



Énoncé Statistique descriptive

IUT STID, 1^{ère} année

Devoir 2

17 janvier 2008

Nom : _____

Attention ! Toutes les questions doivent être effectuées sur la feuille d'énoncé! **Aucune autre copie ne sera acceptée.**

Par ailleurs, il sera **tenu compte des justifications et de la rédaction** des réponses dans la notation.

Note : Sauf indication contraire, les notations utilisées sont celles introduites en cours.

1 Liaison entre deux variables qualitatives (environ 7,5/20 de la note finale)

Le Tableau 1 donne le nombre de cas de chacune des trois maladies, fièvre aphteuse, PPCB et trypanosomose, observés sur les bovins de trois régions de l'Éthiopie.

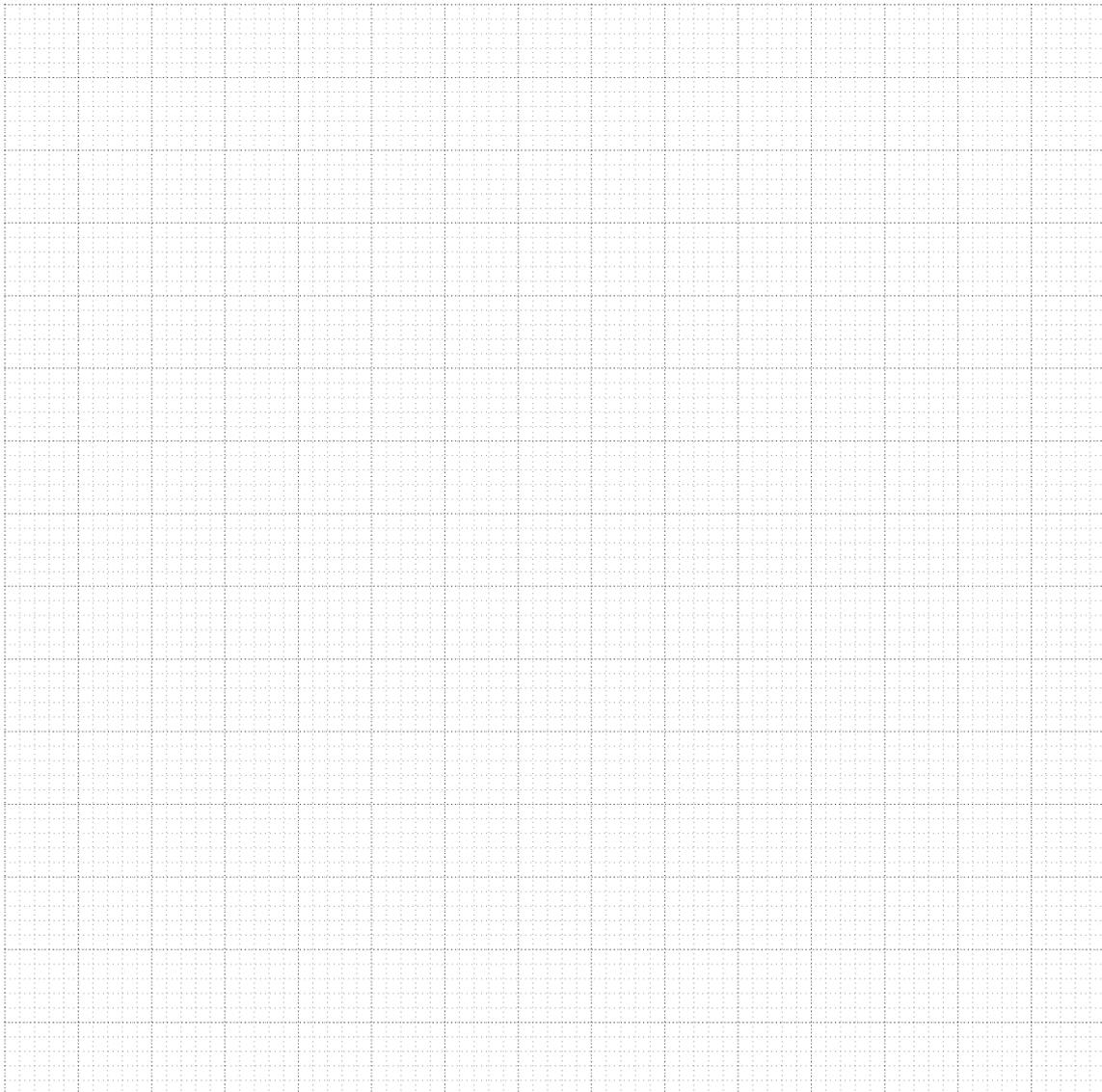
TAB. 1 – Répartition des maladies bovines dans les trois régions éthiopiennes

