

# UN NUEVO PARADIGMA EN LA ESTADÍSTICA: LOS MÉTODOS ROBUSTOS

**Palabras clave:** Métodos estadísticos robustos, observaciones atípicas, modelo de regresión lineal.  
**Key words:** Robust statistical methods, outliers, linear regression model.

■ Víctor J. Yohai

Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, UBA  
Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, UBA  
CONICET

[vyohai@dm.uba.ar](mailto:vyohai@dm.uba.ar)

## ■ MI INFANCIA Y ADOLESCENCIA

Nací en el año 1939 en el seno de una familia judía sefardita. Mis padres habían emigrado de Turquía, y a pesar de haber nacido y vivido hasta a la adultez en Turquía, prácticamente desconocían el idioma turco. Ellos eran descendientes de los judíos españoles que habían sido expulsados de España en el año 1492, y como la mayoría de los judíos de ese origen hablaban el ladino o judeo-español que es un idioma muy cercano al que se hablaba en España en 1492 matizado con algunas palabras en turco y en hebreo. Eso facilitó su inserción en Argentina ya que su idioma difería muy poco del español hablado en Buenos Aires. Mi padre nunca fue a la escuela, de modo que era prácticamente analfabeto aunque podía hacer eficientemente los cálculos que surgían del ejercicio de su actividad como comerciante. En cambio, mi madre tenía una instrucción primaria completa impartida por la así llamada *Alliance Francaise Universelle*. Ésta era una institución francesa que proveía enseñanza primaria y se-

cundaria a los judíos sefarditas que habitaban los países con una importante población de ese origen como Turquía, Marruecos, Bulgaria o Grecia. La primera actividad de mi padre cuando llegó a Buenos Aires, fue la de vendedor ambulante. Sin embargo, al poco tiempo consiguió establecer una tienda, cuya actividad más importante era la venta de telas. A pesar de su escasa educación mi padre hizo que este comercio resultase un próspero negocio.

Mi educación primaria y secundaria se desarrolló en colegios públicos. La primaria en la escuela Santa Fe en el barrio de Saavedra y la secundaria en el Colegio Nacional Julio A. Roca en el barrio de Belgrano. Analizando retrospectivamente la educación recibida, considero que la escuela primaria fue excelente, especialmente en los últimos tres años a cargo del mismo maestro, un hombre de una extensa cultura que nos despertó interés en matemáticas, ciencias naturales, lectura de libros y redacción, etc. Por el contrario la enseñanza en la escuela secundaria dejaba mucho que desear.

Por un lado la mayoría de los profesores eran muy mediocres y además una parte importante del tiempo estaba dedicado al adoctrinamiento político, algo muy propio de las presidencias de Perón. En las clases de castellano había que leer el libro *La Razón de mi Vida* y en la materia Cultura Ciudadana se enseñaba la doctrina peronista. Hubo una sola excepción a la mediocridad general de nuestros profesores secundarios, la profesora de Matemáticas que nos enseñó esta materia desde segundo a quinto año. A pesar de no ser especialmente carismática ejercía su cátedra con seriedad y exigencia. Esto no la hacía popular entre los alumnos, ya que para aprobar su materia se requería mucha dedicación. Aquellos alumnos que teníamos una mayor predisposición para las matemáticas seguíamos sus clases con interés y dedicación, lo que nos permitió terminar el colegio secundario con una buena formación en dicha materia. Esto determinó en gran medida mi elección de la Licenciatura en Ciencias Matemáticas como carrera universitaria.

## ■ 1. ELECCIÓN DE UNA CARRERA UNIVERSITARIA

Sin embargo, hubo otra razón para esta elección. Durante mis estudios secundarios y a pesar de la mediocridad de los profesores, fueron varias las áreas del conocimiento que me interesaron, pero la Matemática tenía una característica que la distinguía de otras ciencias. Sus enunciados eran verdades absolutas que no requerían para su validez la verificación experimental como en las Ciencias Naturales. Tampoco dependían de juicios de valor u opiniones subjetivas como en las Ciencias Sociales. En las Ciencias Sociales como Historia, Economía o Sociología podían convivir teorías diferentes y la aceptación de una u otra muchas veces respondía a ideologías y no a factores objetivos. En las ciencias naturales la verificación de una teoría es generalmente experimental y por lo tanto pueden surgir dudas sobre si los experimentos que la validaron fueron correctamente realizados. En cambio el criterio de verdad de una teoría matemática era únicamente lógico y ajeno a toda consideración ideológica, moral, religiosa o experimental. En mi inseguridad adolescente la Matemática era la única ciencia que me daba garantía de que sus enunciados eran verdades absolutas y esa fue una de las razones que me hicieron elegirla para mis estudios universitarios.

Otra razón para elegir esta disciplina es el profundo sentido de belleza que se encuentra en la matemática. Esa belleza se manifiesta en los métodos de demostración, la armonía y profundidad de algunos de sus resultados y en la misma experiencia de hacer matemáticas. Resulta muy difícil transmitir el sentido de la belleza de la Matemática. Cito un pensamiento del matemático húngaro Paul Erdős

“Preguntar por qué los números son bellos es como preguntar por qué es bella la novena sinfonía de Beethoven. Si uno no se da cuenta solo por qué es bella, nadie se lo podrá explicar.”

Búsqueda de verdades absolutas y un profundo sentido estético de belleza. Éstas fueron las principales razones que me llevaron a estudiar Matemáticas. Puedo decir que disfruté plenamente todos los años que estudié Matemáticas en la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires. Además de estudiar las diferentes teorías en los libros, desde muy temprano se nos planteaban en los cursos problemas a resolver que eran verdaderos desafíos a nuestra inteligencia y que nos sumergían en el mundo maravilloso de esta ciencia. Quisiera agradecer y recordar a mis maestros que tanta importancia tuvieron en mi formación en esta etapa de mi vida. En particular debo mencionar a Micha Cotlar y Gregorio Klimovsky que me enseñaron a entender, amar y disfrutar de la Matemática.

