Université Abdelmalek Essaâdi
Faculté Polydisciplinaire de Tétouan
LEF Sc. éco. & Gestion
S6

# Régression Simple
## Exercices

1

## Exercice 1: (Mesure d'efficacité de la force de vente)

Au cours d'un mois donné, le représntant d'une société commercialisant du matériel de bureau a visité 56 entreprises réparties dans sept départements.

Le tableau suivant indique, département par département, le nombre de visites réalisées de même que les commandes enregistrées pendant la période correspondante mesurées en milliers de dirhams.

2

| Département *(i)* | Nombre de visites *(Xi)* | Commandes *(Yi)* |
|---|---|---|
| 1 | 2 | 23 |
| 2 | 3 | 27 |
| 3 | 5 | 28 |
| 4 | 9 | 39 |
| 5 | 10 | 39 |
| 6 | 12 | 45 |
| 7 | 15 | 51 |

3

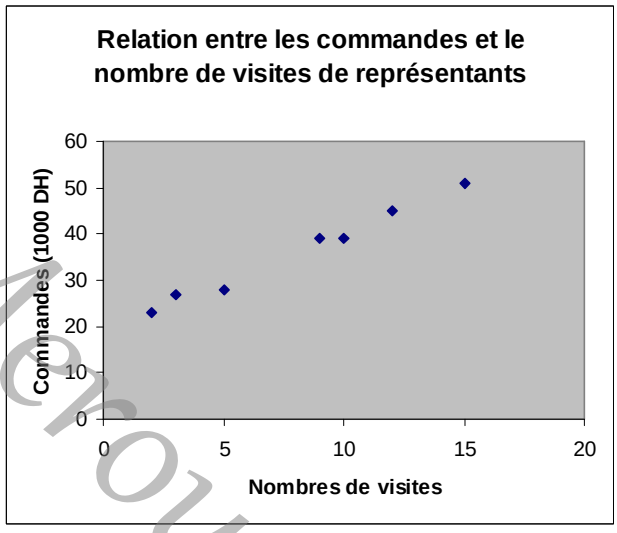
# Questions:

1. Représenter graphiquement le nuage des points et donner le modèle de régression y=ax+b par la méthode des moindres carrées. Interpréter le résultat.
2. Calculer les différents dispersion selon la loi des écarts.
3. Déterminer le coefficient de détermination et le coefficient de corrélation.
4. Représenter l'analyse de la variance et le test F
5. S'assurer à l'aide d'un test T de Student que a est significativement différente de zéro.
6. Déterminer l'intervalle de confiance du paramètre *a*.
7. Prévision de Y pour la valeur X=20 et l'intervalle de confiance de cette prévision.

4

# Solution 1:

**1.-**




5

| $i$ | $X_i$ | $Y_i$ | $X_iY_i$ | $X_i^2$ | $X_i-\overline{X}$ | $(X_i-\overline{X})^2$ | $Y_i-\overline{Y}$ | $(X_i-\overline{X})(Y_i-\overline{Y})$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 23 | 46 | 4 | -6 | 36 | -13 | 78 |
| 2 | 3 | 27 | 81 | 9 | -5 | 25 | -9 | 45 |
| 3 | 5 | 28 | 140 | 25 | -3 | 9 | -8 | 24 |
| 4 | 9 | 39 | 351 | 81 | 1 | 1 | 3 | 3 |
| 5 | 10 | 39 | 390 | 100 | 2 | 4 | 3 | 6 |
| 6 | 12 | 45 | 540 | 144 | 4 | 16 | 9 | 36 |
| 7 | 15 | 51 | 765 | 225 | 7 | 49 | 15 | 105 |
| Total: | 56 | 252 | 2313 | 588 | 0 | 140 | 0 | 297 |
| Moy. | 8 | 36 | 330,4 | 84 | 0 | 20 | 0 | 42,4 |

Les cinq premiers colonnes du tableau détaillent les calculs nécessaires pour obtenir $a$ qui s'élève ici à 2,12. En effet,

