

## TD 2 : Statistiques bivariées

### 1 Décomposition de la variance

On considère un caractère quantitatif  $X$  et un caractère qualitatif  $Y$  à  $p$  modalités.

On notera :

- $n_i$  l'effectif du groupe  $i$  (individus tels que  $y_k = i$ ),
- $n$  l'effectif total,
- $\bar{x}_i$  la moyenne du caractère  $X$  calculée sur le groupe  $i$  (qu'on appellera moyenne conditionnelle du groupe  $i$  pour le caractère  $X$ ),
- $\bar{x}$  la moyenne du caractère  $X$ .
- $\sigma_i^2(X)$  la variance du caractère  $X$  calculée sur le groupe  $i$  (qu'on appellera variance conditionnelle du groupe  $i$  pour le caractère  $X$ ),
- $\sigma^2(X)$  la variance du caractère  $X$ .

1. Exprimer  $n$  en fonction des  $n_i$ .
2. Rappeler l'expression de  $\bar{x}_i$  et  $\bar{x}$ , puis de  $\sigma_i^2$  et  $\sigma^2$ .
3. Montrer que  $\sum_{i=1}^p \sum_{j=1}^{n_i} (x_{i,j} - \bar{x}) = 0$ .
4. En déduire que  $\sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_i) = 0$ . A-t-on aussi  $\sum_{j=1}^{n_i} (x_{i,j} - \bar{x}) = 0$ ?
5. Montrer que  $\sigma^2(X)$  peut s'écrire comme somme de la moyenne des variances conditionnelles et de la variances des moyennes conditionnelles (attention aux pondérations).

### 2 Etude de la répartition de l'offre de soins au sein des régions

Nous souhaitons étudier l'offre de santé dans les différentes régions de France Métropolitaine en 2013. Nous nous intéressons à deux indicateurs, le nombre de médecins et le nombre de lits d'hôpitaux, qui sont disponibles sur le site de l'Insee.

Nous disposons des variables suivantes :

- Région : **reg**,
- Nombre de médecins : **nbMed**,
- Nombre de médecins en milliers : **nbMedM**, sert uniquement dans la Section 2.4,
- Nombre de lits d'hôpitaux : total **nbHtot**, dans les hôpitaux publics **nbHpub** et dans les hôpitaux privés **nbHpri**.

Tous les tableaux de données, graphiques et sorties figurent en Annexe. Précisez dans la justification la ou les Tables qui vous ont servi à chaque question.

## 2.1 Présentation du jeu de données

1. Pour décrire la population, indiquez
  - La population étudiée :
  - L'individu statistique :  
Donnez un exemple :
  - La taille de la population :
2. Pour décrire le jeu de données, indiquez le type des variables :

Variable	Quantitative	Qualitative
<b>reg</b>		
<b>nbMed</b>		
<b>nbMedM</b>		
<b>nbHtot</b>		
<b>nbHpub</b>		

3. Quel est le nombre total de lits d'hôpitaux ? Et le nombre total de médecins ?
4. La région qui a le moins de médecins correspond-elle à celle qui a le moins de lits d'hôpitaux ?
5. Donnez le ratio nombre de lits par médecin au niveau de la région Ile de France et sur la Province (définie comme la France métropolitaine privée de l'Ile de France).

## 2.2 Offre de lits d'hôpitaux et effet secteur

(extrait de l'examen de décembre 2014)

Dans cette partie, on s'intéresse uniquement aux nombres de lits dans les hôpitaux et à leur répartition entre les secteurs public et privé. Les Tables à remplir sont les Tables 4 et 5.

