

Statistiques

Licence 1, Semestre 1

Économie et Gestion

Université de Bourgogne

Cours de Oğuzhan AKGÜN, MCF

Thèmes des séances de travaux dirigés

Fiche de TD n°1	Graphes et Tableaux (Séances 1 et 2)
Fiche de TD n°2	Indicateurs Numériques (Séance 3, 4 et 5)
Contrôle Continu	Graphes, Tableaux et Indicateurs Numériques (Séance 6)
Fiche de TD n°3	Indicateurs Numériques & Forme d'une Distribution (Séances 7 et 8)
Fiche de TD n°4	Probabilité et Variables Aléatoires (Séance 9 et 10)

Fiche de TD n°1

Graphes et Tableaux

Exercice 1 Le tableau suivant présente les valeurs prises par x_i et y_i , $i = 1, 2, \dots, N$ avec $N = 5$.

i	1	2	3	4	5
x_i	9	4	2	8	3
y_i	5	0	4	-3	-1

Calculer les valeurs des expressions suivantes :

1. $\sum_{i=1}^N x_i$,
2. $\sum_{i=2}^N x_{i-1}$,
3. $\sum_{i=1}^N x_i - 10$,
4. $\sum_{i=1}^N (x_i - 10)$,
5. $\sum_{i=1}^N (x_i + y_i)(x_i - y_i)$,
6. $\sum_{i=2}^N x_{i-1}y_i$.

Exercice 2 Indiquer la population, l'échantillon, l'unité statistique et les types des variables étudiées dans les cas suivants.

1. Afin d'étudier la santé et les comportements alimentaires des étudiants français, une équipe de scientifiques a distribué un questionnaire aux étudiants rouennais.¹ Avec ce questionnaire, ils ont collecté des informations de 1 744 étudiants, sur trois variables **imc** (l'indice de masse corporelle, kg/m²), **sexe** (femme ou homme) et **regime** (=1 si l'étudiant a déjà fait un régime et =0 sinon).
2. Dans une étude concernant l'effet de la démocratie sur la richesse des pays du monde,² les chercheurs utilisent l'indice des droits politiques de Freedom House (**droit**, entre 0 et 1) et le PIB par habitant (**pib**, dollars) de 150 pays du monde.

Exercice 3 Indiquer et créer les graphiques les plus adaptés pour les données suivantes.

1. Les parts de marché des navigateurs webs les plus populaires en France en mars 2021, présentées dans le tableau suivant :

Chrome	Safari	Firefox	Edge	Samsung Internet	Autres
56.38	21.58	8.82	5.19	3.66	4.37

2. L'évaluation de la part de marché de Chrome entre avril 2021 et mars 2022 en France, présentée dans le tableau suivant :

11/21	12/21	01/22	02/22	03/22
58.31	57.45	56.79	57.28	56.45

1. Grigioni et al. (2007).
2. Acemoglu et al. (2008).

Exercice 4 Le tableau suivant présente les nombres (les fréquences) d'espèces menacées parmi les vertébrés en 2022 en France.³

TABLE 1 – Nombre d'espèces menacées parmi les vertébrés en France, par classes

Mammifères	Oiseaux	Reptiles	Amphibiens	Poissons	Total
4	33	3	3	11	54

1. En utilisant les fréquences données dans le tableau, créer un graphique à barres pour comparer les espèces menacées en France.
2. Représenter les données dans un tableau, en utilisant les fréquences relatives. Interprétation.

