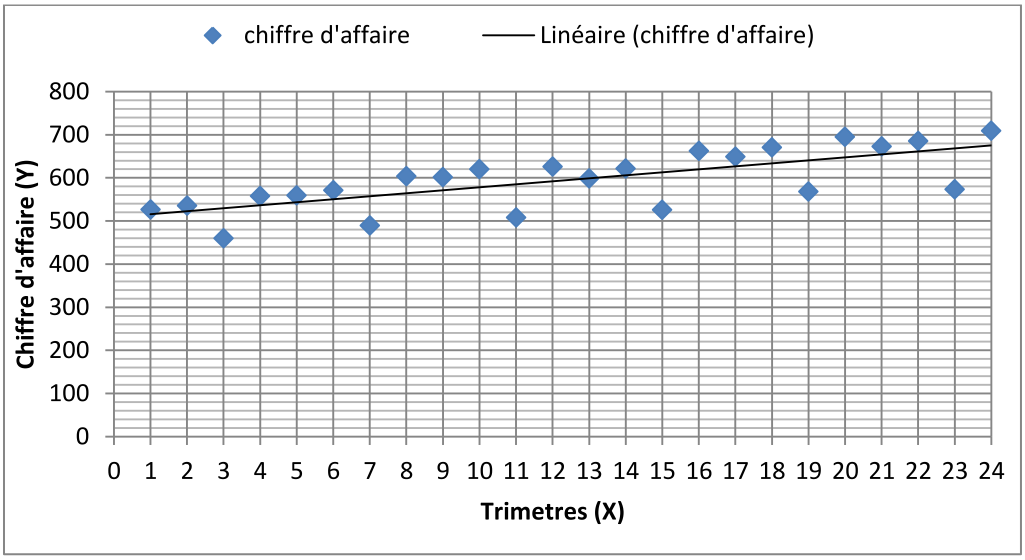
Solution de la série n°1

Exercice n°1

1. Représentation graphique des données



Conclusion :

- Lorsque Y augmente Y augmente également (X et Y évolue dans le même sens).
- Le nuage de point peut être approximé par une droite représentant la tendance générale de la courbe.

2. Le calcul du coefficient de corrélation linaire ($r_{x\,y}$)

X	Y	$X_t - \bar{X}$	$Y_t - \bar{Y}$	$(X_t - \bar{X}) * (Y_t - \bar{Y})$	$(X_t - \bar{X})^2$	$(Y_t - \bar{Y})^2$	$\hat{Y}$	$Y_t - \hat{Y}$	$(Y_t - \hat{Y})^2$	$(\hat{Y} - \bar{Y})^2$
1	526,76	-11,5	-68,6	788,9	132,25	4705,96	515,78	10,98	120,5	6332,97
2	235,08	-10,5	-60,28	632,91	110,25	3633,67	522,7	12,38	153,2	5279,47
3	459,68	-9,5	-135,68	1288,96	90,25	18409,06	529,62	-69,9	4886	4321,74
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
24	709,28	11,5	113,92	1310,12	132,25	12977,76	658,62	50,66	2268,4	4001,82
Somme				7957,04	1150	107342,19			52286,8	55056,07

$\bar{X} = \frac{1}{n} * \sum X_t = 12,5$; $\bar{Y} = \frac{1}{n} * \sum Y_t = 595,36$

$$r_{x\,y} = \frac{\sum (X_t - \bar{X}) * (Y_t - \bar{Y})}{\sqrt{\sum (X_t - \bar{X})^2} * \sqrt{\sum (Y_t - \bar{Y})^2}} = \frac{7957,04}{\sqrt{1150} * \sqrt{107342,19}} = 0,716$$

$r_{x\,y} > 0$ est proche de 1, donc il y a une relation linéaire forte entre x et y.

3. Estimation des paramètres $\hat{b}$ et $\hat{a}$

$$\hat{b} = \frac{\sum(X_t - \bar{X}) * (Y_t - \bar{Y})}{\sum(X_t - \bar{X})^2} = \frac{7957,04}{1150} = 6,92$$

$$\hat{a} = \bar{Y} - \hat{b}\bar{X} = 595,36 - (6,92) * (12,5) = 508,86$$

$\hat{Y} = 508,86 + 6,92\,X$ (1)

Pour tracer la droite de régression il suffit de remplacer les valeurs de X dans (1) et de calculer les valeurs de $\hat{Y}$ qui y correspond.

