

Instituto Nacional de Estadística y Geografía

# Memoria del XXVI Foro Nacional de Estadística

2011



INSTITUTO NACIONAL  
DE ESTADÍSTICA Y GEOGRAFÍA

Obras complementarias publicadas por el INEGI sobre el tema:

Memoria del Foro Nacional de Estadística 1993 a 2010 (Volumen del Núm. VIII al XXV).

Catalogación en la fuente INEGI:

310.4 Foro Nacional de Estadística (26º : 2011 : Villahermosa, Tabasco, México).  
Memoria del XXVI Foro Nacional de Estadística / Instituto Nacional de Estadística y Geografía.-- México : INEGI, c2012.

iii, 98 p.

ISBN 978-607-494-391-7

“Universidad Juárez Autónoma de Tabasco. Villahermosa, Tabasco, del 21 al 23 de septiembre de 2011”

1. Estadística - Alocuciones, Ensayos, Conferencias. I. Instituto Nacional de Estadística y Geografía (México). II. Universidad Juárez Autónoma de Tabasco. III. Asociación Mexicana de Estadística.

Si requiere más información sobre esta obra, favor de contactarnos a través de:

Centros de consulta y comercialización (consulte el domicilio en Internet)

Larga distancia sin costo: 01 800 111 46 34

[www.inegi.org.mx](http://www.inegi.org.mx)

[atencion.usuarios@inegi.org.mx](mailto:atencion.usuarios@inegi.org.mx)

Síguenos en:   

# Presentación

El XXVI Foro Nacional de Estadística se llevó a cabo en la ciudad de Villahermosa, Tabasco del 21 al 23 de septiembre de 2011 teniendo como anfitrión a la Universidad Juárez Autónoma de Tabasco (UJAT). La organización de este evento estuvo a cargo de la Dirección Académica de Ciencias Básicas-UJAT en conjunto con la Asociación Mexicana de Estadística.

En esta ocasión, el volumen de memorias está integrado por dos secciones: en la primera sección se publican los resúmenes *in extenso* de algunas contribuciones libres presentadas en este evento, en la segunda sección se incluyen los resúmenes cortos de los trabajos que formaron parte del programa académico del evento. La primera sección está compuesta por 7 trabajos mientras que en la segunda sección, se encuentran los resúmenes de 4 Conferencias Magistrales, 64 contribuciones libres presentadas tanto en exposición oral como en carteles y finalmente 5 trabajos ganadores del Concurso Nacional de Estadísticos Jóvenes.

Los resúmenes *in extenso* se sometieron a un proceso de revisión académica y de estilo con el fin de asegurar que el abordaje estadístico y la redacción fueran correctos, así como el que se mostrara un mínimo de originalidad o aspectos novedosos en la metodología, en los resultados y/o en las aplicaciones presentadas. Los resúmenes cortos corresponden a aquellos trabajos que fueron sometidos y resultaron aceptados por el Comité de Programa.

Agradecemos a todos los autores por su participación y por la calidad de los trabajos que se presentaron. A nombre de la Asociación Mexicana de Estadística, agradecemos a la Universidad Juárez Autónoma de Tabasco por el apoyo, entusiasmo y dedicación en la organización de este Foro, así como al Instituto Nacional de Estadística y Geografía por el apoyo en la publicación del presente volumen.

## **El Comité Editorial**

Belem Trejo Valdivia  
Edilberto Nájera Rangel  
Eduardo Gutiérrez Peña



# Índice general

## Sección I. Resúmenes *in extenso*

<b>Algunos factores que influyen en el aprendizaje en un primer curso en una facultad de ciencias. . . . .</b>	<b>7</b>
<i>Yoanna Arenas Martínez, Hortensia Reyes C., Hugo Cruz Suárez, Flaviano Godínez Jaimes</i>	
<b>Estimación de totales y medias en el muestreo por bola de nieve y en presencia de probabilidades de ligas heterogéneas y asociadas con la variable respuesta. . . . .</b>	<b>13</b>
<i>Martín H. Félix Medina</i>	
<b>Metodología de observación de estados afectivos durante una situación de aprendizaje . . . . .</b>	<b>19</b>
<i>Nery Sofía Huerta-Pacheco, Genaro Rebolledo-Méndez, Sergio Hernández-González, Francisco Javier Landa-Torres</i>	
<b>Desarrollo de un módulo de pruebas de Estadística No Paramétrica con Java y R . . . . .</b>	<b>25</b>
<i>Francisco Javier Landa-Torres, Sergio Hernández-González, Genaro Rebolledo-Méndez, Nery Sofía Huerta-Pacheco</i>	
<b>Modelo para estimar el riesgo de inundaciones en la cuenca Grijalva Usumacinta . . . . .</b>	<b>31</b>
<i>Ana Laura Reyes Olvera, Eduardo Gutiérrez González</i>	
<b>Elección de variables meteorológicas relacionadas con ozono en la estación Plateros . . . . .</b>	<b>39</b>
<i>Sara Rodríguez R., Hortensia Reyes C., Humberto Vaquera H., Paulino Pérez R.</i>	

<b>Generalización de la condición de la matriz <math>Z</math> para la caracterización del BLUP de <math>X\beta + Zu</math> . . . . .</b>	<b>45</b>
<i>Fernando Velasco Luna, Mario Miguel Ojeda Ramírez</i>	

## Sección II. Resúmenes cortos

<b>Conferencias Magistrales . . . . .</b>	<b>55</b>
<b>Contribuciones Libres . . . . .</b>	<b>59</b>
<b>Trabajos del Concurso Nacional de Estadísticos Jóvenes (patrocinado por SAS) . . . . .</b>	<b>97</b>

## Sección I

### Resúmenes *in extenso*



# Algunos factores que influyen en el aprendizaje en un primer curso en una facultad de ciencias

Yoanna Arenas Martínez<sup>a</sup>, Hortensia Reyes C.<sup>b</sup>, Hugo Cruz Suárez<sup>c</sup>  
*Facultad de Ciencias Físico-Matemáticas, BUAP*