## ■ 2. MI EXPERIENCIA EN EL EQUIPO DE MODELOS MATEMÁTICOS DE LA ECONOMÍA DIRIGIDO POR OSCAR VARSAVSKY

Luego de finalizar la Licenciatura en Matemática en marzo de 1962 decidí que quería realizar una carrera académica en Matemáticas, para lo cual debería completar estudios de Doctorado. La decisión de en qué área de las Matemáticas iba a realizar mi tesis me generó un conflicto interior de difícil resolución. Por un lado la Matemática pura me atraía especialmente por la libertad de elegir los problemas a investigar basado solamente en el interés matemático, sin considerar sus posibles aplicaciones. Pero por otro lado sentía la responsabilidad de

trabajar en temas que fueran útiles a la sociedad en su conjunto y no solo a la comunidad matemática. Es cierto que muchos de los temas que se desarrollaron como ramas de la Matemática pura, con el tiempo, resultaron ser herramientas formidables para resolver importantes problemas prácticos. Un ejemplo es la Teoría de Números que se desarrolló como una rama totalmente abstracta de la Matemática sin tener en consideración ninguna aplicación. En el presente, la Teoría de Números se ha convertido en una herramienta fundamental para el desarrollo de la Criptografía. Sin embargo no siempre ocurre así, y las aplicaciones, cuando se encuentran, pueden aparecer luego de muchos años. En ese momento yo necesitaba que esa utilidad social estuviese garantizada y que surgiese en un tiempo corto.

En 1962, año de mi graduación, volvió al Departamento de Matemática de la Facultad de Ciencias Exactas y Naturales el Dr. Oscar Varsavsky quien había iniciado su carrera científica obteniendo un Doctorado en Química. Pero al poco tiempo se interesó en la Matemática y en particular en la Topología, destacándose rápidamente por su notable capacidad para la investigación en Matemáticas. Esto le valió su nombramiento como Profesor Titular del Departamento de Matemáticas. Sin embargo luego de su retorno su interés cambió radicalmente, y su mira fue puesta en la aplicación de la Matemática para resolver los problemas económicos de nuestro país. Con este propósito inició un proyecto destinado a construir un modelo matemático de la economía argentina. Este modelo se iba a desarrollar con ayuda de la computadora Mercury-Ferranti (conocida popularmente con el sobrenombre de Clementina) instalada, hacía poco tiempo, en el Instituto de Cálculo creado por el Dr. Manuel Sadosky. El objeti-

vo de este modelo era evaluar políticas económicas de manera de poder elegir aquellas que fuesen más convenientes para el país. Al conocer mi interés por la matemática aplicada, el Dr. Varsavsky me entusiasmó para que integrara su equipo. Como los objetivos de este proyecto coincidían con mi interés para usar la matemática como una herramienta para resolver problemas útiles a la sociedad, me integré al grupo del Dr. Varsavsky. El grupo de trabajo lo integraban otros matemáticos y un economista. Los matemáticos eran Mario Malajovich, Nélica Lugo y Roberto Frenkel y el economista Arturo O'Connell. En realidad O'Connell tenía dos títulos, de Licenciado en Economía y en Matemática, pero su papel en el grupo era el de economista. Al poco tiempo de ingresar al grupo surgió un ofrecimiento de una beca para estudiar Economía en el *Centre d'Etudes de Programmes Economiques* (CEPE) en París. El Dr. Varsavsky me convenció que los conocimientos que podía adquirir en el CEPE podrían ser muy útiles para mi labor en el grupo, razón por la cual en septiembre de 1962 partí para París.

Allí me encontré con una ciencia como la Economía que aunque utiliza la Matemática como herramienta, tiene base eminentemente empírica, y donde la decisión de si una cierta política económica es buena o mala depende muchas veces del sistema de valores individuales de cada persona. Es decir, me tropecé justamente con los problemas que quería evitar cuando decidí estudiar Matemática. Sin embargo mi estadía en el CEPE tuvo un aspecto positivo ya que por primera vez asistí a un curso de Estadística con una base matemática, ya que en la Licenciatura en Matemáticas, en ese entonces, no se dictaban cursos similares. Aunque el curso no era de muy alto nivel, despertó mi interés ya que en-

contré que la Estadística hacía uso de la Matemática para resolver problemas reales que podían provenir de distintas ramas de la Ciencia y la Tecnología, y eso era justamente lo que yo estaba buscando.

Al volver a Buenos Aires, le comenté al Dr. Varsavsky el interés que me había despertado la Estadística como así también la decepción que me había causado la Economía. El Dr. Varsavsky nunca había estudiado Estadística, pero tenía una amplia información de los diferentes usos que se podían dar a muchos de los métodos de esta disciplina. De manera que al conocer mi interés por la Estadística me recomendó que estudiara el así llamado método de Componentes Principales, ya que pensaba que podría ser útil para analizar los resultados del modelo económico que estábamos desarrollando.

El método de Componente Principales es una técnica estadística desarrollada por el estadístico suizo Hotelling (1933). El objetivo de este procedimiento es disminuir el número de variables que describen los objetos que se analizan perdiendo la menor cantidad de información posible. El Dr. Varsavsky pensó que dado que un estado de la economía debe ser descripto por un alto número de variables, reducir este número con poca pérdida de información podía simplificar el análisis de los resultados del modelo que se estaba construyendo. El tema me pareció fascinante y como la forma de medir la *pérdida de información* que daba Hotelling en su trabajo original (y que se encontraba en todos los trabajos a los que tuve acceso) no me convenía, busqué una definición alternativa que se acercara más al significado que, para mi intuición, tenía este concepto. De manera que desarrollé una nueva formulación del concepto de *pérdida de infor-*

*mación* y reformulé el problema de reducir el número de variables en forma óptima usando el nuevo concepto. Mi sorpresa fue que obtuve la misma solución que la obtenida originariamente por Hotelling. Es decir dos formas distintas de definir *pérdida de información* llevaban a la misma forma de reducir la dimensión de los datos. Este fue mi primer trabajo de investigación en Estadística realizado con una formación en Estadística, por demás primaria e incompleta.

