

TD1

Effectuer les représentations graphiques suivantes du poly de cours en respectant bien les recommandations :

- Diagramme circulaire de « Groupe sanguin » (variable qualitative nominale) (page 10),
- Diagramme en bâtons de « Âge » (variable quantitative discrète) (page 11),
- Histogrammes de « Poids » version 1 et version 2 (variable quantitative continue regroupée en classes) (page 11).

TD2

Exercice 1 :

Effectuer les représentations graphiques du cours suivantes :

- Fonction de répartition empirique de « Nombre d'enfants » (variable quantitative discrète)
- Polygone des fréquences cumulées de « Poids » version 2 (variable quantitative continue regroupée en classes)

Exercice 2 (Extrait DSI 2024-2025) :

On s'intéresse au « Nombre d'arbres » X plantés sur les parcelles d'un lotissement. Les données obtenues sont les suivantes :

Nombre d'arbres	1	2	4	5
Nombre de parcelles	3	7	5	5

1. Donner la taille n de l'échantillon observé et le type de la variable X .
2. Représenter graphiquement la distribution statistique de X .
3. Quel est le mode de X ?
4. Calculer la valeur moyenne de X .
5. Calculer la variance et l'écart-type de X .
6. Donner dans un tableau les fréquences cumulées associées aux valeurs de X .
7. Rappeler la définition de la fonction de répartition empirique de X et la tracer.
8. Déterminer, graphiquement **et** également grâce aux valeurs calculées à la question 6, la médiane de la variable X .
9. Déterminer, graphiquement **et** également grâce aux valeurs calculées à la question 6, le premier et le troisième quartiles (Q_1 et Q_3 respectivement) de la variable X .
10. Que signifie la valeur du 3^{ème} quartile pour la variable étudiée ?
11. Quel est le pourcentage de parcelles sur lesquelles on a planté moins de 3 arbres ?

Exercice 3 (Extrait DSI 2024-2025) :

On présente ci-dessous les résultats d'une enquête concernant 200 loyers annuels (en milliers d'euros) dans un quartier d'une ville.

Loyers	]4 ; 6]	]6 ; 8]	]8 ; 9]	]9 ; 10]	]10 ; 15]	]15 ; 25]
Effectifs	20	45	40	45	26	24

1. Soit X la variable « Loyer annuel ». Quel est le type de la variable X ?
2. Dessiner l'histogramme associé à cette distribution. On prendra $a = 2$ comme amplitude de référence.
3. Calculer une valeur approchée de la moyenne de X et préciser l'unité de mesure du résultat.
4. Calculer une valeur approchée de la variance de X et préciser l'unité de mesure du résultat.
5. Calculer les fréquences cumulées associées à chaque intervalle.
6. Construire le polygone des fréquences cumulées de X .
7. Déterminer graphiquement **et** à partir des valeurs calculées à la question 5 dans quel intervalle se situe la médiane de X .
8. Montrer que la droite permettant de déterminer plus précisément la médiane a pour équation : $y = 0,2x - 1,275$.
9. Quelle valeur de la médiane pouvez-vous en déduire ?
10. Que signifie cette valeur pour les loyers de cet échantillon ?
11. Déterminer la valeur du 3^{ème} quartile (noté Q_3) de X .

Exercice 4 :

En 1996, une enquête concernant les 100 salariés d'une entreprise contient, entre autres, les informations reportées dans le Tableau 1 ci-dessous. On y trouve :

- Le salaire annuel net de l'année 1995 noté X . On notera pour alléger les calculs $Y = X/1000$.
- Le sexe codé ainsi : 1 Homme, 2 Femme.
- la catégorie socio-professionnelle (csp) codée ainsi : 1 Ouvrier non qualifié, 2 Ouvrier qualifié, 3 Employé, 4 Professions intermédiaires (techniciens, agents de maîtrise), 5 Cadre.
- Le nombre de jours d'absence au travail durant les 100 premiers jours ouvrables de l'année 1995, noté Z .

