

Statistiques : Enoncé du TP 1 Régression linéaire

L'exercice utilise les indices de productions agricoles et industrielles suivants :

Année	Production agricole	Production industrielle
1944	100	100
1945	60	50
1946	76	84
1947	74	99
1948	90	113
1949	93	122
1950	102	128
1951	98	143
1952	103	145
1953	110	146
1954	117	159
1955	118	172
1956	112	188
1957	115	204
1958	116	213
1959	121	220
1960	134	242
1961	130	254

1. Lancer Excel
2. Saisie des données :
 - (a) Cliquer (avec le bouton gauche de la souris) dans la cellule A1, taper 100, déplacer le curseur dans la cellule A2, taper 61 et ainsi de suite pour l'ensemble des indices de production agricole. Faites de même dans la colonne B pour les indices de production industrielle.
 - (b) En cliquant sur 1 en début de première ligne, sélectionner cette ligne puis, en cliquant sur l'icône **Insertion** située en haut de l'écran, puis **ligne**, insérer une ligne dans laquelle vous taperez "agri" et "indus" respectivement en tête des deux colonnes.
 - (c) En cliquant sur A, sélectionner la première colonne et insérer une colonne avant les deux précédentes. Taper "Année" en première ligne et 1944 en deuxième ligne. Sélectionner la portion de la colonne située entre les lignes 2 et 19 en maintenant le bouton de gauche de la souris enfoncé, puis, en cliquant sur l'icône **Edition** puis **Recopier**, puis **Série**, on fait apparaître une fenêtre **Série de données** dans laquelle on sélectionnera **colonnes** puis **linéaire** et enfin **1** pour le pas.
3. Calculs statistiques :
 - (a) En B20, calculer la somme des nombres figurant dans la colonne B en tapant la formule "**=SOMME(B2 :B19)**" et en B21 leur moyenne : taper = puis cliquer sur la cellule B20 et compléter la formule en tapant / puis **18**).
 - (b) Recommencer en colonne C. Enfin en A20 et A21, taper les mots "Somme=" et "Moyenne=" respectivement, les passer en gras (icône **G** en haut de l'écran) et centrer à droite.
4. Représentation du nuage : Après avoir sélectionné les colonnes B et C entre les lignes 2 et 19, tracer le nuage de points à l'aide de l'**assistant graphique** (icône multicolore située en haut de l'écran) : choisir **nuage de point** puis vérifier la plage de données et le choix de colonnes. Choisir un titre de graphique et nommer les axes des X et des Y. Disposer le graphique **en tant qu'objet dans la feuille** (on pourra le déplacer avec la souris). S'il n'est pas satisfaisant, il suffit de cliquer dessus du droit, de choisir **effacer** et de recommencer !

On se propose à présent de calculer les deux droites de régression de l'indice de production industrielle sur l'indice de production agricole et réciproquement, puis de les superposer à ce nuage de points.

5. Calcul de la première droite des moindres carrés :
 - (a) Dans les colonnes D, E et F, taper en ligne 1 les titres “agri*indus”, “agri*agri” et “indus*indus”.
 - (b) En D2, taper = puis cliquer sur la cellule B2, taper * puis cliquer sur la cellule C2. On obtient ainsi le produit des indices de production agricole et industrielle pour l’année 1944, c’est-à-dire 10 000. Pour compléter automatiquement les autres lignes de la colonne D (et ne pas retaper cette formule 18 fois!), sélectionner cette colonne entre les lignes 2 et 19, puis cliquer sur **Edition**, puis **Recopier** puis **En bas**.
 - (c) Procéder de la même façon pour compléter les colonnes E et F.
 - (d) Compléter enfin les lignes 20 et 21, en recopiant, vers la droite cette fois, les fomules donnant la somme de la colonne et sa moyenne.
 - (e) En vous plaçant dans une cellule vide (par exemple A24), calculer l’aide des 5 moyennes obtenues en ligne 21, la pente a de la droite des moindres carrés de indus sur agri (on utilisera les formules données dans le cours). Puis, dans une autre cellule, calculer son ordonnée à l’origine b et dans une troisième cellule le coefficient de corrélation linéaire r_{xy} .
 - (f) Dans la colonne G, calculer les valeurs approchées (on dit aussi *valeurs prédites*) de l’indice de production industrielle (image des indices de production agricole par cette droite des moindres carrés) : attention, au moment de recopier votre formule, comme les références aux deux cellules où a et b ont été calculés doivent rester fixes, il faudra les saisir avec des “\$” ; par exemple A24 devra être remplacée dans la formule par \$A\$24 puisqu’il s’agit d’une référence absolue.
6. Calcul de la deuxième droite des moindres carrés : recommencer les calculs des deux dernières questions (e) et (f) pour obtenir la droite des moindres carrés de l’indice de production agricole sur l’indice de production industrielle cette fois ainsi que les valeurs prédites de l’indice de production agricole.
7. Superposition des droites au nuage :
 - (a) Cliquer sur le graphique obtenu en 4, et sur l’**assistant graphique**, et à l’étape **2 sur 4**, choisir l’onglet **série** et ajouter la série des valeurs approchées de la colonne G comme **valeurs Y**. Commenter les résultats obtenus.
 - (b) Toujours sur le même graphique, ajouter la série des valeurs approchées de l’indice de production agricole. Commenter.
8. Sensibilité des résultats par rapport à des modifications des données : après avoir recopier vos résultats (calculs et graphique) sur une nouvelle feuille, changer les données de l’année 1945 en (60, 100) puis en (100, 50) et observer dans chacun des cas comment les deux droites de régression se déplacent. Qu’en pensez-vous? Etudiez ainsi l’influence des données éloignées.
9. Etude de la dispersion : en utilisant les colonnes C et G, calculer en colonne H les résidus ε_i et leur somme DT . Que vaut leur moyenne? Pouvait-on le prévoir? Représenter sur un même dessin, le nuage des résidus et la droite horizontale d’ordonnée 0. Calculer dans deux cellules vides, les quantités DA et DR et vérifier la formule $R^2 = \frac{DR}{DA}$.