Université Claude Bernard - Lyon 1 **Semestre automne 2022-2023**

L2 - Informatique UE : Statistiques pour l'Informatique

Correction des exercices supplémentaires : Feuille de TD 2

Exercice 2.6 **(Partiel 2021)**

Le tableau suivant donne l'âge, exprimé en année, des 20 arbustes à fleurs d'une petite jardinerie

0.6	1.4	1.3	1	0.8	3.7	1.2	0.6	2.2	1.5
2.4	2.9	1.9	1.1	1.8	2.8	2.1	3.1	2.5	0.9

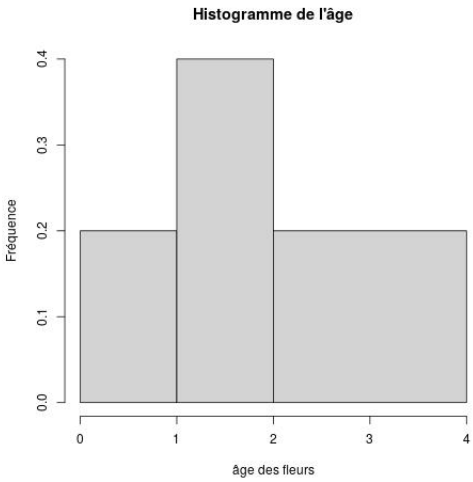
1. Tracer l'histogramme de l'âge des fleurs de cette jardinerie avec les intervalles d'abscisses $[0,1[$, $[1,2[$ et $[2,4]$. On désigne ensuite ces trois groupes par fleurs de moins d'un an, entre 1 et 2 ans et plus de 2 ans.

   On commence par calculer les effectifs de chaque intervalle $N([0,1[) = 4, N([1,2[) = 8$ et $N([2,4]) = 8$. L'effectif total est $N = 20$.

   La hauteur de l'histogramme est $h([a,b[) = \frac{N([a,b[)}{N(b-a)}$.

   On obtient : $h([0,1[) = \frac{4}{20} = 0.2, h([1,2[) = \frac{8}{20} = 0.4$ et $h([2,4]) = \frac{8}{20*(4-2)} = 0.2$.

   On obtient, l'histogramme suivant :



2. (Attention : les questions suivantes utilisent le chapitre 3) Le jardinier a réparti ses fleurs, en trois groupes, les jaunes, les roses et les autres couleurs. Il trouve que parmi ses fleurs de moins d'un an, 25% sont jaunes et 50% sont roses. Parmi ses fleurs de 1 à 2 ans, 25% sont jaunes et 25% sont roses. Enfin, parmi ses fleurs de plus 2 ans, la moitié est jaune et la moitié est rose.

   Quelle est la proportion de fleurs roses dans le magasin ?

   Notons I, II, III les évènements avoir moins d'un an, entre 1 et 2 ans et plus de 2 ans. La question précédente donne les probabilités (empiriques ou fréquences) $P(I) = 0.2, P(II) = 0.4, P(III) = 0.4$

   On note J l'évènement "être jaune", R l'évènement "être rose" et A l'évènement être d'une autre couleur.

   L'énoncé donne les probabilités conditionnelles : $P(J|I) = .25, P(R|I) = .5, P(J|II) = .25, P(R|II) = .25, P(J|III) = .5, P(R|III) = .5,$

   Par la formule des probabilités totales, on obtient la proportion de fleurs roses comme la probabilité d'être rose dans le modèle ci-dessus :

$$P(R) = P(R|I)P(I) + P(R|II)P(II) + P(R|III)P(III) = (.5)(0.2) + (.25)(0.4) + (.5)(0.4) = 0.4$$

3. (1 point) Un client a acheté une fleur jaune, quelle est la probabilité qu'elle ait plus de 2 ans ?

   Par la formule de Bayes :

$$P(III|J) = \frac{P(J|III)P(III)}{P(J|I)P(I) + P(J|II)P(II) + P(J|III)P(III)} = \frac{(.5)(0.4)}{(.25)(0.2) + (.125)(0.4) + (.5)(0.4)} = \frac{2}{3}.$$

Exercice 2.7 **(QCM 2020)** Sept amis de L1 d'informatique se demandent si leur moyenne y sur l'année de L1 est déterminée en terme d'une relation affine simple par leur note x au cours d'informatique obligatoire du premier semestre. Voici le tableau de leurs notes.

étudiants	1	2	3	4	5	6	7
Note d'informatique x	9	8	18	10	16	16	13
Note sur l'année y	11	12	17	13	15	14	14

Calculons les moyennes empiriques de x et y puis la droite de régression linéaire de la note sur l'année par rapport à la note d'informatique. On arrondira

Les moyennes empiriques sont

$$\overline{x} = \frac{9+8+18+10+16+16+13}{7} \simeq 12.86$$

$$\overline{y} = \frac{11+12+17+13+15+14+14}{7} \simeq 13.71$$

Par la formule du cours, on trouve d'abord $a = \frac{cov(x,y)}{cov(x,x)}$. Ici, la covariance empirique non-biaisée est

$$cov(x,y) = \frac{1}{6}\sum_{k=1}^{7}(x_k - \overline{x})(y_k - \overline{y}) = 7.11905 \simeq 7.12$$

et la variance empirique non-biaisée est

$$var(x) = cov(x,x) = \frac{1}{6}\sum_{k=1}^{7}(x_k - \overline{x})^2 = 15.47619 \simeq 15.48,$$

on obtient le quotient :

$$a = \frac{7.11905}{15.47619} \simeq 0.46.$$

Ensuite, on obtient

$$b = \overline{y} - a\overline{x} = 13.71428 - 0.46 * 12.85714 \simeq 7.80.$$

La droite de régression est donc $y = 0.46x + 7.80$