Statistiques descriptives univariées et bivariées

Exercice 1 ($\star$)

- On a $\overline{x} = \dfrac{12+13+15+19+21+22+24+23}{8} = \dfrac{126}{8} \simeq 18,63$.
- Selon le modèle de la calculatrice, la variance est calculée ou l'écart-type ou les deux. Si votre calculatrice ne donne que l'écart-type, on en déduit la variance en mettant l'écart-type au carré.
  Avec une Casio, on obtient l'écart-type $\sigma(X) \simeq 4,39$ et on en déduit la variance $V \simeq 19,27$.
- Calcul de la médiane M_e. L'effectif total est pair (il vaut 8) donc la médiane est égale à $\dfrac{19+21}{2} = 20$.

Exercice 2 ($\star$)

En utilisant la calculatrice, on obtient :
- On a $\overline{x} \simeq 78,76$.
- On a $\sigma(X) \simeq 0,45$ et donc $V(X) =\simeq 0,20$.
- On a $Q_1 = 78,4$
- On a $Q_2 = M_e = 78,8$.
- On a : $Q_3 = 79,3$
- Calcul de l'écar interquartile : $Q_3 - Q_1 = 0,9$.

Exercice 3 ($\star\star$)

Pour calculer la moyenne $\overline{x}$ de cette série, on concentre les valeurs de chaque classe en leurs centres, puis on calcule la moyenne de la série discrète ainsi obtenue. On obtient :

$$\overline{x} = 2\times 0,0108 + 5\times 0,0481 + 7\times 0,2103 + 9\times 0,3245 + 11\times 0,28 + 13\times 0,095 + 15\times 0,0288 + 17\times 0,0024 = 9,4458$$

Le professeur annonce $9,4$. La différence vient de ce que les valeurs des classes ont été concentrées en leurs centres. C'est bien entendu une approximation qui, pour être commode, ne rend pas compte de la réalité (tous les étudiants ayant eu une note entre 0 et 4 n'ont pas eu 2...). Cela dit, l'écart entre ce qui est annoncé et le résultat trouvé n'est pas significatif, ce qui montre que l'approximation réalisée est pertinente.

Calcul de la médiane (notée M_e).
Pour cela donnons le tableau des fréquences cumulées :

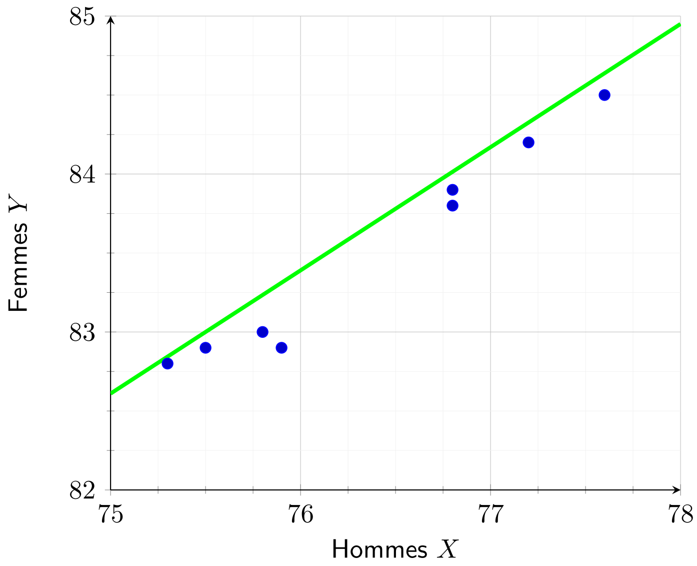
Notes	$[0;4[$	$[4;6[$	$[6;8[$	$[8;10[$	$[10;12[$	$[12;14[$	$[14;16[$	$[16;20[$
Fréquences	$0,0108$	$0,0589$	$0,2692$	$0,5937$	$0,8737$	$0,9687$	$0,9975$	$0,9999$

La première classe atteignant ou dépassant $0,5$ est $[8;10[$.

Exercice 4 ($\star\star\star$)

1. On a : $\overline{x} \simeq 76,36$ et $\overline{y} = 83,5$.
2. On a $V(X) \simeq 0,63$ et $V(Y) = 0,4$. Puis, $\sigma_X \simeq 0,79$ et $\sigma_y \simeq 0,63$.
3. On a : $Cov(X,Y) \simeq 0,49$.
4. On a donc :
   $$r(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \simeq 0,976$$
   $r(X,Y)$ étant proche de 1, un ajustement linéaire est pertinent.

   La calculatrice nous donne l'équation de la droite de régression linéaire. On arrondit au centième et on obtient :
   $$y = 0,79x + 23,4$$
5. On a le graphique suivant :



Exercice 5 ($\star\star\star$)

1. On a $\overline{x} = 2005,5$ et $\overline{y} = 2,25$.
2. On a $V(X) = 5,25$ et $V(Y) = 0,55$. Puis, $\sigma_X \simeq 2,29$ et $\sigma_Y \simeq 0,74$.
3. On a $Cov(X,Y) = 1,6$.
4. On a donc :
$$r(X,Y) = \frac{Cov(X,Y)}{\sigma_x \sigma_Y} \simeq 0,941$$
$r(X,Y)$ est assez proche de 1. Un ajustement linéaire semble donc ici pertinent.
5. Commençons par écrire le tableau correspondant :

X	2002	2003	2004	2005	2006	2007	2008	2009
$\exp(Y)$	$2,01$	$4,95$	$7,39$	$11,02$	$12,18$	$16,44$	$20,09$	$20,09$

On trouve alors $r(X,\exp(Y)) \simeq 0,99$. Ainsi, $r(X,\exp(Y))$ est très proche de 1, et en tout cas plus proche de 1 que $r(X,Y)$: un ajustement linéaire ici s'impose.

La calculatrice nous do„e la droite de régression linéaire de $\exp(y)$ en x (arrondie au centième) :
$$\exp(y) = 2,75x - 5493,09$$
et donc
$$y = \ln(2,75x - 5493,09)$$