En marzo de 1966 ocurrió el golpe de estado encabezado por el General Onganía y al poco tiempo la intervención de la Universidad de Buenos Aires. Esta intervención fue especialmente cruenta en la Facultad de Ciencias Exactas y Naturales donde se produjo una brutal acción de la policía en el episodio conocido como *noche de los bastones largos* y donde varios de los profesores de la Facultad fueron golpeados. Yo, que en ese entonces tenía un cargo Jefe de Trabajos Prácticos con Dedicación Exclusiva, me adherí a la renuncia masiva de docentes de la Facultad como protesta a la intervención de la Universidad y a la brutalidad policial.

Luego de mi renuncia el Dr. Varsavsky me invitó a continuar mi carrera en la Universidad Central en Caracas, Venezuela, país donde él había sido convocado para continuar su trabajo sobre modelos matemáticos de la economía. Allí permanecí hasta septiembre del año 1967. Durante ese período me dediqué a estudiar Estadística en forma autodidáctica. Sin embargo iba creciendo en mi la convicción de que para realizar seriamente investigación en Estadística debía completar mi formación haciendo un doctorado en algún centro de excelencia en esa disciplina.

### ■ 3. MIS ESTUDIOS DE DOCTORADO EN ESTADÍSTICA EN LA UNIVERSIDAD DE CALIFORNIA, BERKELEY

Para concretar estos deseos, expresé mi interés de realizar un doctorado en Estadística a dos de mis ex - compañeros de la Facultad que estaban haciendo un Ph.D. en el Departamento de Estadística del campus de Berkeley de la Universidad de California, Alejandro de Acosta y Pedro Fernández. Como mis compañeros estaban altamente conceptuados, su recomendación bastó para que me aceptaran como estudiante de doctorado y para que me otorgaran un cargo de *teaching assistant* que me permitiera financiar mis estudios. Debo mencionar que en ese momento el Departamento de Estadística de Berkeley estaba conceptuado como uno de los centros de Estadística Teórica más prestigiosos del mundo con eminentes profesores ya consagrados como Erich L. Lehmann, Lucien Le Cam, David Blackwell y Michel Loève. También pertenecían al Departamento algunos profesores jóvenes como Peter Bickel que ya se perfilaban con un futuro brillante.

Durante el primer año de mi estadía en Berkeley disfruté muchísimo del programa de doctorado, ya que todas las materias cursadas me resultaron sumamente interesantes. Lamentablemente cuando comenzaba mi segundo año, mi madre me escribió que estaba con una gran depresión provocada principalmente por el fallecimiento de mi padre ocurrido hacía poco tiempo y me manifestaba su deseo de que volviese lo antes posible a Buenos Aires. Como no quería volver sin completar el Ph.D., decidí hacer un esfuerzo para lograr este objetivo en el menor tiempo posible. El profesor Roy Radner que tenía un nombramiento conjunto en los Departamentos de Estadística y Economía, conociendo

que yo había estudiado Economía y trabajado en Modelos Económicos me propuso un tema para mi tesis que combinaba Estadística con Economía. Aunque el tema no me entusiasmaba demasiado, me pareció que era factible obtener los resultados requeridos en el lapso que yo me había fijado, por lo que decidí aceptar su propuesta. En efecto, al cabo de dos años de mi estadía en Berkeley completé mi Ph.D. y emprendí mi regreso a Buenos Aires en marzo de 1970.

### ■ 4. RETORNO A LA ARGENTINA, EL DESCUBRIMIENTO DE LA ESTADÍSTICA ROBUSTA

Cuando volví a Buenos Aires donde todavía gobernaba el régimen militar del General Onganía y debido a mi renuncia en 1966, mi nombre figuraba todavía en las listas negras de la Universidad de Buenos Aires. Sin embargo, el Dr. Miguel Herrera que en ese entonces era el Director del Departamento de Matemática de la Facultad de Ciencias Exactas de la Universidad Nacional de La Plata se mostró muy interesado en que me incorporase a su Departamento, ya que allí la Estadística era un área de vacancia. Como no estaba especialmente interesado en continuar investigando con el tema de mi tesis durante un tiempo me dediqué a revisar la literatura estadística buscando algún problema cuya resolución requiriese utilizar de lleno mi entrenamiento matemático y que simultáneamente tuviera una utilidad práctica para el análisis estadístico de datos reales. Así llegué a encontrar un artículo de Peter Huber (1964) que me fascinó. Allí se formulaba de manera magistral la teoría de la estimación robusta para un modelo de medición. En este artículo se describían las debilidades de la teoría estadística clásica de estimación que supone que cuando se toma una muestra de mediciones de

una magnitud, los errores siguen una distribución normal. A partir de esta hipótesis se deduce que el estimador óptimo de la medida real está dado por el promedio de las mediciones. Huber mostraba que el promedio era peligrosamente inestable, es decir si la distribución de los errores se apartaba un poco de la normalidad, el promedio podía comportarse muy ineficientemente. Más aún bastaba que una sola observación hubiese sido obtenida bajo alguna situación atípica, por ejemplo bajo un mal funcionamiento del instrumento de medición, para que el promedio se arruinara completamente. Huber proponía utilizar estimadores que fueran eficientes bajo el modelo normal pero también cuando esta distribución estuviese cerca de la normal. Además, el estimador debería ser poco sensible a la presencia, en la muestra, de una pequeña fracción de observaciones atípicas conocidas en Estadística como *outliers*. Los estimadores con estas propiedades eran bautizados como *estimadores robustos*. Huber definía en este trabajo una nueva familia de estimadores, los M-estimadores, que son robustos y encontraba dentro de esta familia el estimador que optimizaba un criterio que combinaba la eficiencia bajo el modelo normal con las propiedades de robustez antes mencionadas.

La lectura de este trabajo tuvo una influencia que duró durante toda mi carrera científica, ya que la mayoría de los trabajos que realicé tuvieron como objetivo la búsqueda de estimadores robustos para diversos modelos estadísticos, como así también el estudio de sus propiedades. Mi primer trabajo sobre estimación robusta que fue publicado en *Annals of Statistics* (1974) estuvo vinculado con la extensión de los M-estimadores que Huber había propuesto para el modelo de medición al modelo de regresión lineal.