6

$$a = \begin{cases} \dfrac{\sum XY - n\overline{X}\overline{Y}}{\sum X^2 - n\overline{X}^2} = \dfrac{2313-(7)(8)(36)}{588-(7)(64)} = 2{,}12 \\ \dfrac{Cov(X,Y)}{Var(X)} = \dfrac{\sigma_{XY}}{\sigma_X^2} = \dfrac{42{,}4}{20} = 2{,}12 \end{cases}$$

$$b = \overline{Y} - a\overline{X} = 36 - (2{,}12)(8) = 19$$

Compte tenu de la valeur du paramètre *b*, égal à 19, l'équation de la droite qui représente le mieux les relations entre le nombre de visites *X* et le montant des commandes *Y* est:

$$Y = 2{,}12\ X + 19$$

7

Ce résultat peut être interprété de la façon suivante:

- en l'absence de visite, le montant des commandes d'un département s'élèverait à 19 000 DH;
- chaque visite d'un représentant amène une masse de commandes supplémentaires d'environ 2120 DH.

8

## 2.- Lois des écarts:

- La loi des écarts permet de relier l'erreur associée à l'hypothèse nulle et l'erreur associée à l'hypothèse "$Y$ dépend de $X$".
- L'erreur attachée à l'hypothèse nulle est mesurée par la *dispersion* totale des $Y_i$, c'est-à-dire par la somme des carrés des écarts des $Y_i$ par rapport à la moyenne $\overline{Y}$:

$$\text{Dispersion totale} = \sum \left(Y_i - \overline{Y}\right)^2$$

9

- Dans le cas étudié, l'erreur de l'hypothèse nulle s'élève à 638:

| Observation | $X_i$ | $Y_i$ | $\hat{Y}_i$ | $Y_i-\overline{Y}$ | $(Y_i-\overline{Y})^2$ | $\hat{Y}_i-\overline{Y}$ | $(\hat{Y}_i-\overline{Y})^2$ | $\hat{Y}_i-Y_i$ | $(\hat{Y}_i-Y_i)^2$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 23 | 23,27 | -13 | 169 | -12,73 | 162 | 0,27 | 0,07 |
| 2 | 3 | 27 | 25,39 | -9 | 81 | -10,61 | 112,57 | -1,61 | 2,59 |
| 3 | 5 | 28 | 29,64 | -8 | 64 | -6,36 | 40,45 | 1,64 | 2,69 |
| 4 | 9 | 39 | 38,12 | 3 | 9 | 2,12 | 4,49 | -0,88 | 0,77 |
| 5 | 10 | 39 | 40,24 | 3 | 9 | 4,24 | 17,98 | 1,24 | 1,54 |
| 6 | 12 | 45 | 44,49 | 9 | 81 | 8,49 | 72,08 | -0,51 | 0,26 |
| 7 | 15 | 51 | 50,85 | 15 | 225 | 14,85 | 220,52 | -0,51 | 0,02 |
| Total: | | | | | 638 | | 630,09 | | 7,94 |

10

- L'erreur attachée à la seconde hypothèse, ou encore dispersion résiduelle est donnée par $e^2$, somme des carrés des écarts entre les observations $Y_i$ et les valeurs estimées $\hat{Y}_i$ par le modèle:

$$\text{dispersion résiduelle} = \Sigma(\hat{Y}_i - Y_i)^2$$

- Dans le tableau précédent, il apparaît que l'erreur associée au modèle est très faible avec $e^2$=7,9.

11

- La différence entre la dispersion totale et la dispersion résiduelle correspond à la dispersion expliquée par le modèle de régression, compte tenu du fait que

$$(Y_i - \overline{Y})^2 = (\hat{Y}_i - \overline{Y})^2 + (\hat{Y}_i - Y_i)^2$$

On en tire la décomposition suivante:

$$\sum(Y_i - \overline{Y})^2 = \sum(\hat{Y}_i - \overline{Y})^2 + \sum(\hat{Y}_i - Y_i)^2$$

relation connue sous le nom de loi des écarts, nous pouvons écrire:

$$\text{dispersion expliquée} = \Sigma(\hat{Y}_i - \overline{Y})^2$$

Donc on a:

dispersion totale=dispersion expliquée+dispersion résiduelle.

Pour le problème considéré, la dispersion expliquée s'élève à 630,09.

