



## Libro de Resúmenes

9, 10 y 11 de Octubre de 2019  
MENDOZA - ARGENTINA



**UNCUYO**  
UNIVERSIDAD  
NACIONAL DE CUYO

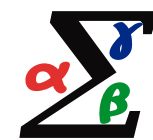
80 años



FACULTAD DE  
**CIENCIAS  
AGRARIAS**



The International  
**Biometric Society**



Grupo Argentino  
de Biometría

XXIV Reunión Científica del Grupo Argentino de Biometría  
XXIV Reunión Científica del Grupo Argentino de Biometría 9,10 y 11 de octubre de  
2019 : Mendoza-Argentina : libro de resúmenes / compilación de Joaquín Llera ... [et al].  
- 1a ed. - Mendoza : Universidad Nacional de Cuyo, 2021.  
Libro digital, PDF

Archivo Digital: descarga y online  
ISBN 978-987-575-218-4

1. Biometría. I. Llera, Joaquín, comp. II. Título.  
CDD 510.72

## Editor

Grupo Argentino de Biometría (GAB) – Región Argentina de la International  
Biometric Society.

## Compiladores

Joaquín A. Llera

Nora B. Martinengo

María Claudia Morelli

María Eugenia Martinez

Eva Marien Maure Russó

ISBN 978-987-575-218-4



La International Biometric Society (IBS) es una sociedad internacional que promueve el desarrollo y la aplicación de la teoría y los métodos estadísticos y matemáticos en las biociencia, incluidas la agricultura, la ciencia biomédica y la salud pública, la ecología, las ciencias ambientales, la silvicultura y disciplinas afines.

La Región Argentina de la IBS llevó a cabo la XXIV Reunión Científica del Grupo Argentino de Biometría (GAB) los días 9, 10 y 11 de Octubre de 2019, organizada por el Grupo Argentino de Biometría y la Facultad de Ciencias Agrarias de la Universidad Nacional de Cuyo (Mendoza).

El encuentro tuvo como objetivo proveer un ámbito propicio para el intercambio y la discusión de ideas motivadoras e inspiradoras para la comunidad de profesionales especializados en Biometría, difundir los resultados de su trabajo, promover y fortalecer la generación de vínculos entre instituciones y reforzar la formación de recursos humanos.

Las actividades programadas fueron: conferencias de reconocidos biometristas, mini-cursos, exposición de trabajos científicos, mesa de análisis de datos, concurso de jóvenes biometristas y la Asamblea Anual Ordinaria del GAB.

## Comisión directiva Región Argentina (GAB)

Presidente: Pablo D. Reeb

Vicepresidente: Silvia Sühling

Secretario: Guillermo Sabino

Tesorero: Maria Inés Urrutia

Vocal Titular: Susana Filippini

Vocal Titular: Gustavo Giménez

Vocal Suplente: María del Carmen Romero

Vocal Suplente: Erika Raña

## Comité Científico

Balzarini, Mónica

Bramardi, Sergio

Casanoves, Fernando

Cendoya, Gabriela

Di Rienzo, Julio

Filippini, Susana

Macchiavelli, Raúl

Reeb, Pablo

Winzer, Nélica

## Comisión Organizadora Local

Joaquín A. Llera

Nora B. Martinengo

María C. Morelli

María E. Martinez

Eva M. Maure Russo

Laura L. Cánovas

Eje  
**OTROS**

Resumen Extendido



## Contenido

### RESUMEN EXTENDIDO

¡Feliz cumpleaños! Una experiencia de clase sobre conceptos básicos del diseño experimental.....	1
Aplicación de ANOVA en la optimización del análisis de las concentraciones de cannabinoides en aceites medicinales de cannabis por cromatografía HPLC.....	4
ASOCIACIÓN DE OBESIDAD CENTRAL CON PRESIÓN ARTERIAL ELEVADA EN ESTUDIANTES UNIVERSITARIOS.....	7
Avances en la aplicación de los números efectivos de especies para evaluar la diversidad de artrópodos: caso viñedo de Mendoza.....	11
Estimación de la temperatura corporal del <i>Liolaemus acostai</i> (lagarto) mediante modelos con errores en las variables.....	15
Evaluación de biomarcadores hematológicos en poblaciones expuestas a arsénico en agua de consumo.....	18
Medición de la variabilidad entre estudios en meta-análisis.....	21
Pronóstico de Estadísticas Vitales haciendo uso del Análisis de Datos Simbólicos..	25
Regresión por cuantiles multivariada para modelar el efecto de cuestionarios online sobre las evaluaciones tradicionales.....	28
Selección de atributos en clasificación supervisada. Una mejora considerando una corrección de la entropía condicional.....	32
UNA APLICACIÓN DE MODELOS DE REGRESIÓN SIMBÓLICA PARA VARIABLES DE INTERVALOS.....	35

# ¡Feliz cumpleaños! Una experiencia de clase sobre conceptos básicos del diseño experimental

CENDOYA, G., MONTERUBBIANESI, G., DINON, A. y SAN MARTINO, S.

Facultad de Ciencias Agrarias, Universidad Nacional de Mar del Plata (UNMdP), Argentina.

mcendoya@mdp.edu.ar

## Resumen

*En la Facultad de Ciencias Agrarias, la asignatura Estadística y Diseño se encuentra en el segundo año de las carreras que se ofrecen. Este año, en dicha asignatura, se llevó a cabo una clase participativa con el objetivo de motivar el interés de los alumnos y facilitar la comprensión de contenidos sobre diseño experimental. En esa clase se planteó una problemática concreta y se discutió sobre los medios y pasos necesarios para poder resolverla, planificando un experimento mediante el uso de los principios básicos del diseño. A continuación, se realizó el experimento planteado, mediante dos alternativas: diseño completamente aleatorizado y en bloques completos aleatorizados. Se pudo cumplir con los objetivos inicialmente propuestos, mostrando buena disposición por parte de los estudiantes. Por falta de tiempo, no se pudo retomar el experimento en clases posteriores para analizar los datos obtenidos. Sin embargo, se discutió de manera individual con algunos de los estudiantes sobre la conveniencia de usar uno u otro diseño y sobre otros factores que pudieron haber afectado el desarrollo de la experiencia y la obtención de resultados confiables. Se espera poder repetir este tipo de experiencia en cursadas futuras ya que resultó muy provechosa.*

## Palabras Clave

Educación, aprendizaje, participación, ciencias agrarias, estadística.

## Introducción

La asignatura Estadística y Diseño se ubica en el segundo año de los planes de estudio de las carreras de grado que se ofrecen en la Facultad de Ciencias Agrarias de la UNMdP. El cuerpo docente a cargo de la misma busca permanentemente motivar a los estudiantes a construir y afianzar conocimiento desde actividades complementarias a las clases teóricas y prácticas convencionales. Cuando el estudiante es el protagonista en el proceso de enseñanza-aprendizaje con la guía del docente, creemos que hay mayor interés y ganas de aprender, y en consecuencia hay mejores logros.

En particular, durante el curso del presente año, 2019, se desarrolló una actividad con el fin de lograr una mejor comprensión de los conceptos básicos del diseño experimental, sus principios (repetición, aleatorización y control local) y la planificación de un experimento de acuerdo a un diseño completamente aleatorizado o en bloques completos aleatorizados (Kuehl, 2001; Cátedra Estadística y Diseño, 2019). En la misma, el docente guía buscó estimular a los estudiantes a tomar el rol de los investigadores y así buscar qué preguntas plantearse, y encontrar respuestas, a fin de

diseñar de la mejor manera posible un experimento. Posteriormente, los estudiantes llevaron a cabo el experimento en el aula en la forma que propusieron, registrando los datos correspondientes para el análisis para, así, intentar dar solución al problema inicialmente planteado.

### **Material y Métodos**

Luego de la primera clase introductoria al diseño experimental, se planificó una segunda clase donde los estudiantes, a partir de un problema sencillo, pudieran diseñar y realizar una experiencia en aula en un tiempo relativamente corto. Un experimento citado por Dean y Voss (1999) fue tomado como base para generar la actividad. El experimento citado (“experimento de los globos”) tenía por objetivo comparar los efectos de cuatro colores de globos sobre el tiempo de inflado de los mismos. Así, el caso fue adaptado y, en dos comisiones de trabajos prácticos, se presentó a los estudiantes el problema de si “la dificultad en el inflado” de los globos podía depender del color de los mismos. Se acotó el problema a la comparación de tres colores de globos y, promoviendo la participación de los estudiantes, se definieron, colectivamente y con precisión, el objetivo, los tratamientos, la variable respuesta y se identificaron otros factores que podían influir en la variable medida.

### **Resultados**

Los estudiantes identificaron los tratamientos: los tres colores de los globos (azul, amarillo y rojo, en una comisión, y lila, amarillo y rojo, en la otra comisión) y, luego de discutir varias alternativas sobre cómo medir la dificultad en inflar un globo, definieron como variable respuesta el tiempo de inflado hasta alcanzar un determinado tamaño. También discutieron cómo garantizar que todos los globos alcanzaran el mismo tamaño final, e identificaron como otras fuentes de variación a controlar a las personas que inflarían (por ejemplo, por diferencias en su capacidad pulmonar) y a las personas que cronometrarían el tiempo, así como también destacaron la importancia de un posible efecto del orden de inflado de los globos. De esta manera, en cada comisión, se les propuso que diseñaran el experimento y luego que lo realizaran, y entonces definieron dos tipos de diseño. Una posibilidad fue un diseño completamente aleatorizado (DCA), en el cual 12 estudiantes (infladores) se dispusieron a inflar los globos y otro estudiante cronometraba a todos. Para la aleatorización de los colores a los infladores se dispuso de una bolsa con un total de 12 bolillas (cuatro bolillas de cada uno de los tres colores a comparar). Así, cada uno de los 12 estudiantes fue tomando al azar una bolilla de la bolsa sin reponerla luego de la extracción. La otra posibilidad de diseño que se consideró fue la de un diseño en bloques completos aleatorizados (DBCA). Para este caso, otros 4 estudiantes (infladores) actuaron cada uno como un bloque. Para cada inflador se aleatorizó el orden en que inflaría los globos de los tres colores, también utilizando una bolsa, pero ahora conteniendo tres bolillas, una de cada color. Cada inflador fue cronometrado por otro estudiante. Otros estudiantes registraban los tiempos de inflado. Así todos los estudiantes participaron activamente de la experiencia. Se usó un aro de Telgopor® para fijar el tamaño al que debía llegar el globo. En ambos diseños, se realizaron entonces 4 repeticiones de cada color de globo. La actividad se llevó a cabo durante una hora.



## Discusión

La actividad propuesta permitió, por un lado, que los estudiantes discutieran, grupalmente, los aspectos a tener en cuenta en la definición de un problema, la relación que debe existir entre el objetivo que se plantea y la experimentación que surge para el logro de dicho objetivo. Asimismo, proponer y analizar fuentes de variación a controlar. Por otro lado, la realización del ensayo a cargo de ellos mismos y en el aula, permitió advertir la importancia de cada uno de los aspectos que se habían señalado *a priori* y que debían ser cuidados durante el desarrollo del experimento. También se planteó la necesidad de que hubiese sido importante realizar alguna prueba piloto a partir de la cual ajustar algunas cuestiones. Para completar, se podría haber solicitado a los estudiantes, por grupos, un breve informe de lo realizado.

Si bien esta actividad fue realizada luego de una primera clase introductoria sobre el diseño experimental, que incluye el porqué de realizar y planificar un experimento, los principios del diseño experimental y la presentación de los dos diseños básicos (DCA, DBCA) puede pensarse para próximos cursos como una actividad para la clase introductoria de la temática, a partir de la cual presentar las definiciones y los conceptos básicos. El análisis de los datos obtenidos en estos experimentos se pensó retomar en la clase correspondiente, y hacer una puesta en común con todos los resultados para dar un mejor cierre a toda esta experiencia. Esto no fue posible por falta de tiempo, ya que hubo suspensión o reducción de clases por diferentes motivos. De todas maneras, se pudo discutir con algunos de los estudiantes en forma individual sobre lo obtenido y la mejor adecuación de uno u otro diseño, así como sobre otros factores que pudieron haber influido en la experiencia y que inicialmente no se tuvieron en cuenta. Se espera poder repetir esta experiencia o alguna similar en años posteriores.

## Referencias

- Cátedra Estadística y Diseño (2019) "Apunte teórico" <http://fca.mdp.edu.ar/cv/index.php>
- Dean, A.; Voss, D. (1999). Design and Analysis of Experiments. New York. Springer-Verlag.
- Kuehl, R. O. (2001). Diseño de experimentos. Segunda ed. Thomson International.

