



ANNEE UNIVERSITAIRE 2020-2021

### Projet de fin de semestre

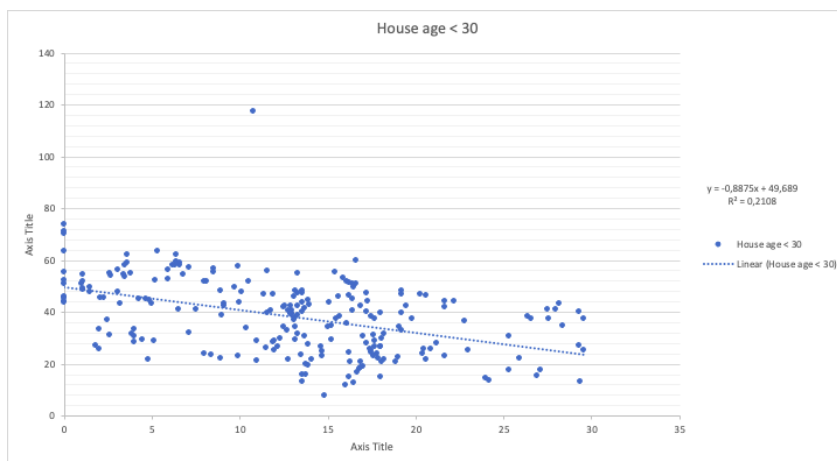
Groupe 5 : AZYADI MAKHLOUF / AMIR OUNOUNA / YASSER SAKHRAOUI / DANIEL TALHI / DAC-THIEN TRAN

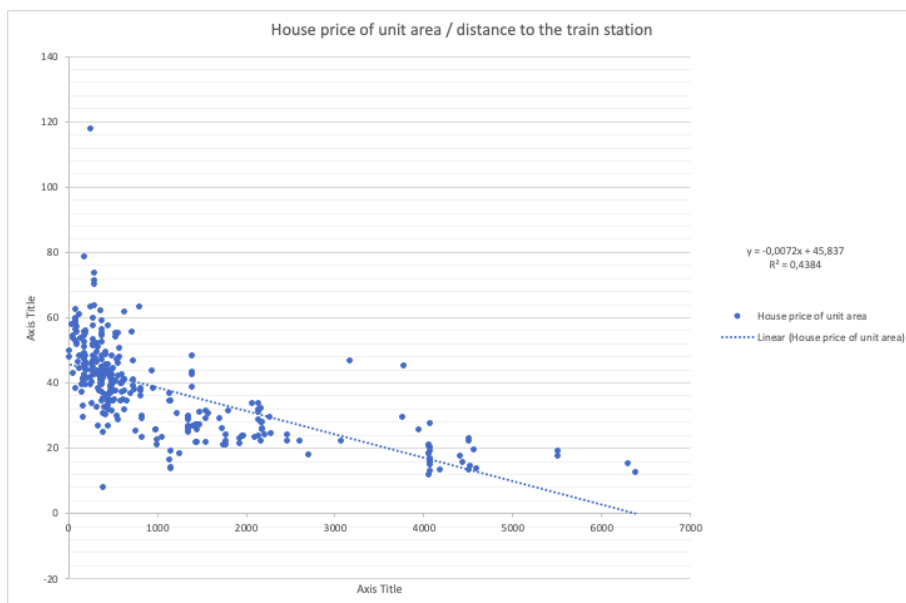
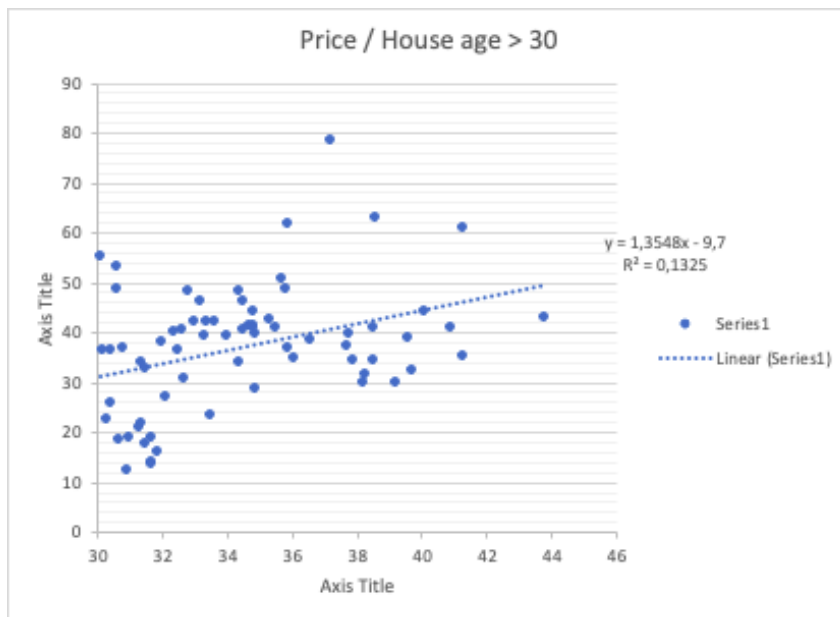
Nous avons effectué un test de nullité de paramètre  $\alpha = 0$ , et de niveau  $(1 - \alpha) = 95\%$ . Les p-valeurs obtenue grace à l'utilitaire d'analyse on constate qu'elles sont inférieures à 5% alors il y a une corrélation linéaire entre le prix des maisons et les autres variables. On observe un échantillage de taille  $n = 300$  avec des couples  $(X_i, Y_i)$ . On fixe notre  $Y_i$  sur le (House price of unit area) et on varie nos variable explicatives (House age, Distance to the nearest MRT (Mass Rapid Train) station, Nb of convenience stores)