Flaviano Godínez Jaimes<sup>d</sup>  
*Unidad Académica de Matemáticas, UAGro*

## 1. Introducción

En las carreras de ciencias hay muchos alumnos que reprueban los primeros cursos. Generalmente en estos cursos el profesor expone los temas donde a partir de un conjunto de axiomas se construye una serie de proposiciones verdaderas usando la lógica matemática y la teoría de conjuntos. Para evaluar, el profesor deja una lista de ejercicios, tipo examen, para que los alumnos se preparen. Cuando los alumnos tienen dudas en sus desarrollos o resultados pueden ir a consultar al profesor antes de la evaluación. En la Facultad de Ciencias Físico Matemáticas de la Benemérita Universidad Autónoma de Puebla (FCFM-BUAP), se ha observado que los alumnos de los primeros semestres están en grupos numerosos y quizá esto los afecta, porque no tienen el tiempo suficiente para adquirir confianza con el profesor y preguntar sus dudas. Otro factor importante para el aprendizaje del alumno es la cantidad de horas de estudio, fuera de las horas de clase, que dedican a la materia (para resolver ejercicios, revisión de material visto en clase o búsqueda de más bibliografía para estudiar). Otros factores son el cambio de nivel medio superior a superior, autoestima baja, falta de madurez ante las decisiones de su propio futuro, etc.; ver Reyes et al. (2011).

---

<sup>a</sup>guadalupeyam@hotmail.com

<sup>b</sup>hreyes@fcfm.buap.mx

<sup>c</sup>hcs@fcfm.buap.mx

<sup>d</sup>fgodinezj@gmail.com

El aprendizaje de los alumnos en cualquier nivel y en especial en la FCFM-BUAP es complejo, por ello se han llevado a cabo investigaciones sobre las causas que lo afectan; ver Capilla et al. (2010) y Hernández et al. (2010). Este trabajo se enfoca en encontrar algunos factores que influyen en la calificación del curso de Matemáticas Elementales (ME). Esta materia es fundamental en la licenciatura de la FCFM-BUAP y se imparte en primer semestre. Además, este trabajo contribuirá a encontrar algunos factores que afecten la permanencia de los alumnos de primer semestre en la FCFM-BUAP. El problema no tiene solución inmediata, pero da la posibilidad de hacer conciencia en los alumnos y profesores que están relacionados en estos temas, sobre la dificultad que se tiene al involucrarse en un problema real en la educación. En el año escolar 2011-2012 la materia de Matemáticas Elementales tuvo un total de 66 aprobados, 84 reprobados y 3 individuos con valores perdidos. Aproximadamente el 56.9% de los estudiantes reprobaron ese curso lo que muestra la difícil situación que se vive en este curso.

## 2. Materiales y métodos

El instrumento utilizado para evaluar el aprovechamiento de los alumnos en años anteriores consta de 21 preguntas y se aplicó a 93, 59 y 72 alumnos de nuevo ingreso de las licenciaturas de Matemáticas (M), Matemáticas Aplicadas (MA) y Actuaría (A) de la FCFM-BUAP respectivamente. La encuesta se realizó un mes después de haber ingresado y ya varios alumnos habían desertado, por lo que sólo se encuestó a los alumnos presentes ese día. Los alumnos encuestados en total fueron 153; 69 de M, 46 de MA y 38 de A.

Se usó un modelo de regresión lineal múltiple para determinar qué variables ayudan a explicar la *Calificación en ME al finalizar el curso*, pues la calificación es la forma en la que los profesores evalúan el aprovechamiento y desempeño de los alumnos. Las variables independientes fueron redactadas como preguntas de opciones múltiples, enseguida se listan y en paréntesis se da el número de opciones de cada una: 1.-Escolaridad del papá (6), 2.- Escolaridad de la mamá (6), 3.- Profesión del papá (4), 4.- Profesión de la mamá (4), 5.- Tipo de escuela de procedencia (2), 6.- Horas de trabajo semanal (4), 7.- Dinero disponible para diversión (3), 8.- Lugar de procedencia (3), 9.- Licenciatura fue tu primer opción (2), 10.- Te gusta el ambiente de estudio y compañerismo de tu facultad (2), 11.- Conceptos adquiridos en la preparatoria te ayudan a comprender mejor los nuevos conceptos (2),

---

12.- Horas de estudio de ME (4), 13.- Horas de estudio de otras materias (4), 14.- Horas de estudio para examen de ME (4), 15.- Qué problemas enfrentas al estudiar (2), 16.- Cuántos ejercicios resuelves para examen de ME (4), 17.- Cómo se te hace más fácil estudiar (2), 18.- Tienes confianza en tu profesor para resolver dudas (2), 19.- Asesorías te ayudan a contestar mejor tu examen (2), 20.- Número de compañeros influye negativamente en tu aprendizaje (2), 21.- Forma de estudiar (2), 22.- Son interesantes los temas de ME (2), 23.- Te gusta la clase de tu profesor (2), 24.- Qué porcentaje entiendes (3).

### 3. Resultados

Las respuestas fueron analizadas en el paquete estadístico SPSS versión 18 (Stephenson y Gillel (2010)).

Se encontró que los alumnos tienen malos hábitos de estudio pues la mayoría no estudia la materia (38 %) y si lo hacen estudian a lo más 2 horas (54.3 %). También se encontró que los alumnos no entienden lo que el profesor les expone en clase (54.3 %) y que no hacen ejercicios o no tienen apuntes (21.7 %).

Como las variables independientes eran principalmente categóricas y el número de observaciones era pequeño, se identificaron las variables dependientes con la *Calificación en ME* para ser incluidas posteriormente en el modelo. Una forma fue categorizando la calificación en tres: 5-6, 7-8 y 9-10 y evaluando la posible dependencia entre ésta y la variable categórica de interés vía un estadístico tipo Chi-cuadrado. Se encontró que no se rechazó la independencia entre la calificación categorizada y las horas de estudio ( $p$ -valor = .932).

Cuando la variable independiente fue dicotómica, la asociación se determinó comparando las medias de la *Calificación en ME* en los dos grupos definidos por las dos categorías de la variable de interés con una prueba de  $t$  para muestras independientes.

En el Cuadro 2 se muestra que la variable dicotómica *primera opción* estuvo asociada con la *Calificación en ME* ya que se rechaza la igualdad de medias ( $p$ -valor = 0.001). Un resultado similar se encontró para la variable *confianza profesor* ( $p$ -valor = 0.022).