1. Calculez les distributions marginales et complétez les ligne et colonne **Total** des Tables 4 et 5.
2. Calculez (a1) et (a2) dans la Table 4 et donnez l'interprétation de (a2).
3. Calculez les autres valeurs manquantes dans les Tables 4 et 5.
4. Quelle est la région qui a le plus de lits dans les hopitaux privés ? Est-ce la même région qui a la plus grande proportion de lits dans les hôpitaux privés ?
5. Donnez l'interprétation des chiffres en gras dans les Tables 4 et 5 (région Bretagne, première colonne).
6. Comparez la distribution des lits dans le secteur public à la distribution marginale du nombre de lits. Indiquez leurs modes, min, max et donnez une région sur-représentée ainsi qu'une région sous-représentée.
7. Calculer les coefficients de variation pour les variables **nbHtot** et **nbHpub**.
8. Quelle est la série la plus dispersée ?
9. Peut-on affirmer que le secteur a un lien avec le nombre de lits ? Appuyez votre réponse sur un indicateur que vous calculerez et dont vous donnerez le nom.

### 2.3 Lien entre le nombre de lits et le nombre de médecins

On étudie maintenant le lien entre le nombre de lits **nbHtot** et le nombre de médecins **nbMedM**.

1. En observant le nuage de points (a), dites s'il est justifié de faire la régression linéaire correspondante.
2. Ecrivez le modèle de régression associé et donnez une estimation par MCO des paramètres.
3. Ecrivez le modèle de régression associé au nuage de points (b) et donnez une estimation par MCO des paramètres.
4. A partir des valeurs des paramètres calculées dans les 2 questions précédentes, déduisez un indicateur de la qualité d'ajustement dont vous donnerez la définition et l'interprétation.

### 2.4 Répartition des médecins

Dans cette partie, on utilise la variable **nbMedM** avec pour objectif de déterminer si la répartition des médecins sur le sol français est égalitaire.

1. Donnez une représentation graphique de la courbe de Lorenz.
2. Calculez l'indice de Gini correspondant. La répartition du nombre de médecins est-elle égalitaire entre les régions ?
3. Sur la base de cet indicateur, peut-on décider quelles sont les régions où il faudrait augmenter le nombre de médecins ? Si non, quelle information supplémentaire est à utiliser ?

## 3 Indices

Un indice de Laspeyres des prix à la consommation est obtenu à partir de trois groupes de produits dont on donne les pondérations (ou coefficients budgétaires) dans le tableau suivant

Type de produit	Pondération
Alimentation	2 516
Produits manufacturés	4 438
Services	3 046

1. Au cours de l'année, les prix de l'alimentation augmentent de 2,5%, ceux des produits manufacturés de 0,2% et ceux des services de 4,6 %. Déterminer la hausse globale des prix à la consommation au cours de l'année.
2. On suppose que, l'année suivante, les prix de l'alimentation croissent de 1 %, ceux des produits manufacturés baissent de 2% et ceux des services croissent de 3%. Déterminer l'évolution globale des prix pendant la seconde année, en utilisant les mêmes pondérations que précédemment.
3. Déterminer l'évolution globale des prix sur l'ensemble des deux années.

## Annexe

Reg	nbHtot	nbHpub	nbHpri	nbMed	nbMedM
Alsace	7 930	5 438	2 492	6 533	7
Aquitaine	14 079	8 011	6 068	11 551	12
Auvergne	5 932	4 117	1 815	4 121	4
Bourgogne	7 464	5 169	2 295	4 758	5
Bretagne	13 090	8 784	4 306	10 061	10
Centre	9 997	6 624	3 373	6 835	7
Champagne-Ardenne	6 274	4 040	2 234	3 773	4
Corse	1 192	657	535	941	1
Franche-Comté	4 770	3 786	984	3 571	4
Île-de-France	46 343	27 715	18 628	47 491	47
Languedoc-Roussillon	10 803	5 991	4 812	9 845	10
Limousin	3 794	2 634	1 160	2 524	3
Lorraine	10 176	6 478	3 698	7 166	7
Midi-Pyrénées	10 903	6 411	4 492	10 151	10
Nord-Pas-de-Calais	17 509	10 304	7 205	12 491	12
Basse-Normandie	6 658	5 004	1 654	4 301	4
Haute-Normandie	6 576	4 365	2 211	5 060	5
Pays de la Loire	13 224	8 194	5 030	10 224	10
Picardie	7 427	5 669	1 758	4 959	5
Poitou-Charentes	6 791	5 028	1 763	5 267	5
Provence-Alpes-Côte d’Azur	21 428	11 624	9 804	20 172	20
Rhône-Alpes	25 256	16 434	8 822	21 432	21
<b>France métropolitaine</b>	<b>257 616</b>	<b>162 477</b>	<b>95 139</b>	<b>213 227</b>	<b>213</b>

