

Analyse bivariée (1)

STA101 TD1

2023-2024

Exercice 0. On a résumé la variable $Vx9$ (force du vent à 9h) du jeu de données ozone¹ selon les indicateurs suivants :

effectif	112
moyenne	-1.21
mediane	-0.87
variance	6.93
ecart-type	2.63
minimum	-7.88
maximum	5.20
étendue	13.07
Q(1/4)	-3.28
Q(3/4)	0.69
ecart inter-quartile	3.97
asymétrie	-0.17
aplatissement	-0.54
cv	-2.17

1. Interpréter chacun de ces indicateurs
2. Construire la boîte à moustaches associée

Exercice 1. Soient X et Y deux variables quantitatives mesurées sur les mêmes individus.

1. Rappeler les formules de la covariance de X et Y et du coefficient de corrélation linéaire entre X et Y
2. On pose $Z = aX + b$ et $T = cY + d$. Calculer le coefficient de corrélation linéaire entre Z et T

Exercice 2.

1. Rappeler la différence entre le coefficient de corrélation de Pearson et le coefficient de corrélation de Spearman.
2. Pour chacun des cas suivants, représenter les nuages de points et calculer ces deux coefficients.

(a)

X	1	2	3	4	5
Y	1	2	4	8	16

(b)

X	1	10	100	1000	10000
Y	0	1	2	3	4

(c)

X	1	1	2	3	3	50
Y	1	3	2	1	3	50

Exercice 3. On s'intéresse au lien entre l'âge et l'indice de masse corporelle chez des garçons Néerlandais. On observe les valeurs suivantes :

1. Ce jeu de données contient 112 mesures météo effectuées à Rennes en 2001. Source <https://r-stat-sc-donnees.github.io/ozone.txt>

age	imc
1.51	15.78
0.19	17.61
0.27	15.62
0.76	17.89
1.51	17.22
0.15	14.90
1.48	16.76
1.53	16.85
0.92	17.49
0.81	17.15

1. Représenter le nuage de points
2. Calculer le coefficient de corrélation linéaire
3. En supposant la normalité des deux variables, tester la nullité du coefficient au risque $\alpha = 5\%$
4. Conclure sur le lien entre l'age et l'indice de masse corporelle

NB : le quantile d'ordre 0.975 de la loi de student à 8 degrés de liberté est 2.306 (cf formulaire)

Exercice 4. On s'intéresse au lien entre le maximum d'ozone mesuré sur une journée dans la ville de Rennes et le maximum d'ozone mesuré la veille. Sur un échantillon de 112 observations, on a obtenu un coefficient de corrélation empirique de 0.68.

1. Construire un intervalle de confiance de niveau 95% pour le coefficient de corrélation théorique ρ
2. Pour quelle(s) raison(s) les conditions d'application du test sont-elles valides ici?

NB : le quantile d'ordre 0.975 d'une loi normale centrée-réduite vaut 1.96