

Ejercicio 5

Equipo

2024-03-26

5. Modelos lineales generalizados para datos categóricos

La base de datos Preg5.csv contiene información sobre el nivel de satisfacción (Sat) de un conjunto de individuos que rentan una vivienda. El interés es identificar si entre los factores que definen este nivel están: el tipo de vivienda (Type), la percepción sobre su influencia en las decisiones sobre el mantenimiento de la vivienda (Infl) y el contacto que tienen con el resto de inquilinos (Cont).

- i. En este caso todas las covariables son categóricas, así que se puede obtener una gráfica que describa las frecuencias relativas para los tres niveles de satisfacción considerando cada cruce Type-Infl-Cont. Presente esta gráfica y comente lo que se observa.
- ii. Ajuste un modelo logístico multinomial considerando todas las posibles interacciones entre Type, Infl y Cont. Este modelo tiene demasiados parámetros y es de interés buscar si es posible considerar un modelo más simple. Ajuste un modelo logístico multinomial que incluya a las tres covariables Type, Infl y Cont, pero sin considerar interacciones. Realice una prueba de hipótesis para argumentar si es plausible considerar el modelo más simple. Complemente esto con medidas como AIC o BIC.

Nota computacional. Para los modelos ajustados en vglm no es posible usar directamente el paquete multcomp, pero está disponible una función anova() que permite comparar dos modelos, siempre que estos estén anidados, por ejemplo: `anova(mod1, mod2, test = "LRT", type = "I")` o equivalentemente `lrtest(mod1, mod2)`.

- iii. Considerando las covariables del modelo ajustado en ii) y notando que la variable Sat puede ser considerada como ordinal, ajuste un modelo logístico acumulativo (cumulative logit) sin considerar el supuesto de proporcionalidad. También ajuste un modelo que asuma el supuesto de proporcionalidad y dado que este último está anidado en el primero, realice una prueba de hipótesis para analizar si es plausible asumir este modelo más sencillo. Complemente su decisión con medidas como AIC o BIC.
- iv. Usando el AIC o BIC, seleccione sólo un modelo entre los ajustados en ii) y iii). Con ese modelo, trate de interpretar los resultados. Por ejemplo, puede apoyarse de una gráfica como la realizada en i), pero donde se presenten las probabilidades estimadas para cada nivel de satisfacción. Por simplicidad sólo analice el efecto que se observa al considerar la variable Infl, cuando se asume que la persona renta una vivienda tipo Apartment y tiene un nivel de contacto con otros inquilinos como High.

i) Gráfica de frecuencias relativas

ii) Modelo logístico multinomial

iii) Modelo logístico acumulativo (cumulative logit)

iv) Selección del modelo e interpretación de resultados