1- Soit a , b , c , d , e et f des constantes, et X et Y des variables aléatoires (v.a.). Montrez que

a) $var(a+bX) = b^2 var(X)$

b) $var(a+bX+cY) = b^2 var(X) + c^2 \operatorname{var}(Y) + 2bc \operatorname{cov}(X,Y)$

c) $\operatorname{cov}(a+bX+cY, d+eX+fY) = be \operatorname{var}(X) + cf \operatorname{var}(Y) + (bf+ce)\operatorname{cov}(X,Y)$

a. Let $Y = a + bX$, then $E(Y) = E(a + bX) = a + bE(X)$. Hence,

$$\begin{aligned} var(Y) &= E[Y - E(Y)]^2 = E[a + bX - a - bE(X)]^2 = E[b(X - E(X))]^2 \\ &= b^2 E[X - E(X)]^2 = b^2 \, var(X). \end{aligned}$$

Only the multiplicative constant b matters for the variance, not the additive constant a.

b. Let $Z = a + bX + cY$, then $E(Z) = a + bE(X) + cE(Y)$ and

$$\begin{aligned} var(Z) &= E[Z - E(Z)]^2 = E[a + bX + cY - a - bE(X) - cE(Y)]^2 \\ &= E[b(X - E(X)) + c(Y - E(Y))]^2 \\ &= b^2 E[X-E(X)]^2 + c^2 E[Y-E(Y)]^2 + 2bc \, E[X-E(X)][Y-E(Y)] \\ &= b^2 var(X) + c^2 var(Y) + 2bc \, cov(X, Y). \end{aligned}$$

c. Let $Z = a+bX+cY$, and $W = d+eX+fY$, then $E(Z) = a+bE(X)+cE(Y)$

$E(W) = d + eE(X) + fE(Y)$

and

$$\begin{aligned} cov(Z, W) &= E[Z - E(Z)][W - E(W)] \\ &= E[b(X-E(X))+c(Y-E(Y))][e(X-E(X))+f(Y-E(Y))] \\ &= be \, var(X) + cf \, var(Y) + (bf + ce) \, cov(X, Y). \end{aligned}$$

2- Indépendance et corrélations

a) Montrez que si les v.a. X et Y sont indépendantes, alors $E(XY) = E(X)E(Y) = \mu_x \mu_y$ avec $\mu_x = E(X)$ et $\mu_y = E(Y)$. Ainsi montrez aussi que $\text{cov}(X,Y) = E(X - \mu_x)(Y - \mu_y) = 0$

b) Montrez que si $Y = a + bX$, avec a et b des constantes, alors $\rho_{xy} = 1$ si $b > 0$ et $\rho_{xy} = -1$ si $b < 0$.

a. Assume that X and Y are continuous random variables. The proof is similar if X and Y are discrete random variables and is left to the reader. If X and Y are independent, then $f(x, y) = f_1(x)f_2(y)$ where $f_1(x)$ is the marginal probability density function (p.d.f.) of X and $f_2(y)$ is the marginal p.d.f. of Y. In this case,

$$\begin{aligned} E(XY) &= \iint xyf(x, y)dxdy = \iint xyf_1(x)f_2(y)dxdy \\ &= (\int xf_1(x)dx)(\int yf_2(y)dy) = E(X)E(Y) \end{aligned}$$

Hence

cov(X,Y)=E[[X- E(X)] [Y- E(Y)]] = E(XY) – E[X E(Y)] – E[E(X) Y] + E[E(X)E(Y)]

= E(XY) – E[X E(Y)] – E[E(X) Y] + E[E(X)E(Y)]

= E(XY) – E[X] E(Y) – E(X) E[Y] + E(X)E(Y)

We can eliminate the last two terms

= E(XY) – E[X] E(Y)

And if we have independence E(XY) = E[X] E(Y)

