

### 1. Lựa chọn đơn vị lập bảng

Trong năm đầu tiên của dự án, do chưa hoàn tất việc khảo sát mẫu nên mới chỉ phát triển mô hình tập trung vào thửa đất đã khảo sát được giá. (trường mẫu: GN\_3)

Ở bước lựa chọn đơn vị lập bảng cần thiết lập phạm vi về mặt không gian, sau đó thực hiện modeling tập trung vào các thửa đất nằm trong phạm vi không gian đó. Theo đơn vị hành chính, theo đơn vị là khu vực định giá, v.v.

Mô hình ở năm đầu tiên được phát triển tập trung vào 1 đơn vị hành chính có giá đã khảo sát được (XTHN).

### 2. Xử lý dữ liệu

Sau chọn đơn vị xử lý dữ liệu đối với các dữ liệu về đặc tính đất, thông tin giá.

1. Tạo Histogram và Box-plot để hiển thị các giá trị bất thường theo IQR và thao tác loại bỏ các giá trị này.

(Chức năng cho phép người dùng có thể trực tiếp loại bỏ các giá trị bất thường)

### 3. Data scaling

#### 3.1. Chuyển đổi biến dummy

Lý do không thực hiện chuyển đổi biến dummy đối với toàn bộ set dữ liệu là vì tùy từng đơn vị lập bảng mà đặc tính phân bố của dữ liệu thực tế sẽ khác nhau. Ví dụ: đường bộ ở Lào chia thành 4 cấp, từ cấp 1 – cấp 4, nhưng ở một số khu vực cụ thể có thể sẽ không có đường cấp 2.

Biến dummy được chia thành biến liên tục (khoảng cách, chiều dài, chiều rộng...) và biến danh nghĩa (hình dạng, mục đích sử dụng đất, phân khu sử dụng đất...). Quá trình chuyển đổi biến dummy là đưa các biến số về dạng biến phạm trù trong phạm vi từ 0 đến 1.

Vì vậy sẽ cần xây dựng chức năng để chia biến liên tục thành các khoảng, sau đó chuyển đổi về dạng phạm trù, và tập hợp các biến nhất định trong biến danh nghĩa để phân loại thành dạng phạm trù.

Ví dụ về mô hình trong giai đoạn 1:

Trường *zoning* bao gồm 5 loại khu vực sử dụng đất ( $N$ ,  $UBb$ ,  $UDb$ ,  $UEa$ ,  $UEb0$ ), trong đó  $UDb$  xuất hiện với tần suất nhiều nhất được thiết lập làm biến số chuẩn (0, 0, 0, 0, 0), tạo được 1 set gồm 4 biến dummy (n-1 biến).

Trường *pave* gồm chia thành 2 loại là *đã trải mặt đường* và *chưa trải mặt đường*, trong đó *chưa trải mặt đường* xuất hiện với tần suất nhiều nhất nên được thiết lập làm biến tiêu chuẩn, từ đó tạo được 1 biến dummy.

Trường *rdGrade* gồm 4 cấp loại đường là 1, 2, 3, 4, trong đó đường loại 3 xuất hiện với tần suất nhiều nhất và được thiết lập làm biến tiêu chuẩn, tạo được 3 biến dummy.

Trường *Eff\_ratio* là thông tin hình dạng, dùng Natural Breaks để chia đều thành 3 khoảng đều nhau (0.25 - 0.83, 0.83 - 0.93, 0.93-1.02), sau đó thiết lập khoảng có tần suất xuất hiện nhiều nhất làm biến số tiêu chuẩn, sẽ thu được 2 biến dummy. Khoảng có tần suất xuất hiện nhiều nhất là *Eff\_ratio\_3* (gần với hình vuông nhất 0.93 - 1.02)

### 3.2. Thay thế bằng giá trị Logarit tự nhiên

Biến phụ thuộc (dữ liệu giá) được chuyển đổi sang giá trị Logarit tự nhiên, sau đó xây dựng bằng hàm log-log.