La variable respuesta, *Calificación en ME*, es continua, por tanto el modelo apropiado es el de regresión lineal. Después de identificar las variables más asociadas, se usaron métodos de selección de variables para determinar que variables permiten explicar la *Calificación en ME*. Los resultados del ajuste fueron:

	Prueba Levene para igualdad de varianzas		Prueba de T para la igualdad de medias						
	F	Sig	t	gl	Sig bilateral	Dif. medias	Error tip. diferencia	95 % Intervalo confianza para la diferencia	
								Inferior	Superior
Varianzas iguales	15.897	.000	-3.07	90	.003	-1.007	.35	-1.773	-.382
Varianzas diferentes			-3.48	79	.001	-1.007	.30	-1.692	-.463

**Cuadro 1:** Prueba de igualdad de medias de la *Calificación en ME* en las categorías de la variable primera opción.

Modelo	Coef. no estand.		Coef. tipifi.	t	Sig	Int. de confianza de 95 % para B	
	B	Error Típ	Beta			Lím. inferior	Lím. superior
Constante	4.840	.429		11.285	.000	3.998	5.69
Prim-Opci	1.146	.343	.328	3.344	.001	.465	1.83
Conf-Profr	.911	.390	.230	2.340	.022	.137	1.69

**Cuadro 2:** Resultados del ajuste del modelo de regresión final.

El modelo ajustado final es:

$$\hat{Y}_i = 4.84 + 1.146X_{1i} + 0.911x_{2i}$$

donde  $X_1 = 1$  si fue la primera opción del alumno fue estudiar alguna de las licenciaturas de Matemáticas, Matemáticas Aplicadas o Actuaría y  $X_1 = 0$  en caso contrario, y  $X_2 = 1$  si tiene confianza en el profesor y  $X_2 = 0$  en otro caso.

Se verificó que los supuestos del modelo se cumplieran, no se rechazó que los residuos tienen distribución normal pues tanto la prueba de *Kolmogorov-Smirnov* como la de *Shapiro-Wilk* producen un  $p$ -valor de 0.000. La gráfica de los valores predichos por el modelo y las variables independientes no muestran problemas.

El modelo encontrado explica el 38.4 % de variación en la calificación de los estudiantes. No se rechaza que ambas variables conjuntamente deben estar en el modelo (p-value=0.001). Para cada variable se rechaza que su verdadero valor sea cero (p-value=0.001 y p-value=0.022). Con un nivel de confianza del 95 %, la calificación es entre 0.455 y 1.827 unidades más en los estudiantes que toman la carrera como primera opción comparado con los que no fue así. Por otro lado, la *Calificación en ME* es entre 0.137 y 1.685 unidades más en los estudiantes que tienen confianza en el profesor para resolver sus dudas de ME comparado con los que no la tienen.

## 4. Conclusiones y discusión

Hay muchas reflexiones que podemos hacer al respecto. Cómo mejorar los métodos de divulgación de la carrera para que los alumnos sepan a qué se enfrentarán cuando ingresan. También se debe apoyar a los alumnos de los primeros semestres con asesorías por parte de grupos de apoyo que siempre tengan espacios físicos adecuados. Otros factores que no hemos discutido y que parecen interesantes en próximas investigaciones están estudiar las desigualdades educativas a partir del análisis de los factores socioeconómicos y culturales que inciden en el ingreso al sistema educativo de nivel superior; ver Serrano y Guzmán (2010).

Las variables que ayudan a explicar la calificación que los alumnos obtuvieron en la materia de Matemáticas Elementales fueron *primera opción* y *confianza profesor*. Es razonable que estas variables estén en el modelo puesto que si los alumnos se encuentran cursando alguna de estas tres licenciaturas es por que les gusta y como consecuencia de esto los alumnos tienen más interés en las clases y en estudiar por su cuenta, de esta manera tienen un mejor aprovechamiento. La confianza que los alumnos le tienen al profesor es muy importante por que les permite preguntar dudas, hacer comentarios y expresar sus ideas sin temor a un comentario negativo del profesor hacia ellos o a que los exhiba en clase ya que estas son unas de las principales causas por las cuales los alumnos no resuelven sus dudas.

El resto de las variables no contribuyeron a explicar la calificación del estudiante. Un primer obstáculo es que son demasiadas preguntas y la mayoría de ellas contienen más de dos categorías lo que ocasiona que al entrar al modelo de regresión incrementa el número de parámetros.

## Bibliografía

- Capilla, J., Reyes, H. y Linares, G. (2010), “Proceso de enseñanza aprendizaje”, *Investigación para el Avance Educativo en Ciencias de la Computación, Dirección de Fomento Editorial* pp. 1–6.
- Hernández, S., Reyes, H., Ibarra, M. y Linares, G. (2010), “Proceso de enseñanza aprendizaje”, *VII Encuentro Participación de la mujer en la ciencia, CIO* pp. 1–10.
- Reyes, H., Cruz, D., Arenas, Y., Cruz, S., Godínez, F. y Ariza, F. (2011), “El uso de la regresión logística para conocer el gusto por las matemáticas en algunas preparatorias públicas de provincia en México”, *II Encuentro sobre didáctica de la estadística, la probabilidad y el Análisis de datos*, pp. 1–12.
- Serrano, O. y Guzmán, C. (2010), “Las oportunidades de ingreso a la licenciatura de la unam: Un análisis de regresión logística multivariado”, *Aportaciones y Aplicaciones de la Probabilidad y Estadística* pp. 117–130.
- Stephenson, G. y Gillel, W. (2010), *Statistical Package for the Social Sciences (SPSS)*, IBM SPSS Statistics 19 para Windows, <http://www.spss.org>.

# Estimación de totales y medias en el muestreo por bola de nieve y en presencia de probabilidades de ligas heterogéneas y asociadas con la variable respuesta\*

Martín H. Félix Medina<sup>a</sup>

*Facultad de Ciencias Físico-Matemáticas de la Universidad Autónoma de Sinaloa*

## 1. Introducción

El Muestreo por Bola de Nieve (MBN) es un método que se ha propuesto para muestrear poblaciones de difícil detección, tales como poblaciones de drogadictos, niños de la calle y trabajadoras sexuales. En este método se selecciona una muestra inicial de miembros de la población de interés, y se les pide a las personas que fueron seleccionadas que nombren a otros miembros de la población objetivo. A las personas que fueron nombradas se les puede pedir que nombren a otras personas, y el proceso de nombramiento puede continuar de esta manera hasta que se satisfaga alguna regla de terminación del muestreo.

