

## RESULTADOS OBTENIDOS:

Describa los resultados obtenidos de su investigación haciendo referencia a los objetivos originales y/o previamente modificados de su propuesta. **La extensión máxima de esta sección es de 5 páginas ( letra tamaño 10, Arial o Verdana).**

Se desarrolló una extensión del test gradiente propuesto por Terrell [Comp. Sci. Stat. 34: 206-215, 2002] al contexto de funciones de inferencia. La motivación fue proponer un mecanismo para la detección de outliers en ecuaciones de estimación generalizadas (GEE) mediante un modelo de salto en la media. Resultados que representan una alternativa a los disponibles en el artículo de Pardo y Hobza [Biometrical J. 56: 838-850, 2014]. Se debe destacar que este ha sido un trabajo desafiante pues en el ámbito de funciones de inferencia, donde GEE corresponde a un caso particular, no se dispone de una función objetivo que permita llevar a cabo la estimación de parámetros lo que dificulta la manera en que se deben realizar extensiones del test gradiente para llevar a cabo test de hipótesis de mayor generalidad. Un aspecto relevante es que también se logró extender la medida de diagnóstico basada en técnicas de eliminación de datos propuesta por Enea y Plaia [Commun. Stat. Theory 46: 1972-1982, 2017]. Estos resultados ofrecen además una alternativa a los procedimientos para detección de outliers reportados en Wei y Fung [Comput. Stat. Data An. 30: 429-441, 1999]. Nuestros resultados se presentan con mayor profundidad en un artículo en conjunto el Dr. Ángelo Gárate (doctor en Estadística por la Pontificia Universidad Católica de Chile) y la académica Cibele Russo (Instituto de Ciências Matemáticas e de Computação - USP, Brasil):

- Felipe Osorio, Ángelo Gárate, Cibele María Russo (2024). The gradient test statistic for outlier detection in generalized estimating equations. *Statistics & Probability Letters* **209**, 110087.

El análisis de diagnóstico a través del método de influencia local usando funciones diferentes que el desplazamiento de verosimilitudes ha sido pobemente explorado. En la propuesta desarrollada en Osorio [Ann. I. Stat. Math. 68: 589-619, 2016] se estudió la insensibilidad de un procedimiento robusto para llevar a cabo el ajuste no paramétrico vía splines penalizados usando métodos basados en eliminación de casos y en el procedimiento de influencia local considerando una versión penalizada de la log-verosimilitud. Diversos autores han desarrollado metodologías robustas con el objetivo de atenuar el efecto de outliers y/o observaciones influyentes en regresión semiparamétrica. En cambio la metodología utilizada en nuestro trabajo se basó en una formulación de datos incompletos para abordar la estimación de parámetros mediante un algoritmo EM penalizado. En el siguiente artículo se presentó el diagnóstico de influencia para una versión robusta del modelo de regresión semiparamétrica introducido por Eilers y Marx [Stat. Sci. 11: 89-121, 1996],

- Felipe Osorio (2016). Influence diagnostics for robust P-splines using scale mixture of normal distributions. *Annals of the Institute of Statistical Mathematics* **68** (3), 589-619.

El diagnóstico de influencia para funciones objetivo con primera derivada no cero en su punto crítico ha sido un tema poco tratado en la literatura de estadística. Utilizando esta perspectiva, llevamos a cabo el diagnóstico de influencia sobre dos medidas para determinar el grado de acuerdo entre dos instrumentos de medición. A saber, el coeficiente de correlación de concordancia propuesto por Lin [Biometrics 45: 225-268, 1989] y la probabilidad de acuerdo definida en Stevens et al. [Stat. Methods Med. Res. 26: 2487-2504, 2017]. Tales resultados, desarrollados en conjunto con la Dra. Carla Leal y el profesor Manuel Galea (Pontificia Universidad Católica de Chile), se encuentran descritos en el trabajo:

- Carla Leal, Manuel Galea, Felipe Osorio (2019). Assessment of local influence for the analysis of agreement. *Biometrical Journal* **61** (4), 955-972.

Para dar cumplimiento a los objetivos, primeramente fue requerido el desarrollo de una extensión del test gradiente al contexto general de estimadores definidos por el óptimo de una función objetivo, también conocidos como extremum estimators. No solamente se realizó una extensión a

este marco de trabajo, sino que también se consideró test de hipótesis no-lineales de la forma  $H_0 : \mathbf{g}(\boldsymbol{\beta}_0) = \mathbf{0}$ , para  $\mathbf{g}$  función continuamente diferenciable. En conjunto con el profesor Federico Crudu (Department of Economics and Statistics, Università di Siena, Italia) se propuso diversas versiones del estadístico gradiente, rotulado como ‘*estadístico de forma bilineal*’, para llevar a cabo test de hipótesis no lineales en extremum estimation. Se estudió la distribución asintótica de la estadística propuesta y mediante dos pequeños estudios de simulación Monte Carlo se pudo notar el buen desempeño del test. Los resultados pueden ser encontrado en el paper:

- Federico Crudu, Felipe Osorio (2020). Bilinear form test statistics for extremum estimation. *Economics Letters* **187**, 108885.