Pour alléger les calculs, on donne les résultats numériques suivant concernant la série Y :

$$y_i = x_i/1000; \quad \sum_{i=1}^{100} y_i = 11338,8; \quad \sum_{i=1}^{100} y_i^2 = 1580681,12$$

Tableau 1

i	Salaire x_i	csp	Absences z_i	sexe	i	Salaire x_i	csp	Absences z_i	sexe
1	58000	1	7	2	51	94000	2	7	1
2	60000	1	8	2	52	94400	3	2	1
3	63000	1	7	2	53	95000	2	4	1
4	65000	1	5	2	54	95500	2	5	1
5	67000	1	7	2	55	96000	2	6	1
6	70000	1	7	1	56	96500	3	0	1
7	71000	1	4	1	57	97000	2	3	1
8	72000	2	6	2	58	97500	2	6	1
9	72500	1	3	1	59	98000	2	4	1
10	73000	1	2	1	60	98600	2	5	1
11	73500	2	7	2	61	99000	2	3	1
12	74000	1	4	1	62	100000	2	2	1
13	75000	2	8	2	63	101000	2	6	1
14	75500	1	5	1	64	101400	2	4	1
15	76000	1	5	1	65	102000	2	3	1
16	76500	2	7	2	66	103000	2	4	1
17	77000	2	5	2	67	103500	2	4	1
18	77500	1	3	1	68	104000	2	3	1
19	78000	1	7	1	69	105000	2	1	1
20	79000	2	6	2	70	105200	2	8	1
21	80000	3	5	2	71	120000	4	4	2
22	80000	1	3	1	72	130000	4	3	2
23	80500	2	4	2	73	130500	4	2	1
24	81000	2	5	2	74	131000	4	2	1
25	81500	3	3	2	75	132000	4	3	1
26	83000	2	9	2	76	133000	4	3	1
27	84000	2	4	2	77	134000	4	4	1
28	84500	3	5	2	78	135000	4	2	1
29	85000	2	7	2	79	137000	4	1	1
30	85500	3	3	2	80	138000	4	2	1
31	86000	3	2	2	81	140000	4	3	1
32	86500	2	6	1	82	142000	4	2	1
33	87000	2	6	2	83	143000	4	3	1
34	87200	2	6	1	84	145000	4	1	1
35	87500	3	5	2	85	146000	4	1	1
36	87600	2	6	1	86	148000	4	1	1
37	88000	2	7	2	87	150000	4	2	1
38	88500	2	6	1	88	152000	4	2	1
39	89000	3	5	2	89	153000	4	3	1
40	89200	2	7	1	90	160000	4	2	1
41	89500	2	4	1	91	170000	5	1	2
42	90000	3	5	2	92	200000	5	2	2
43	90200	2	4	1	93	215000	5	2	1
44	90400	2	3	1	94	220000	5	1	2
45	90600	2	3	1	95	230000	5	0	1
46	91000	2	5	1	96	240000	5	2	1
47	92000	2	3	1	97	260000	5	0	1
48	92500	3	3	1	98	290000	5	0	1
49	93000	2	4	1	99	320000	5	1	1
50	93500	2	6	1	100	360000	5	1	1

Partie 1 :

Donner dans un tableau d'effectifs à double entrée, la répartition des salariés selon le sexe et la catégorie socio-professionnelle ; puis proposer une représentation graphique susceptible d'illustrer le tableau.

Partie 2 :

On s'intéresse d'abord à la distribution de la variable quantitative discrète Z : « Nombre de jours d'absence au travail ». La répartition de cette variable sur l'ensemble des salariés, la sous-population des hommes et la sous-population des femmes est donnée dans le Tableau 2.

Tableau 2

Nombre de jours d'absence a_j	n_j	n_j^H	n_j^F
0	4	4	0
1	9	7	2
2	15	13	2
3	19	16	3
4	14	11	3
5	13	5	8
6	11	8	3
7	11	4	7
8	3	1	2
9	1	0	1
Total	100	69	31

2-a) À partir du Tableau 2 construire le tableau de fréquences de la variable Z sur l'ensemble des salariés et sur les deux sous-populations. On notera f_j la fréquence (relative) de la modalité a_j et f_j^H et f_j^F les fréquences de la modalité a_j pour les hommes et pour les femmes respectivement. Quelle relation lie ces trois fréquences ?

2-b) Calculer à partir du Tableau 2 le nombre moyen de jours d'absence pour l'ensemble des salariés, pour les hommes, pour les femmes. Quelle relation lie ces trois moyennes ?