El modelo de regresión lineal establece que una variable  $Y$ , llamada variable dependiente es explicada en forma lineal por otras variables  $X_1, \dots, X_k$  llamadas variables independientes, en forma lineal. Para esto se supone que  $Y = b_0 + b_1 X_1 + \dots + b_k X_k + e$ , donde los coeficientes  $b_i$ ,  $i=1, \dots, k$  son desconocidos y  $e$  es el error. Supongamos que se tiene una muestra de  $n$  observaciones  $(Y_i, X_{i1}, \dots, X_{ik})$ ,  $i=1, \dots, n$ , que satisfacen el modelo, es decir tales que  $Y_i = b_0 + b_1 X_{i1} + \dots + b_k X_{ik} + e_i$ , y por lo tanto  $e_i = Y_i - b_0 - b_1 X_{i1} - \dots - b_k X_{ik}$ . La teoría clásica asume que los errores  $e_i$  tienen distribución normal, todos con la misma varianza y que, además, son independientes. Bajo estas suposiciones se deriva que el estimador óptimo se obtiene minimizando la suma de los cuadrados de los errores, es decir minimizando  $e_1^2 + \dots + e_n^2$ . Los M-estimadores robustos reemplazan la función cuadrática  $q(e) = e^2$  por otra función  $q$  que crece más lentamente. En ese trabajo (Yohai, 1974) obtuve la distribución asintótica de los M-estimadores del modelo de regresión lineal, es decir la distribución de los estimadores cuando el tamaño de la muestra tiende a infinito. También encontramos la función  $q$  óptima dentro de la familia de funciones definida por Huber (1964).

## ■ 5. MI INGRESO COMO PROFESOR A LA UNIVERSIDAD DE BUENOS AIRES Y MIS PRIMEROS TESISTAS

En el año 1973 con el regreso de la democracia, me invitaron a ingresar como profesor en el Departamento de Matemática de la Universidad de Buenos Aires, invitación que acepté a pesar del importante vínculo afectivo que me unía al Departamento de Matemática de la Universidad de La Plata. Una de las razones para aceptar la propuesta era el hecho de que al vivir en Buenos Aires el viaje de ida y vuelta a La Plata me insumía alrededor de

4 horas. En la Universidad de Buenos Aires me encontré con Ricardo Maronna, quien había terminado la Licenciatura en Ciencias Matemáticas en la Universidad Nacional del Sur y que en ese momento tenía un cargo de Jefe de Trabajos Prácticos. Ricardo había estudiado Estadística en forma autodidáctica y ya había publicado un trabajo en esa disciplina. Le comenté el tema en el que estaba trabajando y le indiqué algunos trabajos (incluyendo el mío propio) para que se interiorizara de la problemática de la Estadística Robusta. Ricardo se entusiasmó con el tema y manifestó su interés en hacer una tesis doctoral en Estadística Robusta. Le propuse que en su tesis estudiara el problema de estimación robusta de posición y matriz de covarianza para datos multivariados, tema que a Ricardo le pareció muy interesante. Durante todo el tiempo que Ricardo trabajó en su tesis no necesitó hacerme ninguna consulta técnica. Luego de algún periodo de tiempo, que lo recuerdo como relativamente corto, me entregó el manuscrito con la tesis terminada. Los resultados obtenidos eran muy importantes y algunos de ellos realmente sorprendentes. Mi trabajo se redujo a corregir algunas erratas y editar el texto. El trabajo fue enviado al *Annals of Statistics* donde fue rápidamente aceptado y publicado (Maronna, 1976). Hoy es uno de los trabajos de robustez que tiene más citas (de acuerdo al Google Scholar: 605 citas). Luego de terminada su tesis continuamos trabajando juntos y ha sido el principal co-autor en toda mi carrera científica. Entre los trabajos publicados con Ricardo puedo mencionar Maronna y Yohai (1979), (1981), (1993), (1995), (2000) y (2008).

A partir de mi nombramiento en la FCEyN de la UBA comencé a dictar diversos cursos sobre temas de Estadística para alumnos de la Licenciatura en Ciencias Matemáticas. Muchos de estos temas me eran

completamente nuevos, ya que dada mi corta estadía en Berkeley no había tenido tiempo de estudiarlos allí. De esta manera estos cursos no sólo sirvieron para que aprendieran los alumnos sino también fueron, para mí, una importante fuente de aprendizaje. Entre otros, dicté cursos de Análisis Multivariado, Series de Tiempo, Análisis de la Varianza, etc. Algunos alumnos de los cursos que dicté en estos años se interesaron en realizar una tesis de doctorado con temas de Estadística. Luego de Ricardo Maronna, los siguientes alumnos de doctorado fueron Ricardo Fraiman y Graciela Boente. Ellos son actualmente reconocidos investigadores que publican en las revistas de Estadística más prestigiosas. Además de interesarse por lo Métodos robustos, ellos también abrieron nuevas líneas de investigación como Métodos no Paramétricos y Datos Funcionales. Dado que Ricardo Fraiman y Graciela Boente se independizaron muy rápidamente, tengo sólo tres trabajos en los que aparecen como coautores. Estos son Bustos, Fraiman y Yohai (1984), Boente, Fraiman y Yohai (1987), y Fraiman, Yohai y Zamar (2001).

