

Hoja de Ejercicios 6 Modelos de Respuesta Binaria

ECONOMETRÍA I. UC3M

1. [W 17.9] Utilice los datos del fichero LOANAPP para este ejercicio. La variable dependiente es *approve*, variable binaria que es igual a 1 si al individuo se le concedió un préstamo hipotecario y 0 en caso contrario. La variable explicativa clave es *white*, una variable ficticia con valor 1 si el solicitante es blanco y 0 en otro caso. Los otros solicitantes en la muestra son negros e hispanos.

Para contrastar la existencia de discriminación en el mercado hipotecario, se puede usar un modelo lineal de probabilidad,

$$approve = \beta_0 + \beta_1 white + \text{otros factores.}$$

- a) Si existe discriminación contra las minorías, y se ha tenido en cuenta adecuadamente la influencia de otros factores, ¿cuál es el signo de β_1 ?
 - b) Efectúe la regresión de *approve* sobre *white*. Interprete el coeficiente de *white*. ¿Es estadísticamente significativo? ¿Es grande en términos prácticos?
 - c) Añada ahora las variables *hrat*, *obrat*, *loanprc*, *unem*, *male*, *married*, *dep*, *sch*, *cosign*, *chist*, *pubrec*, *mortlat1*, *mortlat2* y *vr*. ¿Qué ocurre con el coeficiente de *white*? ¿Existe todavía evidencia estadísticamente significativa de discriminación contra los individuos que no son blancos?
 - d) Incluya ahora una interacción entre la variable ficticia racial y *obrat* (que mide otras obligaciones, como porcentaje de renta). ¿Es estadísticamente significativo el término de interacción?
 - e) En el modelo del apartado (d), ¿cuál es el efecto de ser blanco sobre la probabilidad de que el crédito sea concedido cuando *obrat* = 32, que es aproximadamente el valor medio en la muestra? Obtenga un intervalo de confianza del 95 % para ese efecto.
 - f) ¿Cuántas observaciones tienen *obrat* > 40, esto es, las otras deudas son mayores que el 40 % de los ingresos totales?
 - g) Reestime el modelo anterior, excluyendo las observaciones con *obrat* > 40. ¿Qué ocurre con el estimador y el estadístico *t* de *white*?
 - h) ¿Es el estimador de *white* altamente sensible a la muestra utilizada?
2. Utilice de nuevo los datos del fichero LOANAPP para este ejercicio.
- a) Estime un modelo Probit de *approve* sobre *white*. Calcule la probabilidad estimada de que se apruebe un crédito a personas blancas y a personas no blancas. Compare dichas probabilidades estimadas con las del modelo lineal de probabilidad.
 - b) Incluya ahora las variables *hrat*, *obrat*, *loanprc*, *unem*, *male*, *married*, *dep*, *sch*, *cosign*, *chist*, *pubrec*, *mortlat1*, *mortlat2* y *vr* al modelo probit. ¿Existe evidencia estadística significativa de discriminación contra los individuos que no son blancos?

- c) Estime un modelo Logit con las mismas variables que en el apartado (b). Compare el coeficiente de la variable *white* con el estimado mediante el modelo probit.
- d) ¿Cómo podemos comparar el tamaño de la discriminación entre los modelos probit y logit?
3. [W 17.2] Sea *grad* una variable binaria que toma el valor 1 si un estudiante se gradúa en cinco años y 0 en caso contrario. Sean *hsGPA* y *SAT* el promedio de las calificaciones de bachillerato y el "test de aptitud escolar" para entrar en la universidad, respectivamente. Sea *study* el número de horas semanales que pasa en una biblioteca. Supongamos que, con los datos de 420 estudiantes, hemos obtenido las siguientes estimaciones del modelo logit:

$$\Pr(\text{grad} = 1 | \widehat{\text{hsGPA}}, \text{SAT}, \text{study}) = \Lambda(-1,17 + 0,25\text{hsGPA} + 0,00058\text{SAT} + 0,073\text{study}),$$

donde $\Lambda(z) = \exp(z)/(1 + \exp(z))$ es la función logística. Para $\text{hsGPA} = 3$ y $\text{SAT} = 12000$, calcule la diferencia estimada en la probabilidad de graduarse para un estudiante que pasa 10 horas semanales en la biblioteca respecto a otro que pasa cinco.

4. [W 17.8] Utilice los datos PNTSPRD para este ejercicio.
- a) La variable *spread* indica el diferencial de puntos que las casas de apuestas de Las Vegas daban a priori al equipo favorito para ganar un partido, y *favwin* es una variable binaria que indica si ese equipo efectivamente ganó. Proponemos el siguiente modelo lineal de probabilidad para estimar la probabilidad de que el equipo favorito gane:

$$\Pr(\text{favwin} = 1 | \text{spread}) = \beta_0 + \beta_1 \text{spread}.$$

Explique por qué, si el diferencial a priori de puntos contiene toda la información relevante, esperamos que $\beta_0 = 0,5$.

- b) Estime el modelo del apartado (a) por MCO. Contraste la hipótesis $H_0 : \beta = 0,5$ contra la alternativa $H_0 : \beta \neq 0,5$. Hágalo con los errores estándar habituales y con los que son robustos a la presencia de heteroscedasticidad.
- c) ¿Es *spread* estadísticamente significativa? ¿Cuál es la probabilidad estimada de que el equipo favorito de un partido gane cuando el diferencial de puntos a priori es $\text{spread} = 10$?
- d) Estime ahora un modelo probit para $\Pr(\text{favwin} = 1 | \text{spread})$. Interprete el término constante y contraste si es igual a cero. (Pista: recuerde que $\Phi(0) = 0,5$).
- e) Utilice el modelo probit para estimar la probabilidad de que el equipo favorito gane cuando $\text{spread} = 10$. Compare el resultado con la estimación del modelo de probabilidad lineal del apartado (a).
- f) Añada las variables *favhome*, *fav25*, *und25* al modelo probit y contraste la significación conjunta de esas tres variables usando el contraste de cociente de verosimilitudes. ¿Cuántos grados de libertad tiene la distribución chi-cuadrado?. Interprete este resultado, contestando a la pregunta de si la variable *spread* incorpora toda la información observable antes de que se juegue el partido.