

DEUST GNM - Fiche 5

TRAN Viet Chi (chi.tran@math.univ-lille1.fr, Bureau 316 Bâtiment M3)

page web : <http://math.univ-lille1.fr/~tran/enseignements.html>

Exercice 1 (Conversion)

On a mesuré la température corporelle de 65 singes mâles et 64 singes femelles. La température moyenne des mâles est de 98.105 degrés F avec un écart-type de 0.693. La température moyenne des femelles est de 98.356 degrés F avec un écart-type de 0.679.

1. Quelle est la population ? Donnez sa taille. Quel est le type de la variable "température" ?
2. Quelle est la température moyenne des 129 animaux ?
3. Convertissez les résultats en degrés Celsius sachant que :

$$1 \text{ degré C} = (1 \text{ degré F} - 32) \times \frac{5}{9}$$

Exercice 2 (Air en Lorraine)

Pour juger de la qualité de l'air en Lorraine, une étude a conduit à l'élaboration d'un indice allant de 1 (très bon) à 10 (très mauvais).

Ville \ Indice	[1,4[[4,6[[6,10]
Forbach	74.5	13.7	11.8
Metz	81.1	9.4	9.4
Nancy	81.7	10.1	8.2
Thionville	82.4	8.6	8.9

TAB. 1 – *Pourcentage de jours dans l'année pendant lesquels la qualité était très bonne (1 à 4), moyenne (4 à 6) ou mauvaise (plus de 6), par ville.*

1. Quels sont les individus ? Quelles sont les populations considérées ? Quelle est la variable ? Quel est son type ?
2. Représenter graphiquement les distributions de proportions pour Forbach et Metz.
3. Quel est la proportion des journées pour lesquelles l'indice était inférieur à 3 à Forbach ? supérieur à 5 ?
4. Calculer l'indice moyen pour Nancy et son écart-type. Donner un intervalle de confiance pour l'espérance de l'indice (à Nancy).
5. Dessiner la courbe des fréquences cumulées pour Nancy. Déterminer graphiquement puis calculer la médiane (et les quantiles d'ordre 1 et 3).

Exercice 3 (Fermes canadiennes)

Lors d'une enquête, visant les fermes du Canada, on se sert d'une liste de fermes tirées du recensement de l'agriculture de 1991 pour déterminer l'échantillon. À cause de certaines contraintes relatives à la collecte des données, certaines fermes ont été exclues de la population, soit celles dont les ventes de produits agricoles étaient inférieures à 2 000, celles situées dans les réserves indiennes, les fermes collectives, les pâturages collectifs et les sociétés de portefeuille multiple.

La taille totale de l'échantillon a été établie à 6 000 fermes. La répartition proportionnelle à la densité de la population a été utilisée avec une taille minimale de l'échantillon de 58 fermes par strate. Ce seuil vise à assurer la représentation des fermes aux caractéristiques inhabituelles dans chaque strate. Une caractéristique inhabituelle est un élément rencontré dans au plus 5 p. 100 de la population d'une strate.

Province	Nbre de fermes dans la population	Nbre de fermes dans l'échantillon
Terre-Neuve	504	116
Ile-du-Prince-Édouard	2 115	174
Nouvelle-Écosse	3 310	174
Nouveau-Brunswick	2 680	174
Québec	34 671	781
Ontario	61 021	1 211
Manitoba	22 636	634
Saskatchewan	56 031	1 146
Alberta	52 286	1 067
Columbie-Britannique	14 933	523
Canada	250 187	6 000

TAB. 2 – population et taille de l'échantillon pour chaque province

1. Déterminer la population, l'échantillon. Quelle est la variable ?
2. Donner les distributions en fréquence de cette variable dans la population et dans l'échantillon.
3. Représenter graphiquement ces deux distributions.
4. Comparer ces distributions : se ressemblent-elles ? Qu'en déduire sur la qualité de l'échantillon ?

Exercice 4 (Corail)

Ayant prélevé 54 colonies d'*Alcyonium digitatum* ("corail mou"), on mesure leurs tailles (en mm) et leur poids (en g). Les résultats sont les suivants : **1.** Notre but est de contruire un

Taille	< 2	11	89	47	36	54	74	62	84	55	< 2
Poids	3	23	2657	386	561	375	847	1609	3240	1900	15
Taille	77	65	< 2	74	15	64	56	39	57	80	63
Poids	2211	982	5	3230	92	1795	620	511	872	1586	1243
Taille	2	58	33	40	< 2	84	67	59	8	52	45
Poids	10	1379	295	604	3	2263	1792	786	1403	854	729
Taille	24	105	60	< 2	90	55	60	58	12	50	44
Poids	180	3502	1083	5	4461	1712	1869	5172	25	497	841
Taille	50	62	49	2	107	83	45	65	25	52	
Poids	578	2226	782	11	4389	1449	4480	3680	90	1212	

tableau où les données seront groupées en 5 classes de taille d'égale amplitude et 5 classes de poids d'égale amplitude.