## ■ 6. MI TRABAJO EN COLABORACIÓN CON DOUG MARTIN, PROFESOR DE LA UNIVERSIDAD DE WASHINGTON, SEATTLE

Hasta el año 1980 toda mi actividad científica se desarrolló sin ningún contacto con investigadores de otros centros del exterior. Ese año se llevó a cabo en Mar del Plata la Escuela Latinoamericana de Matemáticas que estuvo dedicada a Matemática Aplicada y que fue organizada por el Dr. Miguel Herrera. Miguel Herrera era un prestigioso matemático teórico dedicado al área de Varias Variables Complejas, pero que en ese momento estaba interesándose en temas de Matemática Aplicada especialmente en Investigación Operativa. Miguel Herrera

me solicitó que organizara una subsección de la escuela dedicada a la Estadística. Con ese motivo invité a varios investigadores extranjeros y en particular a Doug Martín, Profesor de la *University of Washington* en Seattle, EE.UU. quien trabajaba en métodos robustos aplicados a Series de Tiempo. De mi encuentro con Doug surgió una fructífera colaboración científica que duró aproximadamente hasta el año 2000. Durante todo este tiempo, cada año fui invitado por períodos que variaban entre 15 a 60 días al Departamento de Estadística de la *University of Washington*, financiado con los subsidios de investigación que Doug recibía. Esta colaboración se interrumpió en parte cuando Doug decidió crear una empresa dedicada a la producción de software estadístico y como consecuencia disminuyó su dedicación a la investigación. Doug creó una compañía que produjo el programa estadístico SPLUS, el cual marcó una innovación muy importante en la concepción de software estadístico para PC. Pocos años después Doug me invitó a colaborar con el diseño de un paquete de Estadística Robusta para SPLUS. Ésta fue la primera implementación de los métodos robustos en un software de uso masivo. Luego de esta implementación otras compañías que producían software estadístico incorporaron en sus productos paquetes robustos y en particular algunos de los procedimientos de mi autoría. Puedo mencionar los siguientes trabajos realizados en colaboración con Doug Martin: Douglas Martin y Víctor J. Yohai (1984), (1986) y Martin, Yohai y Zamar (1989)

#### ■ 7. MI TRABAJO EN COLABORACIÓN CON RUBÉN ZAMAR, PROFESOR DE LA UNIVERSIDAD DE BRITISH COLUMBIA

Otro colaborador muy importante que he tenido fue Rubén Zamar.

Conocí a Rubén cuando era profesor en la Universidad de Recife, Pernambuco, Brasil. Rubén tenía un título de grado de Contador de la Universidad Nacional de Córdoba y un Master en Estadística del Centro Interamericano de Estadística con sede en Santiago de Chile. En el año 1981 Rubén me solicitó que dirigiera su tesis de Doctorado. Como yo no podía otorgarle financiación en Buenos Aires, le solicité a Doug Martín que dirigiera su tesis en el programa de Ph.D. del Departamento de Estadística en la *University of Washington*. A pesar de no tener estudios avanzados en Matemática, Rubén resultó un brillante estudiante y durante mis visitas a Seattle lo incorporamos a los proyectos de investigación que realizábamos con Doug. Luego de haber obtenido su Ph.D., Rubén se incorporó como *Assistant Professor* al Departamento de Estadística de la Universidad de British Columbia, Vancouver, Canadá, donde actualmente es *Full Professor*. La colaboración científica con Rubén ha sido muy fructífera y se mantiene hasta estos días con frecuentes visitas de Rubén a Buenos Aires y más a Vancouver. Algunos de los trabajos más importantes con Rubén son: Yohai y Zamar (1988), (1993), (1997), (2001), (2004), Martin, Yohai y Zamar (1989), Alqallaf, Van Aelst, Yohai y Zamar (2009) y Danilov, Yohai y Zamar (2012).

#### ■ 8. MI TRABAJO EN COLABORACIÓN CON DANIEL PEÑA, PROFESOR DE LA UNIVERSIDAD CARLOS III DE MADRID

En el año 1992, recibí una invitación del Profesor Daniel Peña, para dar un curso de una semana sobre Métodos Robustos en el Instituto Nacional de Estadística de España, institución destinada a producir las estadísticas oficiales. Daniel se había formado con uno de los estadísticos más importantes del siglo XX,

el Profesor George Box de la Universidad de Wisconsin, EE.UU. La especialidad de Daniel dentro de la Estadística eran las Series de Tiempo y los métodos de Diagnóstico. Los objetivos de los procedimientos de Diagnóstico son similares a los de los procedimientos robustos, por ejemplo eliminar la influencia de los *outliers* en los procedimientos estadísticos. Sin embargo las metodologías son muy diferentes, mientras los métodos robustos están diseñados de manera que los *outliers* tengan una influencia muy limitada, los procedimientos de diagnóstico utilizan los procedimientos clásicos que son altamente influidos por los *outliers*, pero a continuación analizando los resultados de los mismos se puede algunas veces encontrar los *outliers*. Una segunda estimación usando el procedimiento clásico en una muestra donde los *outliers* encontrados son omitidos permite obtener los resultados deseados. Durante mi estadía en Madrid, conversando con Daniel encontramos formas de combinar los procedimientos de diagnóstico con los métodos robustos (ver Peña y Yohai, 1995, 1999). La colaboración con Daniel continuó durante todos estos años, aunque más recientemente esa colaboración disminuyó debido a que Daniel fue designado Rector de la Universidad Carlos III. Otro trabajo importante con Daniel es Muler, Peña y Yohai (2009) donde se proponen estimadores robustos para modelos ARMA.

#### ■ 9. MI TRABAJO EN COLABORACIÓN CON ALFIO MARAZZI, PROFESOR DE LA UNIVERSIDAD DE LAUSANA

Otro colaborador muy frecuente en mis trabajos de investigación es Alfio Marazzi. Alfio se doctoró con Peter Huber en el ETH de Zurich. Alfio fue otro de los especialistas que Doug Martin convocó, con motivo del desarrollo del paquete robusto

para SPLUS en 1998. Aunque yo me había encontrado con Alfio en varios Congresos Estadísticos, fue en esa ocasión donde comenzamos a pensar en una colaboración y para su concreción formulamos varios proyectos de investigación. Desde entonces esta colaboración no ha cesado con continuas visitas de Alfio a Buenos Aires y más a la Universidad de Lausana donde Alfio es Profesor. Con Alfio realizamos, entre otros, trabajos sobre estimación robusta de regresión lineal con errores asimétricos (ver Marazzi y Yohai, 2004), estimación robusta en modelos de regresión cuando la respuesta está transformada (ver Marazzi, Villar y Yohai (2009) y estimación robusta en modelos de regresión con datos censurados y errores asimétricos (ver Locatelli, Marazzi y Yohai, 2011).

