TD UEF03 n^0 7

Statistiques descriptives

1. **(*)** On considère les longueurs de 100 oeufs de coucous (en mm):

22.5	20.1	23.3	22.9	23.1	22.0	22.3	23.6	24.7	23.7
24.0	20.4	21.3	22.0	24.2	21.7	21.0	20.1	21.9	21.9
21.7	22.6	20.9	21.6	22.2	22.5	22.2	24.3	22.3	22.6
20.1	22.0	22.8	22.0	22.4	22.3	20.6	22.1	21.9	23.0
22.0	22.0	22.1	22.0	19.6	22.8	22.0	23.4	23.8	23.3
22.5	22.3	21.9	22.0	21.7	23.3	22.2	22.3	22.8	22.9
23.7	22.0	21.9	22.2	24.4	22.7	23.3	24.0	23.6	22.1
21.8	21.1	23.4	23.8	23.3	24.0	23.5	23.2	24.0	22.4
23.9	22.0	23.9	20.9	23.8	25.0	24.0	21.7	23.8	22.8
23.1	23.1	23.5	23.0	23.0	21.8	23.0	23.3	22.4	22.4

   - Construire un tableau de fréquences pour des classes de longueur 0.5 mm. Illustrer graphiquement les données à l'aide d'un histogramme, puis d'un polygone des fréquences cumulées. En déduire la médiane.
   - Calculer la moyenne et l'écart-type à partir des 100 données, puis à partir des données par classes.

2. **(*)** Les salaires hebdomadaires d'un échantillon de femmes d'une même entreprise en 1974 étaient:

Salaires hebdomadaires (Francs)	Nombre de femmes
moins de 150	1
entre 150 et 200	4
entre 200 et 250	28
entre 250 et 300	42
entre 300 et 350	33
entre 350 et 400	18
entre 400 et 450	13
entre 450 et 600	9
600 et plus	2

   (a) Représenter par un histogramme la répartition des salaires. Construire le polygône des fréquences cumulées et en déduire la médiane et les quantiles à 10% et 90% de cette distribution. Quel est le ratio entre ces deux quantiles (mesure de la répartition des salaires)?

   (b) Calculer la moyenne et l'écart type de cette distribution.

   (c) Que se passe t-il pour la médiane, la moyenne et l'écart-type si chaque femme est augmentée de 50 francs par semaine? Est augmentée de 10%?

3. **(*)** On a demandé à 1000 étudiants et à 1500 étudiantes la couleur de leurs yeux:

Couleur des yeux	Nombre d'étudiants	Nombre d'étudiantes
Bleu	420	650
Vert	120	150
Marron	450	680
Divers	10	20

   Représenter sur un même graphe les deux diagrammes à bâtons des fréquences. Peut-on conclure que les hommes ont des yeux de couleur différente de ceux des femmes?

4. **(*)** 60 % de la population gagne moins que le salaire moyen a dit un politicien. Est-ce posssible? Illustrer votre réponse.

5. **(*)** Calculer la moyenne et l'écart type du nombre de lettres dans les mots de cette phrase.

6. **(*)** Un ensemble d'observations a pour moyenne 20 et pour écart type 4. Quel est l'effet sur la moyenne et sur l'écart type de:

   - On ajoute 3 à toutes les observations.
   - On multiplie toutes les observations par 10.
   - On soustrait 12 à toutes les observations, puis on les divise par 2.

7. **(*)** On s'intéresse aux deux variables P et E, où P est le PNB/habitant (1994, en milliers de \$), et E l'espérance de vie 1994 à la naissance, pour les huit pays suivants:

Pays i	Grèce	Irak	Mali	Vietnam	Egypte	Japon	G-B	Brésil
P_i	6.2	4.1	0.3	0.2	0.6	28.2	17.8	2.8
E_i	77.2	51.5	45	64.5	59.5	79	75.9	66.2

   (a) Représenter le nuage de point associé à cette série statistique (P_i, E_i). Au vu de ce nuage, est-il raisonnable de faire une régression de E en P? Calculer les moyennes et les variances de P et de E. Calculer la covariance σ_{PE}. En déduire une équation de la droite de régression de E en P et la tracer. Calculer le coefficient de corrélation linéaire entre P et E.

   (b) La France a un PNB/hab de 22200\$. D'après les questions précédentes, donner une estimation de l'espérance de vie des Francais.

8. **(**)** Le tableau suivant indique la cote d'une voiture suivant le nombre d'années écoulées. X désigne l'âge de la voiture et Y la cote de la voiture en milliers de francs.

x_i	1	2	3	4	5
y_i	43	33.6	24.75	19.7	14.7

   (a) Représenter graphiquement cette série statistique par un nuage de points. Après avoir déterminer son équation, tracer la droite de régression de Y en X. Calculer le coefficient de corrélation et donner une estimation de la cote d'une voiture de 6 ans.

   (b) On pose $Z = ln(Y)$. Calculer les z_i, puis déterminer l'équation de la droite de régression de Z en X, ainsi que le coefficient de corrélation. Quelle est alors l'estimation de la cote d'une voiture de 6 ans?

9. **(***)** On suppose que X et Y sont deux variables dont on connaît n valeurs $(x_i, y_i)_{1 \leq i \leq n}$. On voudrait effectuer une regression polynômiale de degré 2 de Y sur X, c'est-à-dire que l'on voudrait déterminer les coefficients $(\hat{a}, \hat{b}, \hat{c})$ qui minimise la distance :

$$\sum_{i=1}^{n} (y_i - (ax_i^2 + bx_i + c))^2.$$

   La courbe de régression passe-t-elle encore par le point moyen du nuage? Déterminer l'expression de $(\hat{a}, \hat{b}, \hat{c})$.