Université Grenoble Alpes
L2/STA351 **Frédérique Leblanc**

TD1 : Statistiques descriptives et Rappels de Probabilités

Statistique : Savoir calculer avec la calculette du bac les résumés simples d'un échantillon et savoir représenter. Probabilité : savoir utiliser les axiomes des probabilités, la probabilité conditionnelle, la formule des probabilités totales, les variables de Bernoulli et Binômiale.

Exercice 1.1 : On s'intéresse à la variable "Nature du lieu d'habitation". Le tableau ci-dessous indique les effectifs observés dans un échantillon de taille n.

Nature	Centre Ville	Banlieue	Village	Cité	Autre
Effectifs	87	30	32	30	10

1. Quelle est la variable observée ? Sa nature et ensemble des modalités ?
2. Calculer les fréquences de chaque modalité et proposer un graphique pour les représenter.
3. Cela a-t-il du sens de parler de médiane, moyenne ou tout autre résumé numérique ?

Exercice 1.2 : Le nombre de pannes hebdomadaires d'un appareil IRM a été observé pendant 100 semaines consécutives. On a obtenu les résultats suivants :

Nb Pannes	0	1	2	3	4	5	6	7
Effectifs	42	35	11	5	3	2	1	1

1. Quelle sont la variable étudiée, sa nature et l'ensemble de ses modalités (classées dans l'ordre croissant) ?
2. Calculer la moyenne, la variance empirique et l'écart-type empirique de l'échantillon observé.
3. Faire le tableau de distribution comprenant, modalités, effectifs, fréquences, fréquences cumulées. En déduire les quartiles (quantiles d'ordre $0,25$, $0,5$ et $0,75$: en discret il faut chercher la plus petite modalité à partir de laquelle l'ordre est atteint ou dépassé dans la suite des fréquences cumulées). Quelle est la fréquence empirique de $[2,5]$?
4. Représenter la répartition (fréquences - diagramme en barres) et la fonction de répartition empirique (fréquences cumulées - fonction en escaliers)

Exercice 1.3 : On s'intéresse à la quantité de cuivre trouvée dans la terre d'un terrain. On a pris 90 carottes de mêmes tailles dans le terrain étudié et mesuré le poids de cuivre de chacune (en mg). On a obtenu les résultats suivants :

Poids de cuivre	[12,16[	[16,20[	[20,24[	[24,28[	[28,32[	[32,36[	[36,44[
Effectifs	5	11	16	21	15	12	10

1. Quelle sont la variable étudiée, sa nature et l'ensemble de ses valeurs et les classes envisagées ?
2. Faire le tableau de distribution comprenant, classes, milieux de classes, effectifs, fréquences, fréquences relatives aux amplitudes de classes et fréquences cumulées.
3. Calculer la moyenne, la variance empirique et l'écart-type empirique approximatifs de l'échantillon observé. Pourquoi les résultats sont-ils approchés ?
4. Représenter la répartition (fréquences relatives - histogramme) et la fonction de répartition empirique (fréquences cumulées - fonction linéaire par morceaux). Lire le quantile d'ordre $0,6$ sur l'un de ces graphes

Exercice 1.4 : Pour trois modes de préparation d'un certain vaccin (A, B ou C) on a mesuré la réaction épidermique locale au point d'injection chez $n = 500$ patients. Les observations sont synthétisées dans le tableau à doubles entrées suivant (appelé tableau de contingence) :

Mod. Prép./Réact.	Légère	Moyenne	Ulcération	Abcès	Total
A	13	158	8	1	
B	30	133	6	1	
C	9	129	10	2	
total					

1. Quelle répartition observe-t-on pour la variable réaction (indépendamment du mode de préparation du vaccin) ? La représenter avec un diagramme en barres de couleur noire.
2. Quelle répartition observe-t-on pour la variable réaction dans la sous-popultation des patients ayant reçu un vaccin préparé avec le mode A (cette répartition est appelée "répartition de la variable Réaction conditionnelle au mode de préparation A") ? Ajouter sur le diagramme précédent cette répartition en rouge. Faire de même avec la conditionnelle au mode de préparation B et l'ajouter en vert. Puis pour le mode de préparation C en bleu.
3. Comparer les diagrammes obtenus et les interpréter (la question que l'on se pose est celle de l'effet ou non du mode de préparation sur la réaction).

Exercice 1.5 : Sur un papier dessiner une grille comprenant 6 points régulièrement espacés dans les deux sens (vertical et horizontal par ex tous les cm ce qui produit une grille contenue dans un carré de côté 5cm). Noter A le point en haut à gauche et B le point en bas à droite. Placer ensuite le point R translaté de A de quatre pas à droite et un pas en bas. On part de A et on arrive en B en passant par R et avec les règles suivantes : on n'est autorisé qu'à se déplacer sur les points de la grille soit à droite soit en bas (interdit de monter ou se déplacer à gauche).

1. Combien de trajets différents peut-on trouver pour se rendre de A à B en passant par R.
2. Combien de trajets allant de A à B en passant par R changent de direction deux fois exactement.
3. Quelle est la probabilité de se rendre de A à B en passant par R et en changeant de direction exactement deux fois ?

Exercice 1.6 : Soit X une variable à valeurs dans $\{1, 2, 3, 4\}$ de loi de probabilité uniforme (c'est à dire $P(X = 1) = P(X = 2) = P(X = 3) = P(X = 4)$). On note $A = \{X \in [0, 4]\}$, $B = \{X \in [2, 4[\}$ et $C = \{X \in [0, 2]\}$ trois évènements qui dépendent de X.

1. Décrire les évènements $\bar{A}$, $\bar{A} \cap B$, $A \cup B$ et $B \cap C$. En calculer les probabilités.
2. Calculer la probabilité de B sachant A (notée $P(B|A)$) et celle de B sachant C.
3. Calculer la probabilité de B de façon directe et en utilisant la formule des probabilités totales avec $P(B|A)$ et $P(B|\bar{A})$.

Exercice 1.7 : Dans une forêt 70% des arbres sont des chênes et les autres des hêtres. 40% des arbres sont atteints d'une maladie qui touche un hêtre sur 3. On note C l'évènement "être un chêne" et M celui "avoir la maladie".

1. Faire le tableau à double entrées croisant les variables arbre et maladie et contenant sur chaque case la fréquence d'apparition du couple de modalités correspondante.
2. Faire un arbre pondéré correspondant à ces données.
3. Donner ou calculer les probabilités de :
   (a) être un chêne sain
   (b) être un hêtre malade
   (c) être malade sachant qu'on est un chêne
   (d) être un chêne sachant qu'on est malade

Exercice 1.8 : On sait qu'une certaine opération chirurgicale a 90% de chance de réussir. On réalise cette opération sur 5 patients. Soit X la variable aléatoire indicatrice de réussite pour un patient et Y la variable égale au nombre de réussite parmi 5 patients.

1. Quelles sont les lois de X et Y ?
2. Calculer la probabilité que l'opération rate les 5 fois.
3. Calculer la probabilité que l'opération réussisse au moins 1 fois.