Exercice 5 Le tableau ci-dessous présente la consommation du boeuf (y , en kg par an, par habitant) dans 34 pays en 2020.⁴

TABLE 2 – Consommation du boeuf (kg par an, par habitant) dans 34 pays en 2020

		COL	8.62	ISR	23.21	NOR	12.81	SAU	3.87
AUS	19.44	EGY	7.80	JPN	7.64	NZL	11.55	THA	1.23
BRA	24.45	ETH	2.49	KAZ	19.83	PAK	6.42	TUR	9.63
CAN	17.58	GBR	11.33	KOR	11.85	PER	4.14	UKR	4.64
CHE	13.11	IDN	2.18	MEX	8.97	PHL	3.13	USA	26.20
CHL	20.26	IND	0.53	MYS	5.41	PRY	12.03	VNM	8.25
CHN	4.19	IRN	5.77	NGA	1.31	RUS	10.03	ZAF	11.40

1. Présenter la distribution sous la forme d'un tableau des fréquences pertinentes après avoir regroupé les données dans 5 classes de la même largeur de 5.2 kg.
2. Tracer l'histogramme de cette distribution. Interprétation.
3. Construire le graphe Ogive. Interprétation.

Exercice 6 Le tableau suivant présente les nombres des espèces menacées parmi les vertébrés en 2022 dans 3 régions.⁵

TABLE 3 – Nombre des espèces menacées dans 3 pays, par classes

$x \backslash y$	Mammifères	Oiseaux	Reptiles	Amphibiens	Poissons	Total
Australie	41	55	20	8	20	144
France	4	33	3	3	11	54
États-Unis	28	23	29	16	77	173
Total	73	111	52	27	108	371

3. Sources : The IUCN Red List (Liste rouge de l'Union Internationale pour la Conservation de la Nature). https://stats.oecd.org/Index.aspx?DataSetCode=WILD_LIFE. Accédé le 03/04/22.

4. Sources : OECD-FAO Agricultural Outlook. <https://data.oecd.org/agroutput/meat-consumption.htm>. Accédé le 03/04/22.

5. Sources : The IUCN Red List.

1. Représenter les données en utilisant les fréquences relatives marginales.
2. Calculer les fréquences relatives conditionnelles de y .

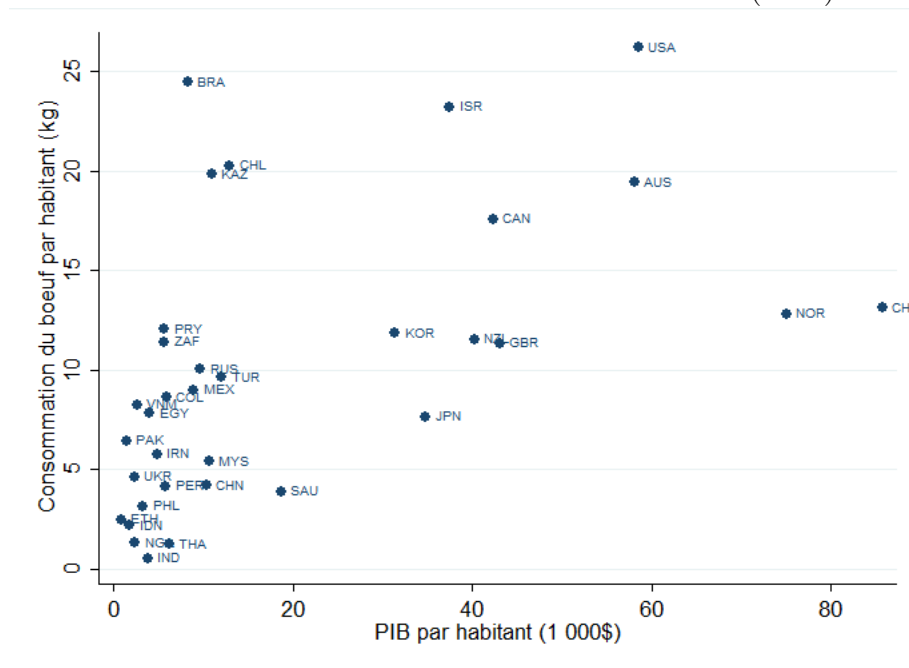
Exercice 7 Considérons les données dans le Tableau 2. Pour étudier la relation entre la consommation du boeuf et la richesse des pays, on donne aussi le PIB (x , en 1 000\$ par habitant) de ces 34 pays en 2020.⁶ Les données sont présentées dans le tableau ci-dessous.

