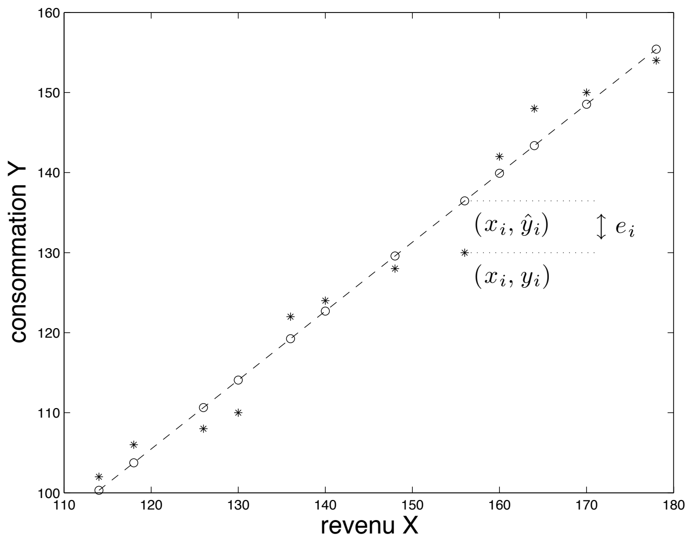
TD1 L3 Econométrie

Rappel : L'estimateur $\hat{\alpha}$ (resp. $\hat{\beta}$) est aussi noté a (resp. b).

I.1°)



2°)a). Sous forme exacte $y_i = \alpha x_i + \beta$.

b) Sous forme aléatoire

$$y_i = \alpha x_i + \beta + u_i,$$

où u_i est une famille de variables aléatoire vérifiant $E(u_i) = 0$ et $(u_i)_{i \leq n}$ deux à deux indépendantes, et la variance $V(u_i)$ est une constante qui ne dépend pas de i.

c) On cherche à minimiser le résidu mais il n'est pas nul *i.e.* les points ne sont a priori jamais sur la droite.

3°) On calcule les moyennes empiriques $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 145$ et $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = 127$.

Par définition, l'estimateur de α est

$$\hat{\alpha} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = 0.8612.$$

On retrouve l'ordonnée à l'origine $\hat{\beta}$ en écrivant que la moyenne $(\bar{x}, \bar{y})$ est sur la droite de régression :

$$\hat{\beta} = \bar{y} - \alpha\bar{x} = 2.1288.$$

Les résidus entre les observations y_i et la prédiction $\hat{y}_i$ s'écrivent $e_i = \hat{y}_i - y_i$.

$$\text{SCR} = \sum_{i=1}^{n} e_i^2 = 115.2685$$

4°) $(\hat{\alpha}, \hat{\beta})$ sont des estimateurs de α, β. Les valeurs e_i sont des instances des variables aléatoires u_i. D'une certaine manière e_i est un estimateur de u_i.

5°) α est le taux marginal de substitution. Pour obtenir l'élasticité il faut se ramener à des pourcentages,

$$\frac{dX/X}{dY/Y} = \alpha \frac{x_i}{y_i} = 0.98$$

6°) On a calculé la somme des carrés des résidus (SCR) et on obtient donc

$$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2} = \frac{115.2685}{10} = 11.5268.$$

Puis $\sigma_{\hat{\alpha}}^2 = \frac{\hat{\sigma}_u^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = 0.0024$ et $\sigma_{\hat{\beta}}^2 = \bar{x}^2 \hat{\sigma}_a^2 + \frac{\hat{\sigma}_u^2}{n} = 51.3247$

7°) On teste si α et β sont significativement non nuls. On sait que $T = \frac{\hat{\alpha} - \alpha}{\hat{\sigma}_\alpha}$ suit une loi de Student $\text{St}(n-2)$.

Pour effectuer le test

$$\begin{cases} H_0 : \alpha = 0 \\ H_1 : \alpha \neq 0 \end{cases},$$

on construit $t_{\text{obs}} = \frac{\hat{\alpha} - 0}{\hat{\sigma}_\alpha} = 17.59$.

Pour $p = 0.05$ on lit dans la table $t_{\frac{p}{2}} = 2.228$. La région d'acceptation (R.ACC) est $[-2.228, 2.228]$. Finalement $t_{\text{obs}} \notin$ R.ACC, donc on rejette H_0 et on conclut que H_1 est significativement vraie.

Le test pour β est le même, $\quad t_{\text{obs}} = \frac{\hat{\beta}}{\sigma_{\hat{\beta}}} = 0.2972 \in$ R.ACC donc la valeur de $\hat{\beta}$ n'est pas significative.

Remarque : la relation linéaire entre x et y est significative si α est significativement non nul.

