

Segundo Informe Artículo: TransModel: un paquete R para el
modelo de transformación lineal con datos censurados
Seminario de Estadística

Andres Felipe Franco Valero
Luisa Fernanda Guantiva Vargas
Juan Andres Valero Sierra

Presentado a:
Profesor Ramon Giraldo Henao



UNIVERSIDAD NACIONAL DE COLOMBIA

Departamento de Estadística
Universidad Nacional de Colombia
Bogotá, Colombia
Mayo de 2022

1. Resumen

El artículo empleado para la elaboración de este informe fue publicado en enero del 2022 en la Revista de software estadístico por Jie Zhou, Jiajia Zhang y Wenbin Lu [ZZL22] elabora un análisis de modelos de transformación lineal, incluido el modelo de riesgos proporcionales y el modelo de probabilidades proporcionales, con censura por la derecha para desarrollar un paquete en el software R llamado `TransModel` que se enfoca en la metodología de Chen, explicando detalladamente su uso y la implementación de dicho paquete para unos datos de cáncer de pulmón de veteranos. El modelo de riesgos proporcionales (PH) (Cox 1992) se ha utilizado ampliamente en muchos campos de investigación, como aplicaciones biomédicas, estudios financieros y estudios epidemiológicos. Sin embargo, el supuesto de proporcionalidad no se cumple y otros modelos, como el modelo de probabilidades proporcionales (PO) (Dabrowska y Doksum 1988), deben usarse como alternativas. Chen et al. (2002) propusieron un procedimiento unificado para el análisis del modelo de transformación lineal, que se reduce al enfoque de verosimilitud parcial en el caso del modelo PH y su validez no depende del supuesto de independencia entre la variable de censura y las covariables. Este modelo no ha sido ampliamente utilizado debido a la falta de funciones fáciles de implementar en un software estadístico común, como R y SAS, pero el procedimiento propuesto es basado en ecuaciones de estimación y se implementa de manera eficiente tanto numérica, como computacionalmente.

2. Metodología

Se realiza la estimación de los parámetros, varianzas y los intervalos de confianza y banda de confianza de la función de supervivencia utilizando la metodología de Chen [CJY02].

Para la estimación de los parámetros se crea un algoritmo iterativo que calcula los coeficientes β y H , luego con estos, se estima una función de supervivencia. Posteriormente se realiza la estimación de varianza de parámetros teniendo en cuenta la condición de regularidad de Chen donde $\hat{\beta}$ es consistente y asintóticamente distribuido normalmente.

Finalmente teniendo en cuenta que la varianza de $H(t)$ no tiene una solución de forma cerrada, el intervalo de confianza y la banda de confianza para la curva de supervivencia se derivan mediante las técnicas de perturbación.

3. Resultados

El paquete `TransModel` ajusta modelos de transformación lineal con datos censurados por la derecha. El modelo **PH** y el modelo **PO** de uso común se incluyen como casos especiales con $r = 0$ y 1 . El paquete proporciona coeficientes, desviación estándar y valor p . Además, predice curvas de supervivencia con su intervalo de confianza y banda de confianza a través de la perturbación. La función principal del paquete es `TransModel` se puede llamar con la siguiente sintaxis:

```
1 TransModel(formula, data, r, CICB.st = FALSE, subset, dx = 0.001, iter.
  max = 100, num.sim = 200)
```

La descripción de los argumentos es:

- **Fórmula:** Una fórmula de supervivencia basada en la función `Surv()`, que contiene tiempo de supervivencia, indicador de censura a la derecha y covariables.
- **Data:** El conjunto de datos con todas las variables necesarias en la fórmula
- **r:** Parámetro en la función de riesgo del término de error.
- **CICB.st:** Si se realizará o no la perturbación del intervalo de confianza y las bandas de confianza de las estimaciones de supervivencia (Por defecto es FALSO).
- **Subset:** Las condiciones utilizadas para seleccionar un subconjunto de los datos.
- **dx:** El límite de tolerancia de la convergencia (Por defecto es 0,001).

- `inter.max`: El número máximo de iteraciones antes de la convergencia (Por defecto es 100).
- `num.sim`: El número de perturbaciones utilizadas (Solo funciona cuando `CICB.st = TRUE`, por defecto su valor es 200)

Como resultado obtenemos:

- **Coefficientes**: Obtenemos los β estimados en el modelo de transformación.
- **vcov**: Matriz de covarianza estimada para los coeficientes.
- **converged**: estado de convergencia y el número de iteraciones utilizadas para la convergencia.
Nota: si obtenemos 0 como resultado significa que el algoritmo converge.

Para obtener las predicciones de las probabilidades de supervivencia usamos la sintaxis:

```
1 predict(object, newdata, new.time, alpha)
```

La descripción de los argumentos es:

- **objeto**: Un objeto devuelto por la función `TransModel`.
- **newdata**: Un vector con valores para cada variable covariable (Si no se especifica, se utilizará 0 para todas las variables).
- **new.time**: Un vector de puntos de tiempo ordenados que se utilizará para el cálculo de la probabilidad de supervivencia. Si es nulo, se utilizarán los puntos de tiempo del conjunto de datos original.
- **alfa**: Nivel de confianza para el cálculo de los intervalos y bandas de confianza de la estimación de supervivencia (Por defecto es 0,05).

Como resultado obtenemos:

- **tiempo**: Puntos de tiempo ordenados en los que se calculan las probabilidades de supervivencia
- **supervivencia**: Probabilidades de supervivencia previstas.
- **low.ci**: El límite inferior del intervalo de confianza.
- **up.ci**: El límite superior del intervalo de confianza.
- **low.cb**: El límite inferior de la banda de confianza.
- **up.cb**: El límite superior de la banda de confianza.