TABLE 4 – PIB (1 000\$ par habitant) dans 34 pays en 2020

		COL	5.89	ISR	37.48	NOR	75.02	SAU	18.69
AUS	58.04	EGY	4.03	JPN	34.81	NZL	40.22	THA	6.20
BRA	8.23	ETH	0.83	KAZ	10.97	PAK	1.45	TUR	12.04
CAN	42.29	GBR	43.02	KOR	31.36	PER	5.79	UKR	2.34
CHE	85.68	IDN	1.81	MEX	8.91	PHL	3.27	USA	58.56
CHL	12.95	IND	3.76	MYS	10.63	PRY	5.67	VNM	2.66
CHN	10.37	IRN	4.88	NGA	2.40	RUS	9.67	ZAF	5.66

1. Présenter la distribution conjointe de ces deux variables sous la forme d'un tableau de contingence après avoir regroupé les données sur le PIB par habitant (x) dans les classes suivantes : $[0; 20)$, $[20; 40)$, $[40; 60)$, $[60; 80)$, $[80; 100)$.
2. Commenter le graphique de nuage suivant.

FIGURE 1 – Consommation du boeuf et le PIB (2020)



6. Sources : World Development Indicators. <https://databank.worldbank.org/source/world-development-indicators>. Accédé le 03/04/22.

Fiche de TD n°2

Indicateurs Numériques

Exercice 1 Soit une variable x mesurée en t et $t+h$. Notons $\theta_{t+h|t}$ son coefficient multiplicateur et $\gamma_{t+h|t}$ son taux de croissance entre les dates t et $t+h$. Montrer que $\gamma_{t+h|t} = \theta_{t+h|t} - 1$.

Exercice 2 Les nombres de nouveaux cas de coronavirus en France (x_t) sont indiqués dans le tableau ci-dessous.⁷

Date	t	x_t
18/11/21	1	28 619
24/11/21	2	9 690
01/12/21	3	8 256

Calculer $\theta_{3|2}$, $\gamma_{3|2}$, $\theta_{3|1}$, $\gamma_{3|1}$.

Exercice 3 Nous lançons une pièce 20 fois et trouvons la suite 1, 0, 1, 1, 1, 1, 1, 0, 1, 0, 1, 0, 0, 0, 0, 1, 1, 0, 1, 0 où 1 représente PILE et 0 représente FACE. Trouver la moyenne empirique et la variance empirique de cette suite.

Exercice 4 Dans cet exercice, nous nous focalisons sur un sous-échantillon des données présentées dans les Tableaux 2 et 4. Les données de 6 pays sur les deux variables, la consommation du boeuf (y , en kg par an, par habitant) et le PIB (x , en 1 000\$ par habitant), sont présentées dans le tableau suivant :

TABLE 5 – Consommation du boeuf et PIB dans 6 pays en 2020

	CAN	GBR	ISR	JPN	KOR	NZL
y	17.58	11.33	23.21	7.64	11.85	11.55
x	42.29	43.02	37.48	34.81	31.36	40.22

1. Calculer la moyenne empirique et la médian de chacune des variables. Interprétation.
2. Calculer Q_1 et Q_3 de chacune des variables. Interprétation.
3. Calculer l'étendue et l'écart inter-quartile de chacune des variables. Interprétation.
4. Calculer la variance empirique, l'écart-type empirique et le coefficient de variation de chacune des variables. Interprétation.
5. Pour chacune des variables, calculer le z-Score. En utilisant le z-Score de y , vérifier que la moyenne et la variance d'un z-Score sont respectivement 0 et 1.

Exercice 5 Pour cet exercice, on dispose d'une base de données françaises de deux séries annuelles entre 2015 et 2020 : le taux de chômage (x_t , la population en chômage divisée par la population en âge de travailler) et le taux d'inflation (y_t , le taux de changement des prix à la consommation).

