



Correction de l'examen de statistique descriptive

Exercice1 : On donne la série des salaires des employés d'une entreprise (exprimés en DH), les données sont présentées dans l'ordre croissant :

2 500	3 200	3 300	4 500	4 800	5 600	5 700	6 200	9 850	11 000	12 500	19 000
-------	-------	-------	-------	-------	-------	-------	-------	-------	--------	--------	--------

1. L'étendu des salaires est $19\,000 - 2\,500 = 16\,500$ DH
2. Le salaire moyen est $88\,150/12 = 7\,345.83$ DH
3. Le salaire médian est $(5\,600 + 5\,700)/2 = 5\,650$ DH

Interprétation : il y a 50% des employés qui ont un salaire supérieur à 5 650 DH et les 50% autres ont un salaire inférieur à 5 650 DH.

4. On donne

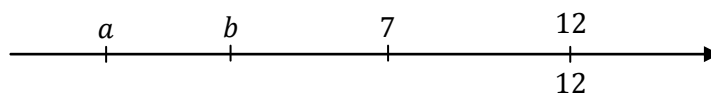
Le premier quartile	3 300 DH
Le troisième quartile	9 850 DH
L'écart interquartile	$9\,850 - 3\,300 = 6\,550$ DH

5. Peut-on dire qu'il existe des valeurs aberrantes dans la série des données ? Expliquez :

Une valeur aberrante selon Tukey s'éloigne de $1.5 \times 6\,550 = 9\,825$ au-dessus de 9 850 et en dessous de 3 300. Les valeurs aberrantes de la série seront à l'extérieur de l'intervalle $[3\,300 - 9\,825, 9\,850 + 9\,825] = [-6\,525, 19\,675]$. Aucune valeur de la série n'est aberrante.

Exercice2 : Dans une famille de cinq enfants l'âge modal est 12 ans, l'âge médian est 7 ans, l'âge moyen est 8 ans et l'étendu de l'âge de ces enfants est 9 ans. Quel est l'âge des deux derniers-nés dans cette famille ?

$n = 5$ impaire, la médiane coïncide avec une valeur observée, la valeur 12 est observée deux fois



$$12 - a = 9 \Rightarrow a = 3 \text{ ans et } (12 + 12 + 7 + 3 + b)/5 = 8 \Rightarrow b = 6 \text{ ans}$$

Exercice3 : On considère deux entreprises E1 et E2 qui emploient chacune 150 salariés. Dans l'entreprise E1 le salaire moyen est égal à 5000 DH avec un écart-type de 1800DH. Tandis que dans l'entreprise E2 le salaire moyen est égal à 8500 DH avec un écart-type de 2500DH. Dans quelle entreprise selon vous le salaire est-il le plus dispersé? Justifiez.

On compare les coefficients de variation des deux entreprises :

Pour E1 : $V_1 = \left| \frac{\sigma_1}{\mu_1} \right| \times 100 = 36\%$ et pour E2 : $V_2 = \left| \frac{\sigma_2}{\mu_2} \right| \times 100 = 29.41\%$

La dispersion des salaires dans l'entreprise E1 est plus élevée que dans l'entreprise E2.

Exercice4 : D'après les statistiques du bureau Census (www.census.gov), aux États-Unis, l'âge médian au premier mariage a augmenté pour les hommes et les femmes. En 2019, l'âge médian au premier mariage était de près de 30 ans pour les hommes et de 28 ans pour les femmes. On donne le tableau des données pour les dix dernières années :

Tableau 1 : Age médian (estimé) au premier mariage des américains

Année	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Âge des hommes H_i	28,2	28,4	28,6	29,0	29,3	29,2	29,5	29,5	29,8	29,8
Âge des femmes F_i	26,1	26,4	26,6	26,6	27,0	27,1	27,4	27,4	27,8	28,0
i	1	2	3	4	5	6	7	8	9	10

Source: U.S. Census Bureau, Current Population Survey, March and Annual Social and Economic Supplements. Nov.2019

Ici on s'intéresse à la variable âge des femmes. Le type de cette variable est

Quantitatif continu

Dans le but de donner une prévision de l'âge médian des femmes américaines au premier mariage pour l'année 2020, on va utiliser les données du tableau ci-dessus pour construire un modèle de régression linéaire. On donne les calculs suivants :

$$\sum_{i=1}^{10} F_i = 270.4 ; \sum_{i=1}^{10} i = 55 ; \sum_{i=1}^{10} F_i^2 = 7315.06 ; \sum_{i=1}^{10} i^2 = 385 \text{ et } \sum_{i=1}^{10} i * F_i = 1503.9$$

1. Compléter le Tableau suivant :

Tableau 2 : modèle de la régression linéaire

Coefficient de covariance entre F_i et i	Coefficient de corrélation entre F_i et i	Equation de la droite d'ajustement $F_i = \hat{a} * i + \hat{b}$
1.67	0.9907	$F_i = 0.2024 i + 25.927$

La droite de la régression linéaire peut expliquée jusqu'à **98,16%** de la variabilité de l'âge médian des femmes au premier mariage.

- L'âge médian prévu par le modèle pour l'année 2020 est **28.15** ans (il correspond à la 11^{ème} période $i = 11$).
- Si on utilise le modèle pour prévoir l'année pour laquelle on atteint un âge médian près de 29ans, on trouvera approximativement l'année **2024**.

Exercice5 :

En 2019 l'indice des dépenses d'un produit A est de 117 (base 100 en 2018). Durant cette période, la quantité consommée de A a augmenté de 20% ; le prix de A a-t-il augmenté ou diminué ? de combien ?

$$I_D(2019/2018) = 1.17 \text{ et } I_Q(2019/2018) = 120\% = 1.20$$

$$\text{La dépense } D = Q * P \text{ et son indice } I_D(2019/2018) = I_Q(2019/2018) * I_P(2019/2018)$$

$$\text{On calcule l'indice du prix } I_P(2019/2018) = \frac{1.17}{1.20} = 0.975$$

$$\tau_P^A(2019/2018) = I_P(2019/2018) - 1 = -2.5\%$$

Le prix de A a diminué entre 2018 et 2019 ; la diminution s'élève à 2,5 %