Nantes Université

**Année 2024/2025**
**Semestre 1- Session 1 / Janvier 2025**

**Unités d'enseignement et de recherche de l'UFR de Sociologie**

Diplôme : Licence Section : Sociologie Année 3ème Cycle : 1er

**Unité SLGSON5U03 : Méthodes sociologiques**

**Enseignement SLGSON5E06 : Introduction à l'analyse multivariée**

**Nom du Professeur responsable** : **Mme Sophie ORANGE**

**Observations des Professeurs : Aucun document autorisé/ usage de la calculatrice autorisé (pas de téléphone portable)**

**Date : mardi 14 janvier 2025**
**Heure : 10h30**
**Lieu : Halle du SUAPS**
**Durée : 2 heures**

**Sujet : Vous répondrez aux questions posées**.

Les données suivantes sont issues de : Cédric Afsa, *Le modèle Logit : Théorie et application*, Insee, Document de travail, 2016. Elles s'appuient sur un échantillon de 13 500 élèves de France métropolitaine entrées en sixième en 1995 passés en troisième générale puis orienté soit en seconde générale (et technologique) soit en seconde professionnelle.

**Tableau 1 – Passage en seconde générale selon la zone de l'établissement**

| | Seconde générale | | Total |
|---|---|---|---|
| | Oui | Non | |
| ZEP | 55,3 % | 44,7 % | 100 % |
| Hors ZEP | 68,8 % | 31,2 % | 100 % |
| Ensemble | 67,3% | 32,7% | 100 % |

**1.** Quel est la différence de taux de passage en seconde générale entre les élèves scolarisés en ZEP et ceux non scolarisés en ZEP ?

**2**. Calculez le rapport de chance (odds-ratios) de passage en seconde générale pour les individus scolarisés en ZEP par rapport ceux non scolarisés en ZEP ? (explicitez votre calcul et exprimez le résultat sous forme d'une phrase rédigée).

**Tableau 2 – Passage en seconde générale selon la zone de l'établissement et le retard en 6e**

| | Élève en retard | | Élève à l'heure ou en avance | |
|---|---|---|---|---|
| | 2de générale oui | 2de générale non | 2de générale oui | 2de générale non |
| ZEP | 24,2 % | 75,8 % | 64,5 % | 35,5 % |
| Hors ZEP | 25,5 % | 74,5% | 74,3 % | 25,7 % |
| Ensemble | 25,3 % | 74,7 % | 73,3 % | 26,7 % |

**3.** Dans le tableau 2, quel est l'écart pour les élèves en retard en 6e du taux de passage en seconde générale entre les élèves scolarisés en ZEP et ceux non scolarisés en ZEP? Et pour ceux à l'heure ou avance ? Comparez ces résultats à celui de la question 1.

**4.** En quoi alors le recours à un modèle de régression logistique est-il justifié pour étudier le lien entre la zone d'éducation et le passage en seconde générale ?

**5.** En quoi l'usage d'un tel modèle est-il possible ?

**Tableau 3 – Passage en seconde générale**

| Variable | Coefficient |
|---|---|
| Constante | 0,806 |
| Appartenance à une ZEP (*ref = non)* | |
| *oui* | -0,382*** |
| Âge à l'entrée en sixième (*ref = à l'heure*) | |
| *en retard* | -2,035*** |
| *en avance* | 1,845*** |
| Sexe de l'élève (*ref = garçon*) | |
| *fille* | 0,400*** |

Seuil de significativité : *** = 1% ; ** = 5% ; * = 10%

**6**. Commentez le tableau 3 en prenant soin entre autres d'indiquer :
- quelles sont respectivement la variable prédite, la variable prédictive principale, les variables de contrôle
- quelle est la situation de référence du modèle
- d'indiquer les effets respectifs des différentes variables (leur sens, leur force relative)

**7.** Calculer la probabilité estimée par le modèle de passage en seconde générale pour la situation de référence

**8.** Calculer sous forme d'odds ratio le coefficient relatif à l'appartenance à une ZEP du tableau 3. Exprimez de la façon la plus rigoureuse possible cet odds-ratio sous la forme d'une phrase.

**Tableau 4 – Passage en seconde générale**

| Variable | Coefficient |
|---|---|
| Constante | -0,177 |
| Appartenance à une ZEP (*ref = non)* | |
| *oui* | -0,063 |
| Âge à l'entrée en sixième (*ref = à l'heure*) | |
| *en retard* | -1,856*** |
| *en avance* | 1,601*** |
| Sexe de l'élève (*ref = garçon*) | |
| *fille* | 0,496*** |
| Milieu social de l'élève (*ref = sans act. prof., non rep.*) | |
| *agriculteurs* | 0,756*** |
| *artis., commerç., chefs d'entrep.* | 0,917*** |
| *cadres, prof. intell. sup.* | 2,352*** |
| *prof. intermédiaires* | 1,427*** |
| *employés* | 0,679*** |
| *ouvriers* | 0,300** |

Seuil de significativité : *** = 1% ; ** = 5% ; * = 10%

**9**. Dans le tableau 4, quelle est la nouvelle variable explicative introduite dans le modèle ? Quel effet a-t-elle sur le passage en seconde générale ? Est-ce surprenant ?

**10.** Dans le tableau 4, le coefficient relatif à l'appartenance à une ZEP est-il interprétable ou non ? et pourquoi ?

**Tableau 5 – Passage en seconde générale**

| Variable | Coefficient |
|---|---|
| Constante | -6,240*** |
| Appartenance à une ZEP (*ref = non)* | |
| *oui* | 0,489*** |
| Âge à l'entrée en sixième (*ref = à l'heure*) | |
| *en retard* | -1,278*** |
| *en avance* | 1,248*** |
| Sexe de l'élève (*ref = garçon*) | |
| *fille* | 0,521*** |
| Milieu social de l'élève (*ref = sans act. prof., non rep.*) | |
| *agriculteurs* | 0,370** |
| *artis., commerç., chefs d'entrep.* | 0,642*** |
| *cadres, prof. intell. sup.* | 1,804*** |
| *prof. intermédiaires* | 1,009*** |
| *employés* | 0,445*** |
| *ouvriers* | 0,244* |
| Niveau en français en 6e (*ref = < moyenne nationale*) | |
| *> moyenne* | 0,061*** |
| Niveau en français en 6e (*ref = < moyenne nationale*) | |
| *> moyenne* | 0,064*** |

Seuil de significativité : *** = 1% ; ** = 5% ; * = 10%

**11.** Calculez sous forme d'odds ratio le coefficient relatif à l'appartenance à une ZEP du tableau 4. Est-il plus grand ou plus petit que 1 ? Qu'en déduire pour son interprétation ? Comparez-le à celui calculé dans la question 6 sur la base du tableau 3

**12.** En quoi à la lecture des tableaux 3, 4, et 5, l'emploi de l'expression « toutes choses égales par ailleurs » est-elle ambigu dans l'interprétation des résultats des modèles de régression logistique dans des formulations du type :
« Toutes choses égales par ailleurs, l'effet de la scolarisation en ZEP sur la probabilité de passer en seconde générale est … »

**13.** En quoi, plus largement, le raisonnement « toute chose égale par ailleurs » qui sous-tend souvent l'usage des modèles de régression logistique est contestable