

María Florencia Statti

Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina
 florencia.statti@ic.fcen.uba.ar

Gran parte de la actividad en robustez concierne al proceso de estimación, pero más allá de desarrollar estimadores robustos, el problema de realizar tests robustos también merece gran atención. De hecho, los tests de hipótesis son parte de la práctica habitual que realiza una persona que trabaja con datos. Por ejemplo, cuando se ajusta un modelo lineal, después del proceso de estimación y a fin de completar el análisis, se suelen hacer tests individuales sobre cada parámetro para verificar si es nulo o no, y así facilitar la interpretación del ajuste realizado. Este trabajo se propone introducir un estadístico robusto que permita contrastar hipótesis que involucren a la componente lineal del modelo.

En general, los test robustos han recibido un tratamiento menos extendido que la estimación robusta. Sin embargo, es sabido que los procedimientos de tests de hipótesis basados en la metodología clásica suelen heredar su sensibilidad a datos atípicos, en el sentido de que una pequeña cantidad de observaciones puede afectar el nivel o la potencia de los tests.

Es así que desarrollar tests de hipótesis que bajo contaminación retengan un nivel de significación estable, es deseable. Los trabajos de Heritier y Ronchetti (1994) y Cantoni y Ronchetti (2001) figuran entre los primeros que van en esta dirección en el campo de modelos paramétricos, el primero en un contexto general, mientras que el segundo está más enfocado a un modelo lineal generalizado. Estos autores también investigan la estabilidad del nivel asintótico bajo contaminación. Más recientemente, Bianco, Boente y Martínez (2006) y Bianco y Martínez (2009) estudian tests robustos en el caso del modelo parcialmente lineal y en el modelo logístico, respectivamente. Maronna et al. (2019) tratan el problema de tests robustos y en particular, se ocupan en el modelo lineal de los tests robustos de tipo Wald.

Consideremos el Modelo Parcialmente Lineal de Índice Simple (MPLIS) en el que se observa un vector $(y, \mathbf{x}, \mathbf{t})$, donde la variable respuesta y se relaciona con los dos vectores de covariables \mathbf{x} y \mathbf{t} mediante la ecuación

$$y = \beta_0^t \mathbf{x} + \eta_0(\boldsymbol{\theta}_0^t \mathbf{t}) + \sigma_0 \epsilon,$$

siendo $\mathbf{x} \in \mathbb{R}^p$ y $\mathbf{t} \in \mathbb{R}^q$, y donde $\beta_0 \in \mathbb{R}^p$, $\boldsymbol{\theta}_0 \in \mathbb{R}^q$ y $\sigma_0 \in \mathbb{R}$ son parámetros desconocidos y la función real univariada continua η_0 también lo es. Además asumiremos que el error ϵ es independiente del vector de covariables (\mathbf{x}, \mathbf{t}) .

Para que el modelo sea identificable, supondremos que $\|\boldsymbol{\theta}_0\| = 1$ y que su primera componente es positiva, ya que por el hecho de que η_0 sea desconocida, sólo la dirección del vector $\boldsymbol{\theta}_0$ puede ser reconocida.

La complejidad intrínseca del modelo que presenta una parte paramétrica y otra no paramétrica, hacen que el estudio de tests de hipótesis se vuelva un mayor desafío. Liang et al. (2010) desarrollan pruebas de hipótesis lineales para los coeficientes lineal e índice simple y proponen un test de bondad de ajuste para la componente no paramétrica. Este trabajo utiliza un método de perfiles que, al basarse en mínimos cuadrados, permite que datos atípicos influyan en la estimación y en consecuencia, en los estadísticos de las pruebas de hipótesis que se consideran allí.

En este trabajo, se proponen pruebas de hipótesis que involucran al parámetro lineal basadas en un estadístico de tipo Wald con el objetivo de que sean resistentes a la presencia de un pequeño porcentaje de observaciones anómalas.

Suponemos que tenemos una muestra aleatoria de vectores $(y_i, \mathbf{x}_i, \mathbf{t}_i) \subset \mathbb{R}^{p+q+1}$ para $i \in \{1, \dots, n\}$, que siguen el modelo antes descripto y el objetivo será decidir entre las hipótesis

$$H_0 : \beta_0 = \beta_* \quad \text{contra} \quad H_1 : \beta_0 \neq \beta_* .$$

Para evaluar el comportamiento de la propuesta se realizaron simulaciones para cuantificar niveles de significación y potencia de los tests, y compararlos con los obtenidos en versiones clásicas.

Gran parte de este trabajo es parte de la tesis de doctorado de la autora bajo la dirección de la Dra. Ana M. Bianco, que se puede descargar en https://web.dm.uba.ar/files/tesis_doc/statti.pdf

Referencias

- [1] Bianco A., Boente G. y Martinez E. (2006) Robust tests in semiparametric partly linear models. *Scandinavian Journal of Statistics*, 33: 435-450.
- [2] Bianco A. y Martinez E. (2009) Robust testing in the logistic regression model. *Computational Statistics and Data Analysis*, 53: 4095 - 4105.
- [3] Cantoni E. y Ronchetti E. (2001) Robust inference for generalized linear models. *Journal of the American Statistical Association*, 96: 1022-1030.
- [4] Heritier S. y Ronchetti E. (1994) Robust Bounded-Influence Tests in General Parametric Models. *Journal of the American Statistical Association*, Vol. 89, No. 427. 897-904.
- [5] Liang H., Liu X., Li R. y Tsai C. L. (2010) Estimation and testing for partially linear single-index models. *The Annals of Statistics*, 38(6): 3811-3836.
- [6] Maronna R. A., Martin R. D., Salibián-Barrera M. y Yohai V. J. (2019) *Robust statistics: theory and methods (with R)*. Second edition - John Wiley Sons, Ltd.