

## IUT STID, 1<sup>ème</sup> année : Statistique Descriptive

Corrigé Examen janvier 2007

### Consignes :

- Les exercices sont indépendants.
- Les réponses aux questions seront faites sur le sujet dans les zones prévues à cet effet.
- N'oublier pas d'écrire votre numéro d'étudiant sur le sujet.

### Numéro étudiant :

### Exercice 1

Voici les tailles en mètre d'étudiants masculins de trois pays :

Pays	Tailles ( $x_i$ )	$\sum x_i$	$\sum x_i^2$
Allemagne	1.79 1.83 2.01 1.92 1.86 1.75 1.86 1.94 1.69 1.87 1.74 1.78 1.93 1.75 1.78 1.80 2.01 2.02 1.70 1.92	36.95	68.4665
France	1.76 1.84 1.93 1.72 1.70 1.91 1.82 1.75 1.67 1.72 1.86 1.72 1.96 1.69 1.90 1.71 1.87 1.69 1.81 1.83 1.68 1.97 1.88	41.39	74.6923
Italie	1.84 1.77 1.79 1.77 1.79 1.76 1.75 1.80 1.75 1.78 1.75 1.77 1.81 1.78 1.73 1.68 1.77 1.70	31.79	56.1687

1. Quels sont les individus étudiés ? les étudiants
2. Quelles sont les variables étudiées ? Préciser le type et leur domaine de définition ou modalités :  
 Soit X la taille des étudiants : variable quantitative, domaine de définition  $[0, +\infty[$ .  
 Soit Y le pays d'origine des étudiants : variable qualitative, modalités  $\{ALL = Allemagne, IT = Italie, FR = France\}$ .
3. Compléter le tableau suivant relatif à la taille des étudiants :

	Allemagne	France	Italie	Total
Moyenne	1.8475	1.7995	1.7661	1.8054
Variance	0.0100	0.0090	0.0013	0.0082

4. Calculer la moyenne des variances conditionnelles Intra-classes (détailler le calcul) :

$$\begin{aligned}
 V_{intra} &= \frac{1}{n} \sum_{j \in \{ALL, FR, IT\}} n_j s^2(x|j) \\
 &= \frac{1}{61} [20 * 0.0100 + 23 * 0.0090 + 18 * 0.0013] = 0.0071
 \end{aligned}$$

5. Calculer la variance des moyennes conditionnelles Inter-classes (détailler le calcul) :

$$\begin{aligned}
 V_{inter} &= \frac{1}{n} \sum_{j \in \{ALL, FR, IT\}} n_j (\bar{x}_{|j} - \bar{x})^2 \\
 &= \frac{1}{61} [20 * (1.8475 - 1.8054)^2 + 23 * (1.7995 - 1.8054)^2 + 18 * (1.7661 - 1.8054)^2] \\
 &= 0.0010
 \end{aligned}$$

6. Calculer le rapport de corrélation (détailler le calcul) :

$$C_{X/Y} = \sqrt{\frac{V_{inter}}{V_{totale}}} = \sqrt{\frac{0.0010}{0.0082}} = 0.3587$$

7. Interpréter le résultat : Comme  $C_{X/Y} > 0.3$  on peut dire qu'il existe un lien moyen entre le pays d'origine et la taille des étudiants.

## Exercice 2

Les données ci-dessous sont les offres de retransmission sportives par chaîne de télévision (TF1, A2, FR3, LA5) exprimées en millièmes pour différentes disciplines lors de l'année 1990 (Foot, Rugby, Tennis, Sport d'hiver, Cyclisme, Golf, Sport automobile, Athlétisme, Autres sports).

Sport	TF1	A2	FR3	LA5
Foot	595	139	104	58
Rugby	0	132	13	16
Tennis	0	294	547	547
Hiver	37	44	69	0
Vélo	0	171	29	153
Golf	0	14	35	0
Auto.	166	2	15	208
Athlé.	1	49	18	0
Autres	201	155	170	18

- Le tableau ci-dessus est un «tableau de profils». Lequel ? Tableau profil colonne
- Quelles sont les variables étudiées ? Préciser le type et leur domaine de définition ou modalités :  
 X : les chaînes de TV, variable qualitative, modalités {TF1, A2, FR3, LA5}.  
 Y : les sports diffusés, variables qualitatives, modalités {Foot, Rugby, ..., Athl, Autres}.
- Le tableau suivant est le tableau construit sous l'hypothèse d'indépendance. Compléter la première colonne :

Sport	TF1	A2	FR3	LA5
Foot	224.00			
Rugby	40.25	40.25		
Tennis	347.00		347.00	
Hiver	37.5			
Vélo	88.25			
Golf	12.25			12.25
Auto.	97.75			
Athlé.	17		17	
Autres	136			

- Quelle particularité possède-t-il compte tenu de la question 1 ? c'est aussi un tableau profil colonne et toutes ses colonnes sont identiques.
- Le tableau suivant est le tableau des contributions de chaque croisement au  $\chi^2$  :

Sport	TF1	A2	FR3	LA5
Foot	?	32.2545	64.2857	123.0179
Rugby	40.2500	209.1444	18.4488	14.6102
Tennis	347.0000	8.0951	115.2738	115.2738
Hiver	0.0067	1.1267	26.4600	37.5000
Vélo	88.2500	77.5928	39.7797	47.5078
Golf	12.2500	0.2500	42.2500	12.25
Auto.	47.6528	93.7909	70.0518	124.3485
Athlé.	15.0588	60.2353	0.0588	17.0000
Autres	31.0662	2.6544	8.5000	102.3824

- (a) Retrouver la valeur manquante (détailler le calcul) :

La contribution au  $\chi^2$  de la case (1, 1) du tableau est  $\frac{(595-224)^2}{224} = 614.4688$

- (b) Calculer :  $\chi^2 = 2660.146$  et indice de Cramer = 0.4708

- (c) Peut-on considérer qu'il y a un lien les deux variables étudiées ? (justifier la réponse)  
 Comme l'indice de Cramer est proche de 0.5 on peut dire qu'il existe un lien entre les chaînes TV et les sports diffusés.