12

## 3.- Coefficients de détermination et de corrélation:

Un premier indicateur de qualité de la représentation consiste à mettre en relation la dispersion expliquée par le modèle et la dispersion totale des données: le coefficient de détermination $R^2$ mesure le pouvoir explicatif du modèle en évaluant le pourcentage de l'information restituée par le modèle par rapport à la qualité d'information initiale:

$$R^2 = \frac{dispersion\ \exp liquée}{dispersion\ totale} = \frac{\sum \left(\hat{Y}_i - \bar{Y}\right)^2}{\sum \left(Y_i - \bar{Y}\right)^2}$$

13

- Avec les données de l'exemple précédent, $R^2=630/638=0,987$, il apparaît que le modèle $Y=2,12X+19$ restitue $98,7\%$ de l'information totale.
- Le coefficient de corrélation est $R$, racine carré du coefficient de détermination. C'est l'indicateur le plus couramment employé.
- On peut le calculer à l'aide de plusieurs formules différentes.

14

- En premier lieu, d'après la définition qui vient d'être donnée, nous avons:

$$R = \sqrt{\frac{\sum\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum\left(Y_i - \bar{Y}\right)^2}}$$

- On montre que R est obtenu également à l'aide des formules suivantes, où $\sigma_X$ et $\sigma_Y$ représentent les écarts-type respectives des $X_i$ et des Yi:

$$R = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \quad \text{et} \quad R = a\,\frac{\sigma_X}{\sigma_Y}$$

15

- Racine carée de $R^2$, c'est-à-dire d'un chiffre au plus égal à 1, $R$ a une valeur absolue également au plus égale à 1.
- $R$ est positif (covariance ou coefficient de régression $a$ positifs) ou négatif (cas inverse).
- Donc $-1 \leq R \leq 1$.
- Un $R$ très élevé en valeur absolue concrétise une relation étroite entre $X$ et $Y$, croissante si $R$ est positif et décroissante, si $R$ est négatif.

16

- Dans l'exemple étudié, $R$=0,994 ce qui indique une relation linéaire presque parfaite sur les données observées.

- Une valeur de $R$ faible en termes absolus caractérise une absence de relation linéaire entre $X$ et $Y$, mais pas nécessairement l'absence de liaison entre les variables.

17

4.- Test $F$:

- La valeur du coefficient de correlation est calculée à partir des données disponibles, les résultats de sept départements dans notre exercice.

- Un coefficient de correlation très élevé, mais obtenu sur peu de données est moins significatif qu'un coefficient plus faible, mais déterminée sur un grand nombre de données.

- A la limite, si nous n'avions que deux observations, $R$ serait égal à 1, mais aucune conclusion ne saurait en être déduite.

18

- Obtenu sur un échantillon de taille réduite, $R$ devrait être rectifié. La formule suivante est utilisée, ou $k$ est le nombre de variables explicatives et $n$ le nombre de données:

$R=1-$ Dispersion résiduelle $\frac{n-1}{n-k-1}$
Dispersion totale

Dans l'exemple, $k=1$ et $n$ le nombre d'obseravtions est $7$.

19

- Le test $F$ (analyse de la variance) permet d'intégrer la taille de l'échantillon dans l'appréciation de la qualité de la représentation:

$$F = \frac{\frac{\sum\left(\hat{Y}_i - \bar{Y}\right)^2}{k}}{\frac{\sum\left(\hat{Y}_i - Y_i\right)^2}{n-k-1}} = \frac{\text{Dispersion expliquée moyenne}}{\text{Dipersion résiduelle moyenne}}$$

• Dans notre exemple, $F=395$. Cette valeur doit être comparée à celle qui est lue dans une table de Fisher-Snédécor pour $k=1$ degré de liberté au numérateur et $n-k-1=7-1-1=5$ au dénominateur à un seuil de confiance $\alpha$.

20

- Pour $\alpha=0,01$, la valeur $F$ théorique lue dans la table est de $16,26$. Il n'y a ainsi qu'une chance sur $100$ de trouver un $F$ observé supérieur à 16,26 lorsque, dans la polpulation totale des observations possibles, aucune relation n'existe entre $X$ et $Y$.