#### ■ 10. MI TRABAJO EN COLABORACIÓN CON MATÍAS SALIBIAN BARRERA, PROFESOR DE LA UNIVERSIDAD DE BRITISH COLUMBIA

Otro asiduo colaborador en mis trabajos de investigación es Matías Salibian-Barrera quien hizo la tesis de Licenciatura bajo mi dirección y luego se doctoró en la Universidad de British Columbia bajo la dirección de Rubén Zamar. Actualmente es *Associate Professor* en esa misma Universidad. En co-autoría con Matías hemos publicado varios trabajos. En los últimos años trabajamos en algoritmos rápidos para el cálculo de estimadores robustos de regresión (ver Salibian Barrera y Yohai, 2008) y métodos robustos para el modelo lineal cuando la variable dependiente está censurada (ver Salibian Barrera y Yohai, 2010). Los datos censurados ocurren por ejemplo en los datos médicos cuando se estudia la sobrevivencia de un paciente luego de algún tratamiento. La censura sobre la sobrevivencia ocurre cuan-

do el paciente deja de concurrir al consultorio y en ese caso el último dato sobre la sobrevivencia del paciente está dado por la última vez que concurrió al consultorio

#### ■ 11. MIS ALUMNOS DE DOCTORADO

Entre los colaboradores locales puedo citar también a mis ex alumnos de doctorado no mencionados anteriormente: Jorge Adrover, Andrea Bergesio, Ana Bianco, Nelly Ferretti, Marta García Ben, Daniel Gervini, Sonia Hernández Alonso, Diana Kelmansky, Elena Martínez, Marcela Svarc, Marina Valdora y Ana J. Villar. Algunos de los trabajos más importantes realizados con ellos son: Ferretti, Kelmansky y Yohai (1991), Adrover, Bianco y Yohai (1994), Bianco, García Ben, Martínez and Yohai (1996), Bianco y Yohai (1996), García Ben, Martínez y Yohai (1999), Ferretti, Kelmansky, Yohai y Zamar (1999), Bianco, García Ben, Martínez y Yohai (2001), Gervini y Yohai (2002), Adrover y Yohai (2002), Hernández Alonso y Yohai (2003), Adrover, Maronna y Yohai (2004), Bianco, García Ben y Yohai (2005), García Ben, Martínez y Yohai (2006), Bergesio y Yohai (2011).

#### ■ 12. OTROS COLOBORADORES DE ARGENTINA

Con Nora Muler, profesora de la Universidad Torcuato di Tella hemos realizado también varios trabajos, especialmente en modelos de series de tiempo, en particular estudiamos métodos robustos para modelos ARCH (ver Muler y Yohai, 2002), para modelos GARCH (ver Muler y Yohai, 2008) y para modelos ARMA (ver Muler, Peña y Yohai, 2009).

En los últimos años comencé a trabajar en colaboración con Mariela Sued, quien realizó un docto-

rado en Matemáticas con una tesis en Probabilidades, pero posteriormente se interesó en investigación en temas de Estadística. Con Mariela he trabajado en problemas de estimación robusta cuando hay datos faltantes (ver Sued y Yohai, 2013). También hemos realizado un trabajo donde se define un nuevo concepto de diferenciabilidad de funcionales estadísticos y se dan aplicaciones para encontrar la distribución asintótica de M-estimadores de un modelo de regresión no lineal (ver Sued y Yohai, 2013).

También trabajé con Oscar Bustos en Series de Tiempo (ver Bustos, Fraiman, y Yohai, 1984 y Bustos y Yohai, 1986) y con Enrique Álvarez en Regresión Monótona (ver Álvarez y Yohai, 2012).

#### ■ 13. RESULTADOS SOBRE ESTIMACIÓN ROBUSTA PARA MODELOS LINEALES

Dentro de la Estadística Robusta realicé trabajo de investigación sobre varios temas entre los cuales podemos mencionar modelos lineales, modelos lineales generalizados, estimación de la matriz de covarianza de datos multidimensionales, series de tiempo, regresión no lineal, estimación con datos faltantes y componentes principales.

Quizás el área de mayor dedicación fue la de los modelos lineales, también conocidos bajo el nombre de modelos de regresión lineal. Como ya ha sido mencionado, en este caso el procedimiento clásico de mínimos cuadrados es el estimador óptimo cuando los errores tienen distribución normal. El primer procedimiento altamente robusto para este modelo de regresión, propuesto primero por Hampel (1975) y luego desarrollado incluyendo su implementación computacional por Rousseeuw (1984), es el estimador

denominado *mínima mediana de cuadrados*. Este estimador es altamente robusto pero muy ineficiente cuando los datos son normales, es decir si los datos son efectivamente normales se requieren muchas más observaciones que en el caso de mínimos cuadrados para obtener la misma precisión. Una de mis contribuciones que considero más importantes fue desarrollar estimadores que simultáneamente fueran altamente robustos cuando las observaciones tienen *outliers* y altamente eficientes cuando los errores fuesen normales sin *outliers*, por ejemplo que tuvieran en este caso una eficiencia del 95%. El significado del 95 % de eficiencia es que para obtener la misma precisión que con el método de mínimos cuadrados bastaría tomar un 5% de observaciones adicionales. Este costo del 5% puede interpretarse como una pequeña prima de seguro que se debe pagar para protegerse de la catástrofe que pueden provocar en el análisis unos pocos datos atípicos. En el año 1984 propuse los MM-estimadores (ver Yohai, 1987) y los tau-estimadores (ver Yohai y Zamar, 1988). Como estas dos clases de estimadores eran de gran complejidad computacional y sólo se podían aplicar para un número no muy grande de variables explicativas, en un trabajo conjunto con Daniel Peña desarrollamos un método alternativo que permite ser aplicado a grandes conjuntos de datos (ver Peña y Yohai, 1993 y 1996). Los últimos trabajos en esta dirección consistieron en tratar de suprimir la prima del seguro o reducirla drásticamente, es decir tener un seguro casi gratis. Esto lo obtuvimos en el año 2002 donde conjuntamente con dos alumnos de doctorado, Daniel Gervini y Sonia Hernández Alonso (ver Gervini y Yohai, 2002 y Hernández Alonso y Yohai, 2003) obtuvimos dos procedimientos para regresión altamente robustos y que tienen eficiencia del 100%. De to-

dos estos procedimientos, los que resultaron más utilizados fueron los MM-estimadores. Estos estimadores fueron incorporados en varios de los programas más utilizados para aplicación de métodos estadísticos. Actualmente, funciones que calculan los MM-estimadores se pueden encontrar en los programas R, SPLUS, SAS y STAT. El Google Académico encuentra alrededor 690 citas del trabajo en el que se introducen estos estimadores. Finalmente, en el año 2013 junto con Ricardo Maronna (ver Maronna y Yohai, 2013) obtuvimos un procedimiento altamente robusto con muy alta eficiencia no sólo para muestras grandes sino también para muestras relativamente pequeñas.

