

EXERCICE 1 : Le tableau ci-contre donne la distribution des revenus déclarés en euros par unité de consommation sur l'année 2011 pour la ville de Paris, celle d'Avignon et celle de Bordeaux. (source : INSEE : <http://www.insee.fr>). N est le nombre de ménages fiscaux.

1. Pour Avignon : La médiane est le cinquième décile, $Me = 14211$; pour évaluer les quartiles Q_1 et Q_3 par interpolation linéaire :

$$Q_1 \text{ est compris entre } D_2 \text{ et } D_3 : Q_1 = \frac{D_2 + D_3}{2} = \frac{5898 + 8612}{2} = 7255 ;$$

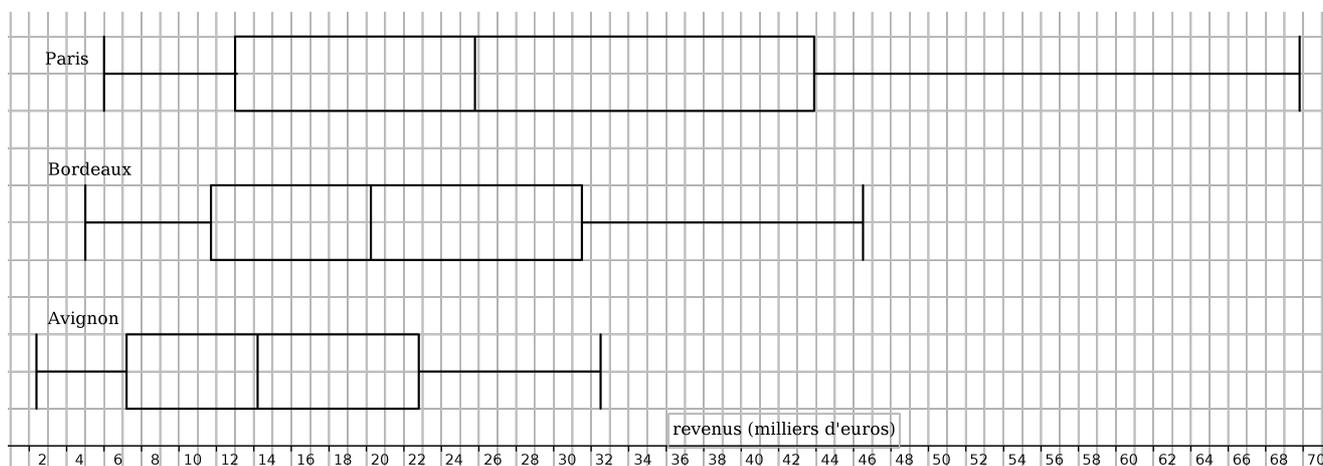
$$Q_3 \text{ est compris entre } D_7 \text{ et } D_8 : Q_3 = \frac{D_7 + D_8}{2} = \frac{20573 + 25078}{2} =$$

22825,5 ;

$$\text{Pour Bordeaux: } Me = 20251 ; Q_1 = \frac{D_2 + D_3}{2} = 11675,5 ; Q_3 = \frac{D_7 + D_8}{2} = 31491 ;$$

$$\text{Pour Paris: } Me = 25830 ; Q_1 = \frac{D_2 + D_3}{2} = 13156,5 ; Q_3 = \frac{D_7 + D_8}{2} = 43938,5.$$

2. Les trois diagrammes de Tukey :



3. L'écart interquartile ε pour chaque ville : Avignon : $\varepsilon = Q_3 - Q_1 = 15570,5$; Bordeaux: $\varepsilon = Q_3 - Q_1 = 19815,5$; Paris: $\varepsilon = Q_3 - Q_1 = 30782$;

4. Comparaison des résultats :

Les petits revenus sont assez proches dans les trois villes ; mais plus les revenus sont élevés, plus les écarts sont importants

5. Pour Avignon : par interpolation linéaire, $D_8 = 25078$ et $D_9 = 32539$, $32529 - 25078 = 7451$ représente 10 % ; $32539 - 30000 = 2539$ représente $2539/7451 * 10 = 3,4$ % ; donc il y a 13,4 % de foyers fiscaux à Avignon dont les revenus sont supérieurs à 30000 €.

Pour Bordeaux : par interpolation linéaire, $D_7 = 28396$ et $D_8 = 34586$, $34586 - 28396 = 6190$ représente 10 % ; $34586 - 30000 = 4586$ représente $4586/6190 * 10 = 7,4$ % ; donc il y a 27,4 % de foyers fiscaux à Bordeaux dont les revenus sont supérieurs à 30000 €.