Félix-Medina y Thompson (2004) desarrollaron una variante del MBN en la que la muestra inicial es una muestra por conglomerados de miembros de la población. Ellos propusieron estimadores máximo verosímiles (EMVs) del tamaño poblacional derivados bajo el supuesto de que la probabilidad de que un individuo sea nombrado por cualquier miembro de un conglomerado en la muestra inicial, la cual denominamos como probabilidad de liga, no depende del individuo nombrado, es decir, que las probabilidades de ligas son homogéneas. Félix-Medina *et al.* (2009) debilitaron el supuesto de homogeneidad y desarrollaron EMVs del tamaño poblacional bajo el supuesto de probabilidades de nominación heterogéneas.

---

\*Trabajo realizado con apoyo del proyecto PROFAPI-2011-25-08.

<sup>a</sup>mhfelix@uas.uasnet.mx

Félix-Medina y Monjardin (2010) consideraron el problema de estimar el total y la media de una variable respuesta, tal como gasto mensual en drogas, edad de inicio de consumo de drogas e ingreso mensual, y propusieron estimadores tipo Horvitz-Thompson de la media y el total derivados bajo el supuesto de probabilidades de ligas homogéneas. Posteriormente, Félix-Medina (2010) propuso estimadores de momentos de estos parámetros, los cuales obtuvo bajo el supuesto de probabilidades de ligas heterogéneas, pero no asociadas con los valores de la variable respuesta. Puesto que en situaciones reales los valores de la variable respuesta generalmente están asociados con las probabilidades de ligas, por ejemplo, individuos muy conocidos tienen mayores gastos en drogas, en este trabajo se proponen estimadores tipo Horvitz-Thompson del total y la media de una variable respuesta para el caso de probabilidades de ligas heterogéneas y asociadas con los valores de la variable respuesta.

## 2. Diseño muestral, notación y modelos probabilísticos

Al igual que en Félix-Medina y Thompson (2004), supondremos que una parte  $U_1$  de la población de interés  $U$  está cubierta por un marco muestral de  $N$  sitios  $A_1, \dots, A_N$ , tales como parques, hospitales o cruceros de calles. De este marco se selecciona una muestra aleatoria simple sin reemplazo de  $n$  sitios, que denotaremos por conveniencia como  $S_A = \{A_1, \dots, A_n\}$ , y que no significa que la muestra esté compuesta por los primeros  $n$  sitios del marco muestral. A las personas de la población de interés que pertenecen, de acuerdo con algún criterio, a cada uno de los sitios seleccionados se les pide que nombren a otros miembros de la población. Como convención, diremos que una persona y un sitio están ligados si cualquiera de los miembros de ese sitio la nombra.

Denotaremos por  $\tau$  el tamaño de  $U$ , por  $\tau_1$  el de  $U_1$ , por  $\tau_2 = \tau - \tau_1$  el de  $U_2 = U - U_1$ , y por  $M_i$  el número de personas en  $A_i$ . Obsérvese que  $\tau_1 = \sum_{i=1}^N M_i$  y que  $M = \sum_{i=1}^n M_i$  es el número de individuos en  $S_0 = \{\text{individuos en sitios } A_i \in S_A\}$ . Los conjuntos de variables  $\{X_{ij}^{(1)}\}$  y  $\{X_{ij}^{(2)}\}$  indicarán la existencia o no de ligas entre individuos y sitios. Así,  $X_{ij}^{(k)} = 1$  si la persona  $u_j \in U_k - A_i$  está ligada al sitio  $A_i$ , y  $X_{ij}^{(k)} = 0$  en otro caso,  $k = 1, 2$ .

Como en Félix-Medina *et al.* (2009), supondremos que las  $M_i$ s son variables aleatorias independientes con distribución Poisson y, que por carecer de información adicional, supondremos que tienen media común  $\lambda_1$ . Por tanto, dado que  $\sum_1^N M_i = \tau_1$ , la distribución condicional conjunta de  $(M_1, \dots, M_n, \tau_1 - M)$  es multinomial con parámetro de tamaño  $\tau_1$

y vector de probabilidades  $(1/N, \dots, 1/N, 1 - n/N)$ . Asimismo, supondremos que dado  $M_i$ , la distribución condicional de  $X_{ij}^{(k)}$  es Bernoulli con probabilidad  $p_{ij}^{(k)} = \Pr[X_{ij}^{(k)} = 1 | M_i] = \exp(\alpha_i^{(k)} + \beta_j^{(k)}) / [1 + \exp(\alpha_i^{(k)} + \beta_j^{(k)})]$ ,  $i = 1, \dots, n$ ;  $j = 1, \dots, \tau_k$ , con  $u_j \notin A_i$ , y  $k = 1, 2$ . Este modelo se conoce como modelo de Rasch. El parámetro  $\alpha_i^{(k)}$  es el efecto del potencial que tiene el sitio  $A_i$  de formar ligas con individuos en  $U_k$  y  $\beta_j^{(k)}$  es el efecto de la susceptibilidad que tiene el individuo  $u_j \in U_k - A_i$  de estar ligado a los sitios. Los efectos  $\beta_j^{(k)}$ s se suponen aleatorios con distribución normal con media cero y varianza  $\sigma_k^2 [N(0, \sigma_k^2)]$  desconocida.

De los supuestos anteriores se sigue que la probabilidad de que un individuo en  $U_k - S_0$  seleccionado al azar esté ligado sólo a los sitios  $A_i$ s con  $i \in \omega \subseteq \Omega = \{1, \dots, n\}$  es  $\pi_\omega^{(k)}(\sigma_k, \alpha^{(k)}) = \int \prod_{i=1}^n \{\exp[x_{\omega i}(\alpha_i^{(k)} + \sigma_k z)] / [1 + \exp(\alpha_i^{(k)} + \sigma_k z)]\} \phi(z) dz$ , donde  $x_{\omega i} = 1$  si  $i \in \omega$ , y  $x_{\omega i} = 0$  en otro caso,  $\alpha^{(k)} = (\alpha_1^{(k)}, \dots, \alpha_n^{(k)})$  y  $\phi(z)$  representa la función de densidad normal estándar.