- Nous sommes ici parfaitement en droit d'admettre la relation linéaire entre $X$ et $Y$, puisque le $F$ calculé est largement supérieur au $F$ théorique. (voir tableau suivant)

21

Analyse de la variance pour la régression (test $F$)

| | Degrés de liberté | Somme des carrés | Carrés moyens | $F$ |
|---|---|---|---|---|
| Régression | $k=1$ | 630,09 = $\Sigma(\hat{Y}_i-\overline{Y})^2$ | 630,09 | 396=630/1,59 |
| Erreur | $n-k-1=5$ | 7,94 = $\Sigma(\hat{Y}_i-Y_i)^2$ | 1,59 | |
| Total | $n-1=6$ | 638= $\sum(Y_i-\overline{Y})^2$ | | |
| $F_{0,01}$=16,26 | | | | |

22

5.- Validité des coefficients:

- Les tests précédents permettent d'avoir une idée de la validité de la régression dans son ensemble. Il importe de connaître également la validité des coefficients du modèle, c'est-à-dire de $a$ dans le cas de la régression linéaire simple.
- Cette validité est vérifiée par le biais du test $t$ et à travers le calcul d'intervalles de confiance.

23

- On définit l'erreur standard sur $a$ comme

$$S_a = \frac{S_{XY}}{\sqrt{\sum X_i^2 - n\overline{X}^2}}$$

Où $S_{XY}$ est l'écart-type des erreurs du modèle avec:

$$S_{XY} = \sqrt{\frac{\sum \left(Y_i - \hat{Y}_i\right)^2}{n-2}}$$

A partir des chiffres de notre exemple, il apparaît que:

$$S_{XY} = \sqrt{\frac{7,94}{5}} = \sqrt{1,59} = 1,26$$

et

$$S_a = 1,26/11,83 = 0,106$$

24

• Si l'on admet que les valeurs à estimer à partir de différents échantillons d'observations suivent une loi de Student d'écart-type $S_a$, nous pouvons évaluer la probabilité que la valeur $a$ soit différente de zéro.

$$t = \frac{a - 0}{S_a} = \frac{a}{S_a}$$

nous donne le nombre d'écarts-type qui séparent la valeur observée de 0.

$t$ mesure ainsi le degré de rareté, dans une population où la valeur de $a$ est 0, d'échantillons d'observations pour lesquels $a=a_0$ (ici $a_0$=2,12).

25

- Dans notre exemple, $t = 2{,}12 / 0{,}106 = 20$, ce qui compré au chiffre lu sur la table de Student pour $n\text{-}k\text{-}1=7\text{-}2=5$ degrés de liberté ($3{,}365$ avec un seuil de confiance de $0{,}01$) paraît très significatif.(Voir table)

6.- L'intervalle de confiance de $a$ est obtenu selon une procédure voisine. Si $t_\alpha$ est le nombre d'écarts-types correspondant au seuil de confiance $\alpha$, il y a une probabilité $(1\text{-}\alpha)$ que la valeur de $a$ soit comprise dans l'intervalle

$$[a\text{-}t_{\alpha/2}S_a;\ a+t_{\alpha/2}S_a].$$

26

- Il y a ainsi 99% de chances que la valeur de $a$ de notre problème soit comprise dans l'intervalle,

  [ 2,12- 4(0,106); 2,12+ 4(0,106) ],

  puisque $t_{0,005}$=4 pour 5 degrés de liberté.

27

7.- Il s'agirait de prévoir quelle serait l'importance des commandes pour un nombre de visites de représentants donné. Ceci peut être réalisé en donnant à $X$, dans le modèle, la valeur choisie.

- Ainsi, $X$=20 visites devraient amener, selon le modèle, 61 400 DH de commandes en moyenne, puisque 61,4=2,12(20)+19.

28

- En fait, il faut tenir compte de ce que le modèle a été construit à partir d'un échantillon de données et qu'il existe de toute façon un certain aléa sur les relations entre $X$ et $Y$.

