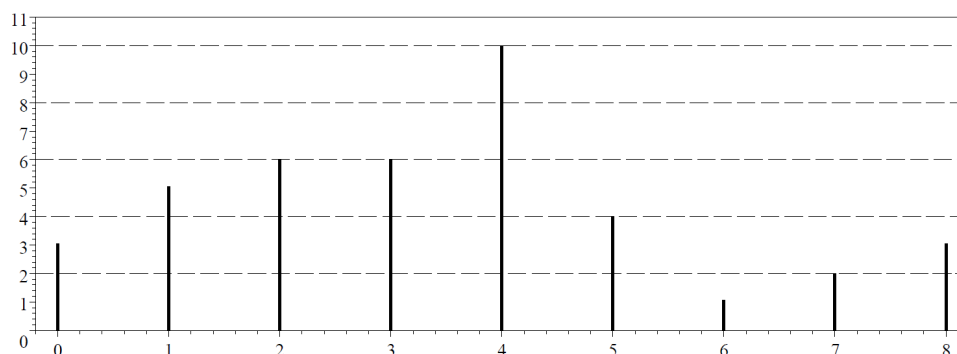


TD Ch 16 – Statistiques

I. Statistique univariée

1. On a demandé à des adolescents de 14 à 18 ans combien de fois ils allaient au cinéma par an. Le diagramme en bâtons ci-dessous présente leurs réponses :



1. Quelle est la population étudiée ? Quel est son effectif total ? Quel est le caractère étudié ?
2. Dresser le tableau statistique de la série en y plaçant les effectifs cumulés.
3. Déterminer l'étendue, le mode, la médiane, la moyenne et l'écart-type de cette série.

2. Dans une entreprise de 90 employés, la répartition des salaires (en milliers d'euros) est donnée par :

Classe de salaire	[5, 0 ; 5, 6[	[5, 6 ; 6, 0[	[6, 0 ; 6, 4[	[6, 4 ; 6, 8[	[6, 8 ; 7, 4[	[7, 4 ; 8, 2[	[8, 2 ; 9, 0[
Effectif	12	12	14	15	17	12	8

1. Construire l'histogramme associé à cette série.
2. Représenter graphiquement les effectifs cumulés, et en déduire une valeur approchée de la médiane et des quartiles.
3. Retrouver par le calcul la valeur (exacte) de la médiane.
4. Calculer la moyenne et l'écart-type de cette série statistique.

3. Lors d'un concours d'entrée à une grande école, deux jurys différents ont attribué aux candidats des notes sur 20, résumées par classes dans le tableau suivant :

Classe de note	[0; 2[	[2; 4[	[4; 6[	[6; 7[	[7; 8[	[8; 9[	[9; 10[	[10; 11[	[11; 13[	[13; 17[
Effectif x (jury A)	0	1	2	6	10	12	18	28	15	8
Effectif y (jury B)	1	2	1	10	14	16	18	24	11	3

On désire harmoniser les notes par *péréquation*. Pour ce faire, on modifie affinement les notes du jury B.

1. On pose donc $z = ay + b$, où a et b sont deux constantes réelles.
 - (a) Donner l'expression de la nouvelle moyenne $\bar{z}$ en fonction de la moyenne d'origine $\bar{y}$.
 - (b) À l'aide de la formule de König-Huygens et de la linéarité de la moyenne, montrer que $V_z = a^2 V_y$.
 - (c) En déduire une expression pour le nouvel écart-type s_z .
2. On souhaite obtenir, après modification, la même moyenne et le même écart-type pour les deux jurys. Comment doit-on choisir les coefficients a et b ?
On donnera l'expression littérale avant de faire l'application numérique.
3. Un candidat du jury B ayant obtenu initialement 9,2 a-t-il maintenant la moyenne ?

4. On considère le loyer mensuel de 1000 appartements parisiens.

Montant des loyers	Effectifs
[500 ; 1000[	30
[1000 ; 1500[	60
[1500 ; 2000[	80
[2000 ; 2500[	90
[2500 ; 3000[	150
[3000 ; 3500[	180
[3500 ; 4000[	150
[4000 ; 5000[	140
[5000 ; 6000[	80
[6000 ; 8000[	40

1. Donner la médiane et les quartiles de cette série, ainsi que l'écart interquartile.
Quel pourcentage des loyers, environ, se situe dans l'intervalle $[Q_1, Q_3]$?
2. Calculer le montant moyen des loyers ($\bar{x}$), puis l'écart-type (s_x).
3. Quel pourcentage des loyers se situe, approximativement, dans l'intervalle $[\bar{x} - s_x; \bar{x} + s_x]$?

5 (Paradoxe de Simpson). Un laboratoire pharmaceutique souhaite comparer l'efficacité de deux médicaments. Deux équipes font des tests, et obtiennent les résultats suivants :

Équipe 1	Succès	Échec	Total
Médicament A	90	10	100
Médicament B	321	40	361

Équipe 2	Succès	Échec	Total
Médicament A	200	60	260
Médicament B	51	23	74

1. Calculer les fréquences de succès des médicaments A et B obtenues par l'équipe 1, puis par l'équipe 2.
2. Quel médicament semble le plus efficace ?
2. On décide de rassembler les résultats des deux équipes. Calculer les nouvelles fréquences de succès des deux médicaments. Conclusion ?

II. Statistique bivariée

6. Le tableau suivant donne la distance de freinage d'un véhicule roulant sur route sèche en fonction de sa vitesse.

Vitesse en km/h	40	50	60	70	80	90	100	110
Distance en m	8	12	18	24	32	40	48	58

1. Représenter cette série par un nuage de point. Y placer le point moyen.
2. Calculer le coefficient de corrélation de cette série double.
3. En utilisant la méthode des moindres carrés, déterminer l'équation de la droite d'ajustement affine de la distance en fonction de la vitesse, et la tracer. On donnera les coefficients arrondis au dixième.
4. Estimer la distance de freinage d'un véhicule roulant à 120 km/h.

7. On s'intéresse dans cet exercice à deux séries statistiques, l'une donnant les revenus des salles de jeux vidéos et l'autre le nombre de doctorats obtenus en informatique aux États-Unis. Les données étudiées sont fournies dans le tableau suivant :

Année	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Revenus (millions \$)	1 196	1176	1269	1 240	1307	1435	1601	1654	1803	1734
Nb doctorats	861	830	809	867	948	1129	1453	1656	1787	1611

1. Représenter le nuage de points correspondant à ces deux séries, et placer le point moyen. Un ajustement affine vous paraît-il justifié ?
2. Calculer les écart-types de ces deux séries.
3. Calculer la covariance ainsi que le coefficient de corrélation linéaire correspondant à ces deux séries, et commenter les résultats.
4. Calculer la droite de régression linéaire et la tracer. Y-a-t-il un lien de cause à effet entre les deux séries étudiées ?