

BTSA 2013 : corrigé de traitement de données, juin

Proposition de corrigé - non officielle. BTSA, toutes options, traitement de données, session 2013, Métropole et Réunion, code 2013-BTS103-NOR-ME-RE. Les exercices et annexes sont traités.

[Sujet officiel](#) · [Annales de ce concours ou examen](#)

Exercice 1 - Départements signalant le frelon asiatique

1. Avec les huit couples de l'énoncé, la calculatrice donne $r(X,Y) \approx 0,991$ et $r(X,\ln(Y)) \approx 0,913$. Les coefficients sont positifs : les deux séries augmentent avec le rang de l'année.

2. Le coefficient de détermination de la régression linéaire de Y en X est $r^2 \approx 0,982$. Environ 98,2 % de la dispersion de Y autour de sa moyenne est expliquée par cet ajustement dans les données observées. Ce chiffre n'est ni une probabilité que le modèle soit vrai ni une garantie de prévision.

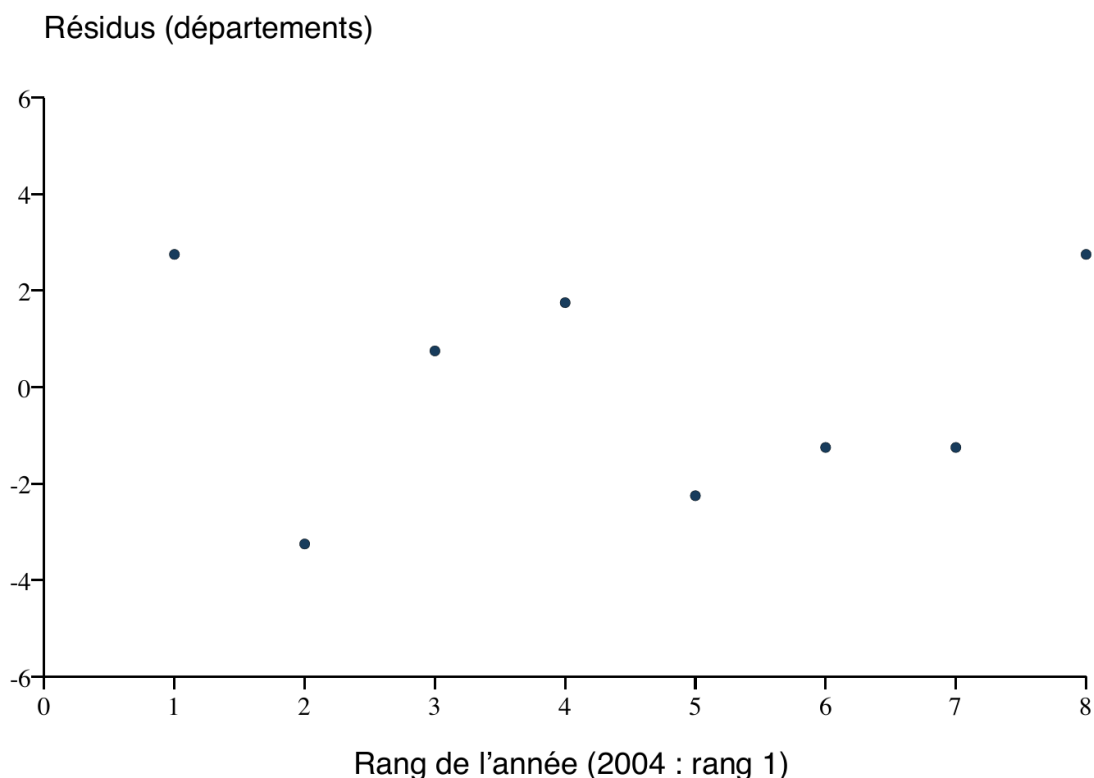
3. La corrélation directe est plus proche de 1 et le premier nuage est mieux aligné. On retient ici l'ajustement de **Y en X**, plutôt que celui du logarithme.

