La régression logistique

Présentation pour le cours SOL6210, Analyse quantitative avancée

©Claire Durand, 2024

Utilisation

- Quand la variable dépendante est nominale ou ordinale
  - Deux types selon la forme de la variable dépendante:
    - binaire pour deux catégories.
    - polytomique pour plusieurs catégories: multinominale ou ordinale.
- Quand les variables indépendantes peuvent être de plusieurs types:
  - Nominales (qualitatives)
  - Ordinales
  - Métriques (quantitatives)
  - Et des interactions de variables, le cas échéant.

Mais encore...

- Comme la variable dépendante prend soit la valeur 0 ou la valeur 1 (régression logistique binaire), la valeur prédite devrait donc se situer entre 0 et 1. C'est pourquoi on transforme la VD en une *fonction de probabilité*.
- Fait partie des modèles utilisant la transformation logarithmique:
  - Modèles loglinéaires: Uniquement des variables nominales ou ordinales
  - Modèles probit ou logit: Variables indépendantes continues
  - Modèles de régression de survie: variable prédite évolue dans le temps; variables indépendantes peuvent aussi varier dans le temps.

Mais encore...

- **Se distingue de la régression linéaire** par le fait que celle-ci demande une distribution normale des variables indépendantes et dépendante. De plus, **si on utilisait la régression linéaire, la valeur prédite pourrait se situer en dehors de l'intervalle 0,1**.
- **Se distingue de l'analyse discriminante** entre autres par le fait que celle-ci demande une distribution multi-normale des variables indépendantes.
- En pratique, lorsque la proportion de zéro ou de 1 se situe à plus de 15%, la régression linéaire donne des résultats fiables en général (voir Cibois).

La variable dépendante

- Variable dépendante: $$\ln\left(\frac{p(y=1|x)}{p(y=0|x)}\right)$$
- Et donc, ***log naturel de la probabilité que y prenne la valeur 1 plutôt que 0 étant donné la valeur des variables indépendantes (x_i)* dans l'équation.**
- On transforme donc la variable dépendante (VD) de telle sorte que l'on prédit la probabilité que Y=1 plutôt que 0, étant donné les valeurs de x (x_1,x_2,x_3,...x_n).
- Il s'agit d'une transformation logarithmique, difficile à interpréter. C'est pourquoi on retransforme par l'inverse (Exp) pour faciliter l'interprétation. On parle alors de **rapport de cote (*odds ratio*)**.

Le processus d'analyse suggéré est le même que pour la régression linéaire

Mais il faut vérifier un peu plus sérieusement la linéarité

- 1. Examiner les relations bi-variées entre chacune des V.I. et la V.D.
  - Vérifier si elles sont linéaires, le cas échéant.
- 2. Faire une régression standard - où on introduit toutes les V.I. dans l'analyse ensemble -- et des régressions bivariées pour chaque V.I. séparément pour comprendre l'effet combiné des variables.
- 3. Établir un modèle et procéder à l'entrée séquentielle (hiérarchique) des variables ou groupes de variables en fonction du modèle.
- 4. Faire une régression parcimonieuse, soit une analyse qui ne garde que les variables significatives, et vérifier si cette analyse diffère significativement du modèle comprenant toutes les variables.
- 5. Examiner la justesse du modèle :
  - χ^2 de maximum de vraisemblance
  - Pourcentage de bonnes prédictions
  - Distribution des résidus.

Les informations

- Justesse du modèle :
  - χ^2 (appelé Chi (Ki) Carré ou Chi (Ki) Deux) du modèle: Est-ce que l'ensemble de variables – ou les variables du bloc entré – est significativement relié à la VD?
  - χ^2 de maximum de vraisemblance (Hosmer et Lemeshow): est-ce que les données sont compatibles avec les relations postulées?
  - Classification: Jusqu'à quel point les informations sur les VI permettent de bien classer les cas dans les catégories de la VD, *étant donné le point de coupure.*
  - Analyse des résidus: proportion de résidus à plus de 3 écarts-types (si n est grand), distribution normale des résidus.