- La prévision de $Y$ doit s'accompagner de la définition d'un intervalle de confiance: à un seuil de confiance $\alpha$, la valeur de $Y$ pour $X=X_0$ est comprise dans l'intervelle

$$\left[\hat{Y}_{X_0} - t_{\alpha/2} S_{XY} \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum X^2 - n\bar{X}^2}} ; \hat{Y}_{X_0} + t_{\alpha/2} S_{XY} \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum X^2 - n\bar{X}^2}}\right]$$

29

- Où, on le rappelle, $S_{XY}$ est l'écart-type des erreurs du modèle. L'intervalle de confiance est d'autant plus important que

- $S_{XY}$ est élevé;
- $n$ est faible;
- $X_i$ est éloigné de la moyenne.

Pour $X_0$=20 et $\alpha$=0,01,

$$Y=61,4\pm4(1,26)\sqrt{\frac{1}{7} + \frac{12^2}{140}}$$

Soit $Y$=61,4±5,9.

La régression linéaire simple nous a permis de présenter les aspects principaux des techniques de régression qui peuvent être utilisées dans l'élaboration de modèles de prévision.

30

# Exercice 2:

- On s'intéresse dans un secteur de production à la relation entre les bénéfices réalisés par les entreprises et le budget annuel qu'elles consacrent à la publicité. 15 observations ont été réalisées:

| Budget de publicité | 15 | 8 | 36 | 41 | 16 | 8 | 21 | 21 | 53 | 10 | 32 | 17 | 58 | 6 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bénéfices | 48 | 43 | 77 | 89 | 50 | 40 | 56 | 62 | 100 | 47 | 71 | 58 | 102 | 35 | 60 |

31

# Questions:

a) On veut établir une régression linéaire entre les deux variables, quelle doit être la variable endogène?

b) On admet l'existence d'une relation linéaire de la forme $y_i = ax_i + b + \varepsilon$ calculez les estimations des coefficients $a$ et $b$.

c) Calculer r l'estimation du coefficient de corrélation R.

d) Précisez l'équation d'analyse de la variance, calculer ses valeurs et en déduire le coefficient de détermination.

e) Sachant que $\hat{\sigma}_\varepsilon^2 = 10{,}155$, procédez à l'estimation des variances de $\hat{a}$ et de $\hat{b}$.

32

## Questions: (suite)

f) Déterminez au seuil de signification de 0,05 , un intervalle de confiance pour a, un intervalle de confiance pour b, et un intervalle de confiance pour $\hat{\sigma}_{\varepsilon}^{2}$.

g) Peut-on affirmer que les coefficients a et b sont significativement différents de 0 pour α=0,05?

h) Déterminez un intervalle de confiance pour le bénéfice prévisible relatif à une entreprise qui consacre un budget de 48 à son programme publicitaire. (α=0,05).

33

## Solution 2:

a) La variable endogène Y correspond aux bénéfices qui sont exprimés en fonction du budget de publicité X.

b) Voir tableau…

$$\hat{a} = \frac{\sum (X_i Y_i) - n\overline{X}\,\overline{Y}}{\sum (X_i^{\,2}) - n\overline{X}^{\,2}}$$

$$\hat{b} = \overline{Y} - \hat{a}\overline{X}$$

34

| $X_i$ | $Y_i$ | $X_i^2$ | $Y_i^2$ | $X_iY_i$ |
|---|---|---|---|---|
| 15 | 48 | 225 | 2304 | 720 |
| 8 | 43 | 64 | 1849 | 344 |
| 36 | 77 | 1296 | 5929 | 2772 |
| 41 | 89 | 1681 | 7921 | 3649 |
| 16 | 50 | 256 | 2500 | 800 |
| 8 | 40 | 64 | 1600 | 320 |
| 21 | 56 | 441 | 3136 | 1176 |
| 21 | 62 | 441 | 3844 | 1302 |
| 53 | 100 | 2809 | 10000 | 5300 |
| 10 | 47 | 100 | 2209 | 470 |
| 32 | 71 | 1024 | 5041 | 2272 |
| 17 | 58 | 289 | 3364 | 986 |
| 58 | 102 | 3364 | 10404 | 5916 |
| 6 | 35 | 36 | 1225 | 210 |
| 20 | 60 | 400 | 3600 | 1200 |
| 362 | 938 | 12490 | 64926 | 27437 |