4. Test d'hypothèse $H_0 : b = 0$ contre $H_1 : b \neq 0$

D'abord il faut calculer la variance de l'erreur :

$$V(\delta_\varepsilon) = \delta_\varepsilon^2 = \frac{\sum \varepsilon_t^2}{n-2} = \frac{\sum (Y_t - \hat{Y})^2}{n-2} = \frac{52286,19}{22} = 2376,64$$

$V(b) = \delta_b^2 = \frac{\delta_\varepsilon^2}{\sum (X_t - \bar{X})^2} = \frac{2376,64}{1150} = 2,06$ donc $\delta_b = \sqrt{2,06} = 1,43$

Sous l'hypothèse H_0 :

$$Tc = \frac{|\hat{b} - 0|}{\delta_b} = \frac{|6,92 - 0|}{1,43} = 4,84$$

La statistique de student tablé $t_{n-2}^{\alpha/2} = t_{22}^{0,05/2} = 2,074$

Donc $Tc > t_{22}^{0,05/2}$ nous rejetons l'hypothèse H_0 et nous acceptons l'hypothèse H_1. Le paramètre b est significativement différent de 0.

L'intervalle de confiance pour le paramètre b est donné comme suit :

$$IC_{\hat{b}} : [\hat{b} - \left(\delta_b * t_{n-2}^{\frac{\alpha}{2}}\right); \hat{b} + \left(\delta_b * t_{n-2}^{\frac{\alpha}{2}}\right)]$$

$$IC_{\hat{b}} : [6,92 - (1,43 * 2,074); 6,92 + (1,43 * 2,074)]$$

$$IC_{\hat{b}} : [3,94; 9,87]$$

5. Test d'hypothèse $H_0 : a = 0$ contre $H_1 : a \neq 0$

La variance de la constante est calculée comme suit :

$$V(a) = \delta_a^2 = \delta_\varepsilon^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{(X_t - \bar{X})^2}\right) = 2376,64 \left(\frac{1}{24} + \frac{12,5^2}{1150}\right) = 421,94$$

donc $\delta_a = \sqrt{421,94} = 20,54$

Sous l'hypothèse H_0 :

$$Tc = \frac{|\hat{a} - 0|}{\delta_a} = \frac{|508,86 - 0|}{20,54} = 24,77$$

Donc $Tc > t_{22}^{0,05/2} = 2{,}074$ nous rejetons l'hypothèse H_0 et nous acceptons l'hypothèse H_1. Le paramètre a est significativement différent de 0.

L'intervalle de confiance pour le paramètre a :

$$IC_{\hat{a}}: [\hat{a} - \left(\delta_a * t_{n-2}^{\frac{\alpha}{2}}\right); \hat{a} + \left(\delta_a * t_{n-2}^{\frac{\alpha}{2}}\right)]$$

$$IC_{\hat{a}}: [508{,}86 - (20{,}54 * 2{,}074); 508{,}86 + (20{,}54 * 2{,}074)]$$

$$IC_{\hat{a}}: [466{,}26\ ;\ 551{,}46]$$

(Question supplémentaire : Tester la significativité globale du modèle)

Afin de tester la significativité globale du modèle nous appliquons le test de Ficher :

$H_0: a = b = 0$ contre $H_1: Il\ existe\ au\ moins\ un\ paramètre\ \neq 0$

$$Fc = \frac{\sum(\hat{Y}-\bar{Y})^2}{\sum(Y_t-\hat{Y})^2/_{n-2}} = \frac{55056{,}074}{52286{,}79/_{22}} = 23{,}16$$

La statistique de Ficher tablé : $F_{(1,n-2)}^{5\%} = F_{(1,22)}^{5\%} = 4{,}30$

$Fc > F_{(1,22)}^{5\%}$ nous rejetons l'hypothèse H_0 et nous acceptons l'hypothèse H_1. Le modèle est globalement significatif.