#### ■ 14. EL LIBRO SOBRE MÉTODOS ROBUSTOS

En el año 2006 conjuntamente con Doug Martin y Ricardo Maronna escribimos un libro sobre Estadística Robusta, en inglés, publicado por la editorial Wiley (ver Maronna, Martin y Yohai, 2006). Nuestro propósito fue escribir un libro que incorporase todos los últimos avances en métodos robustos. El libro tiene dos posibles destinos, por un lado puede ser utilizado como libro de texto para un curso de Estadística Robusta y también para ser consultado por potenciales usuarios de distintas ramas del conocimiento que quieran usar métodos robustos. El libro actualmente de acuerdo al Google Académico tiene más de 1100 citas en trabajos de investigación.

#### ■ 15. COMENTARIOS FINALES

En la Estadística encontré la posibilidad de utilizar mis conocimientos matemáticos, inclusive los que corresponden a aspectos más puros para crear nuevos procedimientos estadísticos de amplia aplicación en las distintas ramas de la ciencia y

la tecnología. Inclusive, en algunos casos tuve que desarrollar nuevas técnicas matemáticas para resolver problemas estadísticos. Este párrafo tiene como principales destinatarios a los graduados en Matemáticas que quisieran usar las matemáticas para resolver problemas reales. A ellos me permito recomendarles que consideren la opción de trabajar en Estadística.

Cuando volví a Argentina, la investigación en Estadística en la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires no existía. Hoy la Facultad cuenta con 5 profesores en esta área, 4 becarios posdoctorales y 7 becarios doctorales. Creo que teniendo ya 75 años, puedo retirarme (aunque no pienso hacerlo inmediatamente) con la satisfacción de tener la convicción que la Estadística en esa Facultad está en muy buenas manos y gozará de muy buena salud durante mucho tiempo.

#### BIBLIOGRAFÍA

- Adrover J., Bianco A., Yohai V.J. (1994) "Efficiency of MM- and t-estimates for finite sample size". *Statistics and Probability Letters*, **19**, 409-415.
- Adrover J., Maronna R.A., Yohai V.J. (2004) "Robust regression quantiles". *Journal of Statistical Planning and Inference*, **122**, 187-202.
- Adrover J., Yohai V.J. (2002) "Projection estimates of multivariate location". *Annals of Statistics*, **30**, 1760-1781.
- Alqallaf F., Van Aelst S., Yohai V.J., Zamar R.H. (2009) "Propagation of outliers in multivariate data." *Annals of Statistics*, **37**, 311-331.
- Álvarez E., Yohai V.J. (2012) "M-es-



- timators for Isotonic Regression*". Journal of Statistical Planning and Inference, **142**, 2241-2284, 201
- Álvarez E., Yohai Víctor J. "M-estimators for Isotonic Regression". Journal of Statistical Planning and Inference, **142**, 2241-2284, 201
- Bergesio A., Yohai V.J. (2011) "Projection Estimates for Generalized Linear Models". Journal of the American Statistical Association, **106**, 661-671.
- Bianco A., García Ben M., Martínez E., Yohai V.J. (1996) "Robust procedures for regression models with ARIMA errors". COMPSTAT 96, Proceeding of Computational Statistics. Ed. Albert Prat, pags. 27-38, Physica-Verlag, Heidelberg.
- Bianco A., García Ben M., Martínez E., Yohai V.J. (2001) "Outlier detection in regression models with ARIMA errors". Journal of Forecasting, **20**, 565-579.
- Bianco A., García Ben M., Yohai V.J. (2005). "Robust estimation for regression with asymmetric errors". Canadian Journal of Statistics, **33**, 528-533.
- Bianco A., Yohai V.J. (1996) "Robust estimation in the logistic regression model." En Robust Statistics, Data Analysis and Computer Intensive Methods, Proceedings of the workshop in honor of Peter J. Huber, editor H. Rieder, Lecture Notes in Statistics **109**, 17-34, Springer-Verlag, New York.
- Boente G., Fraiman R., Yohai V.J. (1987) "Qualitative robustness for general stochastic processes". Annals of Statistics, **15**, 1293-1312.
- Bustos O., Fraiman R., Yohai V.J. (1984) "Asymptotic behavior of the estimates based on residual autocovariances for ARMA models." En Robust and nonlinear time series, editores, Franke, Hardle y Martin. Lectures Notes in Statistics, **29**, 26-49, Springer Verlag, New York.
- Bustos O., Yohai V.J. (1986) "Robust estimates for ARMA models." Journal of the American Statistical Association, **81**, 155-168.
- Danilov M., Yohai V. J., Zamar R.H. (2012) "Robust estimation of multivariate location and scatter in the presence of missing data." Journal of the American Statistical Association, **107**, 1178-1186.
- Fasano M.V., Maronna R.A., Sued M., Yohai V.J. (2012) "Continuity and Differentiability of regression M functionals." Bernoulli, **18**, 1284-1309.
- Ferretti N., Kelmansky D., Yohai V.J. (1991) "Robust estimates for ARMA models based on ranks." Communications in Statistics, Theory and Methods, **20**, 3879-3902.
- Ferretti N., Kelmansky D., Yohai V.J., Zamar R.H. (1999) "A Class of Locally and Globally Robust Regression Estimates". Journal of the American Statistical Association, **94**, 174-188.
- Fraiman R., Yohai V.J., Zamar R.H. (2001) "Optimal M-estimates of Location". Annals of Statistics, **29**, 194-223.
- García Ben M., Martínez E., Yohai V.J. (1999) "Robust Estimation in Vector ARMA Models." Journal of Time Series, **20**, 381-399.
- García Ben M., Martínez E., Yohai V.J. (2006) "Robust estimation for the multivariate linear model based on a based on a tau-scale". Journal of Multivariate Analysis, 2006, **97**, 1600-1622.
- Gervini G., Yohai V.J. (2002) "A class of robust and fully efficient regression estimators." Annals of Statistics, **30**, 583-616
- Hampel F.R. (1975) "Beyond Location Parameters: Robust Concepts and Methods". Bulletin of the International Statistical Institute, **46**, 375- 382.
- Hernández Alonso S., Yohai V.J. (2003) "Combining locally and globally robust estimates of regression". Journal of Statistical Planning and Inference, **113**, 633-661.
- Hotelling H. (1933) "Analysis of a Complex of Statistical Variables Into Principal Components". Journal of Educational Psychology, **24**, 417-441 and 498-52.
- Huber P. J. (1964) "Robust estimation of a location parameter". Annals of Mathematical Statistics, **35**, 73-101.
- Locatelli I., Marazzi A., Yohai V. J. (2011) "Robust accelerated failure time regression." Computational Statistics and Data Analysis, **55**, 874-887.
- Marazzi A., Villar Ana J., Yohai V.J. (2009) "Robust response transformations based on optimal prediction." Journal of the American Statistical Association, **104**, 360-370.
- Marazzi A., Yohai V.J. "Adaptively truncated maximum likelihood regression with asymmetric errors." Journal of Statistical Plan-