Les informations

- Pour chaque variable :
  - Test de Wald (se distribue comme le χ^2): jusqu'à quel point **chaque variable** est liée significativement à la VD.
- Pour chaque catégorie des variables qualitatives:
  - Test de Wald (b^2/s.e.^{2}): jusqu'à quel point **chaque catégorie** de variable se distingue significativement de la catégorie de référence.

Nouvelle notion importante

χ^2 de maximum de vraisemblance

- Pour le χ^2 "ordinaire":
  - L'hypothèse nulle est qu'il n'y a pas de relation entre deux variables.
  - L'hypothèse alternative est qu'il y a une relation, que les deux variables ne sont pas indépendantes.
  - **On cherche à rejeter l'hypothèse nulle.**
- Pour le χ^2 de maximum de vraisemblance (Hosmer et Lemeshow, Brown):
  - L'hypothèse nulle (H_0) est qu'il y a une relation entre l'ensemble de variables indépendantes entrées et la variable dépendante.
  - L'hypothèse alternative est que les relations ne sont pas celles qui sont postulées.
  - **On cherche à accepter l'hypothèse nulle.**

Le χ^2 de maximum de vraisemblance

- Lorsque la probabilité du χ^2 de maximum de vraisemblance de Hosmer et Lemeshow est **plus grande** que 0,05, on conclut que les données sont compatibles avec le modèle postulé, **qu'il est vraisemblable que l'ensemble de relations existe dans la population tel que postulé.**

Interprétation des coefficients

- Les coefficients *b* s'interprètent comme pour la régression linéaire, **mais toujours en comparant à la catégorie de référence**, ce qui donnerait, pour la régression logistique:
  - "à chaque augmentation de 1 de la VI, le log naturel de la probabilité que y soit égal à 1 **plutôt qu'à zéro** augmente de b".
- Concrètement, on regarde e^b (la fonction inverse) et on interprète que,
  - à chaque augmentation de 1 de la VI, la "probabilité" que Y=1 plutôt que zéro augmente de e^b.
    - Lorsque la VI est qualitative: Si le cas est dans la catégorie "i" plutôt que dans la catégorie de référence de la VI, la probabilité que Y=1 plutôt que zéro est de e^b.

Interprétation des coefficients

Exemple (voir listing - analyse hiérarchique sur les prédicteurs de l'appui à la Charte des valeurs)

- Ceux qui ont l'intention de voter pour le Parti Québécois ont huit fois plus de chances (voir e^b=8,161) que ceux qui ont l'intention de voter pour le Parti Libéral (catégorie de référence) de se déclarer d'accord avec la Charte des valeurs proposée par le Parti Québécois.
- Cette probabilité baisse à environ 5 fois quand on contrôle pour le malaise par rapport au port de signes religieux et pour l'attitude négative face à l'immigration.
- Attention: plusieurs débats sur cette manière d'interpréter au point où certains suggèrent de ne pas utiliser le Exp(b).

Classification

- La régression logistique donne aussi la proportion de cas bien classés étant donné les valeurs de x et le *point de césure (cutpoint).*
  - Le point de césure est de 0,5 par défaut dans SPSS. Il faut le modifier selon la distribution de la VD à un point optimal (Si 15% de 1, mettre le point de césure à 0,15).
  - Interprétation: Si on catégorise dans les “1" tous les cas qui ont une probabilité plus grande que le point de césure d’être dans les “1" et que l’on catégorise dans les “0" tous ceux qui ont une probabilité plus petite que le point de césure d’être dans les “1", quelle est la proportion de “1" et de “0" qui sont bien classés?

Les résidus

- Tout comme en régression linéaire, on regarde...
  - La proportion de résidus standardisés plus grande que 3.
  - La distribution des résidus (qui doit être normale et avoir la même variance quelles que soient les valeurs de x, soit *homoscédaste*).

En conclusion

- La régression logistique est relativement facile à utiliser.
- Il faut faire **très attention** à l’interprétation (ne pas oublier que l’on compare toujours à la catégorie de référence).
- La présence d’informations sur la classification peut être un avantage de l’analyse (par rapport à la régression linéaire) dans certains cas.
- Le deuxième avantage sur la régression linéaire est la possibilité d’entrer des variables de plusieurs types comme variables indépendantes et de modéliser facilement les interactions.