4. La droite des moindres carrés est $\hat{y} = 7x - 8,75$.

5. Les résidus $e = y - \hat{y}$ sont :

x	y	$\hat{y}$	e
1	1	-1,75	2,75
2	2	5,25	-3,25
3	13	12,25	0,75
4	21	19,25	1,75
5	24	26,25	-2,25
6	32	33,25	-1,25
7	39	40,25	-1,25
8	50	47,25	2,75

6. La représentation demandée place le rang en abscisse et le résidu en ordonnée :



Résidus de la régression du nombre de départements sur le rang de l'année

7. Les résidus oscillent autour de zéro sans tendance monotone nette. Ils restent compris entre $-3,25$ et $2,75$ départements : cela conforte l'ajustement sur les années observées. Toutefois, sa première valeur ajustée est négative, ce qui n'a pas de sens pour un dénombrement. Il faut donc éviter de prolonger ce modèle sans limite.

8. Puisque 2004 a le rang 1, **2013 a le rang 10**. Le modèle donne $7 \times 10 - 8,75 = \mathbf{61,250}$ départements, soit environ **61 départements**. C'est une estimation extrapolée selon le sujet, et non un bilan réel de présence du frelon en 2013.

Exercice 2 - Longueur de l'abdomen des abeilles

1. $U = (L - 7)/0,9$. Par linéarité, $E(U) = (E(L) - 7)/0,9 = \mathbf{0}$. Pour la variance, soustraire une constante ne change rien et diviser par 0,9 divise la variance par $0,9^2$: $V(U) = 0,9^2/0,9^2 = \mathbf{1}$. Comme L est normale, U est normale centrée réduite.

2. $P(L < 6) = \Phi((6 - 7)/0,9) = \Phi(-1,1111...) \approx \mathbf{0,133}$ au millième.

3. L'intervalle centré contenant 95 % des longueurs est $7 \pm 1,96 \times 0,9$, soit $[5,236 ; 8,764]$ mm, donc **[5,24 ; 8,76] mm** au centième. Il s'agit d'un intervalle de valeurs de L , pas d'un intervalle de confiance d'une moyenne estimée.

4. Garder les 20 % les plus grands revient à chercher le 80e centile : $\Phi((I_0 - 7)/0,9) = 0,80$. Le quantile normal vaut environ 0,84162, donc $I_0 = 7 + 0,9 \times 0,84162 \approx \mathbf{7,76}$ mm.

Exercice 3 - Masse des pots de miel

1. Moyenne et variance estimées

Les centres des quatre classes sont **117,5 ; 122,5 ; 127,5 ; 132,5 g**. La moyenne estimée est $(20 \times 117,5 + 100 \times 122,5 + 120 \times 127,5 + 60 \times 132,5)/300 = \mathbf{126,167 \text{ g}}$.

En remplaçant chaque masse par son centre de classe, la variance descriptive vaut $\sum n_i(c_i - \bar{x})^2/300 = \mathbf{18,222 \text{ g}^2}$. L'estimateur corrigé de la variance de la population utilise le diviseur 299 : $s^2 \approx \mathbf{18,283 \text{ g}^2}$.

La phrase « uniformément réparties dans chaque classe » mérite une précision. Si elle est prise au sens d'une véritable loi uniforme continue sur chacune des classes de largeur 5 g, il faut ajouter leur variance interne $5^2/12 = \mathbf{2,083 \text{ g}^2}$. La variance descriptive devient alors $20,306 \text{ g}^2$ et, avec la même correction d'effectif, l'estimation devient $\mathbf{20,373 \text{ g}^2}$. Le calcul par centres est l'approximation usuelle des données groupées ; il sous-estime cette dispersion interne. La moyenne est identique dans les deux traitements.

2. Moyenne d'échantillon et intervalle à 99 %

2a. Sous le modèle normal du sujet et pour des pots prélevés indépendamment, la moyenne $\bar{X}$ suit une normale de moyenne μ et de variance $\sigma^2/300$, soit d'écart type $\sigma/\sqrt{300}$. Les paramètres de la population restent inconnus ; les estimations précédentes servent au calcul de l'intervalle.

2b. Pour un grand échantillon, l'intervalle à 99 % est approximativement $\bar{x} \pm \mathbf{2,575829 \times s/\sqrt{300}}$.

Avec la convention par centres et la variance corrigée $18,283 \text{ g}^2$, il donne **[125,531 ; 126,803] g**. Si l'on tient explicitement compte de la dispersion uniforme à l'intérieur des classes, avec $20,373 \text{ g}^2$, on obtient **[125,495 ; 126,838] g**. Les deux résultats sont donnés avec leur convention, afin de ne pas faire passer des masses connues seulement par classes pour des mesures individuelles exactes.