2. Calculer les distributions marginales.
3. Construire l'histogramme correspondant. Indiquer la classe modale.

4. Même question pour les données de poids.
5. Construire les courbes de fréquences relatives cumulées pour les deux variables (regroupées en classes). En déduire pour chaque variable (graphiquement et par le calcul) la médiane.
6. Calculer le coefficient de corrélation (à partir du regroupement en classes).

Exercice 5 (Test de médicaments)

Dans le cadre d'une étude de médicaments pour le soulagement des symptômes du rhume, on considère trois type de médicaments (A, B et C). On étudie leur action Y sur une population de 230 individus. On donne à chaque individu l'un des trois médicaments et on lui demande de qualifier son action après 2 jours ("aucune", "faible" ou "efficace").

Y \ X	A	B	C
Aucune	42	35	31
Faible	25	20	30
Efficace	20	12	15

1. Quelle est la population? Donner sa taille. Quelles sont les variables étudiées et leurs types?
2. Quelle est la proportion d'individus pour lesquels les médicaments ont été efficaces?
 3. Parmi les individus ayant pris le médicament A, quelle est la proportion d'individus pour lesquels son action est faible?
 4. Parmi les individus pour lesquels les médicaments n'ont aucune action, quelle est la proportion d'individus ayant pris le médicament C?
 5. Représenter graphiquement les distributions de proportions des trois distributions conditionnelles de Y sachant les modalités de X. Les deux caractères X et Y peuvent-ils être considérés comme indépendants?
 6. On décide d'éliminer les médicaments pour lesquels moins de 20% des individus déclarent qu'il est efficace. Quel(s) médicament(s) retient-on au vu des résultats?

Exercice 6 (Notes à la fac)

Dans une université, on a relevé la série du bac (variable X) et les notes en économie (variable Y) pour les 653 étudiants inscrits en première année d'économie.

X \ Y	[0,7[[7,10[[10,12[[12,15[[15,20]
ES	102	45	49	67	12
S	34	26	99	56	45
L	34	12	5	1	3
STT	45	17	1	0	0

1. Définir la population, sa taille, le type des variables étudiées.
2. Représenter graphiquement la distribution des proportions par série du baccalauréat.
3. Représenter graphiquement la distribution des proportions des notes d'économie des étudiants de première année.
4. Parmi les bacheliers ES, quelle est la proportion d'étudiants ayant obtenu la moyenne en économie?
5. Quelle est la proportion d'étudiants de première année ayant obtenu la moyenne?
6. Quelle est la note moyenne des étudiants de première année?
7. Quelle est la note moyenne des bacheliers ES? des bacheliers S?

8. Parmi les étudiants ayant obtenu une note en économie supérieure ou égale à 15, quelle est la proportion de bacheliers L ? de bacheliers S ? de bacheliers ES ?
9. Représenter graphiquement la distribution de proportions par série du baccalauréat des étudiants ayant obtenu une note en économie comprise entre 7 et 10. Quelle est le mode de cette distribution ?
10. Représenter graphiquement la distribution des proportions des notes d'économie des bacheliers ES. Quelle est la classe modale ?
11. Quelle est la note médiane des étudiants de premier cycle ?
12. Quelle est la note médiane des bacheliers STT ?

Exercice 7 (Nombre de pièce et d'habitants)

Dans une population composée de 110 ménages, on considère deux caractères statistiques : le nombre X de pièces que comporte l'habitation du ménage et le nombre Y d'enfants dans le ménage. **1.** Définir la population, sa taille et le type des variables étudiées.

$X \setminus Y$	0	1	2	3	4	5
1	6	4	1	0	0	0
2	3	11	10	5	1	0
3	1	3	16	13	4	1
4	0	1	3	15	8	4

2. Calculer le nombre moyen d'enfants des ménages habitant un 2-pièces.
3. Calculer le nombre moyen de pièces dans la population étudiée.
4. Calculer la covariance de X et Y , ainsi que le coefficient de corrélation.
5. Les deux caractères sont-ils liés ?

Exercice 8 (Age et consultation)

On étudie la variable X (nombre de consultations chez le médecin en janvier) en fonction de la variable Y (âge). **1.** Déterminer la population, sa taille et lister les variables observées et leurs

$X \setminus Y$	$[0,10[$	$[10,25[$	$[25,50[$	$[50,70]$
1	10	6	3	0
2	4	7	4	1
3	2	4	9	7

types.

2. Calculer la moyenne et l'écart-type de la variable X .
3. Calculer la moyenne et l'écart-type de la variable Y .
4. Calculer la covariance, puis la corrélation de X et Y .
5. Calculer la médiane de Y .
6. Déterminer et représenter les distributions conditionnelles de Y sachant $X = 1$ et $X = 3$. X et Y sont-elles liées ?
7. Parmi les individus ayant consulté 3 fois, quelle est la proportion des plus de 50 ans ?
8. Quelle est la proportion des patients de plus de 25 ans ?
9. Parmi les moins de 10 ans, quelle proportion a visité le médecin une fois ?
10. Déterminer la proportion des patients de moins de 30 ans ?