Thus cov(X,Y) = E[X] E(Y) – E[X] E(Y) = 0

b. If $Y = a + bX$, then $E(Y) = a + bE(X)$ and $\text{cov}(X, Y) = E[X - E(X)][Y - E(Y)] = E[X - E(X)][a + bX - a - bE(X)] = b\,\text{var}(X)$ which takes the sign of b since var(X) is always positive. Hence,

$$\text{corr}(X, y) = \rho_{xy} = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} = \frac{b\,\text{var}(X)}{\sqrt{\text{var}(X)\text{var}(Y)}}$$

but $\text{var}(Y) = b^2\,\text{var}(X)$ from problem 1a. Hence, $\rho_{XY} = \frac{b\,\text{var}(X)}{\sqrt{b^2(\text{var}(X))^2}} = \pm 1$ depending on the sign of b.

3- Soit $X = -2, -1, 0, 1, 2$ avec des probabilités données par $P(X = x) = 1 / 5 = 0.2$. De plus, assumez une relation quadratique parfaite entre X et Y, soit $Y = X^2$.

a) Montrez que $\mathrm{cov}(X, Y) = E(X^3) = 0$.

b) Déduisez que $\rho_{XY} = corr(X, Y) = 0$

Zero Covariance Does Not Necessarily Imply Independence. Let $X = -2, -1, 0, 1, 2$ with $\Pr[X = x] = 1/5$. Assume a perfect quadratic relationship between Y and X, namely $Y = X^2$. Show that $\mathrm{cov}(X, Y) = E(X^3) = 0$. Deduce that ρ_{XY} = correlation $(X, Y) = 0$. The simple correlation coefficient ρ_{XY} measures the strength of the *linear* relationship between X and Y. For this example, it is zero even though there is a perfect *nonlinear* relationship between X and Y. This is also an example of the fact that if $\rho_{XY} = 0$, then X and Y are not necessarily independent. $\rho_{xy} = 0$ is a necessary but not sufficient condition for X and Y to be independent. The converse, however, is true, i.e., if X and Y are independent, then $\rho_{XY} = 0$, see problem 2.

Zero Covariance Does Not Necessarily Imply Independence.

X	P(X)
−2	1/5
−1	1/5
0	1/5
1	1/5
2	1/5

$$E(X) = \sum_{X=-2}^{2} XP(X) = \frac{1}{5}[(-2) + (-1) + 0 + 1 + 2] = 0$$

$$E(X^2) = \sum_{X=-2}^{2} X^2P(X) = \frac{1}{5}[4 + 1 + 0 + 1 + 4] = 2$$

and $\mathrm{var}(X) = 2$. For $Y = X^2$, $E(Y) = E(X^2) = 2$ and

$$E(X^3) = \sum_{X=-2}^{2} X^3P(X) = \frac{1}{5}[(-2)^3 + (-1)^3 + 0 + 1^3 + 2^3] = 0$$

In fact, any odd moment of X is zero. Therefore,

$$E(YX) = E(X^2.X) = E(X^3) = 0$$

6- Pour la régression simple avec une constante $y_i = \beta_1 + \beta_2 x_{i2} + \varepsilon_i$ vérifiez que les propriétés de l'estimateur des MCO suivantes tiennent.

$\sum_{i=1}^{n} \hat{\varepsilon}_i = 0$, $\sum_{i=1}^{n} \hat{\varepsilon}_i x_{i2} = 0$, $\sum_{i=1}^{n} \hat{\varepsilon}_i \hat{y}_i = 0$ et $\sum_{i=1}^{n} \hat{y}_i = \sum_{i=1}^{n} y_i$

The equivalent properties in matrix notation

$\iota'\hat{\varepsilon} = 0$ $x_2'\hat{\varepsilon} = 0$ $\hat{y}'\hat{\varepsilon} = 0$ and $\iota'\hat{y} = \iota' y$

From the OLS defining equation and orthogonality condition we have

$X'\hat{\varepsilon} = X'(y - X\hat{\beta}) = 0$

In the case where $X = \begin{bmatrix} \iota & x_2 \end{bmatrix}$ we have

$$\begin{bmatrix} \iota & x_2 \end{bmatrix}'\hat{\varepsilon} = \begin{bmatrix} \iota' \\ x_2' \end{bmatrix}\hat{\varepsilon} = 0$$