### 3. Estimadores máximo verosímiles de los tamaños poblacionales

Félix-Medina *et al.* (2009) proponen estimar los parámetros  $\alpha^{(k)}$  y  $\sigma_k$  mediante EMVs condicionales  $\hat{\alpha}^{(k)}$  y  $\hat{\sigma}_k$  dado el número  $R_k$  de individuos en  $U_k - S_0$ ,  $k = 1, 2$ , que están ligados a algún sitio en la muestra. Los valores de estos estimadores los obtienen maximizando numéricamente la correspondiente función de verosimilitud condicional. Asimismo, sugieren los siguientes EMVs condicionales de  $\tau_1$  y  $\tau_2$ :  $\hat{\tau}_1 = (M + R_1) / [1 - (1 - n/N) \hat{\pi}_\emptyset^{(1)}(\hat{\sigma}_1, \hat{\alpha}^{(1)})]$  y  $\hat{\tau}_2 = R_2 / [1 - \hat{\pi}_\emptyset^{(2)}(\hat{\sigma}_2, \hat{\alpha}^{(2)})]$ , donde  $\hat{\pi}_\emptyset^{(k)}(\hat{\sigma}_k, \hat{\alpha}^{(k)})$  es un estimador de la probabilidad  $\pi_\emptyset^{(k)}(\sigma_k, \alpha^{(k)})$  de que un individuo de  $U_k - S_0$ , seleccionado aleatoriamente, no esté ligado a algún sitio  $A_i \in S_A$ . Finalmente, proponen estimar  $\tau$  mediante  $\hat{\tau} = \hat{\tau}_1 + \hat{\tau}_2$ .

### 4. Estimadores Horvitz-Thompson del total y la media poblacional

Supóngase que se desea estimar el total y/o la media de una variable respuesta  $y$ , tal como gastos médicos mensuales, gastos mensuales en drogas o edad de inicio de consumo de drogas. Sea  $y_j^{(k)}$  la variable respuesta asociada con el  $j$ -ésimo elemento de  $U_k$ ,  $j = 1, \dots, \tau_k$ ,  $k = 1, 2$ .

Luego,  $Y_k = \sum_{j=1}^{\tau_k} y_j^{(k)}$  y  $\bar{Y}_k = Y_k/\tau_k$  representan el total y la media del subconjunto  $U_k$ ,  $k = 1, 2$ . Similarmente,  $Y = Y_1 + Y_2$  and  $\bar{Y} = Y/\tau$  representan el total y la media de  $U$ . Como en el enfoque tradicional del muestreo, supondremos que los  $y_j^{(k)}$  son cantidades fijas, esto es, no aleatorias.

Las probabilidades condicionales de inclusión basadas en modelo dados los sitios  $A_i$  en  $S_A$  están dadas por  $\pi_j^{(1)}(\alpha^{(1)}, \sigma_1, \beta_j^{(1)}) = 1 - (1 - n/N) \prod_{i=1}^n (1 - p_{ij}^{(1)})$  si  $j \in U_1$  y  $\pi_j^{(2)}(\alpha^{(2)}, \sigma_2, \beta_j^{(2)}) = 1 - \prod_{i=1}^n (1 - p_{ij}^{(2)})$  si  $j \in U_2$ . Estas probabilidades no se conocen porque dependen de parámetros desconocidos. Una estimación de  $\pi_j^{(k)}(\alpha^{(k)}, \sigma_k, \beta_j^{(k)})$  se puede obtener reemplazando los parámetros por estimaciones de ellos. Estimadores de  $\alpha^{(k)}$  y  $\sigma_k$  se presentan en Félix-Medina *et al.* (2009). En este trabajo derivamos un estimador o predictor de  $\beta_j^{(k)}$  maximizando la función de densidad de probabilidad conjunta del vector  $\mathbf{X}_j^{(k)} = (X_{1j}^{(k)}, \dots, X_{nj}^{(k)})$  y el efecto aleatorio  $\beta_j^{(k)}$ . Este procedimiento conduce a que un estimador  $\hat{\beta}_j^{(k)}$  de  $\beta_j^{(k)}$  está dado como la solución de la siguiente ecuación:

$$\sum_{i=1}^n x_{ij}^{(k)} - \sum_{i=1}^n \frac{\exp[\hat{\alpha}_i^{(k)} + \beta_j^{(k)}]}{1 + \exp[\hat{\alpha}_i^{(k)} + \beta_j^{(k)}]} - \frac{1}{\hat{\sigma}_k^2} \beta_j^{(k)} = 0 \quad \text{si } j \in U_k - S_0, k = 1, 2, \text{ y}$$

$$\sum_{i \neq i'}^n x_{ij}^{(1)} - \sum_{i \neq i'}^n \frac{\exp[\hat{\alpha}_i^{(1)} + \beta_j^{(1)}]}{1 + \exp[\hat{\alpha}_i^{(1)} + \beta_j^{(1)}]} - \frac{1}{\hat{\sigma}_1^2} \beta_j^{(1)} = 0 \quad \text{si } j \in A_{i'} \in S_A, i' = 1, \dots, n.$$

Así, estimadores tipo Horvitz-Thompson (EHT) basados en modelo de  $Y_k$ ,  $k = 1, 2$ , y  $Y$  están dados por  $\hat{Y}_k = \sum_{j \in S_k^*} y_{kj} / \hat{\pi}_j^{(k)}(\hat{\alpha}^{(k)}, \hat{\sigma}_k, \hat{\beta}_j^{(k)})$ ,  $k = 1, 2$ , y  $\hat{Y} = \hat{Y}_1 + \hat{Y}_2$ . Similarmente, EHT de las medias  $\bar{Y}_k$  and  $\bar{Y}$  están dados por  $\hat{\bar{Y}}_k = \hat{Y}_k / \hat{\tau}_k$ ,  $k = 1, 2$ , y  $\hat{\bar{Y}} = \hat{Y} / \hat{\tau}$ .

## 5. Estudio Monte Carlo

Se consideraron dos poblaciones de  $N = 150$  valores de  $M_i$ s. En la Población I los valores se generaron mediante una distribución Poisson con media 8.0, mientras que en la Población II mediante una distribución Binomial negativa truncada en cero con media 8 y varianza 24. Las probabilidades  $p_{ij}^{(k)}$  de ligas se obtuvieron mediante el modelo Rasch descrito en la Sección 2 con  $\alpha_i^{(k)} = c_k / (M_i^{1/4} + 0.001)$ ,  $c_1 = -5.7$  y  $c_2 = -6.5$ . Los  $\beta_j^{(k)}$ s se generaron mediante la distribución  $N(0, 1)$ . Los valores de estos parámetros se fijaron de tal manera que los promedios de  $p_{ij}^{(1)}$  y  $p_{ij}^{(2)}$  fueron aproximadamente 0.04 y 0.03. Dado que se usó una muestra

**Cuadro 1:** Sesgos relativos y raíces cuadradas de errores cuadráticos medios relativos de estimadores de totales y medias. Resultados basados en 10000 muestras.

