

TD 8 : Régression

Le PIB et l'âge du président sont-ils liés ?

Année	Âge du président de la république (*)	PIB de la France (**)
1978	52	836,517
1979	53	864,22
1980	54	878,142
1981	65	888,842
1982	66	912,238
1983	67	925,864
1984	68	941,105
1985	69	954,753
1986	70	977,734
1987	71	1002,507
1988	72	1048,768
1989	73	1092,547
1990	74	1121,042
1991	75	1132,209
1992	76	1149,1
1993	77	1138,905
1994	78	1162,435
1995	63	1181,849
1996	64	1194,884
1997	65	1217,64
1998	66	1259,055
1999	67	1299,51
2000	68	1348,801
2001	69	1377,067
2002	70	1393,687

* Âge en fin d'année du président de la république en exercice (Giscard né en 1926, Mitterrand né en 1916 et Chirac né en 1932).

** PIB de la France en milliards d'euros calculé avec les prix de 1995. Source INSEE.

I. Introduction à la régression

Un statisticien croit déceler une liaison statistique méconnue et de la toute première importance pour la compréhension de l'évolution de l'économie entre le PIB (produit intérieur brut) de la France et l'âge du président ! Suivons-le dans ses développements impérissables.

a) Copiez et collez les données sur Excel.

b) Calculez pour les deux dernières variables, la moyenne, la variance ($V(X)$) et la variance modifiée (s^2), l'écart type ($ET(X)$) et l'écart type modifié (s).

Rappel :

- Formules :

$$V(X) = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \text{ et } s_X^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

$$ET(X) = \sqrt{V(X)} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}} \text{ et } s_X = \sqrt{s_X^2} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$$

- Sous Excel, la variance est donnée par la fonction `VAR.P()`, la variance modifiée est donnée par `VAR()`; l'écart type par `ECARTYPEP()` et l'écart type modifié par `ECARTYPE()`.

c) Calculez la covariance ($cov_{X,Y}$) et le coefficient de corrélation (ρ) entre le PIB et l'âge du président, de deux façons différentes :

- sans utiliser les formules Excel (`COVARIANCE()` et `COEFFICIENT.CORRELATION()`) dédiées à ce calcul (*c'est-à-dire* en refaisant le calcul).

- avec les formules Excel `COVARIANCE()` et `COEFFICIENT.CORRELATION()`.

Rappel :

$$cov(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n}$$

$$\rho_{X,Y} = \frac{cov(X, Y)}{\sqrt{V(X) \cdot V(Y)}}$$

d) Le statisticien suppose que l'âge du président « explique » le montant du PIB. Écrire l'équation que notre statisticien va essayer d'estimer avec la technique de la régression. (A partir de la formule $y = a \cdot x + b$, remplacez les variables x et y par les variables pertinentes)

e) Calculez la droite de régression (trouvez a et b)

Rappel :

$$a = \frac{cov(X, Y)}{V(X)}$$

$$b = \bar{Y} - a \cdot \bar{X}$$

f) Calculez pour chaque année le PIB que l'on peut prévoir à l'aide d'une telle régression. Calculez pour chaque année l'erreur de prévision.

- Analyse de la variance :

(Vous avez appris les notions d'analyse de la variance dans le cadre du cours sur l'ANOVA. On peut aussi utiliser ces concepts pour étudier la pertinence d'une droite de régression pour représenter les données.)

g) Calculez la somme des carrés des écarts à la moyenne $\sum_{i=1}^n (X_i - \bar{X})$ pour le PIB (abréviation SC_{Total}), le PIB prévu (ou $SC_{Modèle}$) et pour l'erreur de prévision (ou SC_{Erreur}).

(On pourra utiliser la fonction Var.P qu'on multipliera par l'effectif : $SC_{truc} = n \cdot V(truc)$).

Vérifiez que $SC_{Total} = SC_{Modèle} + SC_{Erreur}$.

- Calcul du R2 :

i) Calculez le R2, un indicateur très important pour juger de la qualité d'une régression : $R^2 = SC_{Modèle} / SC_{Total}$.

On appelle le R2 le % de variance expliquée. Pourquoi ?

- Significativité des paramètres :

La pente de la droite de régression est l'équivalent de la moyenne de la pente de tous les points. Le paramètre a est donc une sorte de moyenne. Pour savoir si la variable X explique bien Y, on se demande si la pente est bien significativement différente de 0. Pour cela, il faut connaître une sorte de variance de la pente.

h) On cherche à connaître la significativité des paramètres : La variance du paramètre a se calcule ainsi :

$$s_a^2 = \frac{SC_{Erreur}}{(n-2) \cdot SC_X} \quad \text{où X représente l'âge du président}$$

Calculez, à partir de ces formules la variance s_a^2 et l'écart type s_a .

i) Pour tester si le paramètre a est significativement différent de 0, on fait l'hypothèse

$$H_0 : a=0$$

Et on regarde si on peut la rejeter. Pour cela on utilise la statistique $T = |a| / s_a$ qui suit une loi de Student à n-2 degrés de liberté.

En utilisant la fonction « Loi.Student » (test bilatéral), déterminez la probabilité de nullité du paramètre α .

Concluez : la corrélation est-elle significative ?

j) Vérifiez avec Excel. Dans le menu *Outils*, Allez dans le menu *Macros complémentaires*, Cochez *Utilitaires d'analyse* (si la case n'est pas déjà cochée).

Une fois la case cochée, sélectionnez *Utilitaires d'analyse*, dans le menu *Outils*. Choisir *Régression linéaire*.

Procédez à la régression.

Retrouvez les paramètres que vous avez calculés.

II. Prédiction :

Avec votre grand modèle macroéconomique, prévoyez l'évolution du PIB jusqu'en 2006, en supposant que Chirac ne meurt pas.

Pour 2007, envisagez soit l'élection de Giscard d'Estaing (né en 1926), soit celle d'Olivier Besancenot (né en 1974).

Que se passe-t-il dans ce modèle si Besancenot est élu ?

III. Réfutation :

1) Sur quoi repose selon-vous cette illusion statistique ?
Pourquoi peut-on dire qu'elle est due aux « années Mitterrand » ?

2) Procédez à une nouvelle régression à deux variables explicatives :

Dans le menu *Outils*, sélectionnez *Utilitaires d'analyse*. Choisir *Régression linéaire*.

Choisir le PIB comme variable explicative et l'âge du président et une variable représentant l'écoulement du temps (par exemple l'année).

Dans la feuille de résultats :

- a) Trouvez le R2 (on prendra le premier)
- b) Trouvez la valeur des paramètres a_1 et a_2 .
- c) Trouvez la valeur des écarts types des paramètres a_1 et a_2 .
- d) Trouvez la probabilité de nullité des paramètres a_1 et a_2 .
- e) L'âge du président a-t-il toujours un impact significatif sur le PIB ?

CONCLURE