So $\sum_{i=1}^{n} \hat{\varepsilon}_i = \iota'\hat{\varepsilon} = 0$ and $\sum_{i=1}^{n} x_{i2}\hat{\varepsilon}_i = x_2'\hat{\varepsilon} = 0$

With regards to $\sum_{i=1}^{n} \hat{\varepsilon}_i \hat{y}_i = 0$

$\sum_{i=1}^{n} \hat{\varepsilon}_i \hat{y}_i = \sum_{i=1}^{n} \hat{\varepsilon}_i (\hat{\beta}_1 + \hat{\beta}_2 x_{i2}) = \hat{\beta}_1 \sum_{i=1}^{n} \hat{\varepsilon}_i + \hat{\beta}_2 \sum_{i=1}^{n} \hat{\varepsilon}_i x_{i2}$ =0 since $\sum_{i=1}^{n} \hat{\varepsilon}_i = \iota'\hat{\varepsilon} = 0$ and $\sum_{i=1}^{n} x_{i2}\hat{\varepsilon}_i = x_2'\hat{\varepsilon} = 0$

Where $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_{i2}$

Using the fact that $\hat{\varepsilon}_i = y_i - \hat{y}_i$ we can sum both sides to get

$$\sum_{i=1}^{n} \hat{\varepsilon}_i = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \hat{y}_i$$

But since $\sum_{i=1}^{n} \hat{\varepsilon}_i = 0$,

then $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i$

Dividing both sides we also get

$$\frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{n} \hat{y}_i}{n}$$

7- Pour la régression avec une constante suivante $y_i = \beta_1 + \varepsilon_i$ avec $\varepsilon_i \sim i.i.d.(0,\sigma^2)$, vérifiez que l'estimateur des moindres carrés de β_1 est tel que $\hat{\beta}_1 = \overline{y}$, que $\mathrm{var}(\hat{\beta}_1) = \sigma^2 / n$ et que la somme des résidus au carré est donnée par $\sum_{i=1}^{n} \overline{\overline{y}}_i^2 = \sum_{i=1}^{n} (y_i - \overline{y})^2$ Il y avait un erratum avec le = sur le y qu'il faut ajouter à la question.

The OLS problem is given by

$$\min_{\beta_1 \in \mathbb{R}} \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_1)^2 = \sum_{i=1}^{n} (y_i - \beta_1)(y_i - \beta_1) = \sum_{i=1}^{n} (y_i y_i - y_i \beta_1 - \beta_1 y_i + \beta_1^2)$$

Minimizing with respect to β_1

The first order condition (FOC) is:

$$\frac{\partial \left(\sum_{i=1}^{n} (y_i^2 - 2 y_i \beta_1 + \beta_1^2) \right)}{\partial \beta_1} = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_1) = 0 \text{ once we equalize it to 0}$$

So this gives us the OLS defining equation $-2 \sum_{i=1}^{n} (y_i - \hat{\beta}_1) = 0$

Solving give us $\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i}{n} = \overline{y}$.

Averaging the model $y_i = \beta_1 + \varepsilon_i$ gives us

$$\underbrace{\frac{1}{n} \sum_{i=1}^{n} y_i}_{\equiv \overline{y}} = \frac{1}{n} \sum_{i=1}^{n} \beta_1 + \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i = \frac{1}{n}(n\beta_1) + \underbrace{\frac{1}{n} \sum_{i=1}^{n} \varepsilon_i}_{\equiv \overline{\varepsilon}} = \beta_1 + \overline{\varepsilon}$$

Hence from $\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i}{n} = \overline{y}$ we can write the above equation as $\hat{\beta}_1 = \beta_1 + \overline{\varepsilon}$