Este tipo de resultados resultan muy interesantes pues permiten llevar a cabo test de hipótesis en contextos de gran generalidad, tales como  $M$ -estimación, funciones de estimación, el método de momentos generalizados y estimación máximo  $Lq$ -verosímil, y ofrecen un mecanismo para la identificación de observaciones atípicas así como la posibilidad de definir residuos estandarizados a través del modelo de salto en la media.

Debemos destacar que, aunque usualmente el análisis de influencia local ha sido utilizado para la detección de observaciones atípicas y/o influyentes, algunos pocos trabajos ha utilizado este enfoque de diagnóstico como un mecanismo para criticar la construcción del modelo estadístico. Por ejemplo, en Galea et al. [Biometrical J. 47: 691-706, 2005] y Osorio et al. [Comput. Stat. Data An. 53: 1249-1263, 2009] se consideran esquemas de perturbación específicos para la mejora del modelo propuesto. En el siguiente paper:

- Felipe Osorio, Manuel Galea, Claudio Henríquez, Reinaldo Arellano-Valle (2023). Addressing non-normality in multivariate analysis using the  $t$ -distribution. *AStA Advances in Statistical Analysis* **107**, 785-813.

Se utiliza diversas herramientas para la crítica del supuesto de normalidad en el contexto de análisis multivariado. En conexión con la temática del proyecto, se considera un particular esquema de perturbación que permite la mejora del modelo propuesto desde la perspectiva de diagnóstico. Es destacable que en aquél trabajo se evaluó también la adecuacida del esquema de perturbación según la propuesta delineada por Zhu et al. [Ann. Stat. 35: 2565-2588, 2007].

En el artículo

- Jonathan Acosta, Felipe Osorio, Ronny Vallejos (2016). Effective sample size for line transect sampling models with an application to marine macroalgae. *Journal of Agricultural, Biological, and Environmental Statistics* **21** (3), 407-425,

se extendió los resultados reportados por Vallejos y Osorio [Spat. Stat. 9: 66-92, 2014]. Adicionalmente, se propuso un procedimiento de estimación eficiente para el modelo de regresión espacial cuya matriz de covarianza sigue una estructura parcialmente lineal, esto permitió extender la definición del tamaño muestral efectivo (ESS) para modelos que presentan estructura serial la que es inherente a problemas donde se desarrolla muestreo por transectas. Este es un tópico de investigación desafiante y permite analizar problemas en contextos tales como modelos para datos espaciales así como en modelos con efectos mixtos.

Es importante resaltar que las temáticas abordadas en cada unos de los trabajos anteriores dan origen a múltiples problemas desafiantes que merecen ser investigados. Por ejemplo, actualmente nos encontramos trabajando en el diagnóstico de influencia local en el contexto de regresión ridge considerando como medida de influencia la divergencia de Kullback-Leibler y la distancia de Frèchet. Un aspecto relevante de aquél trabajo es que la metodología propuesta permite la detección de outliers que generan colinealidad. También estamos concentrados en estudiar con mayor detalle la pérdida de invarianza del test de forma bilineal, así como posibles soluciones de este aspecto.

## Productos generados por el proyecto

Los resultados obtenidos a raíz de este proyecto son reportados en los siguientes trabajos:

6. Osorio, F., Gárate, Á., Russo, C.M. (2024). The gradient test statistic for outlier detection in generalized estimating equations. *Statistics & Probability Letters* **209**, 110087.
5. Osorio, F., Galea, M., Henríquez, C., Arellano-Valle, R. (2023). Addressing non-normality in multivariate analysis using the *t*-distribution. *AStA Advances in Statistical Analysis* **107**, 785-813.
4. Crudu, F., Osorio, F. (2020). Bilinear form test statistics for extremum estimation. *Economics Letters* **187**, 108885.
3. Leal, C., Galea, M., Osorio, F. (2019). Assessment of local influence for the analysis of agreement. *Biometrical Journal* **61** (4), 955-972.
2. Acosta, J., Osorio, F., Vallejos, R. (2016). Effective sample size for line transect sampling models with an application to marine macroalgae. *Journal of Agricultural, Biological, and Environmental Statistics* **21** (3), 407-425.
1. Osorio, F. (2016). Influence diagnostics for robust P-splines using scale mixture of normal distributions. *Annals of the Institute of Statistical Mathematics* **68** (3), 589-619.