$\ln(\text{GN}_3)$  : Sử dụng giá khảo sát

## 4. Lựa chọn biến và phân tích hồi quy từng bước

Danh sách dữ liệu đặc tính đất: *area*(Area), *Rec\_Area*(Minimum bounding rectangular area), *Con\_Area*(Convex hull area), ***Eff\_Ratio***(effective area ratio), *Con\_Ratio*(curvature ratio), *width\_rd*(length of minimum boundary rectangle of parcel), *height\_rd*(depth of minimum boundary rectangle of parcel), *Shape*(shape of the parcel), *Hwratio*(width/depth Ratio), bao gồm dữ liệu Road information by class ngoại trừ thông tin tọa độ), *RoadDst*(distance from the parcel to the nearest road centerline), ***Pave***(pavement of the nearest road), *ori\_rad*(orientation. road\_width, ***rdGrade***, *dst\_railro*(distance from railroad), *dst\_statio*(distance from Vientiane train station), *RdDstStn*(distance from pathway to Vientiane Railway station), *schdst1*(distance from parcel to nearest school), *dis\_hate*(distance to the nearest hate facility), *dis\_pref*(distance to the nearest preferred facility), L1, L2, ***Zoning***, *Zoning\_type*, *ele\_mean*, *Ele\_par*, *EleDiff*, *Slope*

Thông tin giá: *GN\_3* (giá khảo sát), *SV* (giá theo tuyến đường)

Những dữ liệu được liệt kê phía trên là những dữ liệu trọng tâm khi triển khai nghiên cứu trong giai đoạn 1 của dự án. Những dữ liệu này đã bao gồm trong set dữ liệu cung cấp. Thông tin cụ thể hơn vui lòng tham khảo tài liệu mô tả bảng đã được cung cấp kèm theo.

Những đặc tính đất được liệt kê phía trên sẽ được sử dụng làm giá trị  $X_1... X_4$  trong phân tích hồi quy, thông tin giá (GN\_3) được lấy làm giá trị  $Y$ . Thông qua đó sẽ thực hiện phân tích hồi quy từng bước (Stepwise Regression). Phân tích hồi quy từng bước là quá trình phân tích bắt đầu từ 1 biến số độc lập, sau đó lần lượt thêm từng biến số vào để rút ra được mô hình phù hợp nhất với tiêu chuẩn đã thiết lập trước đó.

Tiêu chuẩn thống kê:  $R^2$ : thiết kế giá trị tiêu chuẩn, kiểm chứng ý nghĩa thống kê  $t$  của các biến độc lập riêng lẻ, thiết kế đường cơ sở VIF, kiểm chứng ý nghĩa mô hình tổng thể (thống kê  $F$ ).

Liệt kê các mô hình có  $n$  biến số độc lập phù hợp với các tiêu chuẩn thống kê trên.

Đề xuất các mô hình phù hợp với tiêu chuẩn thống kê.

Ví dụ:

1. Mô hình gồm biến Loại đường, Phân vùng sử dụng đất, (nêu rõ ý nghĩa thống kê, tiêu chuẩn, thể hiện AIC)
2. Mô hình gồm biến Loại đường, Phân vùng sử dụng đất, Tình trạng mặt đường (nêu rõ ý nghĩa thống kê, tiêu chuẩn, thể hiện AIC)
3. Mô hình gồm biến Loại đường, Phân vùng sử dụng đất, Tình trạng mặt đường, Chỉ số hình dạng (nêu rõ ý nghĩa thống kê, tiêu chuẩn, thể hiện AIC)
- .
- .
- n. Mô hình gồm  $n$  biến số (nêu rõ ý nghĩa thống kê, tiêu chuẩn, thể hiện AIC)

So sánh AIC của các mô hình để lựa chọn mô hình hồi quy cuối cùng, sau đó lập bảng so sánh.

Trong danh sách dữ liệu đặc tính đất được liệt kê phía trên, **những dữ liệu được tô màu đỏ** là những dữ liệu được rút ra từ kết quả kiểm chứng ý nghĩa thống kê (giai đoạn 1 dự án).

Nói cách khác, các biến Eff\_Ratio, Pave, rdGrade, Zoning được sử dụng để xây dựng mô hình phân tích hồi quy, rồi từ công thức hồi quy để rút ra hệ số hồi quy của từng biến số để lập thành bảng so sánh.