2-c) Représenter la distribution de la variable Z concernant l'ensemble des salariés à l'aide d'un diagramme en bâtons ; puis, en construire la fonction de répartition empirique. Donner la médiane de la distribution.

2-d) Calculer la variance de Z pour les hommes et pour les femmes. Quelle relation lie les variances des deux sous-populations à la variance du mélange ? En déduire la variance de Z pour l'ensemble des salariés, puis son écart-type.

Partie 3 :

On s'intéresse maintenant à la distribution de la variable quantitative continue X : « Salaire net ». En choisissant comme extrémités de classes 70000, 80000, ..., 250000, on obtient la distribution des salaires donnée dans le tableau suivant :

Tableau 3

Classe	Effectif
≤ 70000	6
]70000, 80000]	16
]80000, 90000]	20
]90000, 100000]	20
]100000, 125000]	9
]125000, 150000]	16
]150000, 200000]	5
]200000, 250000]	4
>250000	4

3a) Représenter graphiquement cette distribution : on tracera l'histogramme, le « début » de la fonction de répartition et le polygone des fréquences. Déterminer graphiquement, à partir du polygone des fréquences et par interpolation linéaire, le salaire médian.

3b) Calculer le salaire moyen à partir Tableau 3 puis à partir du Tableau 1. Comparer.

3c) Calculer à partir du Tableau 1 : le premier et le dernier décile, les quartiles. Comparer le second quartile au salaire médian trouvé en 2a).

3d) Calculer des indicateurs de dispersion des salaires à partir des quantiles de la distribution.

3e) Calculer la variance et l'écart-type de X .

TD3**Exercice 1 (Extrait DST 2024-2025).**

Dans une école préparatoire à un examen, trois méthodes différentes d'apprentissage sont appliquées à 247 étudiants ayant le même niveau initial et répartis en trois groupes d'étudiants. On note X la variable correspondant à la méthode utilisée (parmi 3) et le résultat à l'examen est donné par la variable Y. Les chiffres obtenus sont les suivants :

Y (Résultat à l'examen) X (Méthode)	Admis	Ajournés
Méthode 1	51	29
Méthode 2	38	12
Méthode 3	86	31

1. Donner le type des deux variables étudiées.
2. Quel est le nom statistique de ce tableau ?
3. Ajouter les totaux lignes et colonnes dans le tableau.
4. Quel est le pourcentage d'étudiants admis au concours parmi ceux ayant suivi la méthode 3 ?
5. Quel est le pourcentage d'étudiants ayant suivi la méthode 1 parmi ceux qui sont ajournés au concours ?
6. Calculer l'indice du khi-deux associé au croisement des deux variables X et Y.
7. Comparer la valeur du khi-deux à la valeur seuil 5,99. Que peut-on en conclure ?

Exercice 2 (Extrait DST 2024-2025).

Le tableau ci-dessous représente l'évolution du revenu disponible brut X et de la consommation des ménages Y en euros pour un pays donné sur la période 1992-2001 :

Année	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
Revenu	8000	9000	9500	9500	9800	11000	12000	13000	15000	16000
Consommation (en euros)	7390	8170	8832	8653	8788	9616	10593	11186	12758	13870

On donne également les résumés numériques suivants :

$$\sum_{i=1}^{10} x_i = 112\,800 \quad \sum_{i=1}^{10} y_i = 99\,856 \quad \sum_{i=1}^{10} x_i^2 = 1\,336\,540\,000$$

$$\sum_{i=1}^{10} y_i^2 = 1\,036\,417\,742 \quad \sum_{i=1}^{10} x_i y_i = 1\,176\,479\,900$$

1. Donner le type de chacune des 2 variables.

2. Calculer la moyenne et la variance de Y.
3. Calculer la covariance des variables X et Y. Commenter le signe de cette valeur.
4. La variance de X étant égale à 6 415 600 calculer le coefficient de corrélation linéaire de X et Y. Que peut-on déduire ?
5. Déterminer l'équation de la droite de régression exprimant Y en fonction de X.
6. Interpréter la pente de cette droite.
7. Estimer la consommation en 2003 pour un revenu observé de 17500 euros.