| Maladies Région | Fièvre aphteuse | PPCB | Trypano- -somose | |
|--------------------|--------------------|------|---------------------|--|
| Mana Sibu | 150 | 240 | 210 | |
| Nole Kaba | 140 | 130 | 280 | |
| Anfilo | 220 | 180 | 140 | |
| | | | | |

1. Compléter le Tableau 1 avec, **en bleu**, les effectifs marginaux de la variable « Région » et **en rouge**, les effectifs marginaux de la variable « Maladie ». Ajouter également, en noir, la taille de la population.
2. Calculer, dans le Tableau 1, **en vert**, les distributions de la variable « Maladie » conditionnellement aux modalités de la variable « Région ».
3. Effectuer, sur la Figure 1 ci-dessous, un graphique représentant les distributions de la variable « Maladie » conditionnellement aux modalités de la variable « Région » (plusieurs choix sont possibles).

Commenter ce graphique : les trois régions éthiopiennes sont-elles affectées de la même manière par ces

FIG. 1 – Graphique des distributions de la variable « Maladie » conditionnellement aux modalités de la variable « Région »



trois maladies ?

4. Calculer, dans ce Tableau 2, les effectifs théoriques d'indépendance.

TAB. 2 – Répartition des maladies bovines dans les trois régions éthiopiennes

| Région | Maladies | Fièvre aphteuse | PPCB | Trypano- -somose | |
|--------|----------|--------------------|------|---------------------|--|
| Mana | | | | | |
| Sibu | | | | | |
| Nole | | | | | |
| Kaba | | | | | |
| Anfilo | | | | | |
| | | | | | |

5. Dédurre de la question précédente, la valeur du χ^2 et du C de Cramer. Interpréter ce résultat.

2 Liaison entre une variable qualitative et une variable quantitative (environ 5,5/20 de la note finale)

Le Tableau 3 donne le taux de pauvreté par classes d'âge pour les individus appartenant à un ménage dont le revenu déclaré est positif ou nul et dont la personne de référence n'est pas étudiante (*Source : INSEE, <http://www.insee.fr>*).

1. Quelle est la population ? Sa taille ?

Dans la suite, on considère les sous-populations $\mathcal{P}_1, \dots, \mathcal{P}_8$ induites par les 8 classes d'âges (la variable Y , « classe d'âge », est donc traitée comme une variable qualitative) et la variable $X =$

TAB. 3 – Taux de pauvreté par classe d'âges

| Classe d'âge | Taux de pauvreté | Nombre d'individus dans la classe d'âge (en milliers) | Variance | | |
|-----------------|------------------|---|----------|--|--|
| Moins de 18 ans | 0,1548 | 13 340,85 | | | |
| 18 à 24 ans | 0,1750 | 4 809,21 | | | |
| 25 à 34 ans | 0,0972 | 7 844,84 | | | |
| 35 à 44 ans | 0,1197 | 8 592,78 | | | |
| 45 à 54 ans | 0,1105 | 8 268,12 | | | |
| 55 à 64 ans | 0,0972 | 6 746,58 | | | |
| 65 à 74 ans | 0,0695 | 4 936,18 | | | |
| 75 ans et plus | 0,1141 | 4 611,08 | | | |
| Total | ⊗ | 59 149,64 | | | |

$\mathbb{I}_{\{\text{Être en dessous du seuil de pauvreté}\}}^1$.