Resultados preliminares asociados a este proyecto han sido presentados en las siguientes tesis de pre y posgrado:

- Labalobich, E. (2018). *Estimación maximo Lq-verosímil en modelos lineales generalizados*. Instituto de Estadística, Pontificia Universidad Católica de Valparaíso (para obtener el título de Estadístico).
- Marcela Miranda. *Estimación penalizada en modelos lineales generalizados via Lq-verosimilitud*. Departamento de Matemática, Universidad Técnica Federico Santa María (para obtener el título de Ingeniero Civil Matemático).
- Alonso Ogueda (2018). *Diagnóstico de influencia en regresión ridge basado en una función de divergencia*. Departamento de Matemática, Universidad Técnica Federico Santa María (para obtener el título de Ingeniero Civil Matemático).
- Paula Guerrero (2017). *Influencia local en estimación máximo Lq-verosímil*. Instituto de Estadística, Pontificia Universidad Católica de Valparaíso (para obtener el grado de Magíster en Estadística).
- Diego Estay (2017). *Estimación robusta en modelos de curvas de crecimiento*. Instituto de Estadística, Pontificia Universidad Católica de Valparaíso (para obtener el título de Estadístico).
- Jessenia Cortés (2016). *Leverage generalizado en estimación máximo Lq-verosímil*. Instituto de Estadística, Pontificia Universidad Católica de Valparaíso (para obtener el grado de Magíster en Estadística).
- Jessica García (2015). *Test gradiente en funciones de inferencia*. Instituto de Estadística, Pontificia Universidad Católica de Valparaíso (para obtener el título de Estadístico).
- Agustín Uribe (2015). *Detección de outliers en ecuaciones de estimación*. Departamento de Matemática, Universidad Técnica Federico Santa María (para obtener el título de Ingeniero Civil Matemático).

Es importante destacar que durante el año 2020 en conjunto con los profesores Ronny Vallejos (UTFSM) y Moreno Bevilacqua (UAI) fue publicado el libro:

- Ronny Vallejos, Felipe Osorio, Moreno Bevilacqua (2020). *Spatial Relationships Between Two Georeferenced Variables: with Applications in R*. Springer, Cham.

## Difusión a la sociedad

Algunos de los resultados asociados a este proyecto han sido difundidos en las siguientes charlas:

- “*Toma de Decisiones y Estadística*”. XXIII Semana Nacional de la Ciencia y la Tecnología, programa 1000 Científicos 1000 Aulas. Noviembre 2017.
- “*Aleatoriedade e Modelagem Estatística*” (en portugués). Ciclo de Seminários de Orientación Académica. Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, Brasil. Agosto 2016.
- “*Aplicaciones del test gradiente en funciones de inferencia, diagnóstico de influencia e hipótesis no lineales*”. 7ma V-Escuela del Doctorado en Matemática de Valparaíso. Noviembre 2018 (dictado para potenciales estudiantes del Doctorado en Matemática Consorcio PUCV-UTFSM-UV).

Debemos destacar que nuestro libro *Spatial Relationships Between Two Georeferenced Variables: with Applications in R* publicado por Springer ha servido como un vehículo para diseminar algunos de los resultados desarrollados como parte de este proyecto y ha sido muy usado así como citado por comunidades que llevan a cabo análisis cuantitativos aplicados. En efecto, diversos capítulos del libro han sido usados por profesionales e investigadores de áreas como Ecología, Biología Marina, Climatología y Riesgos Naturales, entre otros (para un listado de referencias que han citado los trabajos producto del proyecto consultar la página <https://github.com/faosorios/FONDECYT-1140580/blob/main/Citations.md>).

Otro medio que ha sido muy efectivo para la diseminación de los resultados asociados al proyecto y que ha sido bastante utilizado y citado, corresponde a la implementación computacional de las metodologías propuestas como parte del mismo en los siguientes paquetes para el ambiente para cálculo estadístico R. A saber:

- Osorio, F. (2019). *heavy: Robust estimation using heavy-tailed distributions*. R package version 0.38.196. URL: [CRAN.R-project.org/package=heavy](https://CRAN.R-project.org/package=heavy)
- Osorio, F. (2024). *MVT: Estimation and testing for the multivariate t-distribution*. R package version 0.3-81. DOI: [10.32614/CRAN.package.MVT](https://doi.org/10.32614/CRAN.package.MVT)
- Osorio, F., Vallejos, R. (2024). *SpatialPack: Tools for assessment of the association between two spatial processes*. R package version 0.4-1. DOI: [10.32614/CRAN.package.SpatialPack](https://doi.org/10.32614/CRAN.package.SpatialPack)

Bibliotecas que en su conjunto han recibido 134 citas, mientras que han sido descargadas en al menos 232 mil oportunidades.