35

$$n = 15$$

$$\overline{X} = \frac{362}{15} = 24{,}13 \Rightarrow \overline{X}^2 = 582{,}26$$

$$\overline{Y} = \frac{938}{15} = 62{,}53$$

$$\hat{a} = \frac{27437 - 15 \times 24{,}13 \times 62{,}53}{12490 - 15 \times 582{,}26} = 1{,}28$$

$$\hat{b} = 62{,}53 - 1{,}28 \times 24{,}13 = 31{,}67$$

$$\hat{Y} = 1{,}28X + 31{,}67$$

36

| $X_i$ | $Y_i$ | $X_i-\bar{X}$ | $(X_i-\bar{X})^2$ | $Y_i-\bar{Y}$ | $(Y_i-\bar{Y})^2$ | $\hat{Y}_i$ | $\hat{Y}_i-\bar{Y}$ | $(\hat{Y}_i-\bar{Y})^2$ | $\hat{Y}_i-Y_i$ | $(\hat{Y}_i-Y_i)^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 15 | 48 | -9,13 | 83,36 | -14,53 | 211,12 | 50,87 | -11,66 | 135,96 | 2,87 | 8,24 |
| 8 | 43 | -16,13 | 260,18 | -19,53 | 381,42 | 41,91 | -20,62 | 425,18 | -1,09 | 1,19 |
| 36 | 77 | 11,87 | 140,90 | 14,47 | 209,38 | 77,75 | 15,22 | 231,65 | 0,75 | 0,56 |
| 41 | 89 | 16,87 | 284,60 | 26,47 | 700,66 | 84,15 | 21,62 | 467,42 | -4,85 | 23,52 |
| 16 | 50 | -8,13 | 66,10 | -12,53 | 157,00 | 52,15 | -10,38 | 107,74 | 2,15 | 4,62 |
| 8 | 40 | -16,13 | 260,18 | -22,53 | 507,60 | 41,91 | -20,62 | 425,18 | 1,91 | 3,65 |
| 21 | 56 | -3,13 | 9,80 | -6,53 | 42,64 | 58,55 | -3,98 | 15,84 | 2,55 | 6,50 |
| 21 | 62 | -3,13 | 9,80 | -0,53 | 0,28 | 58,55 | -3,98 | 15,84 | -3,45 | 11,90 |
| 53 | 100 | 28,87 | 833,48 | 37,47 | 1404,00 | 99,51 | 36,98 | 1367,52 | -0,49 | 0,24 |
| 10 | 47 | -14,13 | 199,66 | -15,53 | 241,18 | 44,47 | -18,06 | 326,16 | -2,53 | 6,40 |
| 32 | 71 | 7,87 | 61,94 | 8,47 | 71,74 | 72,63 | 10,1 | 102,01 | 1,63 | 2,66 |
| 17 | 58 | -7,13 | 50,84 | -4,53 | 20,52 | 53,43 | -9,1 | 82,81 | -4,57 | 20,88 |
| 58 | 102 | 33,87 | 1147,18 | 39,47 | 1557,88 | 105,91 | 43,38 | 1881,82 | 3,91 | 15,29 |
| 6 | 35 | -18,13 | 328,70 | -27,53 | 757,90 | 39,35 | -23,18 | 537,31 | 4,35 | 18,92 |
| 20 | 60 | -4,13 | 17,06 | -2,53 | 6,40 | 57,27 | -5,26 | 27,67 | -2,73 | 7,45 |
| 362 | 938 | | 3753,73 | | 6269,73 | | | 6150,13 | | 132,01 |