# Aplicación de ANOVA en la optimización del análisis de las concentraciones de cannabinoides en aceites medicinales de cannabis por cromatografía HPLC

Bajda L. <sup>★</sup>, Lac Prugent V. <sup>★</sup>, Amaro M. <sup>\*</sup>

<sup>\*</sup> PROBIEN, Instituto de Investigación y Desarrollo en Ingeniería de Procesos, Biotecnología y Energías Alternativas; CONICET-UNCO

<sup>★</sup> Cátedra de Estadística Aplicada, Facultad de Economía y Administración; UNCO

leonardo.bajda@probien.gob.ar

## Resumen

*Este trabajo, enmarcado en la puesta a punto de la técnica cromatográfica para la solicitud de la aprobación de un Servicio Tecnológico de Alto Nivel, tuvo el objetivo de optimizar el cálculo de la concentración de cannabinoides de interés en muestras de aceites de cannabis medicinal, analizadas por HPLC. Surgió el interrogante, a partir de la bibliografía consultada, cuáles longitudes de onda utilizar en la estimación de las concentraciones de dichos compuestos. Para dar respuesta a este interrogante, se compararon las medias de los compuestos: CBG; THC; CBD y CBN, para las siguientes longitudes de onda: 200; 214; 220 y 272 nm. Se realizó un ANOVA por compuesto y, debido a la no existencia de normalidad de los errores, se utilizó el ANOVA de Kruskal-Wallis. Se concluye que no existen diferencias estadísticamente significativas para la estimación de las concentraciones medidas en las distintas longitudes de onda.*

**Palabras clave:** ANOVA; aceites medicinales; cannabis; HPLC; cannabinoides

## Introducción

Existe una falencia regional en la disponibilidad de un análisis que determine y cuantifique cannabinoides presentes en aceites terapéuticos de cannabis. Este análisis es fundamental para que el médico pueda prescribir y dosificar adecuadamente el aceite a sus pacientes, sabiendo con certeza qué cannabinoides y en qué concentraciones se encuentran, permitiendo una correcta administración para cada patología. Es por esto que el objetivo planteado en este trabajo fue la optimización del cálculo de la concentración de cannabinoides de interés en muestras de aceites de cannabis medicinal, analizadas por la técnica de cromatografía líquida de alta performance (HPLC). Se estudiaron los cannabinoides más abundantes en los aceites de cannabis, que presentan un amplio rango de propiedades terapéuticas conocidas según bibliografía, siendo los mismos: tetrahidrocannabinol (THC); cannabidiol (CBD); cannabinol (CBN) y cannabigerol (CBG). Se fijaron cuatro longitudes de onda distintas, seleccionadas según métodos propuestos en diversos trabajos científicos consultados, en la que se pretendió conocer, a través del Análisis de la Varianza si las concentraciones varían o no en las longitudes de onda propuestas.

## Materiales y métodos

El equipamiento empleado para la detección y cuantificación de los cannabinoides analizados en muestras de aceite medicinal de cannabis, las cuáles fueron donadas por Asociaciones Civiles para esta puesta a punto, fue un HPLC Agilent 1260, con un detector de Arreglo de Diodos (DAD), que recoge la respuesta cromatográfica (en unidades de mili absorbancia), en las longitudes de onda seleccionadas para este estudio: 200; 214; 220 y 272 nm. Para el cálculo de las concentraciones de los cannabinoides, utilizadas en el ANOVA, se empleó la técnica del estándar externo en la que se diseñó una curva de calibración para cada compuesto, a partir de estándares grado HPLC. Se aplicó el análisis de la regresión lineal simple para obtener la recta estimada que relaciona la concentración del cannabinoide con la respuesta del detector. Las curvas se realizaron por triplicado, en un rango entre 0.1 y 0.001mg/ml (zona de linealidad) y arrojando, en todos los casos, un  $R^2$  superior a 0.997. Estos datos se obtuvieron a través del software Chemstation.

Con los datos de concentración hallados, se aplicó un ANOVA paramétrico que permite contrastar la hipótesis nula de que las medias de diferentes tratamientos no difieren entre sí, contra la hipótesis alternativa de que al menos una difiere. Este método, supone que los errores son variables aleatorias independientes con distribución normal, media cero y variancia constante para todos los niveles del tratamiento, en nuestro caso las longitudes de onda. Para el diagnóstico de los supuestos se realizó el Test de Levene (homogeneidad de variancias), se visualizó en los gráficos Q-Q plots si la distribución de los errores era normal, luego se testeó la normalidad con el Test de Shapiro-Wilk. Posteriormente, se aplicó el ANOVA no paramétrico de Krustal Wallis, éste permite comparar las medianas de un conjunto de k muestras independientes, cuando los errores no poseen una distribución normal. Estos análisis se realizaron mediante el software Infostat 2018.

## Resultados

Del ANOVA paramétrico para cada compuesto analizado no se rechazó la hipótesis de igualdad de medias de concentraciones para las distintas longitudes de onda. Para verificar el supuesto de que los errores siguen una distribución normal, se visualizó el gráfico de Q-Q plots, en el cual se observa la no normalidad de la distribución de los residuos. Se testeó la misma con el test de Shapiro-Wilk, en el cual se rechazó la hipótesis de normalidad. Por lo tanto, se decidió realizar el ANOVA no paramétrico de Kruskal Wallis. De este último test, no se rechazó la hipótesis nula de igualdad de concentraciones para las distintas longitudes de onda elegidas. Es decir que no existen diferencias estadísticamente significativas entre sí. En la tabla nº 1, se muestra el resultado de la prueba de Kruskal Wallis, con un nivel de significancia;  $\alpha=0.05$  (5%), en la que testearon las siguientes hipótesis:

Ho) las medianas son iguales para cada longitud de onda

H1) al menos una mediana es diferente al resto.

Tabla nº1: resultados obtenidos de la prueba de Kruskal Wallis, con un nivel de significancia;  $\alpha=0.05$

	CBG				THC				CBD				CBN			
log de onda	200	214	220	272	200	214	220	272	200	214	220	272	200	214	220	272
N	77	77	77	77	86	86	86	86	46	46	46	46	81	81	81	81
medias	1.21	1.10	1.06	1.19	16.15	16.59	17.25	18.44	1.08	1.09	1.13	1.15	0.27	0.28	0.26	0.26
medianas	1.02	0.93	0.91	0.90	16.33	16.34	17.97	18.14	0.13	0.13	0.13	0.12	0.23	0.22	0.20	0.21
p-valor	0.45				0.48				0.92				0.99			

Se puede observar en esta tabla, que las medias y medianas no difieren en las cuatro longitudes de onda; los p-valor varían en un rango de 0.45 al 0.99, con lo cual la  $H_0$  es no rechazada, es decir, no existe evidencia suficientemente significativa para pensar que las medianas difieren en las distintas longitudes de onda.

### **Discusión**

Se concluyó que no existen diferencias estadísticamente significativas entre las medianas de las concentraciones de cannabinoides en las distintas longitudes de onda. Este resultado implica que la medición de estas concentraciones puede realizarse por cualquiera de las cuatro longitudes, sin necesidad de cuadruplicar la información. Este trabajo aporta a la toma de decisión en las mediciones de las concentraciones, simplificando y optimizando el trabajo cromatográfico a realizar.

### **Agradecimientos**

Se agradece al Instituto de Investigación y Desarrollo en Ingeniería de Procesos, Biotecnología y Energías Alternativas, PROBIEN (CONICET-UNCO) y a las asociaciones que donaron las muestras: Asociación Civil Ciencia Sativa y Asociación Civil Cannabis Medicinal Rio Negro.

### **Referencias**

- Determination of delta-9-tetrahydrocannabinol content of cannabis seizures in Egypt. Ahmed M.A. El-Din M, Abdel-Salam, ElShebiney. Asian Pacific Journal of Tropical Medicine, 2017
- Innovative development and validation of an HPLC/DAD method for the qualitative and quantitative determination of major cannabinoids in cannabis plant material. De Backer, Debrus, Lebrunb, Theunis, Dubois, Decockc, Verstraetec, Hubert, Charlier. Journal of Chromatography B, 2009.
- Cannabis Oil: chemical evaluation of an upcoming cannabis-based medicine. Romano, Hazekamp. Cannabinoids 2013
- Determination of primary active constituents in Cannabis preparations by high-resolution gas chromatography/flame ionization detection and high-performance liquid chromatography/UV detection. Gambaro, Dell'Acqua, Farè; Frolidi, Saligari, Tassoni. Analytica Chimica Acta, 2002.
- Los cannabinoides y el sistema endocannabinoide. Grotenherme. Cannabinoids, 2006.
- Caracterización del aceite de cannabidiol utilizado como anticonvulsivante en el centro nacional de epilepsia. Ortega Serrano. Trabajo de investigación previo a la obtención del Título Profesional de Química Farmacéutica, 2017.
- Diseño y análisis de experimentos. Douglas Montgomery. Editorial Iberoamericana. 2004.
- Introducción a la Estadística. Thomas H. Wonnacott y Ronald J. Wonnacott. Editorial Limusa. 2004.
- Estadística aplicada a la administración y a la economía. 2ª edición revisada. Schaum. Leonard Kazmier. Alfredo Díaz Mata. Editorial Mc Graw Hill. 1993.
- Bioestadística Amigable 3ª edición. Martinez-Gonzalez, Sanchez-Villegas, Toledo-Atucha y Fajardo. 2014.

# ASOCIACIÓN DE OBESIDAD CENTRAL CON PRESIÓN ARTERIAL ELEVADA EN ESTUDIANTES UNIVERSITARIOS

Manni D<sup>1,2</sup>, Martinelli M<sup>2</sup>, Avila O<sup>1</sup> y Contini L<sup>1</sup>.

<sup>1</sup> Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, Santa Fe, Argentina,

<sup>2</sup> Facultad de Ciencias Médicas, Universidad Nacional del Litoral, Santa Fe, Argentina

E-mail: diegomanni@gmail.com

## Resumen

*El objetivo de este estudio es analizar la posible asociación de la obesidad central con presión arterial elevada en estudiantes universitarios. La población analizada fueron estudiantes entre 18 y 22 años de Ciencias Médicas de la Universidad Nacional del Litoral (FCM-UNL). La muestra estuvo constituida por 1240 estudiantes de ambos sexos, a quienes se les realizaron mediciones de circunferencia de cintura (CC) siguiendo procedimientos estandarizados (OMS), para la tensión arterial (TA) se registró el promedio de dos mediciones y se utilizaron los criterios de corte sugeridos por el Consenso Argentino de Hipertensión Arterial (SAC,2018) para su clasificación. También se indagaron por medio de un cuestionario, costumbres como tabaquismo, actividad física, consumo de alcohol y calidad de la dieta que fueron utilizadas como covariables para la obtención de un modelo de regresión logística (RL). Para la construcción del modelo de regresión logística se tuvieron en cuenta todas las variables que se sabe pueden potencialmente modificar o incluso influir sobre la TA. Para una mejor interpretación de los resultados se dicotomizaron todas las variables incluidas en el análisis de RL. Finalmente en el modelo solo se tuvieron en cuenta las variables: actividad física, circunferencia de cintura y sexo.*

**Palabras Clave:** Tensión arterial- obesidad central- Estudiantes universitarios. Actividad Física-Regresión Logística

## Introducción

La hipertensión arterial (HTA) es el principal factor de riesgo de morbimortalidad cardiovascular. La elevada prevalencia y el escaso control de la HTA exigen a las organizaciones gubernamentales y no gubernamentales locales el conocimiento actualizado sobre la prevalencia de la enfermedad. Según la Organización Mundial de la Salud (OMS,2018), en el año 2008 la prevalencia de HTA en el mundo fue de aproximadamente el 40% en adultos mayores de 25 años. En el estudio multicéntrico realizado en Argentina (RENATA 2016), se detectó una prevalencia de HTA de 25,4% en individuos menores de 35 años. (Delucchi, 2017)

Por otro lado, se ha demostrado que el incremento de la tensión arterial (TA) está asociado con el aumento de la obesidad central en un 60-70% en adultos. Es por esto que indicadores de obesidad tales como el índice de masa corporal (IMC) y la circunferencia de cintura (CC) podrían considerarse para evaluar el riesgo a padecer HTA (Nurdiantami, 2018). La utilización de la CC como indicador de obesidad central puede ser una herramienta óptima para la identificación de personas con riesgo de HTA.

El presente estudio tuvo como objetivo analizar la CC y otras variables tales como, tabaquismo, actividad física, sexo, entre otras como posibles variables predictoras del desarrollo de HTA en estudiantes universitarios.

## Material y Métodos

La población analizada fueron jóvenes entre 18 y 22 años estudiantes de Ciencias Médicas de la Universidad Nacional del Litoral (FCM-UNL), la misma estuvo constituida por 1240 estudiantes de ambos sexos de los cuáles 1113 contestaron todas las encuestas. Se midió la CC siguiendo procedimientos estandarizados (OMS) y se utilizó el criterio de riesgo de la International Diabetes Federation (IDF, 2005):  $\geq 90$ cm en hombres y  $\geq 80$ cm en mujeres. Para la TA se registró el promedio de dos mediciones y se clasificó según los criterios de corte sugeridos por el Consenso Argentino de Hipertensión Arterial (SAC, 2018). Mediante encuestas se obtuvo información relacionada al tabaquismo, actividad física, consumo de alcohol y calidad de la dieta que fueron utilizadas como covariables para la obtención de un modelo de regresión logística (RL).

## Análisis estadístico:

Las variables continuas TA y CC se describen por sexo como media  $\pm$  SD y prueba t para muestras independientes, mientras que las variables categóricas fueron descriptas como frecuencias y porcentajes. Las asociaciones entre las variables categóricas fueron realizadas por medio del test chi-cuadrado.

