

## Exercice 1

Soient deux variables quantitatives  $Y$  et  $X$ . On cherche à étudier la relation entre les deux variables à partir des données d'un échantillon de 15 observations tel que présenté dans le tableau ci-dessous.

$Y$	20	15	19	14	13	21	29	17	22	28	24	19	30	18	16
$X$	293	335	312	350	408	252	223	328	211	140	190	318	136	305	369

1. L'observation du tableau montre qu'il existe une corrélation positive entre les variables  $Y$  et  $X$ .
  - a. Vrai ou Faux ?
  - b. Écrivez la formule du coefficient de corrélation entre les deux variables.
  - c. Donnez la valeur du coefficient de corrélation entre les deux variables.
2. L'estimation de la fonction de régression linéaire de  $Y$  en fonction de  $X$  par la méthode des moindres carrés ordinaires (MCO) a permis d'obtenir le modèle estimé suivant :

$$\hat{Y} = 568.95 - 14.30X \quad (1)$$

- a. Ce résultat est-il juste ? Répondez par oui ou non
  - b. Si oui commentez le résultat. Si non, écrivez le bon modèle estimé
  - c. Écrivez la valeur de la statistique de Student pour vérifier si  $X$  est significatif ou pas.
  - d. Peut-on déduire qu'il existe une relation significative entre  $Y$  et  $X$  ? Répondez par vrai ou faux
  - e. Pourquoi ?
3. La théorie stipule que l'estimation par un modèle de régression linéaire simple n'est pas symétrique comme le calcul des corrélations. Pour vérifier cette théorie, on vous demande :
    - a. D'écrire le modèle :  $\hat{X} = a - bY$  avec les coefficients estimés
    - b. La théorie est-elle vérifiée ? Répondez par Oui ou Non
  4. Écrivez la formule de l'estimateur de la variance.
  5. L'estimateur de la variance dans le modèle (1) est égal à 0,006.
    - a. Vrai ou faux ?
    - b. Si cette valeur est fautive, donnez la bonne valeur.
  6. Le coefficient de détermination du modèle (1) estimé est égal à :  $R^2 = 0,07$ .
    - a. Cette valeur est-elle juste ? Oui ou Non
    - b. Si oui, commentez. Si Non, donnez la bonne réponse et commentez.

## Solution 1

Soient deux variables quantitatives  $Y$  et  $X$ . On cherche à étudier la relation entre les deux variables à partir des données d'un échantillon de 15 observations tel que présenté dans le tableau ci-dessous.

$Y$	20	15	19	14	13	21	29	17	22	28	24	19	30	18	16
$X$	293	335	312	350	408	252	223	328	211	140	190	318	136	305	369

1. L'observation du tableau montre qu'il existe une corrélation positive entre les variables  $Y$  et  $X$ .

- a. Faux
  - b. -0.9
2. L'estimation de la fonction de régression linéaire de  $Y$  en fonction de  $X$  par la méthode des moindres carrés ordinaires (MCO) a permis d'obtenir le modèle estimé suivant :
- $$\hat{Y} = 568.95 - 14.30X \quad (1)$$
- a. Faux
  - b.  $\hat{Y} = 37,31 - 0,06X$
  - c.  $t = -9.49$
  - d. Vrai
  - e. Hypothèse et règle de décision
3. La théorie stipule que l'estimation par un modèle de régression linéaire simple n'est pas symétrique comme le calcul des corrélations. Pour vérifier cette théorie, on vous demande :
- a.  $\hat{X} = 568.95 - 4.30Y$
  - b. Oui
4. Écrivez la formule de l'estimateur de la variance
- $$\widehat{\sigma^2} = \frac{\sum_{i=1}^n e_i^2}{n - 2}$$
5. L'estimateur de la variance dans le modèle (1) est égal à 0,006.
- a. Vrai
  - b. Bonne valeur.
6. Le coefficient de détermination du modèle (1) estimé est égal à :  $R^2 = 0,07$ .
- a. Non
  - b. 0,87

## Exercice 2

On cherche à étudier la relation entre les salaires des parents et leurs nombres d'enfants à partir d'un modèle de régression linéaire. Pour ce faire, des données ont été collectées sur un échantillon de 12 individus. Les données obtenues sont inscrites dans le tableau ci-dessous.