2. Expliquer, en justifiant brièvement, quelle est la valeur des moyennes conditionnelles $(\bar{X}_j)_{j=1,\dots,8}$ sur les populations $(\mathcal{P}_j)_j$.

3. Calculer, dans la quatrième colonne du Tableau 3, la valeur des variances conditionnelles $\text{Var}_j(X)$ sur les populations \mathcal{P}_j . On justifiera préalablement, ci-dessous, la méthode de calcul utilisée.

4. En utilisant, au besoin, les colonnes vides du Tableau 3, déterminer les valeurs de $\text{Var}_{\text{inter}}$ et de $\text{Var}_{\text{intra}}$.

¹On rappelle que $\mathbb{I}_A = \begin{cases} 1 & \text{si l'individu est dans } A \\ 0 & \text{sinon} \end{cases}$

En déduire la valeur de $\text{Var}(X)$.

5. Calculer le rapport de corrélation $\eta(X|Y)$. Interpréter cette valeur de manière concrète.

3 Concentration (environ 7/20 de la note finale)

Le Tableau 4, ci-dessous, donne les valeurs de l'effort en recherche et développement (DIRD) par habitant des 23 régions françaises (Outre-mer incluse), rangées par ordre croissant, ainsi que la population (en fréquence par rapport à la population nationale) de chacune de ces régions.

Source : « *La Vie de la Recherche Scientifique* », numéro 371, publication de la FSU.

On considère ici la population nationale et la variable X : « part de DIRD de l'individu considéré ». La deuxième colonne du tableau correspond donc à des fréquences, f_i , et la troisième aux modalités de la variable X , x_i .

1. À quoi correspond concrètement la valeur $f_i x_i$? Calculer ces valeurs dans la quatrième colonne du Tableau 4.

2. Compléter le tableau précédent de manière à obtenir les coordonnées des points de la courbe de Lorenz, (f_i^*, v_i) (vous utiliserez, à votre choix, toutes ou partie des colonnes vides du Tableau 4).

TAB. 4 – DIRD et populations régionales

| Région | Population (%) | DIRD/hab (en millions € par % pop. nat.) | $f_i x_i$ | f_i^* | | v_i |
|----------------------|----------------|---|-----------|---------|--------------|--------------|
| Corse | 0,4 | 32,5 | 13 | | | |
| Nord-Pas-de-Calais | 6,4 | 88,59 | 567 | | ¹ | ² |
| Outre-Mer | 2,9 | 96,55 | 280 | | | |
| Limousin | 1,1 | 107,27 | 118 | | | |
| Champagne-Ardenne | 2,1 | 108,57 | 228 | | | |
| Poitou-Charentes | 2,7 | 109,63 | 296 | | | |
| Bourgogne | 2,6 | 130 | 238 | | | |
| Basse-Normandie | 2,3 | 144,35 | 332 | | | |
| Lorraine | 3,7 | 146,49 | 542 | | | |
| Picardie | 3 | 148 | 444 | | | |
| Pays de la Loire | 5,4 | 154,26 | 833 | 0,326 | | 0,113 |
| Aquitaine | 5 | 218,4 | 1092 | 0,376 | | 0,145 |
| Haute-Normandie | 2,9 | 228,62 | 663 | 0,405 | | 0,163 |
| Centre | 4,1 | 230,24 | | 0,446 | | 0,190 |
| Bretagne | 4,8 | 235,83 | | 0,494 | | 0,222 |
| Alsace | 2,9 | 245,86 | | 0,523 | | 0,243 |
| France-Comté | 1,8 | 278,89 | | 0,541 | | 0,257 |
| PACA | 7,4 | 312,16 | | 0,615 | | 0,323 |
| Auvergne | 2,1 | 312,38 | | 0,636 | | 0,341 |
| Languedoc-Roussillon | 3,9 | 326,92 | | 0,675 | | 0,377 |
| Rhône-Alpes | 9,5 | 438,32 | | 0,770 | | 0,496 |
| Midi-Pyrénées | 4,3 | 646,05 | | 0,813 | | 0,575 |
| Île de France | 18,4 | 812,99 | | 0,997 | | 1,000 |
| Total | 100 | ∞ | | | | |

