baths\_total**, **lot\_size**, **area\_id** (en facteur), **age**, **latitude**, **longitude**, **garage\_spaces**
7. Conserver  $\alpha = 1$  (Lasso), validation croisée avec `lambda.1se`
8. Cliquer sur **Fit Glmnet Model**
9. Examiner les onglets Summary, Coefficients et Diagnostics
10. Télécharger la sortie (étape 7) et examiner le classement CQA
11. Calculer les ajustements RCA (étape 8) avec un score CQA d'environ 5,00
12. Générer une Sales Comparison Grid (étape 9) avec les comparables recommandés

## 16 Configuration requise et résolution des problèmes

### 16.1 Plateformes prises en charge

glmnetUI fonctionne sous **macOS**, **Windows** et **Linux**. RStudio Desktop (2023.06 ou ultérieur) est vivement recommandé — il embarque **pandoc**, nécessaire aux rapports.

### 16.2 Dépendances facultatives

- **LaTeX** (TinyTeX, MiKTeX ou MacTeX) : uniquement pour la génération de rapports PDF. En l'absence de détection, l'option PDF est automatiquement masquée. Installation : `tinytex::install_tinytex()`
- **sysfonts + showtext** : police Roboto Condensed pour les graphiques. Repli automatique sur la police sans empattement du système si elle est indisponible ou hors ligne.
- **earth** ( $\geq 5.3.0$ ) : nécessaire uniquement pour la chaîne d'import depuis earthUI.

### 16.3 Remarques par plateforme

**macOS** : le moins de difficultés. Installer TinyTeX pour le PDF : `tinytex::install_tinytex()`

**Windows** : fonctionne bien avec RStudio. Les machines d'entreprise verrouillées peuvent imposer des restrictions sur le répertoire temporaire — l'application émet alors un avertissement dans la console mais poursuit son exécution sans persistance des paramètres.

**Linux** : peut nécessiter des bibliothèques système : `sudo apt install libcurl4-openssl-dev libssl-dev libxml2-dev libsqlite3-dev libfontconfig1-dev`

### 16.4 Dégradation progressive

| Composant manquant                                     | Comportement                                                                                                                   |
|--------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------|
| Pas de LaTeX                                           | Option PDF masquée. HTML et Word restent disponibles.                                                                          |
| Pas d'accès internet / échec du chargement des polices | Police sans empattement du système utilisée. Message consigné dans la console.                                                 |
| Pas de RSQLite / système de fichiers en lecture seule  | Les paramètres ne sont pas conservés. L'application fonctionne normalement.                                                    |
| Répertoire temporaire non accessible en écriture       | La génération du rapport échoue avec un message d'erreur explicite. Définir TMPDIR vers un emplacement accessible en écriture. |

### 16.5 Résolution des problèmes

« **PDF option not available** » : exécuter `tinytex::install_tinytex()` dans R, puis redémarrer l'application.

« **Settings will not persist** » : vérifier les permissions du répertoire de données utilisateur (macOS : `~/Library/Application Support/R/glmnetUI/`, Linux : `~/.local/share/R/glmnetUI/`, Windows : `%APPDATA%/R/data/glmnetUI/`).

**Port 7879 déjà utilisé** : exécuter `lsof -ti:7879 | xargs kill` (macOS/Linux) ou `Stop-Process -Id (Get-NetTCPConnection -LocalPort 7879).OwningProcess` (Windows PowerShell).

## Références

- FRIEDMAN, Jerome, Trevor HASTIE et Robert TIBSHIRANI (2010). « Regularization Paths for Generalized Linear Models via Coordinate Descent ». In : *Journal of Statistical Software* 33.1, p. 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, p. 1-13. DOI : [10.18637/jss.v039.i05](https://doi.org/10.18637/jss.v039.i05).
- TAY, J. Kenneth, Balasubramanian NARASIMHAN et Trevor HASTIE (2023). « Elastic Net Regularization Paths for All Generalized Linear Models ». In : *Journal of Statistical Software* 106.1, p. 1-31. DOI : [10.18637/jss.v106.i01](https://doi.org/10.18637/jss.v106.i01).