		Población I			Población II				
Esti- mador	Sesgo rel.	Población I			Población II			Esti- mador	Sesgo rel.
		$\sqrt{\text{ecm-rel.}}$	Sesgo rel.	$\sqrt{\text{ecm-rel.}}$	$\sqrt{\text{ecm-rel.}}$	Sesgo rel.	$\sqrt{\text{ecm-rel.}}$		
$\hat{Y}_1$	-0.03	0.08	-0.04	0.08	$\hat{Y}_1$	-0.03	0.04	-0.04	0.04
$\hat{Y}_2$	-0.16(35)	0.26(35)	-0.18(22)	0.25(22)	$\hat{Y}_2$	-0.06(35)	0.17(35)	-0.09(22)	0.15(22)
$\hat{Y}$	-0.06(35)	0.10(35)	-0.08(22)	0.10(22)	$\hat{Y}$	-0.05(35)	0.06(35)	-0.05(22)	0.06(22)
$\tilde{Y}_1$	-0.15	0.15	-0.18	0.19	$\tilde{Y}_1$	0.17	0.17	0.18	0.18
$\tilde{Y}_2$	-0.31	0.32	-0.33	0.34	$\tilde{Y}_2$	0.23	0.24	0.22	0.22
$\tilde{Y}$	-0.19	0.19	-0.22	0.22	$\tilde{Y}$	0.18	0.18	0.19	0.19
$\check{Y}_1$	-0.15	0.15	-0.18	0.19	$\check{Y}_1$	0.17	0.17	0.18	0.18
$\check{Y}_2$	-0.28	0.29	-0.30	0.31	$\check{Y}_2$	0.21	0.21	0.19	0.20
$\check{Y}$	-0.18	0.18	-0.21	0.21	$\check{Y}$	0.18	0.18	0.18	0.18

Notas: El número en paréntesis indica la cantidad de muestras en las cuales el estimador no se obtuvo.  $\hat{Y}_k$  y  $\hat{Y}_k$ , estimadores propuestos.  $\tilde{Y}_k$  y  $\tilde{Y}_k$ , así como  $\check{Y}_k$  y  $\check{Y}_k$ , estimadores propuestos por Félix-Medina y Monjardin (2010).

inicial de tamaño  $n = 15$ , las fracciones de muestreo en  $U_1$  y  $U_2$  fueron aproximadamente 0.4 y 0.3. El valor  $y_j^{(k)}$  de la variable de interés  $Y$  se generó mediante una distribución ji-cuadrada ( $\chi^2$ ) con 2 grados de libertad y parámetro de no centralidad  $\theta_j^{(k)} = 0.4 \exp(\beta_j^{(k)}) / [1 + \exp(\beta_j^{(k)})]$ . El uso de estas distribuciones condujo a que los valores  $y_j^{(k)}$ s y las probabilidades de inclusión fueran positivamente correlacionadas con coeficientes de correlación  $\rho_1 = 0.73$  y  $\rho_2 = 0.71$  para los elementos en  $U_1$  y  $U_2$ , respectivamente.

El estudio de simulación se realizó seleccionando 10000 muestras de cada una de las poblaciones mediante el diseño propuesto por Félix-Medina y Thompson (2004). Los estimadores que se consideraron fueron los estimadores  $\hat{Y}_k$ ,  $\hat{Y}$ ,  $\tilde{Y}_k$  y  $\tilde{Y}$  propuestos en este trabajo, y los dos conjuntos de estimadores  $\{\tilde{Y}_k, \tilde{Y}, \check{Y}_k, \check{Y}\}$  y  $\{\check{Y}_k, \check{Y}, \tilde{Y}_k, \tilde{Y}\}$  propuestos por Félix-Medina y Monjardin (2010), derivados bajo el supuesto de probabilidades de ligas homogéneas y basa-

dos en los EMVs, propuestos por Félix-Medina y Thompson (2004), y en los estimadores bayesianos, propuestos por Félix-Medina y Monjardin (2006), de los tamaños poblacionales, respectivamente. Los resultados del estudio se presentan en el Cuadro 1.

## 6. Conclusiones

De acuerdo con los resultados del estudio de simulación, los estimadores propuestos parecen trabajar aceptablemente bien cuando cada uno de los supuestos se satisfacen o cuando únicamente el supuesto de la distribución Poisson de los  $M_i$ s no se satisface. Estudios adicionales se requieren para determinar la robustez de los estimadores propuestos a desviaciones de otros supuestos. Los desempeños de los estimadores derivados bajo el supuesto de homogeneidad de las probabilidades de ligas no son buenos cuando este supuesto no se cumple.

## Bibliografía

- Félix-Medina, M. (2010), Estimación de totales y medias en el muestreo por bola de nieve en presencia de probabilidades de nominación heterogéneas, *en* “Memorias del XXV Foro Nacional de Estadística”, pp. 85–90.
- Félix-Medina, M. y Monjardin, P. (2006), “Combining link-tracing sampling and cluster sampling to estimate the size of hidden populations: a Bayesian assisted approach”, *Survey Methodology* **32**, 187–195.
- Félix-Medina, M. y Monjardin, P. (2010), “Combining link-tracing sampling and cluster sampling to estimate totals and means of hidden human populations”, *Journal of Official Statistics* **26**, 603–631.
- Félix-Medina, M., Monjardin, P. y Aceves-Castro, A. (2009), Link-tracing sampling: estimating the size of a hidden population in presence of heterogeneous nomination probabilities, *en* “Proceedings of the Section on Survey Research Methods of the American Statistical Association”, pp. 4020–4033.
- Félix-Medina, M. y Thompson, S. (2004), “Combining cluster sampling and link-tracing sampling to estimate the size of hidden populations”, *Journal of Official Statistics* **20**, 19–38.