Para evaluar el odd ratio (OR) a padecer HTA se realizó una regresión logística binaria y se informan los valores estimados de los OR con su IC 95%. Como variable dependiente se utilizó la HTA, mientras que el resto de las variables observadas en el estudio se incorporaron como posibles variables regresoras.

## Resultados

De los 1113 estudiantes analizados (edad media de 19,5 años), 828 (66,7%) fueron mujeres.

Las variables TA y CC fueron categorizadas teniendo en cuenta los criterios antes mencionados. Los resultados de su distribución por sexo y el resto de las covariables se muestran en la tabla 1.

Tabla 1: Distribución por sexo de las variables analizadas. La columna p indica diferencia de proporciones para cada grupo

		Sexo (%)		p
		Masculino	Femenino	
TA	Normal	158 (42,7)	567 (76,4)	< 0,05
	TA ≥ 130/ 85	215 (58,1)	173 (23,3)	< 0,05
CC	Normal	290 (23,3)	571 (46,0)	>0,05
	Riesgo	72 (5,8)	139 (11,2)	>0,05
Tabaco	No Fuma	326 (78,7)	666 (80,4)	>0,05
	Ex Fumador	35 (8,5)	44 (5,3)	<0,05
	Fumador Actual	31 (7,5)	61 (7,4)	>0,05
Calidad de la dieta	Saludable	26 (6,3)	47 (5,7)	>0,05
	Necesita Mejorar	267 (64,5)	568 (68,6)	>0,05
	Poco Saludable	95 (22,9)	140 (16,9)	<0,05
Actividad Física	Alta	184 (44,4)	207 (25,0)	<0,05
	Moderada	155 (37,4)	350 (42,3)	>0,05
	Baja	45 (10,9)	205 (24,8)	<0,05
Consumo de Alcohol	No	309 (87,5)	658 (93,5)	<0,05
	Si	44 (12,5)	46 (6,5)	<0,05

La prevalencia de TA elevada fue significativamente mayor en varones (58,1%) que en mujeres ( $p < 0,05$ ). El 17% de los estudiantes presentaron CC superior a los puntos de corte, sin diferencia significativa entre varones y mujeres. En el análisis se RL se tuvieron en cuenta todas las variables relevadas. Finalmente fueron eliminadas del modelo aquellas covariables que resultaron no significativas en el modelo.

Para una mejor interpretación de los resultados se dicotomizaron todas las variables incluidas en el análisis de RL. Los resultados de la RL se muestran en la tabla 2.

Tabla 2: Valores y significancia de los coeficientes de la Regresión Logística

	$\beta$	Sig.	OR	95% CI para OR	
				Inferior	Superior
Actividad Física	-0,145	0,336	0,865	0,644	1,162
Sexo	1,538	0,000	4,654	3,478	6,228
CC	1,130	0,000	3,094	2,213	4,326



## Discusión

Es preocupante la alta prevalencia de tensión arterial elevada en los estudiantes analizados. La obesidad central, otro factor de riesgo importante de enfermedad cardiovascular, estuvo asociada con mayor OR de tensión arterial elevada en los estudiantes universitarios.

Del análisis del modelo de RL se puede observar que el sexo y la CC influyen significativamente como variables predictoras de Tensión arterial elevada. De éste análisis se puede observar que el hecho de ser hombre aumenta 4,654 veces el OR de padecer Tensión arterial elevada, de manera similar se puede observar que aquellos estudiantes con CC aumentada tendrán un OR 3,094 veces más de desarrollar Tensión arterial elevada. En el modelo analizado se puede observar que la variable actividad física no permite predecir significativamente ( $p=0.336$ ) la Tensión arterial elevada, sin embargo están demostrados los efectos benéficos que ésta presenta en términos de la variable analizada y es por este motivo que se decidió introducirla en el presente modelo. Estos datos ponen de manifiesto la necesidad de implementar estrategias para mejorar el conocimiento de la Tensión arterial en jóvenes a fin de promover cambios en el estilo de vida para prevenir Enfermedades cardiovasculares en el futuro.

## Referencias.

Delucchi, A; Majul, C; Vicario, A; Cerezo, G; Fábregues, G. (2017). Registro Nacional de Hipertensión Arterial. Características epidemiológicas de la hipertensión arterial en la Argentina. Estudio RENATA 2. Rev. Argent. Cardiol; 85:354-360. <http://dx.doi.org/10.7775/rac.es.v85.i4.11061>

Organización Mundial de la Salud (OMS2018). Situación mundial de las enfermedades no transmisibles. Enfermedades no transmisibles. Datos y cifras. Disponible en: <http://www.who.int/es/newsroom/fact-sheets/detail/noncommunicable-diseases>.

Nurdiantami Y, Watanabe K, Tanaka E, Pradono J, Anme T. (2018). Association of general and central obesity with hypertension. Clin. Nut;37:1259-63. <http://doi.org/gddpbd>

OMS. Guía para las mediciones físicas. Parte 3: Guías para la formación e instrucciones prácticas. Sección 4: Guía para las mediciones físicas (Step 2). Disponible en: [http://www.who.int/ncds/surveillance/steps/Parte3\\_Seccion4.pdf](http://www.who.int/ncds/surveillance/steps/Parte3_Seccion4.pdf)

Consenso Argentino de Hipertension Arterial. Sociedad Argentina de Cardiología y Federación Argentina de Cardiología. Sociedad Argentina de Hipertensión Arterial. Rev Argent Cardiol 2018;86-2.

International Diabetes Federation (IDF). 2005. The IDF consensus world wide definition of the metabolic syndrome. Disponible en: <http://www.idf.org/metabolic-syndrome>.



## Avances en la aplicación de los números efectivos de especies para evaluar la diversidad de artrópodos: caso viñedo de Mendoza

Maure Russo, E. M<sup>1</sup>; Morelli, M. C<sup>1</sup>; Quiroga V.N<sup>2</sup>; Llera, J. A<sup>1</sup>; López Plantey, R.<sup>2</sup>; Holgado, M.G<sup>2</sup>

<sup>1</sup>Cátedra de Cálculo Estadístico y Biometría, FCA, UNCUYO, Argentina

<sup>2</sup>Cátedra de Zoología Agrícola, FCA, UNCUYO, Argentina

emaure@fca.uncu.edu.ar

### Resumen

La cuantificación de la biodiversidad, para cualquier nivel de organización, es un instrumento usado para la indagación de problemas en la gestión ambiental. En este sentido se trabaja en el diagnóstico y estudio de la diversidad de artrópodos en ecosistemas naturales y agroecosistemas vitícolas. Con el objetivo de evaluar la diversidad a nivel de una comunidad de artrópodos de un viñedo, a través de los números efectivos de especies se utilizaron datos de un monitoreo realizado en un viñedo de la localidad de Gualtallary. Fueron colocadas 6 estaciones durante 4 semanas con dos tipos de trampas: Pitfall y de intercepción en las parcelas. Se calculó la diversidad empírica y estimada, el intervalo de confianza del 95% y completitud del inventario, para los valores  $q=0,1$  y  $2$  en la comunidad de artrópodos. Los resultados muestran que la diversidad de orden  $q=0$ , o riqueza de especies, resultó con una baja cobertura (56%), mientras que las diversidades de orden  $q=1$ , ME abundantes, y  $q=2$ , ME muy abundantes, estimada y empírica son valores bastante similares. Por lo tanto el uso de números de Hill permite interpretar tanto la riqueza como la abundancia en un marco unificado para la medición y la estimación de la diversidad.

Palabras Clave: Arthropoda, Diversidad, Estimadores, Vitivinicultura

### Introducción

Los artrópodos conforman la mayor parte de la riqueza de especies en todas las escalas espaciales. Participan en procesos decisivos dentro de los sistemas agrícolas como: reciclaje de nutrientes, polinización y control biológico de plagas (Claps et al. 2008). La Organización Internacional de la Viña y el Vino (Plan estratégico de la OIV 2015-2019) manifiesta interés por mantener los servicios ecosistémicos que provee la biodiversidad en los terroirs vitícolas. La cuantificación de la biodiversidad, para cualquier nivel de organización, es un instrumento usado para la indagación de problemas en la gestión ambiental (Pla y Matteucci 2001). En este sentido un equipo de investigadores, de la Cátedra de Zoología Agrícola de la Facultad de Ciencias Agrarias, UNCuyo, trabaja en el diagnóstico y estudio de la diversidad de artrópodos en ecosistemas naturales y agroecosistemas vitícolas (López Plantey et al. 2014). Los análisis de diversidad se han realizado tradicionalmente con los índices de Simpson y la entropía de Shannon-Wiener que presentan ciertas dificultades en la interpretación. El objetivo de este trabajo es evaluar la diversidad a nivel de una comunidad de artrópodos de un viñedo, a través de los números efectivos de especies o “números de Hill”.

## Materiales y Métodos

Los monitoreos se realizaron en un viñedo implantado en 1992 en la localidad de Gualtallary, departamento de Tupungato, una zona rural al pie de la Cordillera de Los Andes, en el noroeste de la Provincia de Mendoza. Para la captura se colocaron 6 estaciones de monitoreo durante 4 semanas con dos tipos de trampas: Pitfall y de intercepción en las parcelas cultivadas con Pinot Noir.

Para el registro de las especies se utilizó el Método Rapid Biodiversity Assessment (RBA) o Evaluación rápida de la biodiversidad, desarrollado por Oliver y Beattie (1993). La unidad de medida utilizada es la llamada especie "morfológica" y se habla de "morfoespecies" (ME).

La diversidad se evalúa a través de los "números de Hill" o números efectivos de especies (en nuestro caso serán ME) que son ampliamente recomendados por sus ventajas en la interpretación (Jost 2012, Moreno 2011, Magurran 2004). Los números de Hill fueron utilizados por primera vez en ecología por MacArthur (1965), desarrollado por Hill (1973) y reintroducido a los ecologistas por Jost (2006, 2007). La generalización propuesta por Hill toma la forma de la siguiente ecuación, definida para  $q \neq 1$ :

$${}^q D = \left( \sum_{i=1}^S p_i^q \right)^{1/(1-q)}$$

En la ecuación, S es el número total de especies y  $p_i$  la abundancia relativa de la i-ésima especie. El parámetro  $q$  determina la sensibilidad de la medida a la abundancia relativa de especies. La diversidad de orden  $q=0$  es la riqueza de especies observada en la muestra,  ${}^0 D = S$ . La diversidad de orden  $q=2$  resulta la recíproca del índice de Simpson,  ${}^2 D = 1/H_{Simpson}$ .

Para  $q=1$  la ecuación es indefinida, pero en el límite cuando  $q \rightarrow 1$  la ecuación resultante es la exponencial de la entropía de Shannon,  ${}^1 D = \exp^{H_{Shannon}}$  (Chao et al 2014).

Las medidas que usan  $q=0$  consideran las especies por igual sin tener en cuenta sus abundancias relativas, mientras que las medidas que usan valores altos de  $q$  descuentan todas menos las especies dominantes (Chao y Jost 2015).

Mediante el paquete `SpadeR` (Chao et al 2016) se obtuvieron las estimaciones de las diversidades de orden  $q=0, 1$  y  $2$ . La función `Diversity` proporciona el perfil de diversidad empírico (observado) y el estimado por Chao y Jost (2015). Para los límites de confianza, superior (LCL) e inferior (UCL), aplica un método bootstrap, en la diversidad estimada, y de máxima verosimilitud, en la diversidad empírica, para la integridad de la muestra de orden  $q$  en el nivel especificado, con un valor de confianza de 0,95.

La completitud del inventario se evaluó mediante la "cobertura de la muestra", que mide la proporción que representan los individuos de cada especie en la muestra, con respecto al número total de individuos. (López-Mejía et al 2017). La completitud del inventario se obtuvo mediante el paquete `iNEXT` (Hsieh y Chao 2019). Los datos se corrieron en el software R versión 3.6.1.

## Resultados

La riqueza total observada resultó en 329 ME, clasificadas en 12 órdenes de artrópodos, siendo Hymenoptera y Lepidoptera los de mayor riqueza con 65 y 52 ME respectivamente. La abundancia total fue de 2215 individuos, los órdenes Hymenoptera con 755 individuos y Coleoptera con 440 individuos. El orden Lepidoptera se encuentra en el sexto lugar respecto a su abundancia siendo el segundo en cantidad de ME como se observa en la Figura 1.

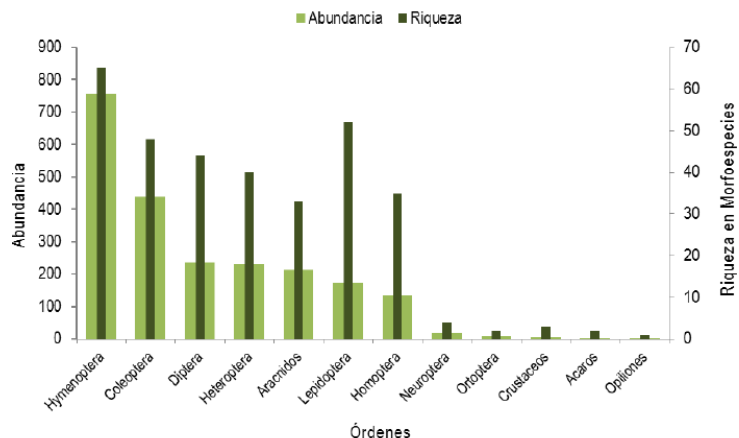


Figura 1: Distribución de los órdenes de artrópodos según sus abundancias absolutas y riqueza en morfoespecies.