37

c)

$$R=\frac{\sum(X_iY_i)-n\bar{X}\bar{Y}}{n\sigma_X\sigma_Y}$$

$$\sigma_X=\sqrt{\frac{1}{n}\sum(X_i-\bar{X})^2}=\sqrt{\frac{3753,73}{15}}=15,82$$

$$\sigma_Y=\sqrt{\frac{1}{n}\sum(Y_i-\bar{Y})^2}=\sqrt{\frac{6269,73}{15}}=20,44$$

$$R=0,989$$

38

d) Dispersion totale:

$$\sum \left(Y_i - \bar{Y}\right)^2 = 6269{,}73$$

Dispersion expliquée:

$$\sum \left(\hat{Y}_i - \bar{Y}\right)^2 = 6150{,}13$$

Dispersion résiduelle:

$$\sum \left(Y_i - \hat{Y}_i\right)^2 = 132{,}01$$

$$6269{,}73 = 6150{,}13 + 132{,}01$$

39

- Le coefficient de détermination est:

$$R^2 = \frac{6137{,}72}{6269{,}73} = 0{,}9789$$

- Ce coefficient est proche de 1, on peut en déduire que la variabilité expliquée par droite de régression est satisfaisante.

40

e) On a $\hat{\sigma}_{\varepsilon}^2 = 10{,}155$

Alors,

$$S_{\hat{a}}^2 = Var(\hat{a}) = \frac{\hat{\sigma}_{\varepsilon}^2}{\sum (X_i - \overline{X})^2} = 0{,}0027$$

et

$$Var(\hat{b}) = \hat{\sigma}_{\varepsilon}^2 \left[ \frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2} \right] = 2{,}2526$$

41

f) Intervalle de confiance pour $\hat{\sigma}_{\varepsilon}^2$

La variable $\frac{\sum \hat{\varepsilon}_i^2}{\sigma_{\varepsilon}^2} = (n-2)\frac{\hat{\sigma}_{\varepsilon}^2}{\sigma_{\varepsilon}^2}$ suit une loi $\chi^2$ à (n-2) degrés de liberté.

Donc, on part de $P\left( A < (n-2)\frac{\hat{\sigma}_{\varepsilon}^2}{\sigma_{\varepsilon}^2} < B \right) = 1 - \alpha$

L'intervalle de confiance pour $\hat{\sigma}_{\varepsilon}^2$ est alors:

$$I = \left[ (n-2)\frac{\hat{\sigma}_{\varepsilon}^2}{B}; (n-2)\frac{\hat{\sigma}_{\varepsilon}^2}{A} \right] = [5{,}336; 26{,}35]$$

42

- L'intervalle pour $a$: $\left[\hat{a} - t_{1-\alpha}\hat{\sigma}_{\hat{a}}; \hat{a} + t_{1-\alpha}\hat{\sigma}_{\hat{a}}\right]$ avec t lue sur la table de Student à *n-2=13* degré de liberté. (*t=2,16*).

$$I = [1{,}166;\ 1{,}391]$$

- Intervalle pour $b$: $\left[\hat{b} - t_{1-\alpha}\hat{\sigma}_{\hat{b}}; \hat{b} + t_{1-\alpha}\hat{\sigma}_{\hat{b}}\right]$

$$I = [28{,}432;\ 34{,}916]$$

43

g) Le t empirique de Student est donné par $\frac{\hat{a}}{\hat{\sigma}_{\hat{a}}}$ , on compare la valeur de ce rapport avec *t=2,16*.

On trouve qu'il est supérieur en valeur absolue à *2,16* pour les deux paramètres *a* et *b*.

Donc ces paramètres sont significativement différents de *0*. La variable exogène contribue bien à expliquer $Y$.

$$P\left(-2{,}16 < t_{(13)} < 2{,}16\right) = 0{,}95$$

44

h)

$$I_{(Y_0)} = \left[ (ax_0 + b) - t_{1-\alpha} S \sqrt{\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum x^2}\right)}; (ax_0 + b) + t_{1-\alpha} S \sqrt{\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum x^2}\right)} \right]$$

$$I_{(Y_{48})} = [(1{,}28 \times 48 + 31{,}67) - 2{,}16 \times 12{,}335; (1{,}28 \times 48 + 31{,}67) + 2{,}16 \times 12{,}335]$$

$$I_{(Y_{48})} = [85{,}45;\ 100{,}65]$$

45

# Références:


- **Exercice 1:**

  Jean-Pierre Vedrine, « Techniques Quantitatives de Gestion », Vuibert gestion.

- **Exercice 2:**

  Kamal Abdelillah, «Sondages et tests Statistiques » Fédala, 1998


46