6. Les prévisions pour l'année 2017 :

L'équation estimé est égale à $\hat{Y} = 508{,}86 + 6{,}92\ X$

Afin de calculer les prévisions nous remplaçons dans l'équation estimé les valeurs de x (25, 26, 27, 28) qui correspond au 1er, 2ème, 3ème et 4ème trimestre de 2017.

$\hat{Y}_{25} = 508{,}86 + 6{,}92\ (25) = 681{,}86$

$\hat{Y}_{26} = 508{,}86 + 6{,}92\ (26) = 688{,}78$

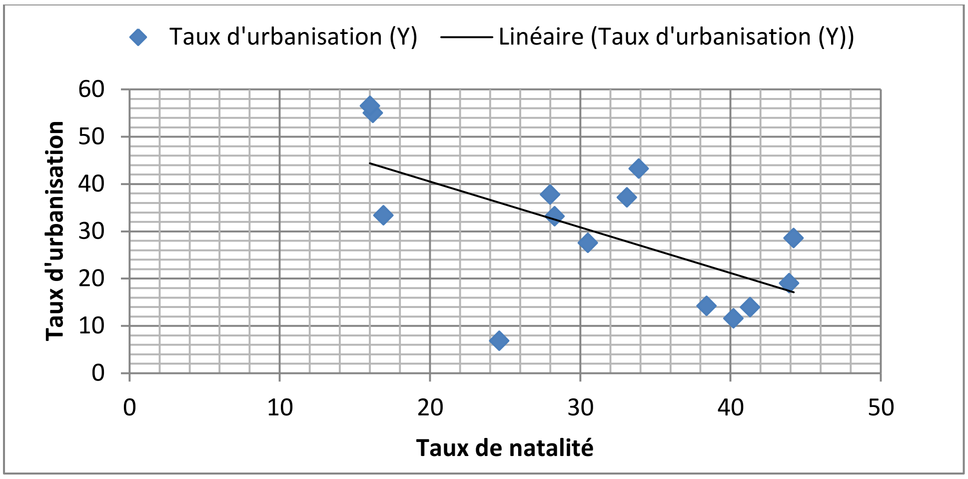
$\hat{Y}_{27} = 508{,}86 + 6{,}92\ (27) = 695{,}7$

$\hat{Y}_{28} = 508{,}86 + 6{,}92\ (28) = 702{,}62$

Donc $\hat{Y}_{2017} = 681{,}86 + 688{,}78 + 695{,}7 + 702{,}62 = 2768{,}96$

Exercice n°2

1. Représentation graphique



2. L'estimation des paramètres $\hat{a}$ et $\hat{b}$ du modèle :

Nous utilisons la même méthode de calcul que nous avons appliqué dans le tableau précédent (exercice 1) pour calculer les quantités suivantes :

$\sum(X_t - \bar{X}) * (Y_t - \bar{Y}) = -1256{,}819$

$\sum(X_t - \bar{X})^2 = 3150{,}96$

$\sum(Y_t - \bar{Y})^2 = 1299{,}15$

Afin de confirmer la relation linaire entre X et Y nous calculons le coefficient de corrélation :

$$r_{x\,y} = \frac{\sum(X_t - \bar{X}) * (Y_t - \bar{Y})}{\sqrt{\sum(X_t - \bar{X})^2} * \sqrt{\sum(Y_t - \bar{Y})^2}} = \frac{-1256{,}819}{\sqrt{3150{,}96} * \sqrt{1299{,}15}} = -0{,}621$$

Il y a une relation linéaire négative entre X et Y.

$\hat{b} = \frac{\sum(X_t - \bar{X}) * (Y_t - \bar{Y})}{\sum(X_t - \bar{X})^2} = \frac{-1256{,}819}{3150{,}96} = -0{,}399$

$\hat{a} = \bar{Y} - \hat{b}\bar{X} = 42{,}99$

$\hat{Y} = 42{,}99 - 0{,}399\, X$

Pour tracer la droite de régression il suffit de remplacer les valeurs de X dans l'équation estimée et de calculer les valeurs de $\hat{Y}$ qui y correspond.