# Metodología de observación de estados afectivos durante una situación de aprendizaje

Nery Sofía Huerta-Pacheco<sup>a</sup>, Genaro Rebolledo-Méndez<sup>b</sup>, Sergio Hernández-González<sup>c</sup>, Francisco Javier Landa-Torres<sup>d</sup>  
*Facultad de Estadística e Informática. Universidad Veracruzana*

## 1. Introducción

En el medio ambiente educativo existen diversos factores que influyen en los estados afectivos de los estudiantes, como por ejemplo, la falta de atención, la no comprensión de la información proporcionada, inquietud por aprender e interpretar la información, entre otros. Estos temas se consideran como efecto de los Estados Afectivos Observables (EAO) que han sido identificados como relevantes en un entorno educativo por computadora Craig et al. (2004). Una posible causa para la aparición de EAO no conducentes, como el aburrimiento, son representados por factores como: falta de transmisión de información en forma clara, información proporcionada en forma repetitiva, actividades que limiten el aprendizaje de los estudiantes o bien, en casos extremos, índices de reprobación por falta de comprensión de los temas de las asignaturas, que tal vez en el futuro esto conlleve a una problemática mayor como la deserción. Sin embargo, al identificar los EAO de los estudiantes durante su desarrollo educativo, se podrá aportar un indicador a los profesores o responsables de asignaturas, de tal forma que se identifique de manera más precisa los comportamientos afectivos presentados en un grupo de estudiantes. Con la finalidad de que estos desarrollen, implementen y utilicen diversas técnicas de apoyo durante el proceso de enseñanza.

---

<sup>a</sup>nehuerta@uv.mx

<sup>b</sup>grebolledo@uv.mx

<sup>c</sup>sehernandez@uv.mx

<sup>d</sup>fco.j.landa@gmail.com

Este trabajo tiene como objetivo comparar la mejora de los observadores independientes, los cuales son dos personas ajenas a los sujetos observación con características similares como la edad, el género, nivel de estudios, entre otros; de acuerdo con la capacitación que se les proporcionó, para obtener fuerzas de concordancias buenas Cohen (1968) y así poder identificar si a mayor porcentaje de Estados Afectivos Observables Conducentes (EAOC) se obtienen mayores Ganancias de Aprendizaje (GA). La relevancia de los EAO en el aprendizaje se debe a que la afectividad es un factor que puede ayudar a la educación, utilizando tecnología en distintos ámbitos. A través de la identificación de EAO se podrá contribuir a la detección precisa de la actitud que tenga un estudiante durante el uso de tecnología educativa y esto conlleve a proponer estrategias para atacar a los EAO que no sean conducentes a un mejor aprendizaje.

## 2. Materiales y métodos

### 2.1. Metodología

En este estudio participaron 28 estudiantes seleccionados a través de un muestreo estratificado proporcional de la asignatura Experiencia Recepcional considerados porque presentaban características similares (edad, nivel de conocimientos y grado de avance) de la Licenciatura en Informática de la Universidad Veracruzana, campus Xalapa, Veracruz, México, a los cuales se les solicitó su participación mediante una forma de consentimiento.

El estudio consistió en que los estudiantes interactuaron con el videojuego de agilidad mental llamado “Brain Challenge”, Gameloft (2011), el cual consta de cinco fases: memorización, visualización, lógica, atención y cálculo. Previamente, a los estudiantes se les explicó cómo utilizar el videojuego y los tiempos que debería de cumplir.

A cada estudiante se les colocaron dos sensores: uno que mide la actividad electrodérmica, el “Q Sensor” de Afectiva, mientras que el otro mide la actividad neuronal, el “MindSet” de NeuroSky.

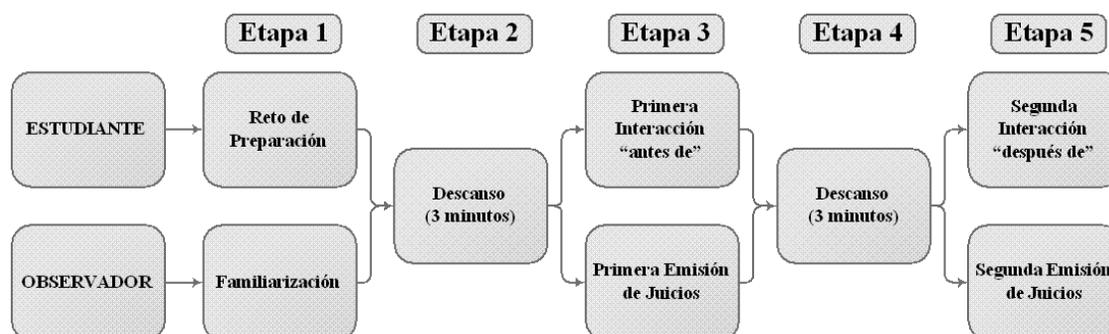
En dicho estudio se llevó a cabo en dos evaluaciones en el ambiente de desarrollo educativo. Una, donde los estudiantes no tenían conocimiento previo de cómo interactuar con el juego (.antes de”) y la segunda, donde los estudiantes ya habían interactuado con el videojuego (”después de”). El lapso de tiempo entre interacciones fue de 20 minutos aproximadamente.

Se contó con dos observadores, los cuales emitieron juicios, de acuerdo con cinco EAO (aburrimiento, frustración, confusión, flujo y neutral) de los estudiantes. Estos juicios afectivos fueron registrados, mediante la observación de videograbaciones y videollamadas. La metodología de observación se basó en trabajos anteriores (ver Craig et al. (2004) y D’Mello et al. (2006)), los cuales consideran que los periodos de observación adecuados para emitir juicios son durante lapsos de 20 segundos.

Durante la emisión de juicios de las videograbaciones, los observadores registraban los EAO al final de cada intervalo de tiempo, donde se le proporcionaba un periodo de cinco segundos para efectuar su opinión acerca del estado afectivo observado. Existen cinco EAO que han sido identificados como relevantes en situaciones de aprendizaje con tecnología educativa: flujo, eureka, frustración, aburrimiento y confusión (ver D’Mello et al. (2006) y McDaniel et al. (2007)).

Sin embargo, en estudios más actuales Rodrigo et al. (2009), se identifica un cambio en la metodología, donde se mantienen los mismos periodos de tiempo ya establecidos para la emisión de juicios, pero con cambios en los estados afectivos a observar, donde se toman a consideración a los estados de aburrimiento, frustración, confusión y flujo como relevantes en la interacción individuo-computadora y se toma en cuenta a un estado neutral para la ausencia de las reacciones afectivas.