Sau khi thực hiện phân tích hồi quy sẽ cho ra kết quả như dưới đây:

1. pave (tình trạng thi công mặt đường)

Có ý nghĩa/Không có ý nghĩa:

Có ý nghĩa ( $p\text{-value} < 0.0001$ )

2. Zoning (Quy hoạch sử dụng đất)

Có ý nghĩa/Không có ý nghĩa:

Zoning\_N: Có nghĩa (p-value < 0.0001)

Zoning\_UBb: Có nghĩa (p-value = 0.006)

Zoning\_UEa: Có nghĩa (p-value < 0.0001)

Zoning\_UEb: Có nghĩa (p-value < 0.0001)

### 3. rdGrade (Loại đường)

Có bao gồm toàn bộ biến dummy không: có

Có ý nghĩa/Không có ý nghĩa:

rdG\_1: Có nghĩa (p-value < 0.0001)

rdG\_2: Có nghĩa (p-value < 0.0001)

rdG\_4: Có nghĩa (p-value < 0.0001)

### 4. Eff\_ratio (Chỉ số hình dạng)

Có ý nghĩa/Không có ý nghĩa:

Eff\_ratio\_1: có nghĩa (p-value < 0.0001)

Eff\_ratio\_2: có nghĩa (p-value = 0.010)

Variable	Coefficient	Std. Error	t-value	P-value	VIF
const	16.48447	0.011477	1436.356	0	5.136066
pave	0.066485	0.014198	4.682723	3.51E-06	1.630081
Zoning_N	-0.16804	0.019583	-8.58133	8.25E-17	1.157754
Zoning_UBb	0.057592	0.020841	2.763396	0.005898	1.121671
Zoning_UEa	0.138235	0.015167	9.114207	1.22E-18	1.250815
Zoning_UEb	0.140779	0.012547	11.22009	1.27E-26	1.289564
rdG_1	0.213916	0.019744	10.83433	4.46E-25	1.422285
rdG_2	0.094589	0.017277	5.474866	6.48E-08	1.296538
rdG_4	-0.4906	0.014177	-34.6059	2.49E-144	1.135481
Eff_ratio_1	-0.0565	0.012633	-4.4721	9.29E-06	1.375996
Eff_ratio_2	-0.03244	0.012565	-2.58181	0.010068	1.371413
Danh sách biến số	Hệ số hồi quy	Độ lệch chuẩn	Ý nghĩa của biến độc lập		Tính đa cộng tuyến

## 5. Xây dựng ma trận

### 5.1. Rút ra hệ số hồi quy và chỉ số hóa

Exp được lấy để lập chỉ mục cho hệ số hồi quy rút ra từ phương trình hồi quy trên và xây dựng ma trận với biến chuẩn được đặt là 1. Ví dụ: biến Zoning có tổng cộng 5 loại dữ liệu trong đó có biến tham chiếu nên xây dựng được ma trận 5x5.

### **5.2. Điều chỉnh tỷ lệ so sánh:**

Khi bảng so sánh được tạo ra, chiều dọc có thể được hiểu là thửa đất chuẩn chuẩn và chiều ngang có thể được hiểu là tỷ lệ so sánh của thửa đất cụ thể cần định giá.

Cần phải có chức năng trên hệ thống cho phép người dùng trực tiếp sửa tỷ lệ so sánh trong bảng so sánh.

### **5.3. Tính giá thửa cụ thể:**

Giá đất của một thửa cụ thể được tính bằng cách so sánh thửa đất đó với 1 thửa đất chuẩn có những đặc tính tương tự nhất với thửa đất cần định giá hoặc nằm gần với thửa đất cần định giá.

Phương pháp tính toán dựa trên ví dụ nghiên cứu trong giai đoạn 1 dự án như sau:

Giá thửa đất chuẩn được chọn để so sánh x Tỷ lệ so sánh (zoning) x Tỷ lệ so sánh (rdG) x Tỷ lệ so sánh (Eff\_ratio) x tỷ lệ so sánh (pave) = giá đất của thửa cụ thể.

Phương pháp, tiêu chuẩn lựa chọn thửa chuẩn so sánh trong tổng số các thửa chuẩn sẽ được chúng tôi cung cấp sau.