La riqueza total de especies, diversidad de orden  $q=0$ , resultó con una baja cobertura (56%) del inventario de lo estimado por el estimador de Chao-Jost (Tabla1). Esta discrepancia entre la riqueza estimada, que calcula 258,78 especies más que la riqueza observada, puede deberse al tiempo en que estuvieron instaladas las trampas de la estación de muestreo o la disposición de las mismas.

Cuadro 1: Diversidad empírica y estimada, intervalo de confianza del 95% y completitud del inventario para los valores  $q=0,1$  y  $2$  en la comunidad de artrópodos del viñedo.

Diversidad	Empírica	LCL	UCL	Diversidad	Estimada	LCL	UCL	Completitud
${}^0D$	329,00	313,18	344,82	${}^0\hat{D}$	587,78	503,31	672,23	56,0%
${}^1D$	98,38	93,34	103,42	${}^1\hat{D}$	113,09	107,30	118,88	93,0%
${}^2D$	45,30	42,22	48,38	${}^2\hat{D}$	46,22	42,01	49,43	99,9%

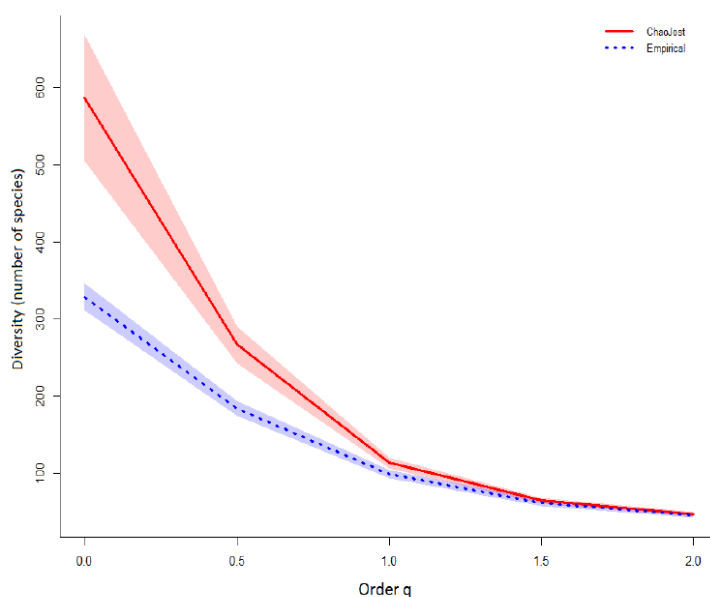


Figura 2: Perfil de diversidad empírica y estimada para la comunidad artrópodos de un viñedo para valores de  $q=0,1,2$ , en la comunidad de artrópodos de un viñedo.

La figura 2 muestra el perfil de diversidad para la comunidad de artrópodos del viñedo. El gráfico es estrictamente decreciente, siempre se cumple que  ${}^0D > {}^1D > {}^2D$ . La pendiente de la curva está reflejando la desigualdad en la distribución de las abundancias relativas.

El número de “ME muy abundantes”, diversidad de orden  $q=2$ , estimado y el valor empírico son valores bastante similares. En tanto las “ME abundantes” o diversidad de orden  $q=1$ , estimada resultó algo superior a la empírica. Sin embargo, ambas medidas de diversidad presentaron completitud mayores al 90%.

Las especies que se encuentran entre valores de  $q$  entre 0 y 1 son las que presentan bajas abundancias. En el caso de un viñedo es deseable que las especies que pueden perjudicar el desarrollo del cultivo se presenten en baja abundancia.

## Discusión

El análisis presentado complementa el estudio de la diversidad del viñedo que se ha realizado anteriormente y facilita la comparación de ensambles de artrópodos entre comunidades.

El uso de números de Hill permite interpretar tanto la riqueza como la abundancia en un marco unificado para la medición y la estimación de la diversidad. Presenta ventajas en la interpretación de los valores obtenidos como números efectivos de especies.

También se recomienda el uso de la completitud que permite cuantificar la integridad de la muestra. Para el caso de estudio, se evidencia la necesidad de rediseñar la estrategia de muestreo.

## Referencias

- Chao, A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K. & Ellison, A.M. (2014) Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecological Monographs*, 84, 45-67.
- Chao, A. & Jost, L. (2012) Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology*, 93, 2533-2547.
- Hill, M. O. 1973. Diversity and evenness: a unifying notation and its consequences. *Ecology* 54:427-432.
- Hsieh T. C., Ma K. H. and Chao A. 2019 iNEXT: iNterpolation and EXTrapolation for species diversity. R package version 2.0.19 URL: <http://chao.stat.nthu.edu.tw/blog/software-download/>.
- Jost, L. 2006. Entropy and diversity. *Oikos* 113:363-375.
- Jost, L. 2007. Partitioning diversity into independent alpha and beta components. *Ecology* 88:2427-2439.
- López Plantey, R.; Holgado, M.; Ortega, A.; Torres, M.L.; Rabinovich, D.; Quiroga, V. et al. 2014. Diagnóstico de la diversidad de artrópodos en un agroecosistema mendocino: caso de un cultivo vitícola. XXXVII Congreso Argentino de Horticultura, Mendoza. Libro de Resúmenes.
- MacArthur, R. H. 1965. Patterns of species diversity. *Biological Review* 40:510-533.
- Magurran, A. 2004. *Measuring Biological Diversity*. Blackwell Science Ltd. 256 p.
- Moreno, C. E. 2001. Métodos para medir la biodiversidad. M&T Manuales y Tesis SEA, vol. 1. Zaragoza, 84 p.
- Moreno, C.E.; Barragán, F.; Pineda, N.; Pavón, N. P. 2011. Reanálisis de la diversidad alfa: alternativas para interpretar y comparar información sobre comunidades ecológicas. *Revista mexicana de biodiversidad*, 82: 1249-1261.
- OIV. Plan Estratégico de la OIV 2015-2019. URL: <http://www.oiv.int/es/la-organizacion-intenacional-de-la-vina-y-el-vino/plan-estrategico-de-la-oiv>
- Oliver, I.; Beattie, A.J. 1993. A possible method for the rapid assessment of biodiversity, *Conservation Biology*, 7, 562-568. URL: <http://links.jstor.org/sici?sici=0888-8892%28199309%29%3A3%3C562%3AAPMFTR%3E2.0.C%3B2-3>
- Pla L.; Matteucci, S. 2001. Intervalos de confianza bootstrap del índice de biodiversidad de Shannon. *Revista de la Facultad de Agronomía LUZ* 18: 222-234.
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>.



# Estimación de la temperatura corporal del *Liolaemus acostai* (lagarto) mediante modelos con errores en las variables

OZAN S<sup>1</sup>, DE LOS RÍOS C<sup>2</sup>, GÓMEZ M<sup>1</sup>, VILLAVICENCIO H<sup>2</sup>.

<sup>1</sup>Facultad de Ingeniería. UNSJ <sup>2</sup>Facultad de Ciencias Exactas Físicas y Naturales. UNSJ

sozan@unsj.edu.ar

## Resumen

*Se estudió la relación de la temperatura corporal (Tc) del Liolaemus acostai, especie endémica de lagartos de Argentina, con la temperatura del suelo (Ts) medida en el lugar donde se observaron los individuos. La experimentación realizada lleva a que los valores medidos de la variable explicativa del problema son obtenidos con poca precisión o no de forma directa. Esto implica realizar el abordaje del problema considerando la relación entre el valor observado de la variable explicativa y el verdadero valor de dicha variable. Lo cual origina tener en cuenta una nueva variable, llamada variable latente, que es el verdadero valor de la variable explicativa (no observada o medida con error). Estos modelos se denominan “modelos con errores en las variables” y cuando la variable no observada (latente) es considerada aleatoria el modelo se llama estructural. En el presente trabajo existen dificultades para obtener mediciones precisas en la Ts. Por este motivo se decidió ajustar la relación entre la Tc y la Ts a través de un modelo con errores en las variables. Así, se plantea un modelo estructural con errores de medición entre la Tc y la Ts, incorporando la relación entre la temperatura observada y la temperatura real del suelo. El análisis de los datos y el cálculo de los estimadores del modelo se efectuaron con el programa R. El análisis de diagnóstico permitió observar un buen ajuste. La metodología propuesta disminuye el error estándar de los estimadores del modelo ajustado con respecto al obtenido en el modelo de regresión lineal simple, permitiendo mayor precisión en las predicciones de la Tc.*

**Palabras Clave:** Modelo estructural. Errores de medición. Estimación.

## Introducción

El presente trabajo trata del análisis del comportamiento de la temperatura corporal (Tc) de *Liolaemus acostai*, especie endémica de lagartos que se puede encontrar en La Laja (San Juan, Argentina), en función de la temperatura del suelo (Ts) medida en el lugar donde se observaron los individuos.

El registro de la Ts, en la mayoría de los casos, se realiza retirando una piedra laja bajo la cual se encuentra el individuo y se efectúa la medición, esto ocasiona diferencias entre la Ts observada en el instrumento y la real temperatura cuando el individuo ocupaba el área correspondiente.

Por este motivo se decidió ajustar la relación entre la Tc y la Ts a través de un modelo estructural con errores en las variables bajo del supuesto de distribución normal.

## Material y Métodos

El modelo estructural con errores en las variables (MEV) que representa la relación entre la variable real o latente y la variable respuesta Y, para una muestra de n observaciones es, (Fuller )

$$\begin{cases} Y_i &= \beta_0 + \beta_1 x_i + e_i \\ X_i &= x_i + u_i \end{cases}$$

Siendo  $(Y_i, X_i)^T$  el vector aleatorio que representa las observaciones de la variable dependiente o respuesta (Y) y de la variable observada (X),  $(e_i, u_i)^T$  corresponde a los errores de medición en la variable respuesta y observada respectivamente y  $x_i$  representa los valores de la variable no observada o latente.

En este caso, se estudia la relación verdadera entre la temperatura corporal del individuo y la temperatura real del suelo para lo cual se plantea el siguiente modelo,

$$\begin{cases} Tc_i = \beta_0 + \beta_1 ts_i + e_i \\ Ts_i = ts_i + u_i \end{cases}$$

Con  $e_i$  y  $u_i$  errores aleatorios independientes,  $e_i \sim N(0, \sigma_e^2)$  y  $u_i \sim N(0, \sigma_u^2)$

Tc es la variable respuesta, temperatura corporal

Ts es el predictor , covariable observada , temperatura del suelo medida

ts es el predictor o covariable de interés , latente , temperatura real del suelo.

Se realizó el análisis exploratorio de la variable Tc y de la covariable, Ts, la cual se consideró que presenta error en su medición teniendo en cuenta las condiciones del terreno y los hábitos comportamentales de esta especie. La normalidad de los datos se verifica con la prueba de Shapiro Wilks. Además para obtener estimaciones de los parámetros del modelo se asume el supuesto  $\delta = \sigma_e^2 / \sigma_u^2 = 1$  para evitar los problemas de identificabilidad del modelo.

## Resultados

Los resultados obtenidos para las estimaciones de los parámetros se muestran en la Tabla 1.

Se realiza un análisis de diagnóstico confirmando la buena calidad del ajuste y la ausencia de puntos atípicos.

Tabla 1. Estimaciones de los parámetros del modelo

	Estimación	Error Estándar	Intervalo de confianza (90%)
Intercepto	11.55	1.347	9.334 - 13.769
Temperatura del suelo	0.672	0.043	0.6 - 0.744

Algunas estimaciones de la Tc calculadas con el MEV y el MRL se muestran en la Tabla 2. La comparación de los resultados obtenidos permite observar mejores estimaciones con el MEV.

Tabla 2. Estimaciones de Tc

Ts (X observada)	Tc (Y observada)	$\widehat{Tc}$ con MRL	$\widehat{Tc}$ con MEV	$\widehat{Tc}$ con MEV **
19.2 °C	20.4 °C	25.25 °C	24.46 °C	23.85 °C
32.6 °C	38.2 °C	33.13 °C	33.47 °C	34.72 °C

\*\*Considerando la estimación de la temperatura real del suelo

## Discusión

La estimación de la temperatura corporal de *Liolaemus acostai* mediante el MEV tiene como primera ventaja el haber contemplado, en la metodología propuesta, la situación del experimento y la forma de tomar las mediciones. Por otro lado, al comparar las estimaciones obtenidas usando el MEV con las del modelo de regresión lineal (Tabla 2.) se observa un menor error estándar en las estimaciones MEV. Esto implica que las predicciones de la temperatura corporal tienen mayor precisión. El análisis de diagnóstico confirma la bondad del ajuste del modelo propuesto para este caso.

