

Examen Final de Microeconometría

Martín González-Rozada

(Fecha de Entrega: lunes 19 de febrero 2024).

1 de diciembre de 2023

1. Primera parte: Logit, Probit, Tobit y Heckman

Haga los ejercicios 1 a 8 del capítulo 11 del libro *The practice of Econometrics* de Ernst R. Berndt.

2. Segunda parte: análisis de supervivencia

La base de datos utilizada para los ejercicios sale de un ensayo controlado aleatorizado (RCT) de tratamientos para perros diagnosticados con linfoma en varias clínicas veterinarias.

El estudio involucró 7 clínicas y 300 perros que, aleatoriamente, fueron asignados a tratamientos de radioterapia, quimioterapia, ambos o ninguno. En este ensayo clínico se incluyeron perros con edades hasta 15 años.

La variable de interés es el tiempo medido en días (meses) desde el diagnóstico hasta la muerte debido a la linfoma. Se considera que perros que murieron de otras causas o no volvieron a las entrevistas de seguimiento son eventos censurados.

Las variables del *dataset* están descritas en la próxima tabla. La base de datos para trabajar es: **linfoma.dta**.

Variable	Descripción	Rango
id	No. de identificación de cada perro	1 – 300
clinic	No. de identificación de cada clínica	1 – 7
age	Edad del perro al momento del diagnóstico	1,4 – 14,4
rad	Radiación	0 =no, 1 = yes
chemo	Quimioterapia	0 =no, 1 =yes
died	muerte debido a linfosarcoma	0 = censored, 1 = died
days	tiempo (en días) desde el diagnóstico a la muerte (o censura)	7 – 1363
months	tiempo (en meses) desde el diagnóstico a la muerte (o censura)	0,45 – 45,5

3. Introducción al análisis de supervivencia

1. Genere una tabla de Kaplan-Meier para los datos mensuales. ¿Qué proporción de perros sobreviven 1 año, 2 años y 5 años? Genere la curva de supervivencia e incluya intervalos de confianza alrededor de dicha curva.
2. Genere la curva de Nelson-Aalen de *hazard* acumulado. ¿En qué punto la función de *hazard* acumulada llega a 1? Genere un gráfico con el *hazard* suavizado. Interprete el gráfico de la función de *hazard*.
3. Genere la curva de Kaplan-Meier de supervivencia para evaluar los efectos de radiación en las expectativas de supervivencia de los perros de la muestra. Genere un gráfico de los tiempos de supervivencia para el grupo de tratamiento y para el grupo de control. Repita el ejercicio para cuando se aplica quimioterapia.
4. Cree una nueva variable que represente cada una de las cuatro combinaciones de tratamiento posibles. Genere un gráfico mostrando cada una de las funciones de supervivencia.
5. Use un test de Wilcoxon (revise la literatura) para determinar si hay diferencias estadísticamente significativas entre las 4 combinaciones de tratamiento.

4. Estimando un modelo de Cox (*Cox proportional hazard model*)

1. Estime un modelo de Cox con **chemo** y **rad** como predictores. ¿Qué efecto tienen los tratamientos en las probabilidades de supervivencia? Use el tiempo definido en días.
2. El estudio fue realizado en 7 clínicas y podría ser una *confounding variable*. Agregue la variable **clinic** al modelo y determine si agregar esta variable tiene algún efecto en los coeficientes de **chemo** y **rad**.
3. Utilice tanto un test de Wald como un test de cociente de verosimilitud para evaluar la significatividad de la variable **clinic** como predictor. Utilice el criterio de información de Akaike para determinar si el modelo que mejor ajusta los datos es el que incorpora la variable **clinic** o el que no la incorpora.
4. Hay alguna evidencia de la interacción entre **chemo** y **rad**? ¿Cuál es el efecto de recibir ambos tratamientos sobre la probabilidad de morir?
5. ¿Cuál es la forma de la hazard en el baseline *baseline hazard*?
6. Realice un análisis estratificado por clínica (**strata(clinic)**) de **chemo**, **rad** y su interacción. Compare este modelo con el que incluye la variable **clinic** como efecto fijo. ¿Cómo difieren estos modelos? ¿Cuál tiene la mayor log-verosimilitud? ¿Son los coeficientes para las variables de tratamiento similares? Explique en sus propias palabras cuáles son los efectos de cada uno de los tratamientos sobre la probabilidad de que un perro sobreviva. Estime un modelo que determine si el efecto de quimioterapia varía de una clínica a la otra.
7. Estime un modelo que use como variable explicativa una variable categórica que represente los cuatro grupos de tratamiento (es decir, cada uno de los dos tratamientos y su interacción) y efectos fijos por clínica y evalúe el supuesto de *proportional hazard* utilizando un gráfico de la función hazard acumulada. ¿Qué puede concluir?
8. Evalúe también el supuesto de *proportional hazard* utilizando gráficos de residuos Schoenfeld escalados y un test estadístico de proporcionalidad basado en los residuos de Schoenfeld (ayuda: vea **schoenfeld** y **scaledsch** en Stata) .

5. Modelos Paramétricos

En este ejercicio tiene que calcular estimaciones para diferentes tipos de modelos de supervivencia paramétricos. Las variables explicativas de los modelos serán: la edad del diagnóstico, su cuadrado y los tratamientos **rad** y **chemo**.

1. Comience calculando:

- la variable **age**²
 - tres variables que representen cada grupo tratamiento: los que reciben solamente quimioterapia, los que reciben solamente radioterapia y quienes reciben ambos.
- a)* Ajuste modelos paramétricos con el tiempo de supervivencia modelado mediante: una distribución exponencial, una Weibull y una distribución log-normal.
- b)* Interprete los coeficientes estimados cuidadosamente.
- c)* Compare cuál de los tres modelos ajusta mejor los datos.
- d)* Calcule la curva de supervivencia para cada grupo de quimioterapia para los modelos que usan distribuciones de Weibull y log-normal.
- e)* Dibuje la hazard en la línea de base para cada grupo de radiación para los modelos que usan distribuciones de Weibull y log-normal.
- f)* Evalúe el ajuste de los modelos que usan distribuciones de Weibull y log-normal.
- g)* Resuma las principales conclusiones de su análisis.