3. Calcule de SCT, SCE, SCR et R^2

$SCT = \sum(Y_t - \bar{Y})^2 = 1299{,}15$

$SCE = \sum(\hat{Y} - \bar{Y})^2 = 501{,}304$

$SCR = \sum(Y_t - \hat{Y})^2 = SCT - SCE = 797{,}84$

$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} = \frac{501,304}{1299,15} = 0,386$

Le coefficient de détermination R^2 représente le degré d'ajustement des observations sur la droite de régression. Pour cette estimation, nous pouvons dire que 38,6% des variations de Y sont expliquées par les variations de X tandis que le reste (1-0,386=0,614 c'est-à-dire 61,4%) sont expliqués par des valeurs résiduelles.

4. Test d'hypothèse $H_0: b = 0$ contre $H_1: b \neq 0$

D'abord il faut calculer la variance de l'erreur :

$$V(\delta_\varepsilon) = \delta_\varepsilon^2 = \frac{\sum \varepsilon_t^2}{n-2} = \frac{\sum (Y_t - \hat{Y})^2}{n-2} = \frac{SCR}{n-2} = \frac{797,84}{12} = 66,48$$

$$V(b) = \delta_b^2 = \frac{\delta_\varepsilon^2}{\sum (X_t - \bar{X})^2} = \frac{66,48}{3150,96} = 0,021 \quad \text{donc } \delta_b = \sqrt{0,021} = 0,141$$

Sous l'hypothèse H_0 :

$$Tc = \frac{|\hat{b} - 0|}{\delta_b} = \frac{|0,399 - 0|}{0,141} = 2,746$$

La statistique de student tablé $t_{n-2}^{\alpha/2} = t_{12}^{0,05/2} = 2,179$

Donc $Tc > t_{22}^{0,05/2}$ nous rejetons l'hypothèse H_0 et nous acceptons l'hypothèse H_1. Le paramètre b est significativement différent de 0.

L'intervalle de confiance pour le paramètre b est donné comme suit :

$$IC_{\hat{b}}: [\hat{b} - \left(\delta_b * t_{n-2}^{\frac{\alpha}{2}}\right); \hat{b} + \left(\delta_b * t_{n-2}^{\frac{\alpha}{2}}\right)]$$

$$IC_{\hat{b}}: [-0,399 - (0,141 * 2,179)\ ;\ -0,399 - (0,141 * 2,179)]$$

$$IC_{\hat{b}}: [-0,715\ ;\ -0,082]$$

5. Test d'hypothèse $H_0: a = 0$ contre $H_1: a \neq 0$

La variance de la constante est calculée comme suit :

$$V(a) = \delta_a^2 = \delta_\varepsilon^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{(X_t - \bar{X})^2}\right) = 66,48\ \left(\frac{1}{14} + \frac{29,79^2}{3150,96}\right) = 23,47$$

donc $\delta_a = \sqrt{23,47} = 4,845$

Sous l'hypothèse H_0 :

$$Tc = \frac{|\hat{a} - 0|}{\delta_a} = \frac{|42,99 - 0|}{4,845} = 8,872$$

Donc $Tc > t_{22}^{0,05/2} = 2,179$ nous rejetons l'hypothèse H_0 et nous acceptons l'hypothèse H_1. Le paramètre a est significativement différent de 0.

L'intervalle de confiance pour le paramètre a :

$IC_{\hat{a}}: [\hat{a} - \left(\delta_a * t_{n-2}^{\frac{\alpha}{2}}\right); \hat{a} + \left(\delta_a * t_{n-2}^{\frac{\alpha}{2}}\right)]$

$IC_{\hat{a}}: [42{,}99 - (4{,}845 * 2{,}179); 42{,}99 + (4{,}845 * 2{,}179)]$

$IC_{\hat{a}}: [32{,}433 ;\ 53{,}548]$

Exercice n°3

1. Estimation des paramètres $\hat{a}$ et $\hat{b}$

X	Y	$X_i * Y_i$	X_i^2
7 9 9 . . . 25	5 4 6 . . . 0	35 36 54 . . . 0	49 81 81 . . . 625
		253	2596

$\bar{X} = \frac{1}{n} * \sum X_t = 15$; $\bar{Y} = \frac{1}{n} * \sum Y_t = 2{,}4$

$$\hat{b} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^{n} x_i^2 - n\bar{X}^2} = \frac{253 - 10(15)(2{,}4)}{2596 - 10\ (15)^2} = \frac{-107}{346} = -0{,}309$$

