

**UE Statistique**

**Semestre 1 de la session 2**

**Matière : Méthodes quantitatives descriptives**

**Sujet de : M. Robin**

**Durée: 2 heures - Aucun document autorisé**  
Seules sont autorisées les calculatrices prévues par le règlement de l'Université  
Les graphiques doivent être tracés avec soin et correctement légendés

**Sujet:** traitez les exercices ci-dessous. Prenez le temps de lire attentivement et complètement chaque énoncé avant de répondre.

**EXERCICE 1 (8 points):**

On considère deux échantillons d'entreprises industrielles françaises, collectés en 2000 et en 2004. L'échantillon de 2000 compte 3524 entreprises, et celui de 2004 en compte 4955. On s'intéresse au secteur d'activité principal de ces entreprises. Le Tableau 1 ci-dessous donne la répartition des entreprises de chaque échantillon par secteur d'activité, en présentant la distribution statistique des effectifs de la variable « secteur ».

**Tableau 1 : répartition des entreprises industrielles enquêtées par secteur**

Secteur	Echantillon de 2000	Echantillon de 2004
	<i>Effectif</i>	<i>Effectif</i>
Textile	475	648
Bois et Papier	386	706
Produits chimiques	374	451
Plastique et Caoutchouc	276	313
Minerais non ferreux	168	273
Métallurgie	607	699
Fabrication de machines	418	439
Composants électriques et électroniques	452	657
Véhicules	192	483
Divers	176	286
<b>TOTAL</b>	<b>3524</b>	<b>4955</b>

**Question 1 (4 points)**

- a. Pour chaque distribution, indiquez la modalité de la variable « secteur » qui présente l'effectif le plus élevé. S'agit-il du même secteur en 2000 et en 2004 ? (1 point)  
Pour chaque distribution, indiquez maintenant la modalité qui présente l'effectif le plus faible. S'agit-il du même secteur aux deux dates ? (1 point)
- b. Pour chaque distribution des effectifs présentée dans le Tableau 1, calculez la distribution statistique des fréquences correspondante (2 points).

**Question 2 (4 points).**

Tracez le diagramme de Pareto correspondant à chaque distribution du Tableau 1 (2 points). Rappelez l'énoncé de la « loi de Pareto », puis utilisez cette « loi » pour donner une interprétation précise de chaque graphique (2 points).

**EXERCICE 2 (5 points):**

Attention, cet exercice ne dépend pas du précédent, et peut donc être traité de manière complètement indépendante.

On considère deux échantillons d'entreprises industrielles françaises, collectés en 2000 et en 2004. Le Tableau 2 ci-dessous donne la moyenne et l'écart-type de trois variables statistiques :

- (i) les dépenses de R&D (en milliers d'euros) par employé
- (ii) le chiffre d'affaires (C.A., en milliers d'euros) par employé
- (iii) la valeur ajoutée (V.A., en milliers d'euros) par employé.

**Tableau 2 : moyenne et écart-type de trois variables statistiques pour chaque échantillon**

Variable	Echantillon de 2000		Echantillon de 2004	
	Moyenne	Ecart-type	Moyenne	Ecart-type
Dépenses de R&D par employé	7.4	20.2	7.7	19.9
Productivité du travail (C.A. par employé)	182.7	684.3	209.7	647.0
Productivité du travail (V.A. par employé)	54.01	100.7	55.1	101.9

**Question 1 (1 point)**

Donnez le nom et la formule de l'indicateur statistique qui permet, à l'aide des seules indications fournies dans le Tableau 2, de mesurer la dispersion de la distribution de chaque variable statistique dans un échantillon donné.

**Question 2 (4 points)**

Calculez la valeur de cet indicateur pour chacune des trois variables statistiques de chaque échantillon présenté dans le Tableau 2 (3 points). Indiquez ensuite, pour chaque variable, si la dispersion s'est accrue ou a diminué entre 2000 et 2004 (1 point).

**EXERCICE 3 (7 points):**

Un fabricant de pièces pour machines-outils souhaite développer un modèle statistique qui lui permette de déterminer le nombre d'heures de travail nécessaires pour produire une quantité donnée de pièces. Le fabricant observe pendant un an 14 niveaux de production différents (vingt pièces, trente pièces, etc.) et note le nombre d'heures de travail requis pour les atteindre. Il organise ensuite cet échantillon par ordre croissant de niveaux de production, en reportant à côté de chaque niveau de production le nombre d'heures de travail requis, comme dans le Tableau 3 ci-dessous.

Tableau 3 : niveaux de production atteints et nombre d'heures de travail requises

Niveau de production atteint (en nombre de pièces produites)	Heures de travail requises
20	50
20	55
30	73
30	67
40	87
40	95
50	108
50	112
60	128
60	135
70	148
70	160
80	170
80	162

**Question 1 (1 point)**

Sur un graphique, tracez le nuage de points représentant les heures de travail requises en fonction du niveau de production atteint.

**Question 2 (6 points)**

- En supposant que la relation suggérée par le graphique soit linéaire, calculez les paramètres  $b_0$  et  $b_1$  du modèle de régression simple  $Y = b_0 + b_1 X + \varepsilon$ , qui relie le nombre d'heures de travail nécessaires  $Y$  au niveau de production atteint  $X$ . Expliquez bien votre raisonnement, en précisant les étapes intermédiaires de vos calculs ainsi que les formules utilisées (4 points).
- D'après le modèle estimé, combien d'heures de travail faudrait-il pour atteindre un niveau de production de 45 pièces ? (1 point)
- Expliquez pourquoi il ne serait pas pertinent de chercher à prédire le nombre d'heures nécessaires pour produire 100 pièces (1 point).