## Referencias

1. Fuller, W. (1987). *Measurement error models*. Wiley, New York.
2. Ozán, N.S.; De Los Ríos, C.; García, G.; Gómez, M.; Morales, E.: *Introducción al Lenguaje R*. [http://www.virtual.unsj.edu.ar/wpv\\_proyecto\\_estadistica\\_aplicada](http://www.virtual.unsj.edu.ar/wpv_proyecto_estadistica_aplicada). ISBN 978-987-42-7950-7. UNSJ. (2018).
3. R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>



## Evaluación de biomarcadores hematológicos en poblaciones expuestas a arsénico en agua de consumo

Cabagna Zenklusen M, Simoniello M F, Amherdt L, Grigolato R, Loteste, A, CONTINI L, AVILA O

Facultad de Bioquímica y Cs. Biológicas-Universidad Nacional del Litoral, Santa Fe, Argentina

mcabagna@fbc.unl.edu.ar

### Resumen

*El arsénico(As) es un metaloide que se encuentra en forma ubicua en la corteza terrestre esta forma inorgánica puede pasar al agua subterránea por disolución de la roca. El consumo de agua con contenidos variables de As ha generado un impacto negativo en la salud humana. Debido a esto, en distintos lugares del mundo se iniciaron estudios que utilizan biomarcadores y anomalías preclínicas para evaluar la susceptibilidad relacionada a factores genéticos, nutricionales y estilo de vida. En este contexto, uno de los objetivos de este trabajo fue caracterizar el impacto que puede tener la exposición crónica de As en el hemograma, en el desarrollo de daño oxidativo, medido a través de la peroxidación lipídica (POL) y en el agrupamiento de eritrocitos en habitantes de cinco localidades de la zona centro-oeste de la Provincia de Santa Fe con contenidos de As superiores a los recomendados, contrastándolos con pobladores de dos localidades con niveles bajos de As en agua subterránea. Nuestro estudio muestra que los residentes de zonas con mayores niveles de As en agua subterránea, sin enfermedades preexistentes, presentan niveles de POL mayores, tienen un riesgo superior a desarrollar acúmulos de eritrocitos, con el consecuente daño en la microcirculación por obstrucción respecto a las poblaciones no expuestas.*

### Palabras Clave

Arsénico, daño oxidativo, microcirculación, riesgo.

### Introducción

El As inorgánico está naturalmente presente en altos niveles en las aguas subterráneas de diversos países, entre ellos la Argentina, Bangladesh, Chile, China, India, México y Estados Unidos (Organización Mundial de la Salud, 2018).

En lo que respecta a la salud humana, estudios desarrollados en numerosos países, entre los que se encuentra la Argentina, demuestran que la ingestión crónica de As a través de agua contaminada, está vinculada con lesiones de piel, enfermedades cardiovasculares, alteraciones hematopoyéticas e inmunológicas, diabetes y al aumento de la incidencia de cáncer en distintos órganos (Abdul y col., 2015). Uno de los mecanismos de acción de la toxicidad del As más estudiado en la actualidad es la

formación de especies reactivas de oxígeno (EROs). En general, las EROs están en equilibrio con las defensas antioxidantes celulares, cuando su producción aumenta (por exposición a sustancias oxidantes como el As) se produce un desbalance que conduce al daño oxidativo. Uno de las posibles formas de cuantificar este daño es a través de la POL, que puede medirse de forma indirecta a través de la determinación de sustancias reactivas con el ácido tiobarbitúrico, TBARS por su sigla en inglés.

Debido al impacto negativo que produce el As en la salud, en distintos lugares del mundo se iniciaron estudios que utilizan biomarcadores y anomalías preclínicas para evaluar la susceptibilidad relacionada a factores genéticos, nutricionales y estilo de vida. En este contexto, entre los objetivos de este estudio se destaca la importancia de caracterizar el impacto que puede tener la exposición crónica a As en el hemograma (determinación bioquímica ampliamente solicitada en los chequeos de rutina), en el desarrollo de daño oxidativo y en el agrupamiento de eritrocitos (patrón que puede estar alterado por daño oxidativo y es medida de daño vascular) en habitantes de cinco localidades de la zona centro-oeste de la Provincia de Santa Fe, Argentina, con contenidos de As superiores a los recomendados, contrastándolos con pobladores de dos localidades con niveles bajos de As en agua subterránea.

## Material y Métodos

Se eligieron siete poblaciones de la zona centro de la provincia de Santa Fe con características socioculturales similares, cinco de ellas presentan registros de agua subterránea con niveles de As superiores a las recomendaciones (Providencia, Humboldt, Santa Clara de Sagüier, Lehmann y Presidente Roca) y dos, Nelson y Laguna Paiva, con concentraciones por debajo de los 50 µg/l, valor recomendado por el Código Alimentario Argentino para el agua de bebida.

En dos encuentros previos con los pobladores de estas siete localidades, se informó sobre los objetivos del proyecto y se los invitó a concurrir al Servicio de Atención Médica de la Comunidad (SAMCo) para acordar el consentimiento informado, realizar la encuesta sobre datos socio demográficos, hábitos alimentarios, enfermedades pre-existentes, tipo de agua de consumo y obtener la muestra biológica (sangre por venopunción). A partir de los datos recabados de lugar de nacimiento, años de residencia en las localidades y la ingesta de agua, se dividió a los individuos en expuestos (personas nacidas y/o que se encuentren habitando la localidad con registros históricos de niveles altos de As en agua subterránea por más de 20 años y que consuman agua no segura) y no expuestos (personas nacidas y que habiten en localidades sin registros históricos de niveles altos de As en agua subterránea y que tomen agua segura). Se tomaron estos criterios ya que existen evidencias que los daños producidos por la ingesta crónica de As pueden aparecer varios años después de la exposición.

Como marcadores biológicos se utilizaron hemograma, evaluación del agrupamiento de eritrocitos y determinación de POL midiendo TBARS. Para este informe se realizó el análisis descriptivo de las variables, para comparar medias se utilizó la prueba t de Student, para el cálculo de las asociaciones estadísticas entre variables cualitativas se usaron pruebas  $\chi^2$ , se estimaron puntualmente y por intervalo de confianza OR. Nivel de significancia adoptado 0.05 (Altman, 1997).

## Resultados

Se trabajó con una muestra de 308 pobladores (228 mujeres y 80 hombres) que residen en 7 localidades de la provincia de Santa Fe, con un rango entre 18 y 65 años de edad.

Del total, 214 pertenecen a cinco localidades con niveles de As en agua subterránea superior a lo recomendado y 94 a dos poblaciones con concentraciones de As dentro de lo recomendado por las normativas locales.

De los 214 residentes de áreas con As en aguas de consumo, se consideran grupo expuesto a 205 personas, basándose en su consumo actual o histórico del metaloide y como grupo no expuesto a As se consideró a los pobladores de Laguna Paiva y Nelson.

No se encontraron diferencias estadísticamente significativas entre los parámetros del hemograma, como tampoco en el patrón de agrupamiento de los eritrocitos (acúmulos) entre ambos grupos. A pesar de esto último, cabe señalar que el porcentaje de personas que presentan acúmulos de eritrocitos y pertenecen al grupo de “no expuestos” es 25,4 (17 de 67) mientras que en las que consumen agua de pozo este porcentaje asciende a 31,8 (41 de 129), aumentando el riesgo potencial de desarrollar problemas en la microcirculación.

En lo que respecta al daño oxidativo, la POL media de los expuestos fue mayor (58,9 vs 17,3  $\mu\text{mol/g Hb}$ ) encontrándose diferencias estadísticamente significativas ( $p < 0.001$ ) entre expuestos y no expuestos a As.

Dentro del grupo expuesto a As, la media de POL fue significativamente mayor (64,6 vs 54,7  $\mu\text{mol/g Hb}$ ;  $p = 0.0280$ ) en los que actualmente continuaban tomando agua de pozo respecto a los que modificaron esta conducta.

Considerando que ciertas enfermedades preexistentes registradas en las encuestas (hipertensión y diabetes) pueden estar relacionadas con el desarrollo de acúmulos de eritrocitos y daño vascular, se decidió excluirlas del análisis. Así, en los 133 individuos sin estas patologías preexistentes, se observaron diferencias estadísticamente significativas en el valor medio de POL entre los que presentaban acúmulos (72,6  $\mu\text{mol/g Hb}$ ) y los que no (58,4  $\mu\text{mol/g Hb}$ ) ( $p = 0.018$ ).

Por último, se dividió el grupo de los expuestos según superaban o no el máximo valor de POL encontrado en los pobladores no expuestos (30,72  $\mu\text{mol/g Hb}$ ) para determinar el riesgo que poseen las personas expuestas a desarrollar alteraciones en el patrón de agrupamiento de eritrocitos, obteniendo un valor de OR de 3,49 con un intervalo de confianza del 95% de (1,02 – 11,7).

## Discusión

A pesar de que existen trabajos previos en los que se vincula a la exposición al As con perturbaciones en los parámetros del hemograma, no se observaron en nuestro estudio. Esta diferencia puede deberse al tipo de dieta de los voluntarios (rica en proteínas, fundamentalmente provenientes de carnes y lácteos) hecho que impacta directamente en el estado nutricional del individuo y podría colaborar con los procesos de metabolización del As (metilación). Uno de los

primeros blancos biológicos del As en los elementos sanguíneos son los eritrocitos y en particular su membrana. La interacción entre los componentes de la membrana eritrocitaria determina sus propiedades físicas que le permiten la adaptación al flujo sanguíneo fundamentalmente en la microcirculación. Hernandez y col. (2015) señalaron que éste deterioro reológico es originado por daño oxidativo en los glóbulos rojos producido por la presencia del As. Estos hallazgos son a nivel de laboratorio, nuestro estudio avala este supuesto y muestra que los residentes de zonas con As, sin enfermedades preexistentes, presentan niveles de POL mayores a los de poblaciones no expuestas y además, tienen un riesgo superior a desarrollar acúmulos de eritrocitos, con el consecuente daño en la microcirculación por obstrucción. Por otra parte, como fuera señalado, es necesario continuar estos estudios que utilizan biomarcadores y anomalías preclínicas con nuevas investigaciones que permitan poner de manifiesto la susceptibilidad relacionada a factores genéticos, nutricionales y estilo de vida. Con el objetivo de describir la realidad argentina vinculada al As y poder generar estrategias en salud adecuadas para nuestra población.

## Referencias

Abdul KS, Jayasinghe SS, Chandana EP, Jayasumana C, De Silva PM. Arsenic and human health effects: A review. *Environmental Toxicology and Pharmacology* 40 (2015) 828–846.

Altman, D. *Practical Statistics for Medical Research*. Chapman and Hall/CRC. (1997).

Código Alimentario Argentino (CAA). Capítulo XII: Bebidas hídricas, agua y agua gasificada. [Internet]. 2012. [citado 2019 Sep 05]; Disponible en: [http://www.anmat.gov.ar/alimentos/codigoa/CAPITULO\\_XII.pdf](http://www.anmat.gov.ar/alimentos/codigoa/CAPITULO_XII.pdf)

Hernández G, Bollinia A, Mengarelli G, Rasia M, Casco C, Ruiz MF, Visconti M, Huarte M, Rubin de Cellis E, Bazzoni G. Protective effect of quercetin against in vitro erythrocyte rheology alterations produced by arsenic. *Clinical Hemorheology and Microcirculation* 59 (2015) 355–364.

## Medición de la variabilidad entre estudios en meta-análisis

TAMUCH S, CUESTA C. Facultad De Cs. Económicas y Estadística, Rosario, Santa Fe, Argentina.

sso.fia@hotmail.com

### Resumen

*El objetivo de este trabajo es presentar diversos métodos que permiten estimar la heterogeneidad entre estudios y sus propiedades. Se prueban estos métodos a partir de los resultados incluidos en una revisión sistemática actual sobre salud materna para el caso de mediciones de tipo continuas. Los resultados muestran que los distintos métodos de estimación de la heterogeneidad afectan moderadamente la estimación del efecto global en el caso en que la variabilidad sea alta o moderada.*

**Palabras Claves:** Revisiones sistemáticas, meta-análisis, heterogeneidad.

### Introducción

El gran crecimiento de publicaciones científicas ha sido un disparador para el desarrollo de la ciencia de la síntesis de la evidencia a lo largo de las últimas décadas. A partir de la necesidad de contar con una herramienta que permita sintetizar, valorar y poner al día toda la información médica existente, surgen las revisiones sistemáticas, que permiten identificar los estudios primarios que dan respuesta a una misma pregunta de interés. Si es posible, se unifican luego los resultados de los estudios seleccionados en la revisión sistemática a través de un meta-análisis. El meta-análisis consiste en la combinación y resumen de los resultados de estudios independientes, con el propósito de llegar a una única conclusión, a partir de métodos estadísticos. Sin embargo, combinar los resultados de todos los estudios relevantes implica cuestionarse acerca de diferencias producidas por características propias de la población objetivo de cada estudio, por la definición o medición de las variables consideradas, como así también por la calidad metodológica derivada de cada diseño de investigación particular. Estas diferencias se manifiestan en la heterogeneidad de las variables estudiadas, la cual se define como una variación extra entre los resultados de los estudios además de la esperada por simple variación aleatoria. En la práctica los métodos para combinar los resultados de los estudios dependen según se tenga en cuenta o no la heterogeneidad entre los mismos. Los modelos de efectos fijos consideran que no hay heterogeneidad entre estudios mientras que los modelos de efectos aleatorios si la tienen en cuenta. La importancia de la determinación del método a utilizar radica en que la misma es relevante para estimar el efecto global a través de todos los estudios.