$\hat{a} = \bar{Y} - \hat{b}\bar{X} = 7{,}039$

$\hat{Y} = 7{,}039 - 0{,}309\, X$

2. Intervalle de confiance pour b et a au seuil de signification de 99%

$\sum(X_t - \bar{X})^2 = 346$

$SCT = \sum(Y_t - \bar{Y})^2 = 42{,}4$

$SCE = \sum(\hat{Y} - \bar{Y})^2 = 33{,}09$

$SCR = \sum(Y_t - \hat{Y})^2 = SCT - SCE = 9{,}31$

La variance de l'erreur : $V(\delta_\varepsilon) = \delta_\varepsilon^2 = \frac{\sum(Y_t - \hat{Y})^2}{n-2} = \frac{SCR}{n-2} = \frac{9{,}31}{8} = 1{,}163$

$V(b) = \delta_b^2 = \frac{\delta_\varepsilon^2}{\sum(X_t - \bar{X})^2} = \frac{1{,}163}{346} = 0{,}00336$ donc $\delta_b = \sqrt{0{,}00336} = 0{,}058$

$IC_{\hat{b}}: [\hat{b} - \left(\delta_b * t_{n-2}^{\frac{\alpha}{2}}\right); \hat{b} + \left(\delta_b * t_{n-2}^{\frac{\alpha}{2}}\right)]$

$IC_b: [-0{,}309 - (0{,}053 * 3{,}355); -0{,}309 - (0{,}053 * 3{,}355)]$

$IC_{\hat{b}}$: $[-0{,}504\ ;\ -0{,}115]$

$V(a) = \delta_a^2 = \delta_\varepsilon^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{(X_t - \bar{X})^2}\right) = 1{,}163\ \left(\frac{1}{10} + \frac{15^2}{346}\right) = 0{,}872$; donc $\delta_a = \sqrt{0{,}872} = 0{,}934$

$IC_{\hat{a}}$: $\left[\hat{a} - \left(\delta_a * t_{n-2}^{\frac{\alpha}{2}}\right); \hat{a} + \left(\delta_a * t_{n-2}^{\frac{\alpha}{2}}\right)\right]$

$IC_{\hat{a}}$: $[7{,}039 - (0{,}934 * 3{,}355); 7{,}039 - (0{,}934 * 3{,}355)]$

$IC_{\hat{a}}$: $[3{,}903\ ;\ 10{,}174]$

Exercice n°4

1. Estimation du modèle

Nous estimons d'abord le modèle avec constante. $\hat{Y} = \hat{a} + \hat{b}\,X$

Nous calculons les paramètres ci-dessous de la même manière que nous les avons calculés dans les exercices précédents.

$\bar{X} = \frac{1}{n} * \sum X_t = 7{,}3267$; $\bar{Y} = \frac{1}{n} * \sum Y_t = 7{,}4977$

$\sum(X_t - \bar{X}) * (Y_t - \bar{Y}) = 5{,}2609$

$\sum(X_t - \bar{X})^2 = 5{,}3568$

$SCT = \sum(Y_t - \bar{Y})^2 = 7{,}53$

$SCE = \sum(\hat{Y} - \bar{Y})^2 = 5{,}167$

$SCR = \sum(Y_t - \hat{Y})^2 = SCT - SCE = 2{,}364$

Il y a une relation linéaire négative entre X et Y.

$\hat{b} = \frac{\sum(X_t - \bar{X}) * (Y_t - \bar{Y})}{\sum(X_t - \bar{X})^2} = 0{,}9820$

$\hat{a} = \bar{Y} - \hat{b}\bar{X} = 0{,}3021$

$\hat{Y} = 0{,}3021 + 0{,}9820\,X$

Nous testons la significativité des paramètres estimés b et a

$H_0: b = 0$ contre $H_1: b \neq 0$

$V(\delta_\varepsilon) = \delta_\varepsilon^2 = \frac{\sum(Y_t - \hat{Y})^2}{n-2} = \frac{SCR}{n-2} = 0{,}084$

$V(b) = \delta_b^2 = \frac{\delta_\varepsilon^2}{\sum(X_t - \bar{X})^2} = 0{,}0158$ donc $\delta_b = \sqrt{0{,}0158} = 0{,}126$

Sous l'hypothèse H_0 :

$Tc = \frac{|\hat{b}-0|}{\delta_b} = \frac{|0,9820-0|}{0,126} = 7,793$

La statistique de student tablé $t_{n-2}^{\alpha/2} = t_{28}^{0,05/2} = 2,048$

Donc $Tc > t_{22}^{0,05/2}$ nous rejetons l'hypothèse H_0 et nous acceptons l'hypothèse H_1. Le paramètre b est significativement différent de 0.

