

Supports à lire avant de commencer les exercices

Site du cours : http://eric.univ-lyon2.fr/~ricco/cours/cours_econometrie.html

A lire en particulier pour cette séance :

- Régression multiple (slides) : http://eric.univ-lyon2.fr/~ricco/cours/cours/Regression_Lineaire_Multiple.pdf
- Test de Chow (slides) : http://eric.univ-lyon2.fr/~ricco/cours/slides/Reg_Multiple_Test_de_Chow.pdf
- Régression multiple (fascicule de cours), chapitres 9 à 13 : http://eric.univ-lyon2.fr/~ricco/cours/cours/econometrie_regression.pdf

Exercice 1 : Nous tentons d'expliquer leur teneur en CO (monoxyde de carbone) des cigarettes à partir de leur composition en goudron (TAR) et NICOTINE. Nous obtenons le modèle suivant avec un échantillon de $n = 24$ cigarettes :

$$CO = 19.88 + 0.892 \times TAR + 0.629 \times NICOTINE$$

Avec $R^2=0.934$ et $SCT = 413.750$

1. Construire le tableau d'analyse de variance. La régression est-elle globalement significative à 5% ?
2. Calculez le R^2 corrigé.
3. Calculez l'estimation de l'écart type des erreurs ($\hat{\sigma}_\varepsilon$).
4. Sachant que $(X'X)^{-1}$ est égal à :

	const	tar	nicotine
const	0.5500	0.0492	-1.2961
tar	0.0492	0.0284	-0.4528
nicotine	-1.2961	-0.4528	7.8421

Calculez l'écart-type du coefficient associé à la variable TAR.

5. Le coefficient de TAR est-il significativement différent de zéro à 5% ?
6. Les coefficients associés à TAR et NICOTINE sont-ils simultanément nuls à 5% ?

Exercice 2 : On souhaite analyser l'influence de l'âge (X1) et du nombre d'années d'études (X2) sur le salaire à l'embauche (Y) dans une grande compagnie. Voici les résultats fournis par DROITEREG.

AGE	ETUDES	CONST
20.736	312.921	-2314.186
17.611	91.817	1618.277
0.240	1258.282	#N/A
5.852	37	#N/A
18530083.276	58581099.699	#N/A

1. Quelle est la taille n de l'échantillon qui a servi à l'élaboration du modèle ?
2. Le modèle est-il globalement significatif au risque 5%

Un membre de la DRH suspecte que l'impact des exogènes sur l'endogène n'est pas le même selon que le candidat soit un homme ou une femme. Il propose donc de faire 2 régressions séparées sur ces mêmes données. Voici les nouveaux résultats.

Hommes

AGE	ETUDES	CONST
14.529	289.542	-1353.040
34.032	161.853	2985.445
0.163	1467.356	#N/A
1.658	17	#N/A
7140077.268	36603249.932	#N/A

Femmes

AGE	ETUDES	CONST
23.803	236.380	-1914.875
16.180	95.015	1570.655
0.274	920.088	#N/A
3.212	17	#N/A
5437813.011	14391556.739	#N/A

1. Il y a combien d'homme et de femmes dans l'échantillon de données.
2. La régression est-elle globalement significative chez les hommes ? Chez les femmes ?
3. Est-ce que les régressions sont significativement différentes dans les deux sous-populations ?

Exercice 3 : On veut expliquer l'accélération (ACCELERATION) d'un véhicule à partir du nombre de cylindres (CYLINDERS), du volume du moteur (ENGINE.DISP), de sa puissance (HORSEPOWER) et de son poids (WEIGHT).

Voici les résultats fournis par la fonction **DROITEREG** d'EXCEL.

DROITEREG				
weight	horsepower	engine.disp	cylinders	constante
4.1315	-10.1781	-0.8946	-0.3460	17.3682
0.6655	1.5944	1.4096	0.6806	2.1443
0.6853	1.6008	#N/A	#N/A	#N/A
24.5040	45	#N/A	#N/A	#N/A
251.1644	115.3118	#N/A	#N/A	#N/A

Par ailleurs, nous avons calculé la matrice $(\mathbf{X}'\mathbf{X})^{-1}$. L'ordre des variables dans la matrice X est : constante, CYLINDER, ENGINEDISP, HORSEPOWER, WEIGHT.

1.7944	-0.4735	1.0895	-0.4714	-0.2677
-0.4735	0.1808	-0.3368	0.0661	0.0215
1.0895	-0.3368	0.7755	-0.3208	-0.1357
-0.4714	0.0661	-0.3208	0.9921	-0.0999
-0.2677	0.0215	-0.1357	-0.0999	0.1728

1. Donner les principaux indicateurs relatifs à cette régression : n – nombre d'observations, p – nombre de variables, R^2 - coefficient de détermination, F de Fisher, ddl - degré de liberté, SCR, SCE et SCT.
2. Construire le tableau d'analyse de variance et tester la significativité globale de la régression à 5% ? Tester également la significativité de chaque coefficient (excepté la constante) à 5%.
3. Tester à 5% l'hypothèse suivante : H_0 , les coefficients de ENGINE.DISP et CYLINDERS sont simultanément égaux à 0 ; contre, H_1 , au moins un des deux est différent de 0¹.
4. Calculer la prédiction ponctuelle et l'intervalle de prédiction pour le véhicule avec les caractéristiques suivantes (cylinders = 6 ; engine.disp = 2.5 ; horsepower = 1 ; weight = 3.282).
5. Tester à 5% l'hypothèse selon laquelle (H_0) : le coefficient de HORSEPOWER est égal à celui de CYLINDERS.

¹ Pour l'inversion d'une matrice 2 x 2, utiliser la technique suivante : $\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$