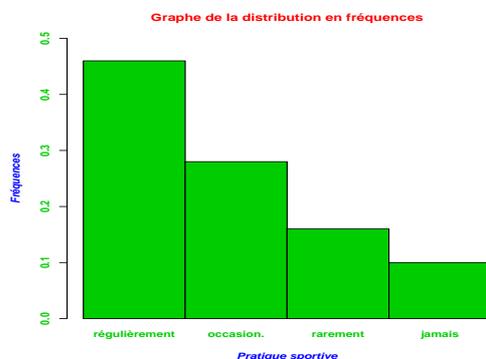


CORRIGÉ

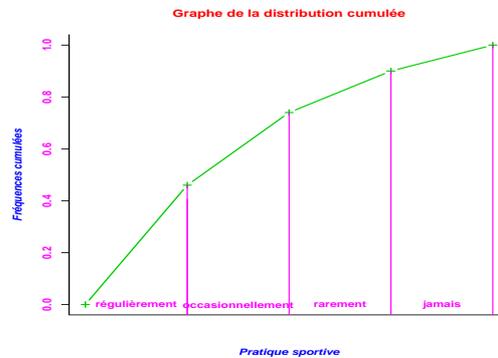
Exercice 1 :

1. Décrire la situation statistique : la population, l'échantillon, la variable et sa nature, et faire une représentation graphique de la distribution (en indiquant soigneusement les axes, leur échelle et leur orientation si nécessaire).
 - **Population** : l'ensemble des personnes de 15 à 25 ans
 - **Echantillon** : de taille 500 personnes (aucune précision sur la manière dont il a été constitué).
 - **Variable** : "Pratique sportive" à 4 modalités {régulièrement, occasionnellement, rarement, jamais} - variable qualitative ordinale.



2. Dans cette situation, s'intéresser à la moyenne de la variable a-t-il du sens : Oui/Non, pourquoi ? Si Oui, calculez-la.
Non, ça n'a pas de sens, la variable n'est pas quantitative.
3. Dans cette situation, s'intéresser au mode a-t-il du sens : Oui/Non, pourquoi ? Si Oui, déterminez-le.
Oui, on peut s'intéresser à la modalité la plus fréquente (la plus répondue). Le mode est régulièrement.

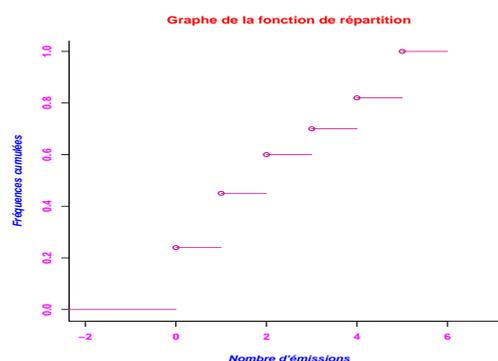
4. Dans cette situation, s'intéresser aux fréquences cumulées a-t-il du sens : Oui/Non, pourquoi ? Si Oui, faites-en une représentation graphique.
Oui, puisqu'il existe un ordre sur les modalités on peut donc faire un cumul selon cet ordre naturel.



Exercice 2 :

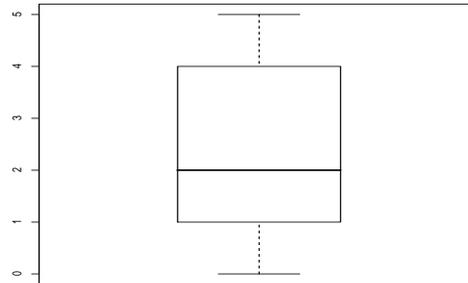
- Décrire la situation statistique.
 - **Population** : les étudiants de l'UPV
 - **Echantillon** : de taille 500 personnes (aucune précision sur la manière dont il a été constitué).
 - **Variable** : "Nombre d'émissions" à 6 modalités observées $\{0, 1, 2, 3, 4, 5\}$ - variable quantitative discrète.
- Donner les fréquences cumulées et en faire une représentation graphique.
Tableau de la distribution et de la distribution cumulée :

Nbémissions	0	1	2	3	4	5	Total
n_k	120	105	75	50	60	90	500
f_k (en %)	0.24	0.21	0.15	0.10	0.12	0.18	1
N_k	120	225	300	350	410	500	
F_k (en%)	0.24	0.45	0.60	0.70	0.82	1	



3. Quels sont la médiane, les 2 quartiles et l'étendue de cette distribution. Faire la représentation graphique appropriée représentant ces 4 indices.