- ning and Inference, **122**, 271-291.
- Maronna R.A (1976) "Robust M-Estimators of multivariate location and scatter." *Annals of Statistics*, **4**, 51-67.
- Maronna R.A., Douglas Martin., Yohai V.J. (2006) "Robust Statistics: Theory and Methods." Wiley, Chichester, 2006.
- Maronna R.A., Yohai V.J. (1979) "Asymptotic behavior of M-estimates for the linear model." *The Annals of Statistics*, **7**, 258-268.
- Maronna R.A., Yohai V.J. (1981) "Asymptotic behavior of general M-estimators for regression and scale with random carriers". *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **58**, 7-20.
- Maronna R.A., Yohai V.J. (1993) "Bias-robust estimates of regression based on projections." *Annals of Statistics*, **21**, 965-990.
- Maronna R.A., Yohai V.J. (2000) "Robust regression with both continuous and categorical predictors." *Journal of Statistical Planning and Inference*, **89**, 197-214.
- Maronna R.A., Yohai V.J. (2008) "Robust Low-Rank Approximation of Data Matrices with Elementwise Contamination." *Technometrics*, **50**, 295-304.
- Martin R.D., Yohai V.J. (1984) "Robustness in time series and estimating ARMA models. Trabajo invitado en *Handbook of Statistics*", editores E. Hannan, P. R. Krishnaiah y M. M. Rao. Elsevier, Amsterdam, 119-125.
- Martin R.D., Yohai V.J. (1986) "Influence functionals for time series." Trabajo especial invitado con discusion en *The Annals of Statistics*, **14**, 781-818.
- Martin R.D., Yohai V.J., Zamar R. H. (1989) "Min-Max Robust Regression." *Annals of Statistics*, **17**, 1608-1630.
- Muler N., Peña D., Yohai V.J. (2009) "Robust estimation for ARMA models." *Annals of Statistics*, **37**, 816-840.
- Muler N., Yohai V.J. (2002) "Robust Estimates for ARCH Processes." *Journal of Time Series*, **23**, 341-375.
- Muler N., Yohai V.J. (2008) "Robust Estimates for GARCH Models." *Journal of Statistical Planning and Inference*, **138**, 2918-2940.
- Peña D., Yohai V.J. (1995) "The detection of influential subsets in linear regression using an influence matrix." *Journal of the Royal Statistical Society, series B*, **47**, 145-156.
- Peña D., Yohai V.J. (1999) "A fast procedure for outlier diagnostics in large regression problem". *Journal of the American Statistical Association*, **94**, 434-445.
- Rousseeuw P.J (1984) "Least Median of Squares Regression." *Journal of the American Statistical Association*, **79**, 871-880.
- Rousseeuw P., Yohai V.J. (1984) "Robust regression by means of S-estimators." En *Robust and nonlinear time series*, editores Franke, Hardle y Martin. *Lectures Notes in Statistics*, **29**, 256-272, Springer Verlag, New York.
- Salibian-Barrera M., Yohai V.J. (2006) "A fast algorithm for S-regression estimates." *Journal of Computational and Graphical Statistics*, **15**, 414-427.
- Salibian-Barrera M., Yohai V.J. (2008) "High breakdown point robust regression with censored data." *Annals of Statistics*, **36**, 118-146.
- Sued M., Yohai V.J. (2013) "Robust location estimation with missing data" *Canadian Journal of Statistics*, **41**, 111-132.
- Yohai V.J. (1974) "Robust estimation in the linear model." *The Annals of Statistics*, **2**, 562-567.
- Yohai V.J. (1987) "High breakdown-point and high efficiency M-estimates for regression". *The Annals of Statistics*, **15**, 642-656.
- Yohai V.J., Zamar R. H. (1988) "High breakdown-point estimates of regression by means of the minimization of an efficient scale." *Journal of the American Statistical Association*, **83**, 406-413.
- Yohai V.J., Zamar R. H. (1993) "A minimax-bias property of the least-a-quantile-estimates." *Annals of Statistics*, **21**, 1824-18.
- Yohai V.J., Zamar R. H. (1997) "Optimal locally robust M-estimates of regression". *Journal of Statistical Planning and Inference*, **64**, 309-323.
- Yohai V.J., Zamar R. H. (2004) "Robust non parametric inference for the median". *The Annals of Statistics*, **32**, 1841-1857.