Ce qui nous donne pour les couples :

- (House age, price) :  $\{0;70,1\}\{0;50,8\}\{0;45,4\}...$ etc
- (Distance to MRT, price) :  $\{292,9978;70,1\}\{338,9679;50,8\}\{274,0144;45,4\}...$ etc
- (Nb of stores, price) :  $\{6;70,1\}\{9;50,8\}\{1;45,4\}...$ etc

Ensuite nos 300 couples sont représentés dans des nuages de points pour chaque variable explicative avec la variable endogène (Price)







Ensuite nous calculons la différence entre les prix observés dans notre base de données et notre  $a(x) + b$  ( pour retrouver notre epsilon )

| Distance jusqu'au MRT (Mass Rapid Train) station | Prix par unité | Distance centrée | Prix centrée | Prix hat | Epsilon |
|--|----------------|------------------|--------------|----------|---------|
| 293,00   | 70,10          | -792,65          | 13,70        | 43,73    | 26,371  |
| 338,97   | 50,80          | -746,68          | -5,60        | 43,40    | 7,401   |
| 274,01   | 45,40          | -811,64          | -11,00       | 43,87    | 1,534   |
| 208,39   | 45,70          | -877,26          | -10,70       | 44,34    | 1,362   |
| 293,00   | 71,00          | -792,65          | 14,60        | 43,73    | 27,271  |

## Estimation MCO au moyen de l'utilitaire d'analyse d'Excel (Cf. Excel Partie 2)

| Les variables exogènes X | L'estimateur de la pente A | L'estimateur de la constante B | La P.Value           |
|--------------------------|----------------------------|--------------------------------|----------------------|
| X1 : House age < 30      | -0,88750134                | 49,6892899                     | 1.68006323286208E-13 |
| X1 : House age > 30      | 1.35480337346078           | -9.7000417386822               | 0.0021085546643828   |
| X2                       | -0.00719267984926122       | 45.8366365856513               | 2.4120137665448E-39  |
| X3                       | 2.66682298565804           | 27.023939707019                | 2.88091779897929E-26 |

## Interpréter le test de nullité et expliquer son intérêt (Cf. Excel Partie 2)

### Test de nullité au niveau 95 % :

L'hypothèse nulle indique que la pente est égale à zéro et l'hypothèse alternative indique que la pente n'est pas égale à zéro.

#### Pour X1 : House age < 30

$H_0 : a=0$  contre  $H_1 : a \neq 0$ .

P.Value = 1.68006323286208E-13.

P.Value <  $\alpha$  avec  $\alpha=5\%$

**Conclusion** : Comme la P.Value est inférieure à 5% alors on rejette l'hypothèse donc il n'y a pas de corrélation linéaire entre les prix des maisons et le fait qu'elles soient récentes (moins de 30ans).

#### Pour X1 : House age > 30

$H_0 : a=0$  contre  $H_1 : a \neq 0$ .

P.Value = 0.0021085546643828

P.Value <  $\alpha$  avec  $\alpha=5\%$

**Conclusion** : Comme la P.Value est inférieure à 5% alors on rejette l'hypothèse donc il n'y a pas de corrélation linéaire entre les prix des maisons et le fait qu'elles soient récentes (30ans et plus).

**Pour X2 :**

$H_0 : a=0$  contre  $H_1 : a \neq 0$ .

P.Value = 2.4120137665448E-39

P.Value <  $\alpha$  avec  $\alpha=5\%$

**Conclusion :** Comme la P.Value est inférieure à 5% alors on rejette l'hypothèse donc il n'y a pas de corrélation linéaire entre les prix des maisons et la distance à la station la plus proches, "Distance to the nearest MRT (Mass Rapid Train) station".

**Pour X3 :**

$H_0 : a=0$  contre  $H_1 : a \neq 0$ .

P.Value = 2.88091779897929E-26

P.Value <  $\alpha$  avec  $\alpha=5\%$

**Conclusion :** Comme la P.Value est inférieure à 5% alors on rejette l'hypothèse donc il n'y a pas de corrélation linéaire entre le prix des maisons et le nombre des convenience stores.

**Intérêt du test de nullité :**

s'il existe une relation linéaire significative entre la variable explicative x et la variable endogène y, la pente ne sera pas égale à zéro.

**Effectuer des tests qui vous semblent pertinents sur la pente (Cf. Excel Partie 2)**

**Test de Fisher (F-test) :**

Le test de Fisher semble pertinent car le coefficient de la pente varie.

Régression multiple :

Grace à l'utilitaire d'analyse d'Excel on obtient une moyenne des carrés de régression égale à 10187.26 et une moyenne des carrés de résidus égale à 91.92. L'utilitaire d'analyse nous donne également  $R^2 = 0.53$ , ce qui signifie que 53% est la part de variabilité Y prix expliqué par les variations des variables age, distance et store. En additionnant les  $R^2$  provenant des régressions linéaires simples on obtient  $\sim 0.81$  soit 81%. On remarque alors le coefficient de régression multiple est différent de la somme des régressions simples. Cela peut s'expliquer puisque les variables s'influencent entre elles, et puisque nous travaillons avec 3 variables l'apport explicatif des variables réunies est plus faible que si l'on prenait chaque variable séparément. Car dans chacune des variables, on retrouve des parties d'informations des autres variables. L'intérêt du test de nullité est de savoir s'il existe une relation linéaire significative entre la variable explicative x et la variable endogène y, la pente ne sera pas égale à zéro.