Taking expectations, we get

$E(\hat{\beta}_1) = \beta_1 + E(\overline{\varepsilon}) = \beta_1$ since $E(\overline{\varepsilon}) = E\left(\frac{1}{n} \sum_{i=1}^{n} \varepsilon_i \right) = 0$ from the fact that $\varepsilon_i \sim i.i.d.(0,\sigma^2)$

$$\mathrm{var}(\hat{\beta}_1) = (\hat{\beta}_1 - E(\hat{\beta}_1))^2 = (\hat{\beta}_1 - \beta_1)^2 = E(\overline{\varepsilon}^2)$$ from $\hat{\beta}_1 = \beta_1 + \overline{\varepsilon}$

$$= E\left(\left(\frac{\sum_{i=1}^{n} \varepsilon_i}{n}\right)^2\right) = \frac{1}{n^2} E\left(\left(\sum_{i=1}^{n} \varepsilon_i\right)^2\right) = \frac{1}{n^2} E\left(\left(\varepsilon_1 + \varepsilon_2 + ... + \varepsilon_n\right)^2\right)$$

$$= \frac{1}{n^2} E\Big(\left(\varepsilon_1 + \varepsilon_2 + ... + \varepsilon_n\right)\left(\varepsilon_1 + \varepsilon_2 + ... + \varepsilon_n\right)\Big)$$

$$= \frac{1}{n^2} E\Big(\left(\varepsilon_1\varepsilon_1 + \varepsilon_2\varepsilon_1 + ... + \varepsilon_n\varepsilon_1\right) + \left(\varepsilon_1\varepsilon_2 + \varepsilon_2\varepsilon_2 + ... + \varepsilon_n\varepsilon_2\right) + ... + \left(\varepsilon_1\varepsilon_n + \varepsilon_2\varepsilon_n + ... + \varepsilon_n\varepsilon_n\right)\Big)$$

$$= \frac{1}{n^2} E\Big(\left(\varepsilon_1\varepsilon_1 + \varepsilon_2\varepsilon_2 + ... + \varepsilon_n\varepsilon_n\right)\Big) \text{ since } \varepsilon_i\varepsilon_j = 0 \ \ \forall i \neq j$$ because of independance $\varepsilon_i \sim i.i.d.(0,\sigma^2)$

$$= \frac{1}{n^2} E\Big(\left(\varepsilon_1^2 + \varepsilon_2^2 + ... + \varepsilon_n^2\right)\Big) = \frac{1}{n^2}\left(\sigma^2 + \sigma^2 + ... + \sigma^2\right) = \frac{1}{n^2}\left(n\sigma^2\right) = \frac{\sigma^2}{n}$$ since $E(\varepsilon_i) = 0$

With the OLS matrix expression for the variance we would have obtained more easily
$\mathrm{var}(\hat{\beta}_1) = \sigma^2(X'X)^{-1} = \sigma^2(\iota'\iota)^{-1} = \sigma^2 / n$ since $X \equiv \iota$
This is much more easy on the eyes notation wise, and less time consuming.

The USS = $\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_1)^2 = \sum_{i=1}^{n} (y_i - \overline{y})^2 = \sum_{i=1}^{n} \overline{\overline{y}}_i^{\,2}$

8- Pour la régression sans constante $y_i = \beta_1 x_{i1} + \varepsilon_i$ avec $\varepsilon_i \sim i.i.d.(0,\sigma^2)$

a) Dérivez l'estimateur des MCO de β_1 et trouvez sa variance.

b) Quelles propriétés de l'estimateur des MCO décrites dans la question 6 tiennent toujours pour ce modèle.

a) $\min_{\beta_1 \in \mathbb{R}} \sum_{i=1}^{n} (y_i - \beta_1 x_{i1})^2 = \sum_{i=1}^{n} (y_i - \beta_1 x_{i1})(y_i - \beta_1 x_{i1}) = \sum_{i=1}^{n} (y_i y_i - y_i \beta_1 x_{i1} - \beta_1 x_{i1} y_i + \beta_1 x_{i1} x_{i1} \beta_1)$

with respect to β_1 the FOC gives us $-2\sum_{i=1}^{n} (y_i x_{i1} - \hat{\beta}_1 x_{i1} x_{i1}) = 0$ once we equalize it to 0

Solving give us $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{n} x_{i1} y_i}{\sum_{i=1}^{n} x_{i1} x_{i1}} = \dfrac{\sum_{i=1}^{n} x_{i1} y_{i1}}{\sum_{i=1}^{n} x_{i1}^2}$.