Salaire ( $Y$ )	510	590	900	900	1200	1420	1500	2000	600	850	1300	2200
Nombre d'enfants ( $X$ )	4	3	3	2	2	1	5	0	5	6	7	8

1. L'estimation de la fonction de régression linéaire des salaires en fonction du nombre d'enfant par la méthode des moindres carrés ordinaires (MCO) a permis d'obtenir le modèle estimé suivant :

$$\hat{Y} = 110,89 + 12,37X$$

- a. Répondez par vrai ou faux : ce résultat est-il juste ?
- b. Si oui commentez le résultat. Si non, écrivez le bon modèle estimé et commentez.

2. La propriété selon laquelle la moyenne des résidus est nulle, n'est pas vérifiée dans ce cas. Vrai ou faux ?
3. Écrivez la formule de l'estimateur de la variance.
4. L'estimateur de la variance est égal à 43085.56. Vrai ou faux ? Si cette valeur est fausse, donnez la bonne valeur.
5. A partir du modèle estimé que vous avez retenu à la question 1), peut-on dire que la variable Nombre d'enfants ( $X$ ) est significatif ? Justifiez votre réponse.
6. Répondez par vrai ou faux : l'intervalle de confiance du coefficient estimé pour la variable  $X$  est donnée par :  $I = [-143,9041 \quad 170,6553]$ .
7. Le coefficient de détermination du modèle estimé est égal à :  $R^2 = 0,05$ .
  - a. Cette valeur est-elle juste ? Vrai ou faux
  - b. Si vrai, commentez. Si faux, donnez la bonne réponse et commentez.
8. On voudrait tester la significativité globale du modèle.
  - a. Quelle est la statistique qui nous permet de faire ce test ?
  - b. Donnez sa valeur pour ce modèle estimé
  - c. Commentez.
9. Répondez par vrai ou faux : l'équation de l'analyse de la variance  $SCT = SCE + SCR$  est donnée par :  $3283542,67 = 11599,2657 + 3271943,4$ . Si faux, donnez la bonne réponse.

## Solution 2

On cherche à étudier la relation entre les salaires des parents et leurs nombres d'enfants à partir d'un modèle de régression linéaire. Pour ce faire, des données ont été collectées sur un échantillon de 12 individus. Les données obtenues sont inscrites dans le tableau ci-dessous.

Salaire ( $Y$ )	510	590	900	900	1200	1420	1500	2000	600	850	1300	2200
Nombre d'enfants ( $X$ )	4	3	3	2	2	1	5	0	5	6	7	8

1. L'estimation de la fonction de régression linéaire des salaires en fonction du nombre d'enfant par la méthode des moindres carrés ordinaires (MCO) a permis d'obtenir le modèle estimé suivant :
  - a. Faux
  - b.  $\hat{Y} = 1112.893 + 13.37563 X$
  - c. Commentaire

2. Faux, la moyenne des résidus est nulle.
3.  $\widehat{\sigma^2} = \frac{\sum_{i=1}^n e_i^2}{n-2}$
4. Vrai
5. Non, la variable n'est pas significative.
6. Vrai
7. Le coefficient de détermination du modèle estimé est égal à :  $R^2 = 0,05$ .
  - a. Faux
  - b.  $R^2 = 0,0036$ .
  - c. Commentaire
8. On voudrait tester la significativité globale du modèle.
  - a. La statistique de Fischer
  - b.  $F(1, 10) = 0,04$
  - c. Commentaire : Le modèle est globalement significatif
9.
  - a. Faux
  - b.  $3283691,67 = 11748,2657 + 3271943,4$

### Exercice 3

On désire analyser l'effet de la maternité précoce (ChildBirth) sur le nombre d'années d'éducation (Year\_Educ) des adolescentes au Bénin à partir de la base EDS-2018. Les résultats des estimations sont présentés dans le tableau ci-dessous.

Interprétez ces résultats et faites-en des recommandations de politiques publiques.

#### Linear regression

Year_Educ	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
ChildBirth	-1.193	.06	-19.89	0	-1.311	-1.075	***
Age	-.09	.004	-25.31	0	-.097	-.083	***
Habite en milieu rural	-1.808	.068	-26.53	0	-1.942	-1.675	***
Constant	6.483	.138	46.86	0	6.212	6.754	***
R-squared		0.120	Nombre d'observations			11756.000	
F-test		550.276	Prob > F			0.000	

\*\*\*  $p < .01$ , \*\*  $p < .05$ , \*  $p < .1$

### Solution 3

- Nombre d'observation
- Significativité de chaque variable et interprétation
- R-square
- Significativité globale
- Conclusion