

1. Données quantitatives du TD-2

- Chargez les données contenues dans le fichier `TD2-vitdist.xlsx`, puis trouvez le moyen créer la variable V^2 .
 - Trouvez le moyen de visualiser le nuage de points $(v_i^2, d_i)_{i=1,n}$.
 - Trouvez le moyen de calculer la valeur du coefficient de corrélation linéaire entre la distance et le carré de la vitesse (résultat de TD).
 - Que pensez-vous de celui entre la distance et la vitesse ?
 - On souhaite estimer les coefficients de la droite de régression (modèle affine). Utilisez **Linear Regression** dans ce but. À partir des sorties possibles de ce composant, trouvez ce qu'il faut y connecter pour visualiser ces coefficients.
 - Le centre du nuage (\bar{x}, \bar{y}) est-il sur la droite de régression ?
 - En mathématiques, comment appelle-t-on une fonction comme celle-ci : $f(x) = -2x^3 + x^2 + 2x$?
 - Testez le composant de régression associé, et discutez ses possibilités.
 - À votre avis, que représente **Root mean square error** ?
- (HW) Trouvez le moyen de créer un nouveau *feature* pour les *résidus* $e_i = \hat{y}_i - y_i$, puis calculez son écart-type.

2. Données qualitatives du TD-2

- Chargez les données contenues dans le fichier `TD2-notes.xlsx`, puis affichez le tableau.
- Croisez la méthode d'enseignement et l'avis du jury à l'aide de **Contingency Table** ou **Sieve Diagram** et commentez les valeurs les de χ^2 et p ?
- Le tableau de données comporte également une variable binaire. Analysez les différents couples de variables qualitatives en termes d'indépendance.
- Réalisez l'analyse des correspondances de ces données.

3. Données Quantitatives

- Chargez les données du fichier `anscombeQuartet.xlsx`, et réalisez une analyse statistique unidimensionnelle (numérique et graphique). Commentez les résultats.
- Visualisez tous les nuages bidimensionnels (x, y) possibles, et commentez-les notamment en termes de liaison possible entre la variable en ordonnée et celle en abscisse.
- Vérifiez vos commentaires à l'aide du composant **Correlations**.
- Régressez les y par les x .

4. Données Qualitatives

- `AchatDordi.xlsx` : peut-on expliquer l'usage prévu par le lieu/site d'achat d'un ordinateur ?
- `titanic.tab` : peut-on expliquer la survie ?

5. Sur-apprentissage de modèle

- À l'aide de **Paint Data**, créez des données bidimensionnelles qui ressemblent à celles de la figure ci-contre.
- Déterminez le polynôme de X qui constitue le meilleur modèle de prévision pour Y . Vous vous appuyerez pour cela sur la *variance résiduelle* et sur l'allure de la courbe \hat{Y} .