TABLE 1 – Données du problème. Source : Insee.

Variable	nbHtot	nbHpub	nbHpri	nbMed
Min	1192	657	535	941
Médiane	9997	5991	3373	6835
Moyenne	22401	14128	8273	18541
Max	257616	162477	95139	213227
Ecart-type	52127.22	32801.93	19347.73	43529.78

TABLE 2 – Statistiques descriptives

nbMedM	Effectif
1	1
3	1
4	4
5	4
7	3
10	4
12	2
20	1
21	1
47	1

TABLE 3 – Répartition des régions en fonction du nombre de médecins (en milliers).  
Source : Insee

Reg	Public	Privé
Alsace	(a1)	(a2)
Aquitaine	57%	43%
Auvergne	69%	31%
Bourgogne	69%	31%
Bretagne	<b>67%</b>	33%
Centre	66%	34%
Champagne-Ardenne	64%	36%
Corse	55%	45%
Franche-Comté	79%	21%
Île-de-France	(b1)	40%
Languedoc-Roussillon	55%	45%
Limousin	69%	31%
Lorraine	64%	36%
Midi-Pyrénées	59%	41%
Nord-Pas-de-Calais	59%	41%
Basse-Normandie	75%	(b2)
Haute-Normandie	66%	34%
Pays de la Loire	62%	38%
Picardie	76%	24%
Poitou-Charentes	74%	26%
Provence-Alpes-Côte d’Azur	54%	46%
Rhône-Alpes	65%	35%
<b>Total</b>		

TABLE 4 – Répartition des hopitaux par type de secteur dans chaque région.

<b>Reg</b>	<b>Public</b>	<b>Privé</b>	<b>Total</b>
Alsace	3%	3%	
Aquitaine	5%	6%	
Auvergne	3%	2%	
Bourgogne	3%	2%	
Bretagne	5%	5%	
Centre	4%	4%	
Champagne-Ardenne	2%	2%	
Corse	0%	1%	
Franche-Comté	2%	1%	
Île-de-France	17%	20%	
Languedoc-Roussillon	4%	5%	
Limousin	(b3)	1%	
Lorraine	4%	4%	
Midi-Pyrénées	4%	5%	
Nord-Pas-de-Calais	6%	8%	
Basse-Normandie	3%	2%	
Haute-Normandie	3%	2%	
Pays de la Loire	5%	5%	
Picardie	3%	2%	
Poitou-Charentes	3%	2%	
Provence-Alpes-Côte d'Azur	7%	(b4)	
Rhône-Alpes	10%	9%	

TABLE 5 – Répartition des lits par région dans les hopitaux publics et privés.