Substituting $y_i = \beta_1 x_{i1} + \varepsilon_i$ gives us $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{n} x_{i1}(\beta_1 x_{i1} + \varepsilon_i)}{\sum_{i=1}^{n} x_{i1}^2}$

Taking the expectation

$$E(\hat{\beta}_1) = E\left(\frac{\sum_{i=1}^{n} \beta_1 x_{i1} x_{i1}}{\sum_{i=1}^{n} x_{i1}^2}\right) + E\left(\frac{\sum_{i=1}^{n} x_{i1} \varepsilon_i}{\sum_{i=1}^{n} x_{i1}^2}\right) = \beta_1 \left(\frac{\sum_{i=1}^{n} x_{i1}^2}{\sum_{i=1}^{n} x_{i1}^2}\right) + \left(\frac{\sum_{i=1}^{n} x_{i1} E(\varepsilon_i)}{\sum_{i=1}^{n} x_{i1}^2}\right) = \beta_1 + 0 = \beta_1$$

since x_{i1} is non-stochastic and since $E\varepsilon_i = 0$ or since $E(X'\varepsilon) = E(x_1'\varepsilon) = E(\sum \varepsilon_i x_{i1}) = 0$

Thus $E(\hat{\beta}_1) = \beta_1 + 0 = \beta_1$

Also from the general OLS matrix formulation, the variance

$$\mathrm{var}(\hat{\beta}_1) = \sigma^2 (X'X)^{-1} = \sigma^2 (x_1' x_1)^{-1} = \frac{\sigma^2}{\left(\sum_{i=1}^{n} x_{i1}^2\right)}$$

b) Because we don't have a constant $\sum_{i=1}^{n}\hat{\varepsilon}_i \neq 0$ generally

* Prove this by a counter example

$y = \beta_1 x_1 + \varepsilon$

$\begin{bmatrix}6\\4\end{bmatrix} = \begin{bmatrix}3\\1\end{bmatrix}\beta_1 + \varepsilon$ estimated by ols would give us $\hat{\beta}_1 = \left(\begin{bmatrix}3 & 1\end{bmatrix}\begin{bmatrix}3\\1\end{bmatrix}\right)^{-1}\begin{bmatrix}3 & 1\end{bmatrix}\begin{bmatrix}6\\4\end{bmatrix} = \frac{1}{10}\cdot 22 = 11/5$

$\sum_{i=1}^{n}\hat{\varepsilon}_i = \sum_{i=1}^{n}(y_i - x_{i1}\hat{\beta}_1) = (6 - 3\cdot\frac{11}{5}) + (4 - 1\cdot\frac{11}{5}) = 1.2 \neq 0$ You could use any numbers that shows inequality.

* From the FOC orthogonality condition we have that $0 = X'\hat{\varepsilon} = x_1{}'\hat{\varepsilon} = \sum \hat{\varepsilon}_i x_{i1}$

* $\sum_{i=1}^{n}\hat{\varepsilon}_i \neq 0$ generally implies that $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n}(\hat{y}_i + \hat{\varepsilon}_i) = \sum_{i=1}^{n}\hat{y}_i + \underbrace{\sum_{i=1}^{n}\hat{\varepsilon}_i}_{\neq 0\, generaly}$

* $\sum_{i=1}^{n}\hat{\varepsilon}_i \underbrace{\hat{y}_i}_{x_{i1}\hat{\beta}_1} = \sum_{i=1}^{n}\hat{\varepsilon}_i x_{i1}\hat{\beta}_1 = \hat{\beta}_1 \underbrace{\sum_{i=1}^{n}\hat{\varepsilon}_i x_{i1}}_{=0} = 0$