7. Sources : Ritchie et al. (2020). Accédé le 03/04/22.

TABLE 6 – Taux de chômage et taux d’inflation en France (2015-2020)

	2015	2016	2017	2018	2019	2020
Taux de chômage (x_t)	10.35	10.05	9.41	9.02	8.41	8.01
Taux d’inflation (y_t)	0.04	0.18	1.03	1.85	1.11	0.48

1. Pour chacune des variables, calculer la série qui représente son taux de croissance annuel.
2. Calculer la moyenne du taux de croissance annuel de x_t .

Exercice 6 Cet exercice est consacré à l’étude de la relation entre la consommation du boeuf (y , en kg par an, par habitant) et le PIB (x , en 1 000\$ par habitant) en utilisant les données présentées dans le Tableau 5.

1. Calculer la covariance empirique entre y et x .
2. Calculer le coefficient de corrélation entre y et x . Interprétation.

TABLE 7 – Consommation du boeuf et PIB dans 4 pays en 2020

country	AUS	CHE	NOR	USA
y	19.44	13.11	12.81	26.20
x	58.04	85.68	75.02	58.56

Le tableau ci-dessus présente un autre sous-échantillon des données des Tableaux 2 et 4. Calculer le coefficient de corrélation entre y et x en utilisant cet échantillon et comparer le résultat avec celui de la question précédente. Conclusion(s).

Fiche de TD n°3

Indicateurs Numériques & Forme d'une Distribution

Exercice 1 En utilisant le tableau de fréquences obtenu dans la correction de l'Exercice 5 de la fiche de TD n°1, trouver une valeur approximative pour la moyenne et l'écart-type de la variable y (consommation du boeuf, kg par an, par habitant).

Nous utilisons les données présentées dans le Tableau 2 de cet exercice et nous trouvons $\bar{y} = 10.04$. Comparer la valeur approximative que vous avez trouvée avec cette valeur exacte. Quelle est l'erreur d'approximation en pourcentage par rapport à la valeur exacte ?

Exercice 2 Le tableau ci-dessous contient les données sur le pourcentage d'entreprises ayant une femme à la tête de la direction (x) dans 62 pays.⁸

TABLE 8 – Pourcentage d'entreprises ayant une femme à la tête de la direction

ALB	18.1	EST	22.5	LTU	30.7	RWA	21.8
ARG	8	GMB	9.6	LUX	7.4	SRB	18.2
ARM	19.1	GEO	16.5	MLT	11.3	SLE	15.9
AZE	16.5	GRC	17.2	MDA	18.6	SVK	22.9
BLR	21.1	GTM	18.5	MNG	38.9	SVN	18.8
BEL	17.9	HUN	22.1	MNE	15	SUR	11.9
BOL	26.3	ITA	15.3	MAR	5.4	TJK	6.6
BIH	16.6	JOR	3.1	MOZ	15.6	TUN	10.4
BGR	28.8	KAZ	26	NER	10.6	TUR	3.9
TCO	12	KEN	18.1	MKD	21.3	UKR	17.7
COL	18.9	XKX	2.7	PRY	19.6	URY	10.6
HRV	27	KGZ	32.9	PER	19.9	UZB	12.4
CYP	8.2	LAO	43.1	POL	27.8	PSE	0.9
CZE	16.1	LVA	32.6	PRT	14	ZMB	12.9
ECU	22.9	LBN	5.9	ROU	17.2		
EGY	6.3	LBR	20.4	RUS	24.3		