Remarque : On ne demande pas de détailler toutes les opérations effectuées pour toutes les cases du tableau mais simplement de préciser, ci-dessous, les opérations effectuées pour le calcul de ¹ et ² (et des éventuels calculs intermédiaires figurant dans d'autres colonnes du tableau pour cette ligne).

3. La Figure 2, ci-dessous, est la courbe de Lorenz de la variable « PIB régional par habitant » pour la population française. Que permet, concrètement, de mesurer cette courbe? Interpréter la forme de cette figure.

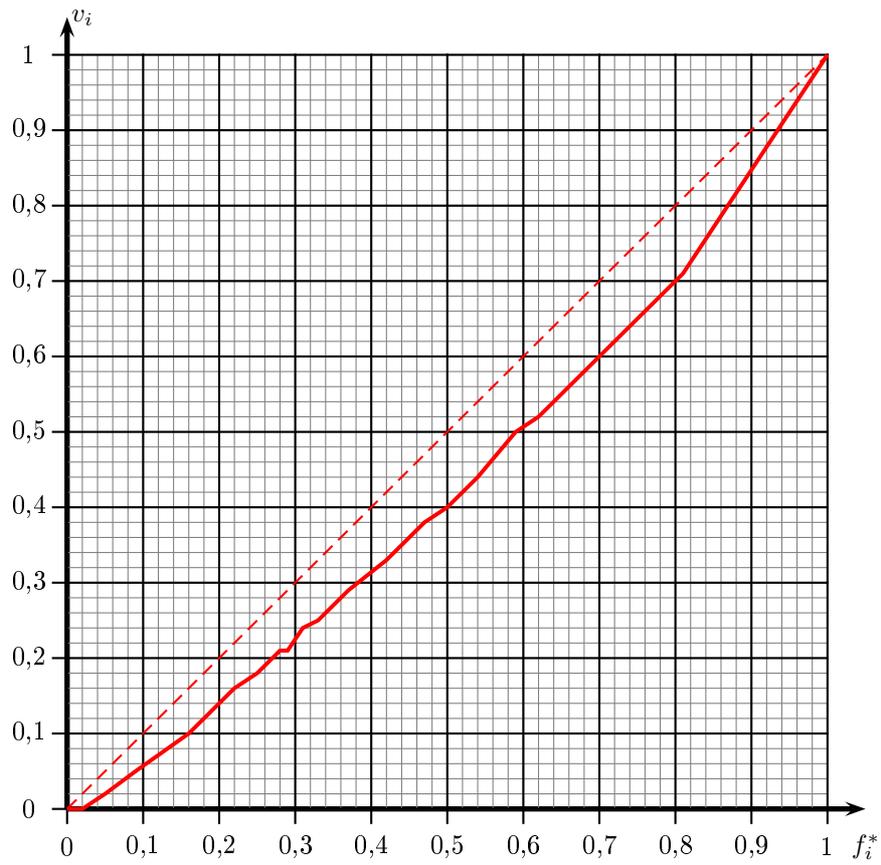


FIG. 2 – Courbe de Lorenz du PIB régional par habitant sur la population française

4. Quel pourcentage de la population nationale, située dans les régions les plus riches, concentre 50% du PIB

régional? (On fera apparaître les traits de construction sur le graphique.)

5. Compléter la Figure 2 avec la courbe de Lorenz de la DIRD par habitant. Que peut-on dire en comparant ces deux courbes?

6. Déterminer l'indice de Gini de la DIRD par habitant. Interpréter.