

Travaux Pratiques : Analyse Exploratoire des Données avec le Titanic

Objectif : Utiliser Python pour mener des analyses statistiques descriptives et inférentielles sur le dataset Titanic.

Instructions:

1. Exploration des données

- 1.1 Chargez le jeu de données Titanic.
- 1.2 Affichez les 5 premières lignes pour avoir un aperçu des colonnes.
- 1.3 Utilisez les méthodes *info()* et *describe()* pour analyser les types de données et les principales statistiques descriptives.
- 1.4 Identifiez les colonnes contenant des valeurs manquantes et proposez des stratégies pour les gérer.

2. Statistiques descriptives

- 2.1 Calculez la moyenne, la médiane et l'écart-type de l'âge des passagers.
- 2.2 Analysez la répartition des classes (`Pclass').
- 2.3 Réalisez un histogramme des âges. Commentez la symétrie et la dispersion de cette distribution.

3. Loi normale

- 3.1 Utilisez le test de Shapiro-Wilk pour vérifier si les âges suivent une loi normale.
- 3.2 Comparez la p-value obtenue avec un seuil de signification de 0,05. Concluez sur la normalité des âges.

4. Test d'hypothèse

4.1 Formulez l'hypothèse nulle (H_0) et l'hypothèse alternative (H_1) pour tester si les âges des survivants et des non-survivants diffèrent significativement.

4.2 Réalisez un test de Student pour comparer les âges moyens des survivants et des non-survivants.

4.3 Interprétez les résultats en vous basant sur la p-value.

5. Statistiques inférentielles

5.1 Effectuez un test du Chi-deux pour étudier la relation entre la survie et la classe des passagers.

5.2 Interprétez la p-value pour déterminer si la classe influence les chances de survie.

5.3 Réalisez une analyse de la variance (ANOVA) pour étudier l'effet de la classe sur le prix du billet (`Fare`).

5.4 Interprétez les résultats de l'ANOVA.

6. Régression linéaire

6.1 Entraînez un modèle de régression linéaire pour prédire le prix du billet (`Fare`) en fonction de l'âge et de la classe.

6.2 Analysez les coefficients du modèle pour comprendre l'impact de l'âge et de la classe sur le tarif.

7. Conclusion

7.1 Résumez vos observations :

- Les principales tendances observées dans les âges et les classes.
- L'impact de la classe sur la survie et le tarif.
- La pertinence du modèle de régression.

Travail à rendre

- Rédigez un Notebook structuré comprenant :

- Résultats : Exposition des calculs, graphiques et résultats des tests.

- Interprétation : Analyse des résultats et discussion des implications.
- Conclusion : Synthèse et perspectives.