1. Tracer l'histogramme de cette distribution après avoir regroupé les données dans les classes suivantes : $[0; 10)$, $[10; 20)$, $[20; 30)$, $[30; 40)$, $[40; 50)$.
2. Trouver le mode de la distribution en utilisant les résultats de la question précédente.
3. Après avoir réordonné les données dans l'ordre croissant, nous trouvons les valeurs suivantes : $x_{(1)} = 0.9$; $x_{(7)} = 6.3$; $x_{(8)} = 6.6$; $x_{(16)} = 11.3$; $x_{(17)} = 11.9$; $x_{(31)} = 17.2$; $x_{(32)} = 17.7$; $x_{(46)} = 21.3$; $x_{(47)} = 21.8$; $x_{(55)} = 27$; $x_{(56)} = 27.8$; $x_{(62)} = 43.1$. Calculer et interpréter les statistiques suivantes :
 - (a) Q_1 , Q_2 , Q_3 .
 - (b) L'étendue et l'écart inter-quartile.
 - (c) $q_{0.1}$, $q_{0.5}$, $q_{0.9}$.

8. Sources : Enterprise Surveys. <https://databank.worldbank.org/source/enterprise-surveys>. Accédé le 02/10/22.

4. Définissons $\tilde{x}_i = x_i - \bar{x}$. On donne les statistiques suivantes :

$$\begin{aligned} \sum_{i=1}^{62} x_i &= 1072.8; & \sum_{i=1}^{62} x_i^2 &= 23092.58; \\ \sum_{i=1}^{62} \tilde{x}_i^3 &= 19982.32; & \sum_{i=1}^{62} \tilde{x}_i^4 &= 1157437.82. \end{aligned}$$

Calculer et interpréter les statistiques suivantes :

- (a) La variance, l'écart-type et le coefficient de variation.
 - (b) m_3 et m_4 .
5. Tracer le diagramme en boîte. Interprétation.

Exercice 3 Le tableau de fréquences ci-dessous présente le nombre de pays du monde par l'âge de début de l'école primaire (x années). Les valeurs prises par la variable sont montrées avec C_l , $l = 1, \dots, L$.⁹

TABLE 9 – Âge de début de l'école primaire (années)

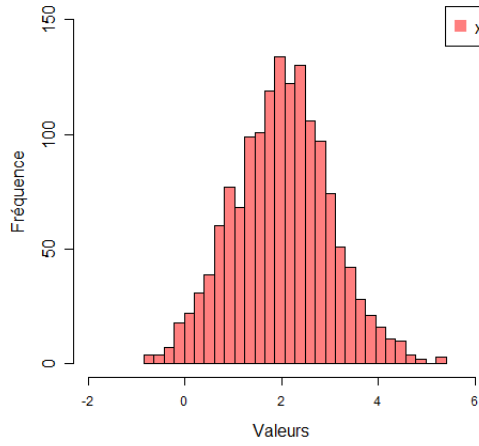
C_l	Fréquence
5	25
6	151
7	34
Total	210

1. Présenter la distribution sous la forme d'un tableau des fréquences pertinentes.
2. Trouver la moyenne, le mode et l'écart-type de cette variable.
3. Présenter la distribution à l'aide d'un graphique à barres. La distribution est-elle symétrique ?
4. Calculer la mesure d'asymétrie de Pearson et le coefficient d'aplatissement de Pearson. Conclusion(s).

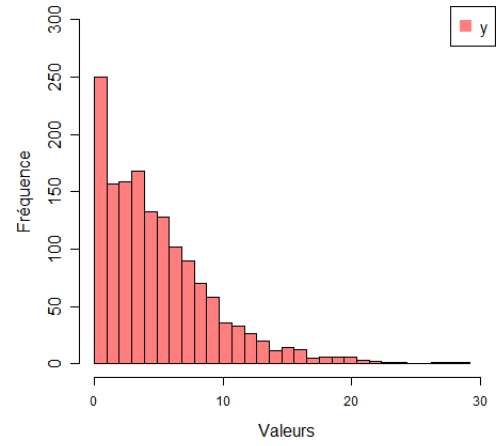
Exercice 4 Nous observons deux variables x_i et y_i , $i = 1, \dots, 1500$, dont les histogrammes avec $L = 30$ sont présentés dans Figure 2(a) et Figure 2(b). En utilisant ces échantillons, nous trouvons les statistiques suivantes :