### Materiales y Métodos

El modelo de efectos fijos, para la estimación del efecto global, asume que los estudios que participan del meta-análisis provienen de una única población homogénea.

El modelo de efectos fijos que se postula es:  $\hat{\theta}_k = \mu_{FE} + \varepsilon_k$ ,  $\varepsilon_k \sim N(0, \sigma_k^2)$  donde  $\theta_k$  la estimación del efecto de la intervención en el estudio  $k$  y  $\mu_{FE}$  el efecto global de la intervención en la población.

El "Método de Variancia Inversa" propone estimar el efecto global  $\mu_{FE}$  a partir  $\hat{\theta}_k$  y de  $\hat{\sigma}_k^2$

$$\hat{\mu}_{FE} = \frac{\sum_{k=1}^K \hat{\theta}_k / \hat{\sigma}_k^2}{\sum_{k=1}^K 1 / \hat{\sigma}_k^2} = \frac{\sum_{k=1}^K w_k \hat{\theta}_k}{\sum_{k=1}^K w_k}, \text{ donde } w_k = 1 / \hat{\sigma}_k^2$$

El modelo de efectos aleatorios, en cambio, supone que cada estimación del efecto puede descomponerse en dos componentes de variancia mediante un proceso jerárquico de dos niveles.

Se plantea el modelo de efectos aleatorios como:  $\hat{\theta}_k = \mu_{RE} + \varepsilon_k$  donde  $\delta_k$  representa un efecto aleatorio para el estudio  $k$  (cuya variancia es  $\tau^2$ ). El estimador es:

$$\hat{\mu}_{RE} = \frac{\sum_{k=1}^K w_k^* \hat{\theta}_k}{\sum_{k=1}^K w_k^*} \text{ con } w_k^* = 1 / (\tau^2 + \sigma_k^2).$$

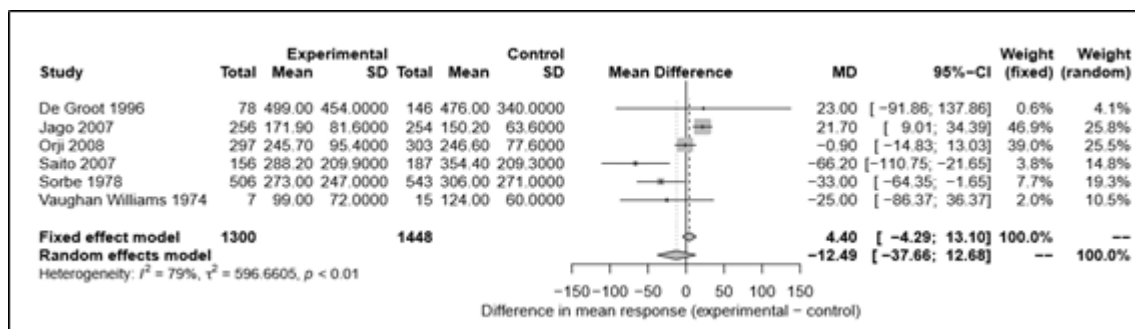
Algunos de los estimadores de heterogeneidad entre estudios ( $\tau^2$ ) son: Estimador de DerSimonian-Laird (DL), Estimador de Hunter-Schmidt (HS), Estimador de Hedges (HE), Estimador de Máxima Verosimilitud (ML), Estimador de Máxima Verosimilitud Restringida (REML).

## Resultados

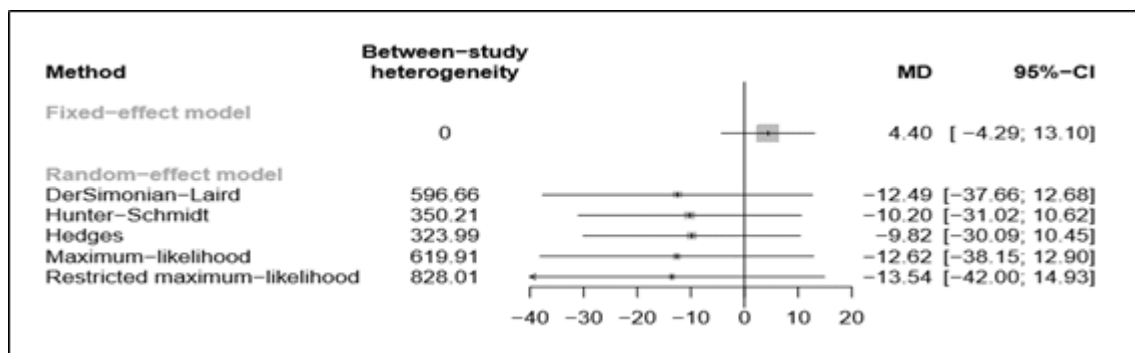
Se probaron los estimadores de heterogeneidad antes mencionados y su impacto en el efecto global, en la revisión sistemática "Oxitocina profiláctica en la tercera etapa del parto para prevenir la hemorragia postparto" (Westhoff G, Cotter AM, Tolosa JE.).

A modo de ejemplo se muestra la Figura 1 donde se compara el efecto de la oxitocina profiláctica contra los alcaloides ergóticos siendo la variable respuesta la cantidad de sangre perdida durante el parto (ml). La particularidad de este caso es el resultado de la prueba de heterogeneidad, que resulta significativa, con un grado de heterogeneidad alta del 79%. Esto significa que la magnitud del efecto de la oxitocina profiláctica en la cantidad de sangre perdida durante el parto con respecto a los alcaloides no fue la misma en los distintos estudios. La estimación global de la cantidad de sangre perdida durante el parto, en el modelo de efectos aleatorios, fue menor (-12.49 ml; IC<sub>95%</sub>: -37.66, 12.68) con la oxitocina que con alcaloides. En cambio, en el modelo de efectos fijos ese efecto global resultó 4.40ml; (IC<sub>95%</sub>: -4.29, 13.10), favoreciendo el tratamiento con alcaloides. Es decir, hay una notoria diferencia entre utilizar efectos fijos o aleatorios y en ello impacta la estimación de la componente de variancia entre estudios. La estimación puntual para esta componente es similar con todos los métodos propuestos, aunque su precisión es diferente (Figura 2).

**Figura 1:** Oxitocina vs. alcaloides en la cantidad de sangre perdida durante el parto (ml).



**Figura 2:** Comparación de los métodos de estimación para la heterogeneidad.



## Conclusiones

Las diferencias entre los estimadores de la heterogeneidad entre estudios pueden explicarse en cierta medida por los diferentes esquemas de ponderación que utilizan. La importancia de la determinación del método a utilizar radica en que la misma es relevante para estimar el efecto global a través de todos los estudios. Además, en la práctica el uso de los métodos de estimación alternativos está limitado por su disponibilidad en el software utilizado.

## Bibliografía

- Matthias Egger, G. S. (2001). *Razones, potenciales y promesa de revisiones sistemáticas*.
- Veroniki, A. A. (2015). *Métodos para estimar la variancia entre estudios y su incertidumbre en el meta-análisis*. Wiley Online Library, 55-79.
- Vietchtbauer, W. (2005). *Sesgo y eficiencia de los estimadores de variancia meta-analítica en el modelo de efectos aleatorios*. *Journal of Educational and Behavioral Statistics Fall 2005, Vol. 30, No. 3.*, 261-293.
- Westhoff G, C. A. (Noviembre de 2018). *Oxitocina profiláctica para la tercera etapa del trabajo de parto para prevenir la hemorragia posparto (Revisión)*. Obtenido de <https://www.cochranlibrary.com/>
- Zwinderman, T. J. (2017). *Meta-Analysis moderno*. Springer.



# Pronóstico de Estadísticas Vitales haciendo uso del Análisis de Datos Simbólicos

Martínez C. <sup>(1)(2)</sup>, Salas A. <sup>(1)(2)</sup> y Mallea A. <sup>(2)</sup>

(1) Facultad de Ciencias Exactas, Físicas y Naturales. Universidad Nacional de San Juan

(2) Facultad de Filosofía, Humanidades y Artes. Universidad Nacional de San Juan

ceciliaevelyn99@gmail.com

## Resumen

En general, una observación clásica es un simple valor que no posee variabilidad inherente, y por ello solamente se analiza la variación entre observaciones. En contraste, el análisis de datos simbólicos (ADS) no sólo considera la variación entre los distintos valores sino también la variación interna de cada observación. Es por ello que este tipo de análisis nos permite una nueva manera de procesar información de diversas clases o grupos. Una serie temporal es el resultado de la observación de los valores de una variable a lo largo del tiempo en intervalos regulares. Las series temporales de intervalo (STI) son un tipo de series donde cada observación se describe mediante un intervalo o rango de valores. Como cada observación está representada mediante un dato simbólico, en este caso, un intervalo, lo natural es abordarlas desde la perspectiva del ADS. Un indicador como la tasa de natalidad, podemos pensarlo variando en un rango de valores, en un periodo de un año. El conocimiento de la predicción del intervalo de valor mínimo y máximo para el periodo siguiente permite pronosticar el rango en que se encontrará. En este sentido, se busca presentar una herramienta potente del análisis simbólico para el pronóstico de estadísticas vitales. Poder predecir esta tasa abre puertas a nuevos conocimientos y desarrollos de estrategias, en vistas de poder comprender fenómenos tales como el crecimiento de la población mundial y la transición demográfica.

**Palabras Clave: Análisis Simbólico - Series Temporales - Tasa de Natalidad.**

## Introducción

Dado que los hechos vitales son todos los acontecimientos relacionados con el comienzo y

fin de la vida de los individuos y con los cambios en su estado civil que pueden ocurrir durante su existencia, tener nuevas alternativas para llevar adelante análisis más precisos de estos es un buen panorama para trabajos a futuro. Conocer la magnitud de los cambios en la población, medidos mediante tasas de natalidad y mortalidad y otras, son fundamentales para poder determinar políticas y atender las demandas de servicios de salud. En este marco, considerar los métodos del ADS para obtener predicciones de tales valores y así también una descripción de los mismos, considerando su variabilidad inherente, abre nuevas posibilidades de estudio y análisis.



## Material y Métodos

Consideramos como fuentes al Instituto Nacional de Estadísticas y Censos (INDEC) y la Dirección de Estadísticas e Información de Salud (DEIS).

Para realizar análisis estadístico de individuos descritos con variables simbólicas es necesario tener en cuenta que las variables simbólicas acarrearán de forma implícita una variabilidad ([4]). Teniendo en cuenta esta premisa, se llevó a cabo un análisis descriptivo desde la perspectiva del ADS de la variable para conocer su comportamiento, es decir, que para realizar este análisis se generó una tabla simbólica y de esta manera se obtuvo para cada año un rango de valores en los cuales se considera todos los valores de la variable en tal año, no sólo los valores mínimo y máximo. El método adaptado considerado para la predicción fue el Alisado Exponencial Simple (AES) ([2]). El error considerado es el error cuadrático medio cometido por el método considerado en el periodo de interés (RECEM).

## Resultados

Los resultados fueron obtenidos por medio del paquete RSDA y time series del lenguaje de programación, Software R ([3]). Para el método de predicción adaptado se hizo uso de algoritmos desarrollados para correr en Octave (versión libre de Matlab) ([1]).

De los resultados del análisis descriptivo, podemos mencionar que la tasa de natalidad tiene una asimetría a la derecha, lo que indica que la mayoría de las tasas son menores a su media [14.37455, 28.73358]. Esto es, la variable a medida que pasan los años ha ido reduciendo su valor gradualmente. En la STI, que abarca los valores que toma la variable entre los años 1980 y 2016, se observa un comportamiento decreciente muy marcado, a pesar de algunos cambios que se observan muy particulares en los años 1984 al 1985 y del 1990 al 1991. A su vez, no hay una componente estacional en este caso, ya que no se observa un patrón de cambio recurrente a través de los años. Por otro lado, para poder medir la potencia de nuestro método de predicción, hicimos la predicción de los valores máximos y mínimos del año 2017, haciendo uso del resto de los valores de los cuales se dispone. Luego comparamos nuestros resultados en términos del error cuadrático medio. En el año 2017 se registró una tasa de [11.7, 21.1].

*Tabla 1: Predicción para los valores máximos y mínimos del año 2017*

Predicción 2017		
Modelo	Mínimo	Máximo
AES (Método clásico)	13,10006	21,71175
AES (Método adaptado STI)	13,1	21,0

*Tabla 2: Error cuadrático medio escalado*

RECEM		
Modelo	Mínimo	Máximo
AES (Método clásico)	1,2183	1,0824
AES (Método adaptado STI)	1.2182	0.99825

*Tabla 3: Error porcentual*

Error Porcentual		
Modelo	Mínimo	Máximo
AES (Métodos clásicos)	11,96632479%	2,8992891%
AES (Método adaptado STI)	11,96581197%	0.473933649%

De los resultados obtenidos podemos ver que claramente el método adaptado predice mejor que el clásico, el máximo. En el caso del mínimo parece ser que es un poco mejor que el clásico. Pero, teniendo en cuenta que no se calcularon las tasas para Catamarca en los años 1983 a 1989, podemos suponer que quizás este hecho puede haber afectado al resultado global de nuestro estudio. En este sentido, podemos considerar que el método adaptado ha presentado buenos resultados y una opción interesante para considerar a la hora de hacer predicciones para variables de esta naturaleza. Se ha demostrado en ([2]) que se obtienen mejores resultados al hacer predicciones con datos simbólicos por medio de los métodos adaptados, ya que estos son más robustos al considerar la variabilidad inherente. Aunque para nuestro conjunto de datos la diferencia entre los valores obtenidos por ambos métodos parece ser pequeña en lo global, considerando la naturaleza de la variable analizada cualquier diferencia implica miles de personas y tal hecho puede ser determinante a la hora de tomar decisiones en políticas de salud, por ejemplo.