8°) Avec la même valeur de $t_{\frac{p}{2}}$ que pour la question précédente on construit

$$\text{IC}(\alpha) = [\hat{\alpha} \pm t_{\frac{p}{2}} \sigma_{\hat{\alpha}}] = [0.75, 0.97]$$

$$\text{IC}(\beta) = [\hat{\beta} \pm t_{\frac{p}{2}} \sigma_{\hat{\beta}}] = [-13.83, 18.09].$$

9°) Par définition $R^2 = \frac{\text{cov}(X,Y)^2}{V(X)V(Y)}$. La formule de calcul usuel est

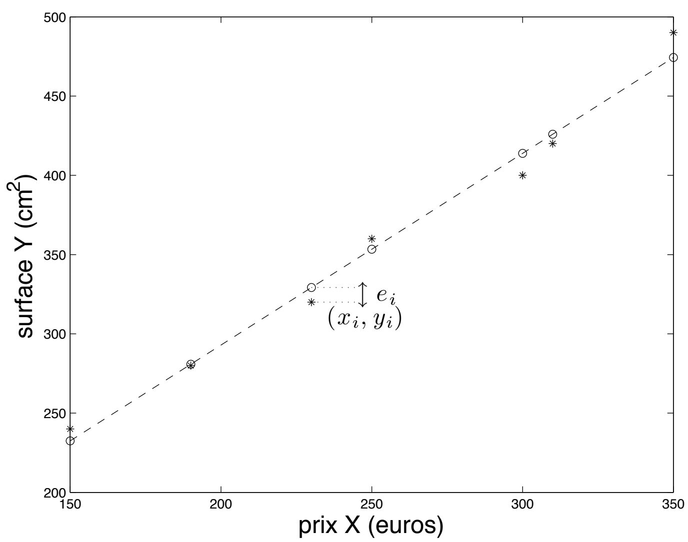
$$R^2 = R^2 = \frac{(\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y})^2}{(\sum_{i=1}^n x_i^2 - n\bar{x}^2)(\sum_{i=1}^n y_i^2 - n\bar{y}^2)} = 0.9687$$

Remarque : pour établir cette égalité on utilise $\sum_{i=1}^n (x_i - \bar{x})^2 = (\sum_{i=1}^n x_i^2) - n\bar{x}^2$, formule à comparer avec l'égalité $V(X) = E(X^2) - E(X)^2$. On a également $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = (\sum_{i=1}^n x_i y_i) - n\bar{x}\bar{y}$.

10°)

$$\boxed{\begin{array}{rr} \hat{y} = 0.86x + 2.13 & R^2 = 0.97 \\ & (1) : 51.32 \end{array}}$$

II.A.



On calcule pour les estimateurs des coefficients des deux régressions $\hat{\alpha} = 1.21$, $\hat{\beta} = 51.08$ et pour la relation inverse, $\hat{\alpha}' = 0.8147$ et $\beta' = -37.8452$.

II.B. 1°) La formule de calcul donne $R^2 = R^2 = \frac{(\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y})^2}{(\sum_{i=1}^n x_i^2 - n\bar{x}^2)(\sum_{i=1}^n y_i^2 - n\bar{y}^2)} = 0.9852$. On peut aussi vérifier

$$R^2 = \hat{\alpha}\hat{\alpha}' = 1.21 \times 0.8147.$$

2°) Il faut calculer σ_u^2, l'estimateur de la variance de u. Vu qu'on a déjà calculé R^2 on peut utiliser l'égalité

$$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^n e_i^2}{n-2} = \frac{1-R^2}{n-2}\left(\sum_{i=1}^n y_i^2 - n\bar{y}^2\right) = 88.81.$$

On en déduit les variance des estimateurs de α et β , $\sigma_{\hat{\alpha}}^2 = 0.0044$ et $\sigma_{\hat{\beta}}^2 = 269.39$. On construit les tests à partir de la variable T

$$T = \frac{\hat{\alpha} - \alpha}{\sigma_{\hat{\alpha}}} \sim \text{St}(n-2)$$

$$\begin{cases} H_0 : \alpha = 0 \\ H_1 : \alpha \neq 0 \end{cases} , \quad t_{\text{obs}} = \frac{\hat{\alpha}}{\hat{\sigma}_\alpha} = 408.065$$

On lit dans la table $t_{\frac{p}{2}} = 2.365$, la région d'acceptation de H_0 est

$$[-t_{\frac{p}{2}}, t_{\frac{p}{2}}] = [-2.365, 2.365] \not\ni t_{\text{obs}},$$

donc on peut rejeter H_0, α est significativement non nul.