- $\bar{x} = 2$, $s_x^2 = 1$, $ETE(x) = 6.25$, $EIQ(x) = 1.32$, $m_3(x) = 0.056$, $m_4(x) = 2.97$,
 - $\bar{y} = 5$, $s_y^2 = 18.42$, $ETE(y) = 29.15$, $EIQ(y) = 5.25$, $m_3(y) = 114.13$, $m_4(y) = 1987.47$.
1. Pour chacun des cas suivants, calculer $\bar{z}$, s_z^2 , $ETE(z)$, $EIQ(z)$, $m_3(z)$, $m_4(z)$, la mesure d'asymétrie de Pearson et le coefficient d'aplatissement de Pearson :
 - (a) $z_i = 2x_i$,
 - (b) $z_i = -y_i$.
 2. Commenter les histogrammes de ces transformations des variables x_i et y_i , présentés dans Figure 2(c) et Figure 2(d) en les comparant avec les histogrammes des variables originales.

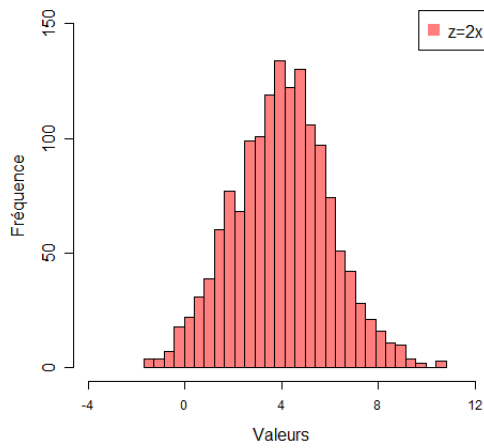
9. Sources : World Development Indicators. <https://databank.worldbank.org/source/world-development-indicators>. Accédé le 02/10/22.



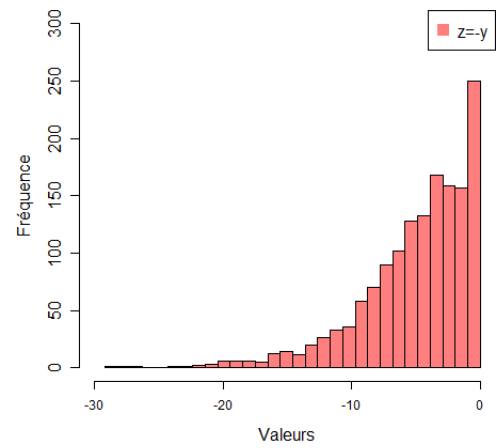
(a) Histogramme de x_i



(b) Histogramme de y_i



(c) Histogramme de $z_i = 2x_i$



(d) Histogramme de $z_i = -y_i$

FIGURE 2 – Histogrammes de x_i , y_i et ses transformations

Fiche de TD n°4

Probabilité et Variables Aléatoires

Exercice 1 Une expérience en trois étapes a trois résultats possibles à la première étape, deux résultats possibles à la seconde étape et quatre résultats possibles à la troisième étape. Combien de résultats possibles existe-il pour l'expérience considérée dans son ensemble ?

Exercice 2 De combien de façons peut-on sélectionner trois éléments parmi six,

1. quel que soit l'ordre,
2. en tenant compte de l'ordre de tirage.

Exercice 3 À partir d'un jeu ordinaire de 52 cartes, dénombrez le nombre de mains de 3 cartes contenant

1. 3 as (ou brelan d'as),
2. 3 cœurs,
3. un brelan.

Exercice 4 Considérer l'expérience qui consiste à lancer trois fois une pièce de monnaie.

1. Énumérer les résultats possibles de l'expérience.
2. Quelle est la probabilité de chaque résultat possible ?