- *Médiane* : 2
- *1er quartile* : 1
- *3ème quartile* : 4



4. Pour le quartile Q_1 , donner la proportion d'individus dont la réponse est \leq et \geq à Q_1 . Quelles propriétés doit-on alors vérifier pour valider ce quartile Q_1 .

$$Prop(rep \leq 1) = 0.45 \text{ et } Prop(rep \geq 1) = 0.76$$

On vérifie que $0.45 \geq 0.25$ et $0.76 \geq 0.75$

5. Calculer la moyenne de cette distribution.

$$moy = (120 * 0 + 105 * 1 + 75 * 2 + 50 * 3 + 60 * 4 + 90 * 5) / 500 = 2.19$$

6. Quelle est la distance adéquate pour évaluer la dispersion des observations autour de la moyenne, notée $disp(\bar{x})$. Calculer cette dispersion.

La distance adéquate est la différence au carré.

$$disp = (120 * (0 - moy)^2 + 105 * (1 - moy)^2 + 75 * (2 - moy)^2 + 50 * (3 - moy)^2 + 60 * (4 - moy)^2 + 90 * (5 - moy)^2) = 1666.95$$

7. Sans faire de calculs, que pouvez-vous dire de $disp(0)$, $disp(1)$ et $disp(2)$ comparativement à $disp(\bar{x})$?

Toutes ces dispersions seront plus grandes que $disp(\bar{x})$ (qui est la plus petite possible). De plus, la moyenne étant égale à 2.19, plus on s'en éloigne, plus la dispersion augmente. Ainsi

$$disp(0) > disp(1) > disp(2) > disp(\bar{x})$$

8. Donner la valeur de la variance et de l'écart-type.

$$\sigma^2 = 1666.95 / 500 = 3.33 \text{ et } \sigma = \sqrt{3.33} = 1.83$$

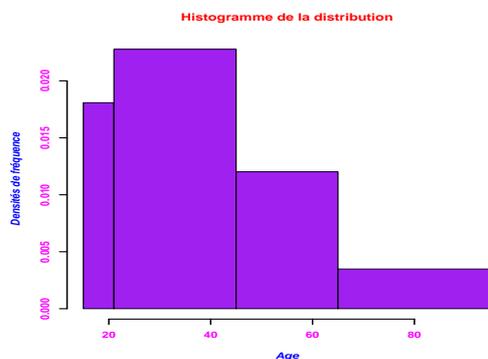
9. Le skewness de cette distribution a été calculé mais mal noté. Et on hésite entre les valeurs 0.3093 et 0.0093. Laquelle de ces 2 valeurs vous paraît la plus vraisemblable ? Pourquoi ?

La distribution étant très dissymétrique, le skewness n'est pas proche de 0. La valeur la plus vraisemblable est 0.3093.

Exercice 3 :

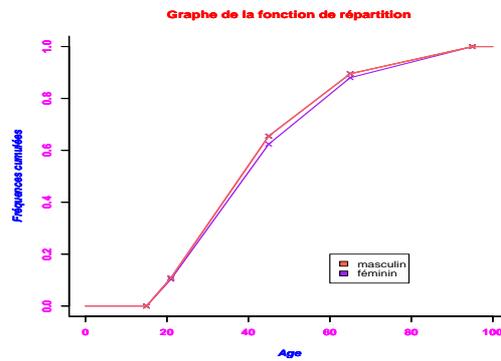
- Préciser la(les) variable(s) de cette situation et sa(leur) nature.
 2 variables :
 - *Âge* : variable quantitative continue regroupée ici en 4 classes
 - *Sexe* : variable dichotomique à 2 modalités {masculin, féminin}.
C'est une variable qualitative nominale.
- Parmi les hommes consommateurs, quel est le pourcentage de ceux compris entre 21 et 45 ans ? Parmi les consommateurs de plus de 65 ans, quel est le pourcentage de femmes ?
 - $141/258 = 0.547 = 54.7\%$
 - $29/56 = 0.518 = 51.8\%$
- Construire l'histogramme de la distribution de l'âge pour les hommes.

Tps	[15; 21[[21; 45[[45; 65[[65; 95[
f_k	0.109	0.547	0.240	0.105
F_k	0	0.109	0.655	0.895
a_k	6	24	20	30
$d_k \times 100$	1.882	2.277	1.202	0.349



- Calculer moyenne et variance de l'âge des hommes consommateurs d'alcool.
 - $moy = (28 * 18 + 141 * 33 + 62 * 55 + 27 * 80) / 258 = 10727 / 258 = 41.6$
 - $\sigma^2 = (28 * (18 - moy)^2 + 141 * (33 - moy)^2 + 62 * (55 - moy)^2 + 27 * (80 - moy)^2) / 258 = 298.33$
- Construire ensuite l'intervalle de dispersion empirique (à 95 %) centré autour de la moyenne de l'âge des hommes consommateurs.
 $[moy - 2 * \sigma ; moy + 2 * \sigma] = [7.03 ; 76.12]$

6. Sur un même graphique, tracer les 2 fonctions de répartition chez les hommes et chez les femmes de l'âge des individus. Commenter.



La fonction de répartition des femmes est très légèrement décallée à droite (ou autrement dit celle des hommes est très légèrement au dessus de celle des femmes). Cela signifie une très légère tendance des femmes consommatrices d'alcool à être plus âgées que les hommes.