### Referencias.

- [1] ARROYO GALLARDO, Javier (2008), "Tesis para la obtención del título de doctor: Métodos de Predicción para Series Temporales de Intervalos e Histogramas". Departamento de Organización Industrial Escuela Técnica Superior de Ingeniería (ICAI). Universidad Pontificia Comillas. Madrid.
- [2] ARROYO, Javier; GONZÁLES RIVERA, Gloria; MATÉ, Carlos (2010), "Forecasting with interval and histogram data. Some financial applications". Working Papers 201438, University of California at Riverside, Department of Economics.
- [3] DEWAYNE R. Derryberry (2014), "Basic data analysis for time series with R". John Wiley & Sons.
- [4] DIDAY, Edwin; NOIRHOMME FRAITURE, Monique(2008), "Symbolic data analysis and the SODAS software". John Wiley & Sons.
- [5] HYNDMAN, Rob J., KOEHLER Anne B., ORD J. Keith, SNYDER Ralph D.(2008), Forecasting with Exponential Smoothing". Springer Series in Statistics.
- [6] MOORE, R. E. (1996) "Interval Analysis". Prentice Hall, Englewood Cliffs, N.J.

# Regresión por cuantiles multivariada para modelar el efecto de cuestionarios online sobre las evaluaciones tradicionales

Vaira S, Contini L

Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, Santa Fe - Argentina

stella.vaira@gmail.com

## Resumen

En este trabajo de investigación se aprovecharon las posibilidades educativas que tiene el entorno virtual que ofrece la Universidad Nacional del Litoral (UNL), a través de la plataforma Moodle con el objetivo de aplicar dos cuestionarios online de preguntas con múltiples opciones a los alumnos que se encuentran en primer año de la carrera Licenciatura en Nutrición (LN), en la asignatura Matemática General y explicar mediante regresión por cuantiles la relación entre los resultados de estos cuestionarios con las evaluaciones tradicionales. El estudio se llevó a cabo en el año 2018, con todos los alumnos que se encontraban cursando Matemática General de la carrera de LN (FBCB – UNL). La muestra estuvo compuesta por  $n = 155$  alumnos. Las estimaciones de los coeficientes (con el valor  $p$  asociado) de las regresiones por cuantiles ( $\hat{\beta}(\tau); \tau \in (0,1)$ ), para cada una de los modelos de regresión fueron calculadas y se obtuvo que la influencia del segundo parcial tradicional ( $X_2$ ) se mantiene a lo largo de toda la distribución, en cambio el segundo cuestionario online es significativo para la RC 0,20 y 0,90; de lo que surge que la predicción de las notas obtenidas por los alumnos es diferente dependiendo de si han obtenido notas bajas o altas.

## Palabras Clave

Cuestionarios online, Educación matemática, Regresión por cuantiles

## Introducción

La educación superior que se está construyendo viene caracterizada por al menos dos factores predominantes: (i) la creciente incorporación de las TIC (Internet, software especializado) en los procesos tanto de enseñanza como de aprendizaje, (ii) la existencia de un interés generalizado entre las instituciones por reforzar un enfoque aplicado de la matemática para hacer más visibles las notables competencias profesionales que deben tener los alumnos en un entorno social con fuerte predominio tecnológico (Vaillant, 2012; Marcelo, 2013). Las TIC han producido importantes cambios en las sociedades modernas y tienen una fuerte influencia sobre el entorno universitario, puesto que promueven la aparición tanto de nuevas necesidades formativas como de nuevas

posibilidades metodológicas. Como resultado de esta influencia, emergen nuevos modelos de formación en los que el papel desarrollado por profesores y estudiantes difiere parcialmente del tradicionalmente establecido durante los últimos años. Las TIC ofrecen nuevas formas de comunicación, colaboración y participación en los procesos educativos/formativos (Llinares, 2013). Si bien la regresión por mínimos cuadrados ordinarios (MCO) es uno de los métodos estadísticos más utilizados, la regresión por cuantiles (RC) (Koenker y Bassett, 1978) permite estimar los efectos marginales de las variables explicativas en distintos puntos de la distribución de probabilidad de la variable en estudio, con lo que es posible explorar diferentes aspectos de la relación entre la variable dependiente y las variables independientes. Hay motivos para la RC, la variable dependiente es multimodal, y al desconocer el motivo de la multimodalidad, no es posible un análisis por subgrupos o la búsqueda de un diseño estadístico estratificado, además interesa determinar lo que predice la mediana o algún otro cuantil particular, ya que se trata de las notas de los alumnos y permitiría modelar diferentes niveles de respuesta, tanto los causantes de las bajas notas, las notas moderadas o las notas altas. En este trabajo de investigación se aprovecharon las posibilidades educativas que tiene el entorno virtual que ofrece la Universidad Nacional del Litoral (UNL), a través de la plataforma Moodle con el objetivo de aplicar dos cuestionarios online de preguntas con múltiples opciones a los alumnos que se encuentran en primer año de la carrera Licenciatura en Nutrición (LN), en la asignatura Matemática General y explicar mediante un modelo estadístico la relación entre los resultados de estos cuestionarios con las evaluaciones tradicionales.

## Material y Métodos

El estudio se llevó a cabo en el año 2018, con todos los alumnos que se encontraban cursando Matemática General de la carrera de LN (FBCB – UNL). La muestra estuvo compuesta por  $n = 155$  alumnos de los cuales 41 eran varones, lo que representa el 23% del grupo. El dictado de la asignatura estaba organizado en 5 comisiones de aproximadamente 30 alumnos cada una. Los dos cuestionarios online se implementaron antes de cada parcial tradicional, es decir en la modalidad presencial y escrito. En total y al finalizar el cuatrimestre se obtuvieron 4 variables: 3 consideradas como explicativas y una (el último parcial tradicional) como variable a explicar, ya que los resultados de esta definen las condiciones finales de los alumnos: Regular, Libre o Promocionado. Se utilizó regresión por cuantiles multivariada con la que se compararon los efectos de las variables regresoras entre diferentes estimaciones por cuantiles, considerando a  $\tau = \{0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9\}$ , las variables fueron: Y: notas del segundo parcial tradicional, X1: resultados del primer cuestionario online, X2: notas del primer parcial tradicional, X3: resultados del segundo cuestionario online. Los cuestionarios online se diseñaron utilizando el entorno virtual de la Facultad. Los datos fueron procesados con el software R, utilizando el paquete “quantreg”, y al momento de calcular los valores p del estadístico asociado a la significancia de los coeficientes se utilizó el método de bootstrap.

## Resultados

La distribución de notas del segundo parcial de los 155 alumnos que cursaron Matemática General en el primer cuatrimestre de 2018, resultó ser fuertemente sesgada, con algunos “picos” a lo largo de la variable que se quiere modelar, este fue uno de las razones que motivó la utilización de regresión por cuantiles.

La Tabla 1 muestra las estimaciones de los coeficientes (con el valor p asociado) de las regresiones por cuantiles ( $\hat{\beta}(\tau)$ ;  $\tau \in (0,1)$ ), para cada una de los modelos de regresión. La influencia del segundo parcial tradicional ( $X_2$ ) se mantiene a lo largo de toda la distribución, en cambio el segundo cuestionario online es significativo para la RC 0,20 y 0,90, en ambos casos con una pendiente entre 1,4 y 1,75 respectivamente tiene una influencia positiva, pero al comparar estas dos pendientes no resultaron estadísticamente diferentes, pero esta información es útil para explicar y reorganizar el cuestionario online al momento de diagnosticar resultados.

**Tabla 1:** Para las diferentes regresiones por cuantiles múltiples, se muestran las estimaciones puntuales y el valor p asociado.

<u>Cuantil</u>	$X_1$ (cuestionario 1)	$X_2$ (parcial tradicional)	$X_3$ (cuestionario 2)
0,1	0,605 (0,484)	0,436 ( $<10^{-3}$ )	0,608 (0,281)
<b>0,2</b>	<b>0,158 (0,868)</b>	<b>0,488 (<math>&lt;10^{-3}</math>)</b>	<b>1,476 (0,048)</b>
0,3	0,975 (0,323)	0,516 ( $<10^{-3}$ )	0,295 (0,714)
0,4	0,358 (0,643)	0,583 ( $<10^{-3}$ )	0,382 (0,535)
0,5	0,475 (0,453)	0,611 ( $<10^{-3}$ )	0,271 (0,577)
0,6	0,672 (0,178)	0,594 ( $<10^{-3}$ )	0,328 (0,459)
0,7	0,762 (0,212)	0,557 ( $<10^{-3}$ )	0,506 (0,239)
0,8	0,670 (0,465)	0,589 ( $<10^{-3}$ )	1,053 (0,082)
<b>0,9</b>	<b>0,638 (0,689)</b>	<b>0,533 (<math>&lt;10^{-3}</math>)</b>	<b>1,714 (0,043)</b>

## Discusión

Para cada caso se obtuvieron estimaciones significativas del coeficiente del efecto de la variables regresoras  $X_2$  y  $X_3$ , en particular en los cuantiles 0,2 y 0,9, logrando valores de pseudos  $R^2$  cercanos a 0,46, de lo que surge que la predicción de las notas obtenidas por los alumnos es diferente dependiendo de si han obtenido notas bajas o altas. Este resultado es de particular importancia pues, poder predecir rendimiento de los alumnos permitiría al grupo de docentes-investigadores generar estrategias en las prácticas de enseñanza, reforzando acciones en el grupo de alumnos con bajo rendimiento y potenciando a aquellos con alto rendimiento. Una propuesta a futuro puede ser utilizar estas estimaciones como estimaciones a priori y realizar un análisis bayesiano, separar muestras y validar ambos métodos.

## Agradecimientos

El presente trabajo se realizó en el marco del proyecto de investigación PI50120150100053LI “Educación Matemática e Interdisciplinariedad. Eje principal de trabajo: la Modelación”, subsidiado por la Universidad Nacional del Litoral.

## Referencias

Marcelo, C. (2013). Las tecnologías para la innovación y la práctica docente. Revista Brasileira de Educação v. 18 n. 52

Vaillant, Denise; Marcelo, Carlos. Ensinando a ensinar. As quatro etapas de uma aprendizagem. Curitiba: Editora Universidade Tecnológica Federal do Paraná, 2012.

Llinares, S. (2013). Innovación en la educación matemática: más allá de la tecnología. Modelling in Science Education and Learning Volumen 6(1), No. 1. Instituto Universitario de Matemática Pura y Aplicada.

Koenker, R.; Bassett, G. (1978). “Regression Quantiles”. *Econometrica*, 46:1, pp. 33-50.

Koenker, R.; Hallock, K. (2001). “Quantile Regression: An Introduction”. *Journal of Economic Perspectives*, 15, pp. 143–156

Koenker, R. (2006). Quantile Regression in R: a vignette. CRAN: <http://cran.r-project.org>.

## Selección de atributos en clasificación supervisada. Una mejora considerando una corrección de la entropía condicional

ROMERO, MARÍA DEL CARMEN<sup>1</sup> y CLAUSSE, ALEJANDRO<sup>2</sup>

<sup>1</sup>Facultad de Ciencias Económicas, Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina

<sup>1</sup>PLADEMA, Facultad de Ciencias Exactas, Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina

romero@econ.unicen.edu.ar

### Resumen

*La alta dimensionalidad genera escenarios que conducen al replanteo de los métodos a utilizar para la obtención de información a partir de un conjunto de datos. En casos de clasificación supervisada, en los cuales el interés se centra en la selección de atributos que distinguen entre clases (o grupos), los métodos convencionales tienen problemas debido al sobreajuste. En Romero, Clausee y Di Rienzo (2013) se propone un método para seleccionar un conjunto de atributos que puedan considerarse importantes en la distinción entre clases en contextos de datos binarios. Los atributos son seleccionados en forma incremental conforme a la relevancia que tienen en la distinción entre clases, la cual se mide con la entropía condicional de la clase dado que se conoce el estado simultáneo de los atributos del conjunto diferencial. El método de máxima verosimilitud subestima la entropía, y esta subestimación crece con la cardinalidad del conjunto diferencial. El objetivo de este trabajo es desarrollar una mejora considerando una corrección por sesgo que tenga en cuenta esta cardinalidad. Para explorar el comportamiento de esta mejora, se generaron escenarios de atributos binarios por simulación con diferente cantidad de observaciones usando el software R. Entre los principales resultados, se encuentra que la corrección por sesgo propuesta mejora el rendimiento del método. En una etapa posterior, se estudiará la posibilidad de brindar un intervalo de confianza para la estimación de la entropía condicional que pueda guiar en la decisión de la cantidad de atributos a seleccionar.*

### Palabras Clave

Selección de atributos – Clasificación supervisada – Entropía – Simulación.