Exercice 5 Supposons qu'un espace-échantillon soit composé de cinq résultats possibles équiprobables : $\omega_1, \dots, \omega_5$. Soient

$$\begin{aligned}A &= \{\omega_1, \omega_2\}, \\B &= \{\omega_3, \omega_4\}, \\C &= \{\omega_2, \omega_3, \omega_5\}.\end{aligned}$$

1. Calculer $P(A)$, $P(B)$ et $P(C)$.
2. Calculer $P(A \cup B)$. Les événements A et B sont-ils mutuellement exclusifs ?
3. Déterminer A^c , C^c , $P(A^c)$ et $P(C^c)$.
4. Déterminer $A \cap C$ et $P(A \cap C)$.

Exercice 6 La distribution de probabilité de la variable aléatoire X est donnée dans le tableau ci-dessous.

x	$p(x)$
2	0.2
4	0.15
5	0.25
7	0.4

1. Les conditions d'une distribution de probabilité sont-elles vérifiées ? Expliquer.

2. Quelle est la probabilité que X soit égal à 5 ?
3. Quelle est la probabilité que X soit inférieur ou égal à 4 ?
4. Quelle est la probabilité que X soit supérieur à 5 ?
5. Calculer l'espérance et la variance de X .
6. Calculer l'écart-type de X .
7. Calculer l'espérance et la variance de X^2 .
8. Vérifier que $V(X) = E(X^2) - [E(X)]^2$.

Exercice 7 La variable aléatoire continue X est uniformément distribuée entre 1.0 et 1.5.

1. Représenter graphiquement la fonction de densité de probabilité.
2. Calculer $P(X = 1.25)$.
3. Calculer $P(1 \leq X \leq 1.25)$.
4. Calculer $P(1 \leq X \leq 1.1)$.
5. Calculer $P(1.2 < X \leq 1.5)$.
6. Calculer $E(X)$ et $V(X)$.

Soit $Z \sim N(0;1)$. Utilisez les valeurs suivantes pour les exercices ci-dessous : $P(Z < 1) = 0.8413$, $P(Z < 1.5) = 0.9332$, $P(Z < 2) = 0.9772$, $P(Z < 2.5) = 0.9938$, $P(Z < 4) \approx 1$.

Exercice 8 Soit $Z \sim N(0;1)$. Calculer les probabilités suivantes.

1. $P(Z \leq 1.5)$
2. $P(1 < Z < 1.5)$
3. $P(0 < Z < 2.5)$
4. $P(Z \leq -1)$
5. $P(Z > -1)$
6. $P(Z > -1.5)$

Exercice 9 Une variable aléatoire est normalement distribuée, avec une moyenne μ égale à 50 et un écart-type σ égal à 5.

1. Quelle est la probabilité que la variable aléatoire prenne une valeur comprise entre 45 et 55 ?
2. Quelle est la probabilité que la variable aléatoire prenne une valeur comprise entre 40 et 60 ?

Exercice 10 Soit $X \sim N(3;9)$ et $Y \sim N(2;4)$. Calculer

1. $P(X < 0)$,
2. $P(-3 < X < 3)$,
3. $P(Y > 2)$,
4. $P(-6 < Y < 4)$.

Références

- Acemoglu, D., Johnson, S., Robinson, J. A., and Yared, P. (2008). Income and democracy. *American Economic Review*, 98(3) :808–42.
- Grigioni, S., Beaucreux, M., Ladner, J., and Déchelotte, P. (2007). P068 perception du poids, modes de consommation et troubles du comportement alimentaire chez 1 744 étudiants de l’agglomération rouennaise. *Nutrition Clinique et Métabolisme*, 21 :78–79. 25e congrès annuel de la Société Francophone Nutrition Clinique et Métabolisme SFNEP 2007.
- Ritchie, H., Mathieu, E., Rodés-Guiraoand, L., Appel, C., Giattino, C., Ortiz-Ospina, E., Hasell, J., Macdonald, B., Beltekian, D., and Roser, M. (2020). Coronavirus pandemic (covid-19). *Our World in Data*. <https://ourworldindata.org/coronavirus>.