Exercice 3 (Extrait DST 2024-2025).

Nous nous intéressons ici à la teneur en vitamine C de différentes variétés de pommes. Soit X la variable « Variété de pommes » prenant les valeurs V1, V2 et V3 , et soit Y la variable « Teneur en vitamine C » (exprimée en mg/100g).

Le tableau ci-dessous contient $n = 17$ mesures de Y réparties selon la variété de pommes :

Variété V1 ($n_1=5$)	Variété V2 ($n_2=7$)	Variété V3 ($n_3=5$)
93,6 95,3 96 93,7 96,2	98,8 98,2 97,8 97,2 97,9 98 97,5	94,6 97,8 98 95 98,9

1. Donner le type des 2 variables X et Y.
2. Calculer la moyenne et la variance de Y sur la totalité des 17 mesures. On donne :

$$\sum_{i=1}^{17} y_i = 1\,644,5 \quad \sum_{i=1}^{17} y_i^2 = 159\,129,41.$$

3. Montrer que la moyenne conditionnelle de Y sachant que $X=V1$ (notée $\bar{y}_{|x=V1}$ ou encore $\bar{y}_1$) est égale à 94,96.
4. Montrer que la variance conditionnelle de Y sachant que $X=V1$ (notée $S_{y|x=V1}^2$ ou encore S_1^2) vaut 1,2344.
5. On donne les moyennes et variances conditionnelles de Y sachant que $X=V2$ et sachant que $X=V3$:

$$\bar{y}_{|x=V2} = \bar{y}_2 = 97,914 ; S_{y|x=V2}^2 = S_2^2 = 0,224 ;$$

$$\bar{y}_{|x=V3} = \bar{y}_3 = 96,86 ; S_{y|x=V3}^2 = S_3^2 = 2,9824.$$

Calculer la moyenne des variances conditionnelles de Y sachant X.

6. Calculer la variance des moyennes conditionnelles de Y sachant X.
7. Calculer le rapport de corrélation de Y sachant X. Commenter ce résultat.

TD4**Exercice 1 :**

On veut savoir si la réussite (R) d'un traitement est indépendante du niveau de tension artérielle (T) du malade. On dispose pour cela de 250 observations réparties comme suit :

	Échec	Succès
Basse	21	104
Élevée	29	96

La réussite du traitement dépend-elle du niveau de tension du malade ? (valeur seuil = 3,84).

Exercice 2 :

Une variable biologique Y est mesurée dans un échantillon de 308 adultes. Cette variable est un indicateur de bonne forme physique : plus elle est élevée, meilleure est la qualité de vie physique du sujet. On a relevé aussi l'âge X des sujets.

On donne :

$$\sum_{i=1}^{308} x_i = 14875 \quad \sum_{i=1}^{308} x_i^2 = 825699 \quad \sum_{i=1}^{308} y_i = 9393 \quad \sum_{i=1}^{308} y_i^2 = 311359 \quad \sum_{i=1}^{308} x_i y_i = 423145$$

1. Calculer la moyenne et la variance de l'âge.
2. Calculer la moyenne et la variance de la variable Y.
3. Calculer le coefficient de corrélation de (X,Y) et les coefficients de la droite de régression. Interpréter les résultats.

Exercice 3.

Dans une expérience concernant la réaction à un signal lumineux, on a réparti de façon aléatoire 20 sujets en cinq groupes équilibrés. Les sujets de chaque groupe sont soumis à une lumière d'une couleur différente. Le temps de réaction des sujets au signal lumineux est mesuré (en secondes) et les résultats sont donnés dans le tableau suivant :

Type de lumière					
	A	B	C	D	E
	10	7	10	10	3
	7	6	11	10	6
	9	9	12	11	6
	6	8	9	10	4
Moyenne	8	7,5	10,5	10,25	4,75

On donne le résultat suivant pour les temps y_i :

$$\sum_{i=1}^{20} (y_i - \bar{y})^2 = 115,2$$

On souhaite savoir s'il existe un effet de lumière sur le temps de réaction.

1. Calculer la moyenne et la variance du temps de réaction pour le groupe total.
2. Calculer les variances conditionnelles.
3. Calculer le rapport de corrélation du temps de réaction suivant le type de lumière. Commenter.