## Introducción

El objetivo de los problemas de clasificación supervisada es, dado un conjunto de observaciones clasificadas en grupos conocidos, descubrir un conjunto de atributos que sean relevantes en la distinción entre grupos y que permitan clasificar posteriormente nuevas observaciones (por ejemplo, identificación de genes asociados con el estado de salud o enfermedad de un paciente respecto a cierta patología). En contextos de alta dimensionalidad (gran cantidad de datos y/o mayor cantidad de atributos que de observaciones) se suelen utilizar métodos filtro que seleccionan atributos por medio de métricas que describen la capacidad que tienen para distinguir entre clases. La mayoría de las propuestas generan el conjunto de atributos de manera incremental donde los atributos nuevos son agregados al conjunto de atributos seleccionados previamente por pasos sucesivos. En Romero, Di Rienzo y Clause (2013) se presentó un método que usa la entropía condicional para seleccionar atributos que pueden considerarse diferenciales en la distinción de grupos en contextos de datos binarios. En este trabajo se propone una mejora al uso de la entropía condicional.

## Material y Métodos

La entropía condicional puede interpretarse como una medida de la reducción de incertidumbre de una variable condicionada al conocimiento de otra. El conjunto de atributos diferenciales se construye de forma incremental: en la primera iteración se calculan las entropías condicionales de la clase respecto de cada atributo tomado en forma individual, resultando el de menor entropía condicional, el primero a ingresar; el segundo atributo a incluir será aquel que, en conjunto con el ya incluido, mejor explique el comportamiento de la clase y así sucesivamente. El algoritmo termina al obtener una entropía condicional nula o cuando ya se ingresaron todos los atributos.

El método de máxima verosimilitud subestima la entropía condicional ya que genera un sesgo (*bias*) que está asociado al tamaño de la muestra. Miller en 1955 (Miller, 1955) propuso una corrección respecto de una única variable. A partir de esta corrección, y evaluando teóricamente el sesgo de la estimación, se replanteó el sesgo a considerar en cada una de las iteraciones (para diferente cardinalidad del conjunto diferencial). El término de corrección del sesgo promedio  $\beta^{(d)}$  en cada iteración  $d$ , para la estimación de la entropía de la clase condicional a los atributos en el conjunto diferencial en un contexto de  $N_{obs}$  observaciones está dado por: 
$$\beta^{(d)} = \frac{1}{\ln(2)} \frac{2^d}{N_{obs}}$$

El rendimiento del método se analizó a través de un caso generado por simulación dado por una clase binaria  $C$  y un conjunto de 5 atributos binarios independientes e idénticamente distribuidos. La clase está asociada con  $X_1$  y  $X_2$ :  $H(C/X_1)=0.67$ ,  $H(C/X_2)=0.84$  y  $H(C/X_1X_2)=0.46$ . Se generaron 10000 muestras para distinta cantidad de observaciones (10, 30, 50, 100 y 1000). Los algoritmos se desarrollaron en el software R (R Core Team, 2018) y se usó la función *entropy* del *package entropy* (Hausser y Strimmer, 2014).



## Resultados

Entre los principales resultados, se observa que la entropía de la clase condicionada al conjunto de atributos diferenciales:

- se acerca a los valores teóricos a medida que aumenta la cantidad de observaciones,
- tiene una mayor subestimación a medida que se incrementa la cantidad de atributos en el conjunto diferencial, y que este sesgo crece ante una baja cantidad de observaciones y
- el sesgo propuesto en este trabajo mejora la estimación de la entropía, obteniéndose valores más cercanos a los teóricos (que sin considerar el sesgo) en todos los casos

## Discusión

Los resultados presentados en este trabajo aportan al desarrollo de un método que permita identificar un conjunto de atributos diferenciales en contextos de atributos binarios. En una etapa posterior, se estudiará la construcción de un intervalo de confianza para la estimación de la entropía condicional de cada iteración del método propuesto que pueda guiar en la decisión de la cantidad de atributos a seleccionar. Entre las líneas futuras de investigación se prevé la evaluación del comportamiento del método en escenarios con atributos dependientes entre sí y con mayor cantidad de atributos que de observaciones.

## Referencias

Hausser, J. y Strimmer, K. *Package Entropy: Estimation of Entropy, Mutual Information and Related Quantities*. Versión 1.2.0. <https://cran.r-project.org/web/packages/entropy/index.html>. 2014.

Miller, G. A. (1955). Note on the bias of information estimates. En H. Quastler, editor, *Information Theory in Psychology II-B*, pag. 95–100. Free Press, Glencoe, IL.

R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Disponible online en <https://www.R-project.org/>. 2018.

Romero, M. del C.; Di Rienzo, J. y Clausse, A. "Selección de atributos en clasificación supervisada. Uso de la entropía condicional". *IV Encuentro Iberoamericano de Biometría y XVIII Reunión Científica del Grupo Argentino de Biometría*. Mar del Plata, Argentina, 25, 26 y 27 de septiembre de 2013.

# UNA APLICACIÓN DE MODELOS DE REGRESIÓN SIMBÓLICA PARA VARIABLES DE INTERVALOS

Carrizo, J., MALLEA, A.

Departamento de Matemática, FFHA, Universidad Nacional de San Juan. Argentina

[jcarrizo@ffha.unsj.edu.ar](mailto:jcarrizo@ffha.unsj.edu.ar), [jorgelinav\\_carrizo@yahoo.com.ar](mailto:jorgelinav_carrizo@yahoo.com.ar)

## Resumen

*El presente trabajo se enmarca dentro de la regresión simbólica para datos de intervalo, que extiende métodos de regresión clásica al caso simbólico. La regresión simbólica es parte del análisis simbólico de datos, nuevo paradigma que surge desde la década del 80 para analizar datos con variabilidad intrínseca. Se aplican métodos de regresión simbólica a datos obtenidos por agregación a partir de la matriz de datos "Murciélagos", que tiene como unidades a murciélagos que viven en Europa y se caracterizan por las variables tipo de especie, longitud de la cabeza, longitud de la cola, longitud del antebrazo y peso. Información más detallada sobre el conjunto de datos se puede encontrar en Douzal-Chouakria et al. (2009). Se comparan los distintos métodos con el error cuadrático medio, a fin de evaluar su performance.*

**Palabras Clave:** Análisis Simbólico de Datos. Variable simbólica de intervalo. Regresión Simbólica.

## Introducción

En el Análisis de Datos clásico, los datos proceden de observaciones únicas, de determinadas variables sobre individuos únicos. Pero cuando los elementos de interés son clases o grupos de algún tipo, hay variabilidad inherente en los datos. Los datos simbólicos, introducidos por E. Diday [1] en la década de los ochenta, proporcionan un marco que permite representar datos con variabilidad mediante nuevos tipos de variables, llamadas simbólicas. Estas variables pueden tomar como valores un conjunto, intervalo y, más generalmente, una distribución.

Una variable simbólica de intervalo es aquella donde a cada entidad le corresponde un intervalo de valores, se asume que los valores dentro de cada intervalo se distribuyen uniformemente a fin de extender estadísticos clásicos a este tipo de variables simbólicas. En este trabajo se presentan adaptaciones del modelo de regresión lineal clásica para variables de intervalo. La metodología empleada forma parte de lo que se denomina *Regresión simbólica* de datos, área de investigación activa desde que Billard y Diday [2] introdujeron el primer enfoque del modelo de regresión a datos con valores de intervalo, trabajando con un modelo de regresión clásica sobre los centros de los intervalos. Posteriormente propusieron estimaciones a partir de modelos independientes para

los límites inferior y superior de los intervalos. Lima Neto y de Carvalho [7], siguiendo el enfoque de Billard y Diday, emplearon los centros y rangos de los intervalos para construir dos regresiones independientes.

Sin embargo, la predicción del límite inferior podría exceder el límite superior previsto. Para evitar esta situación, Lima Neto y de Carvalho [8] consideraron un modelo restringido que obliga a todos los parámetros del modelo de los rangos a ser positivos. Más adelante Lima Neto [9] definió el modelo de regresión simbólica bivariada, basado en la teoría de modelos lineales generalizados.

Si bien existen otros métodos de Regresión Simbólica en la literatura, se mencionan sólo éstos por ser los que se aplican en el presente trabajo.

## Material y Métodos

Los modelos de regresión simbólica de intervalos que se aplican en este análisis son:

*Método del Centro:* los coeficientes del modelo se estiman aplicando el modelo clásico al punto medio de los intervalos,

*Método Mínimo Máximo:* ajusta dos modelos de regresión lineal, para los límites inferior y superior del intervalo,

*Método del Centro y del Rango:* ajusta dos modelos de regresión lineal, para el punto medio y el radio del intervalo,

*Método de Centros y Rangos Restringido:* coeficientes del modelo de centros se estiman como en el modelo anterior y coeficientes del modelo de rango se estiman utilizando el algoritmo de Lawson y Hanson,

*Método Simbólico Bivariado,* se adapta a variables de intervalo el modelo lineal generalizado bivariado. Es un modelo probabilístico que permite hacer inferencia.

Los métodos presentados se aplican a la tabla simbólica de datos, construida tomando como concepto cada especie de murciélago y como variables de intervalo: longitud de la cabeza, longitud de la cola, longitud del antebrazo y peso.

Para los métodos de regresión simbólica se toma coma respuesta peso y el resto como variables explicativas.

*Rendimiento de los modelos.* (Lima Neto y De Carvalho [8]): La evaluación del desempeño de los modelos de regresión lineal descriptos se basan en el error cuadrático medio para límites inferiores y superiores:

$$RMSE_L = \sqrt{\frac{\sum_{j=1}^m (L_{Y(j)} - L_{\hat{Y}(j)})^2}{m}} \quad RMSE_U = \sqrt{\frac{\sum_{j=1}^m (\bar{I}_{Y(j)} - \bar{I}_{\hat{Y}(j)})^2}{m}}$$

## Resultados

Para comparar los distintos métodos se ha utilizado, dentro de la plataforma R, el paquete "iRegression". Este paquete tiene como autores a Lima Neto E., De Souza Filho C. y Marinho P. Contiene algunos métodos de regresión para variables simbólicas de intervalo. La Tabla1, resume la performance de los modelos, basada en sus respectivos errores cuadráticos medios.

Para los modelos simbólicos bivariados se usó la función de enlace identidad obteniendo los mismos resultados en las estimaciones que MinMax para Bivar1 y CRM para Bivar2, sin embargo los errores no coinciden.

Se observa que el método Bivar 2 funciona mejor que los restantes métodos al estimar los coeficientes del límite inferior ya que proporciona el menor valor de  $RMSE_L$ . Mientras que para el límite superior del intervalo se comportan mejor los métodos MinMax o Bivar1.

Tabla 1: **Evaluación del rendimiento de los modelos**

Modelos	$RMSE_L$	$RMSE_U$
CM	4.475	4.763
Min Max	1.644	2.869
CRM	3.741	4.086
CCRM	3.785	4.193
Bivar1	1.644	2.869
Bivar2	1,141	7,49

## Conclusiones

En este trabajo se han presentado algunos modelos de regresión lineal simbólica para variables de intervalo, los cuáles han sido probados usando datos de clases de murciélagos, obtenidas por la especie a la que pertenecen. La idea es mostrar las ventajas de trabajar con regresión simbólica, en el caso de tener clases de individuos, en lugar de entes individuales y por ende intervalos de valores de variables explicativas o de respuesta.

A pesar de todos los modelos propuestos en la literatura, aún no existe un modelo que simultáneamente resuelve todas las limitaciones. Estas observaciones motivan que a futuro se desarrollen metodologías de regresión utilizando la aritmética de intervalos y respuestas de naturaleza aleatoria.

## Referencias

- [1] Bock, H.-H.; Diday, E. (2000): *Analysis of Symbolic Data: Exploratory methods for extracting statistical information from complex data*. Berlin-Heidelberg: Springer-Verlag.
- [2] Billard and Diday (2000). *Regression analysis for interval-valued data*. In: Proc. of IFCS'00, Belgium, pp. 369-374, Springer.
- [3] Billard, L., Diday, E. (2002). *Symbolic Regression Analysis*. In: Proc. IFCS'02, Poland, pp. 281-288, Springer.
- [4] Billard, L., Diday, E. (2006). *Symbolic Data Analysis: Conceptual Statistics and Data Mining*. Wiley.
- [5] De Carvalho, F.A.T., Neto, E.A.L. (2008). *Centre and Range method for fitting a linear regression model to symbolic intervalar data*. Computational Statistics & Data Analysis, 52, pp. 1500-1515.
- [6] De Carvalho, F.A.T., Neto, E.A.L., (2010). *Centre and Range method for fitting a linear regression model to symbolic intervalar data*. Computational Statistics & Data Analysis, 54, pp. 333-347.
- [7] Lima Neto, E. and F. De Carvalho (2008). Centre and Range method for fitting a linear regression model to symbolic interval data. Computational Statistics and Data Analysis 52(3), 15001515.
- [8] Lima Neto, E. and F. De Carvalho (2010). Constrained linear regression models for symbolic interval-valued variables. Computational Statistics and Data Analysis 54(2), 333347.
- [9] Lima Neto, Cordeiro, De Carvalho, 2011. *Bivariate symbolic regression models for interval-valued variables*. Journal of Statistical Computation and Simulation 81 (11), pp.1727-1744