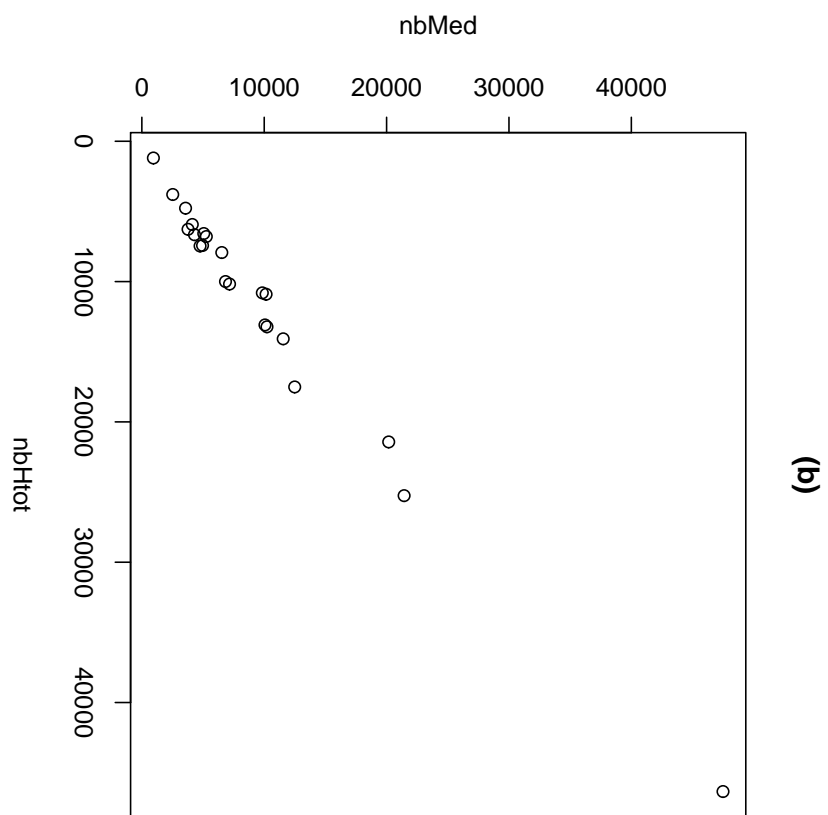
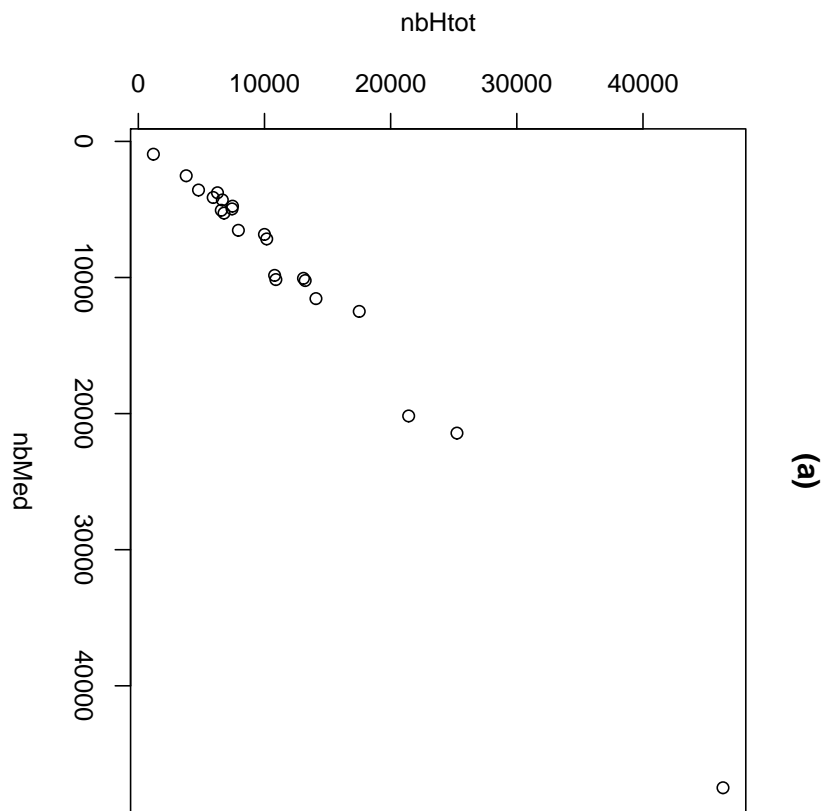


FIGURE 1 – Nuages de points