Pour Paris: par interpolation linéaire, $D_5 = 25830$ et $D_6 = 31591$, $31591 - 25830 = 5761$ représente 10 % ; $31591 - 30000 = 1591$ représente $1591/5761 * 10 = 2,8$ % ; donc il y a 42,8 % de foyers fiscaux à Paris dont les revenus sont supérieurs à 30000 €.

6. Pour Avignon : par interpolation linéaire, $D_3 = 8612$ et $D_4 = 11395$, $11395 - 8612 = 2783$ représente 10 % ; $10000 - 8612 = 1388$ représente $1388/2783 * 10 = 5$ % ; donc il y a 35 % de foyers fiscaux à Avignon dont les revenus sont inférieurs à 10000 €.

Pour Bordeaux: par interpolation linéaire, $D_2 = 9783$ et $D_3 = 13568$, $13568 - 9783 = 3785$ représente 10 % ; $10000 - 9783 = 217$ représente $217/3785 * 10 = 0,6$ % ; donc il y a 20,6 % de foyers fiscaux à Bordeaux dont les revenus sont inférieurs à 10000 €.

Pour Paris: par interpolation linéaire, $D_1 = 6016$ et $D_2 = 10687$, $10687 - 6016 = 4671$ représente 10 %; $10000 - 6016 = 3984$ représente $3984/4671 * 10 = 8,5$ %; donc il y a 18,5 % de foyers fiscaux à Paris dont les revenus sont inférieurs à 10000 €.

EXERCICE 2 : Un entomologiste a fait des relevés sur la taille de 50 courtilières adultes.

33 35 36 36 37 37 37 38 38 38 39 39 39 39 40 40 40 40 40 41 41 41 41 41 41 41 42 42 42 42 42 42 43 43 43 43 44 44 44 44 45 45 45 46 46 47 47 48 48 50.

1. La moyenne $\bar{x} = 41,5$ et l'écart-type $\sigma = 3,58$.

2. L'intervalle $[\bar{x} - 2\sigma ; \bar{x} + 2\sigma] = [34,34 ; 48,66]$ contient toutes les valeurs sauf 33 et 50, soit 48 valeurs sur les 50, ce qui représente $\frac{48 \times 100}{50} = 96$ %.

3. La médiane est la moyenne entre la vingt-cinquième et la vingt-sixième valeur, soit $Me = 41$; Q_1 est la treizième valeur, soit $Q_1 = 39$; Q_3 est la trente-septième valeur, soit $Q_3 = 44$; et le premier décile D_1 est la cinquième valeur, $D_1 = 37$ et D_9 est la cinquante-cinquième valeur, $D_9 = 46$.

4. L'écart interquartile $\varepsilon = Q_3 - Q_1 = 44 - 39 = 5$.

5. Le statisticien Tukey qualifiait d'aberrantes les valeurs d'une série statistique qui se situait à l'extérieur de l'intervalle $[Q_1 - \frac{3}{2} \varepsilon ; Q_3 + \frac{3}{2} \varepsilon] = [31,5 ; 51,5]$. Toutes les valeurs sont dans cet intervalle, donc le pourcentage de valeurs aberrantes est 0 %.

EXERCICE 3

On complète le tableau des résultats à un examen universitaire dans deux filières avec les totaux :

1. Le tableau des fréquences de réussite pour chaque filière :

	Lettres	Sciences	Total
Étudiants reçus	50 %	80 %	75 %
Étudiantes reçues	50 %	80 %	56 %
Total	50 %	80 %	66,4 %

	Lettres	Sciences	Total
Étudiant présents	100	500	600
Étudiants reçus	50	400	450
Étudiantes présentes	400	100	500
Étudiantes reçues	200	80	280
Total présents	500	600	1100
reçus	250	480	730

3. Les deux jugements sont corrects, tout dépend du point de vue ;

a) « Dans chaque matière, les filles réussissent aussi bien que les garçons » ;

b) « L'université est sexiste car les étudiantes réussissent moins bien que les étudiants ».

C'est ce qu'on appelle le paradoxe de Simpson : Le **paradoxe de Simpson** ou **effet de Yule-Simpson** est un paradoxe statistique décrit par Edward Simpson en 1951 et George Udny Yule en 1903, dans lequel le succès de plusieurs groupes semble s'inverser lorsque les groupes sont combinés. Ce résultat qui paraît impossible est souvent rencontré dans la réalité, en particulier dans les sciences sociales et les statistiques médicales.

(voir http://fr.wikipedia.org/wiki/Paradoxe_de_Simpson)