4. Descripción de los datos

Se utilizan los datos de cáncer de pulmón de la Administración de Veteranos como ejemplo para ilustrar el uso del paquete `TransModel`. Los datos provienen de un estudio realizado con pacientes hombres con cáncer de pulmón avanzado e inoperable. Se les administró una terapia estándar o una quimioterapia de prueba. Se registró el tiempo hasta la muerte de 137 pacientes. El objetivo principal era evaluar si la quimioterapia de prueba era beneficiosa. Adicionalmente se registraron algunas covariables para cada paciente. Los datos contienen las siguientes variables:

- **ID**: identificación del paciente.
- **TIME**: Hora de inicio del periodo de observación.
- **Y**: 0 indica el comienzo del periodo de observación o censura y 1 indica la muerte.
- **trt**: tipo de tratamiento.

- celltype: tipo histológico del tumor.
- karno: puntuación de rendimiento de Karnofsky que describe el estado general de los pacientes al comienzo del estudio.
- diagtime: tiempo entre el diagnóstico y el inicio del estudio.
- age: edad del paciente (en años).
- priorthrapy: indica si el paciente ha recibido otro tratamiento antes del que está recibiendo.

Los datos están disponible en el paquete `survival` de R y se disponen de esta manera

tr	celltype	time	status	karno	diagtime	age	prior	squamous	smallcell	adeno
1	squamous	72	1	60	7	69	0	1	0	0
1	smallcell	30	1	60	3	61	0	0	1	0
1	adeno	8	1	20	19	61	10	0	0	1
2	large	105	1	80	11	66	0	0	0	0

5. Código en

Además de ajustar modelos de transformación lineal con datos censurados por la derecha el modelo PH y el modelo PO de uso común se incluyen como casos especiales con $r = 0$ y 1 . El paquete puede proporcionar coeficientes, desviación estándar y valores p. También puede predecir curvas de supervivencia con su intervalo de confianza y banda de confianza si así se requiere.

```

1 library(TransModel)
2 ##### veteran data analysis #####
3 set.seed(100)
4 data(veteran)
5 veteran$celltype <- relevel(veteran$celltype, ref = "squamous")
6 fit <- TransModel(Surv(time, status) ~ karno + factor(celltype), data =
  veteran,
7                   r = 1, CICB.st = TRUE, subset = (prior == 0))
8 fit$coefficient
9 fit$vcov
10 summary(fit)
11 ## r = 0
12 fit_0 <- TransModel(Surv(time, status) ~ karno + factor(celltype), data
  = veteran,
13                    r = 0, CICB.st = TRUE, subset = (prior == 0))
14 summary(fit_0)
15 ## r = 0.5
16 fit_05 <- TransModel(Surv(time, status) ~ karno + factor(celltype),
  data = veteran,
17                    r = 0.5, CICB.st = TRUE, subset = (prior == 0))
18 summary(fit_05)
19 ## r = 2
20 fit_2 <- TransModel(Surv(time, status) ~ karno + factor(celltype), data
  = veteran,
21                    r = 2, CICB.st = TRUE, subset = (prior == 0))
22 summary(fit_2)
23 # Predict survival curves for each tumor type
24 pred1 <- predict(fit, newdata = c(60, 0, 0, 0))
25 pred2 <- predict(fit, newdata = c(60, 1, 0, 0))
26 pred3 <- predict(fit, newdata = c(60, 0, 1, 0))
27 pred4 <- predict(fit, newdata = c(60, 0, 0, 1))
28 # Figure 1

```

```

29 pdf("veteran.pdf")
30 plot(pred1, lty = 1)
31 lines(pred2$time, pred2$survival, type = "s", lty = 2)
32 lines(pred3$time, pred3$survival, type = "s", lty = 3)
33 lines(pred4$time, pred4$survival, type = "s", lty = 4)
34 legend("topright", c("squamous", "adeno", "large", "smallcell"),
35       title = "Tumor Type", lty = 1:4, bty = "n")
36 dev.off()
37 # Figure 2
38 pdf("cicb.pdf")
39 plot(pred1)
40 lines(pred1$time, pred1$low.ci, lty = 2)
41 lines(pred1$time, pred1$up.ci, lty = 2)
42 lines(pred1$time, pred1$low.cb, lty = 3)
43 lines(pred1$time, pred1$up.cb, lty = 3)
44 legend("topright", c("Survival estimates", "95% CI", "95% CB"),
45       lty = 1:3, lwd = 1, bty = "n")
46 dev.off()
47 sessionInfo()

```

Referencias

- [CJY02] Kani Chen, Zhezhen Jin y Zhiliang Ying. “Semiparametric analysis of transformation models with censored data”. En: *Biometrika* 89.3 (ago. de 2002), págs. 659-668. ISSN: 0006-3444. DOI: 10.1093/biomet/89.3.659. eprint: <https://academic.oup.com/biomet/article-pdf/89/3/659/699721/890659.pdf>. URL: <https://doi.org/10.1093/biomet/89.3.659>.
- [ZZL22] Jie Zhou, Jiajia Zhang y Wenbin Lu. “TransModel: An R Package for Linear Transformation Model with Censored Data”. En: *Journal of Statistical Software* 101.9 (2022), págs. 1-12. DOI: 10.18637/jss.v101.i09. URL: <https://www.jstatsoft.org/index.php/jss/article/view/v101i09>.