3°) On construit à partir de R^2 la variable F

$$F = \frac{R^2/1}{(1-R^2)/(n-2)} \sim \text{F}(1, n-2).$$

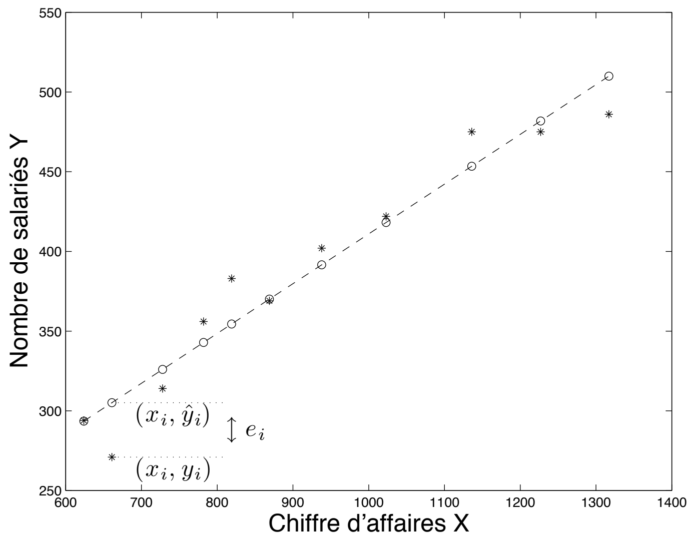
Ici $F_{\text{obs}} = 332.48$. On lit pour $n = 7$ et $p = 0.05$, la valeur critique $t_p = 6.61$ donc $F_{\text{obs}} \geq t_p$ la régression est bonne dans son ensemble.

Remarque : le test de Student pour α et β est bilatéral, le test de Fisher pour R^2 est unilatéral.

4°) On prédit à partir des estimations de α' et β' $0.81 \times 510 - 37.84 = 377.65$ euros

5°) De même à partir de $\hat{\alpha}$ et $\hat{\beta}$, $1.2 \times 400 + 51 = 534.8$ cm^2.

III. 1°)



2°) On trouve $a = \hat{\alpha} = 0.31$ et $b = \hat{\beta} = 98.68$.

3°) $\hat{\sigma}_u^2 = \frac{\sum_{i=1}^n e_i^2}{n-2} = 389.47$,

$R^2 = \frac{(\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y})^2}{(\sum_{i=1}^n x_i^2 - n\bar{x}^2)(\sum_{i=1}^n y_i^2 - n\bar{y}^2)} = 0.937$, $\sigma_{\hat{\alpha}}^2 = \frac{\hat{\sigma}_u^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = 7.3 \times 10^{-3}$.

4°) Ce test vérifie que α est significativement non nul. On teste ainsi l'existence d'une relation linéaire entre x et y.

On construit $t_{\text{obs}} = \frac{\hat{\alpha}}{\hat{\sigma}_\alpha} = 11.54$. La valeur critique pour $n = 11$ observations est $t_{\frac{p}{2}} = 2.262$, donc $t_{\text{obs}} \notin [-t_{\frac{p}{2}}, t_{\frac{p}{2}}]$ et α est significativement non nul.

5°) $0.31 \times 1772 + 98.68 = 652.031$. On prédit 652031 salariés.

6°) On utilise les intervalles de confiance de α et β à $p = 0.05$. On reprend dans la table de Student le même $t_{\frac{p}{2}}$ que pour le 4°) et on obtient les intervalles

$$\text{IC}(\alpha) = [\hat{\alpha} \pm t_{\frac{p}{2}} \sigma_{\hat{\alpha}}] = [0.25, 0.37]$$

$$\text{IC}(\beta) = [\hat{\beta} \pm t_{\frac{p}{2}} \sigma_{\hat{\beta}}] = [40.8, 156.5].$$

En réutilisant ces intervalles il vient

$$\text{IC}(\alpha) \times 1772 + \text{IC}(\beta) = [485.79, 818.27].$$

Donc la valeur fournie ne contredit pas le modèle, elle est dans l'intervalle de confiance.

IV. 1˚) On retrouve dans le tableau $\hat{\alpha} = 0.0589$, $\hat{\beta} = 36.42$, SCT $= 1150$, SCE $= 972.3$ et SCR $= 177.7$. On peut vérifier SCT $=$ SCE$+$SCR. Par définition

$$R^2 = \frac{\text{SCE}}{\text{SCT}} = 1 - \frac{\text{SCR}}{\text{SCT}} = 0.8455.$$

2˚) $\begin{cases} H_0 : \alpha = 0 \\ H_1 : \alpha \neq 0 \end{cases}$, $t_{\text{obs}} = \frac{\hat{\alpha}}{\hat{\sigma}_\alpha} = 5.23$ La valeur critique pour $n = 7$ observations est $t_{\frac{p}{2}} = 2.571$, donc $t_{\text{obs}} \notin [-t_{\frac{p}{2}}, t_{\frac{p}{2}}]$ et α est significativement non nul. On peut commenter que t_{obs} n'est pas très loin de la région d'acceptation.

3˚) On construit à partir de R^2 la variable F

$$F = \frac{R^2/1}{(1 - R^2)/(n-2)} \sim \mathrm{F}(1, n-2).$$

On obtient $F_{\text{obs}} = 27.36$. On lit pour $n = 7$ et $p = 0.05$, la valeur critique $t_p = 6.61$ donc $F_{\text{obs}} \geq t_p$ la régression est bonne dans son ensemble. En fait il y a plusieurs valeurs de I pour une même valeur de V dans le tableau, on pouvait s'attendre à ce que la régression soit mauvaise.