$V(a) = \delta_a^2 = \delta_\varepsilon^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{(X_t-\bar{X})^2}\right) = 0,848$

donc $\delta_a = \sqrt{0,848} = 0,921$

Sous l'hypothèse H_0 :

$Tc = \frac{|\hat{a}-0|}{\delta_a} = \frac{|0,3021-0|}{0,921} = 0,328$

Donc $Tc < t_{22}^{0,05/2} = 2,048$ nous acceptons l'hypothèse H_0. Le paramètre a n'est pas significatif.

Tandis que la constante a n'est pas significative donc le modèle à spécifier doit être sans constante.

Nous estimons donc le modèle : $Y_t = a\,X + \varepsilon_t$

$\sum_{i=1}^{n} x_i y_i = 1653,2481$

$\sum_{i=1}^{n} x_i^2 = 1615,7582$

$$\hat{a} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} = \frac{1653,2381}{1615,7582} = 1,023$$

6. Test d'hypothèse $H_0: a = 1$ contre $H_1: a > 1$:

$SCR = \sum(Y_t - \hat{Y})^2 = 2,3763$

$V(\delta_\varepsilon) = \delta_\varepsilon^2 = \frac{\sum(Y_t-\hat{Y})^2}{\boldsymbol{n-1}} = \frac{SCR}{\boldsymbol{n-1}} = \frac{2,3763}{29} = 0,0819$

$V(a) = \delta_a^2 = \frac{\delta_\varepsilon^2}{\sum x_i^2} = \frac{0,0819}{1615,7582} = 5,0635 * 10^{-5}$ donc $\delta_a = \sqrt{5,0635 * 10^{-5}} = 0,007115$

Sous l'hypothèse H_0 :

$Tc = \frac{|\hat{a}-1|}{\delta_a} = \frac{|1,023-1|}{0,007115} = 3,2857$

La statistique de student tablé $t_{n-2}^{\alpha/2} = t_{28}^{0,05/2} = 2,045$

Donc $Tc > t_{28}^{0,05/2}$ nous rejetons l'hypothèse H_0 et nous acceptons l'hypothèse H_1. L'homme a un salaire plus élevé que sa conjointe.

Exercice n°5

1. L'interprétation des paramètres $\hat{a}_0$ et $\hat{a}_1$

$\hat{a}_0$: Représente l'estimation du salaire moyen des femmes (lorsque l'ouvrier n'est pas un homme).

$\hat{a}_1$: Représente la différence entre le salaire moyen estimé des hommes et celui des femmes.

2. La différence de salaire entre les deux échantillons est-elle significativement différente de 0

Pour répondre à cette question, nous testons la significativité du paramètre a_1 (Test de Student).

$H_0: a_1 = 0$ contre $H_1: a_1 \neq 0$:

$Tc = \frac{|\hat{a}_1 - 0|}{\delta_{a_1}} = \frac{|2{,}12 - 1|}{0{,}36} = 5{,}89$

La statistique de student tablé $t_{n-2}^{\alpha/2} = t_{528}^{0{,}05/2} = 1{,}96$

$Tc > t_{528}^{0{,}05/2}$ nous rejetons l'hypothèse H_0 et nous acceptons l'hypothèse H_1. Le paramètre $a_1$est significativement différent de 0. Donc la différence de salaire dans les deux échantillons est significative.

3. Détermination de $\hat{b}_0$ et $\hat{b}_1$

$\hat{b}_0$: Représente l'estimation du salaire moyen des hommes : $\hat{b}_0 = 14{,}64$

Donc l'estimation du salaire moyen des femmes : $\hat{b}_1 = 12{,}52 - 14{,}64 = -2{,}12$

$\hat{Y} = 14{,}64 - 2{,}12M$