Los periodos de emisión de juicios de los observadores con respecto a las acciones de los estudiantes a evaluar, se llevaron a cabo como lo muestra la Figura 1.



**Figura 1:** Diagrama de Periodos de Observación

Se utilizaron dos medios distintos para la reproducción de rasgos faciales, para así poder registrar los juicios que el observador emitía, los cuales fueron los siguientes:

1. EN VIVO, este medio se llevó acabo a través de una videollamada por Skype<sup>1</sup> emitida por la computadora donde el estudiante interactuaba. Se emitían juicios por los dos observadores al mismo tiempo en otra computadora, donde se consideró el tiempo del contador de la llamada para determinar los periodos de 20 segundos.
2. VIDEOGRABACIÓN, se emitieron juicios a través de una grabación editada en lapsos de 20 segundos con cinco segundos de interrupción de los estudiantes interactuando con el videojuego, los cuales permitieron la evaluación y emisión de juicios tomando en cuenta la hora del video sincronizado con la interacción.

Cabe mencionar que se registraron los juicios emitidos por los dos observadores, en dos etapas distintas: 1) Cuando los observadores no estaban capacitados y 2) Cuando se les había dado capacitación. También se calcularon los índices de kappa Cohen (1968) para determinar la concordancia entre los dos observadores (antes y después de la capacitación), tablas de contingencia de EAO para ver la frecuencia de cada uno con respecto a sus Ganancias de Aprendizaje y un análisis de varianza para distinguir las diferencias entre los puntajes obtenidos antes y después de las interacciones.

### 3. Resultados

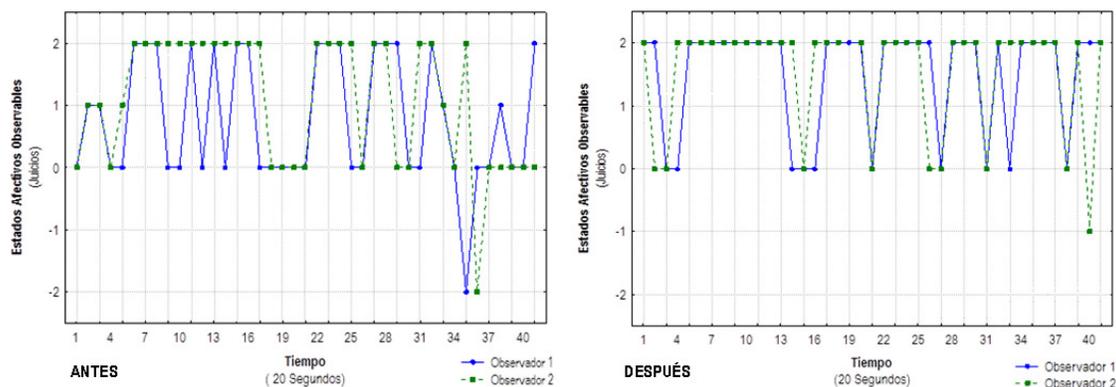
La comparación de los grados de acuerdos entre los dos observadores independientes durante las etapas de evaluación de los 28 estudiantes, mostraron que después de la capacitación existió un incremento considerable en el acuerdo observado emitido referente a los estados afectivos. Asimismo se identificó que existieron diferencias significativas ( $p > 0.05$ ) entre el aprendizaje inicial y final de los estudiantes.

Se observa que la concordancia entre los dos observadores antes de la capacitación se consideraron moderadas con un máximo .56 y con un acuerdo observado del 70.83 %. Asimismo se identificó que elK índice máximo de kappa de los observadores ya capacitados fue de .53 con un acuerdo observado de 82.50 %. (Ver Figura 2)

Con base en lo anterior, se observó que aquellos estudiantes con mayor grado de aprendizaje (392 a 478 puntos) antes de la capacitación fueron aquellos que tuvieron los índices

---

<sup>1</sup>Skype: Software que permite la comunicación vía SMS, voz y video con tecnología VoIP entre dos o más usuarios a través de la red.



**Figura 2:** Concordancia entre los observadores antes y después de la capacitación

de kappa mayores (.48 al .56) al igual que los acuerdos observados (68.29% al 70.83%) y presentaron un mayor porcentaje de estados afectivos conductores ( $EAO > 25\%$ ).

Por otro lado, después de la capacitación se identificó que los estudiantes con mayor grado de aprendizaje (671 a 750 puntos) fueron aquellos que tuvieron los mayores índices de kappa (.53 al .55) al igual que los acuerdos observados (77.78% al 82.50%) y presentaron un 58% de Estados Afectivos Observables Conductores. Cabe mencionar que en la identificación de estados afectivos observables no se presentó distinción por género, por lo cual consideró a los sujetos de estudio como homogéneos.

## 4. Conclusiones

Se observó que los índices de kappa obtenidos antes y después de la capacitación no se consideraron simbólicos y suficientes para representar los EAO de los estudiantes durante esta investigación, dado que al no obtener índices de kappa mayores a 0.60 no se consideran fuerzas de concordancias adecuadas. Cabe destacar que, a pesar de las condiciones de similitud entre los sujetos de estudio, el acuerdo observado después de la capacitación de los dos observadores fue mayor a los índices de kappa obtenidos en el estudio preliminar. Esto quiere decir que los observadores, al conocer la definición de los EAO, obtuvieron mayor número de concordancias de los EAO de los juicios que emitieron y que los juicios emitidos descartaron la presencia de estados afectivos no conductores.

Finalmente se identificó que los estudiantes que obtuvieron mayor grado de aprendizaje presentaron mayor porcentaje de EAO conducentes emitidos por dos observadores capacitados, por lo cual se concluye que a mayor porcentaje de Estados Afectivos Observables Conducentes (EAOC) se obtuvieron mayores Ganancias de Aprendizaje (GA) de la interacción de los estudiantes de la Licenciatura de Informática en el periodo agosto-diciembre 2010, durante la utilización del videojuego de agilidad mental.

Como trabajos a futuro se contemplará la aplicación de este estudio en una población más grande y así poder reconocer patrones de comportamiento definidos por estudiante o usuario de la tecnología en una situación en específico. Por consiguiente, también se mejorará la metodología de capacitación de los EAO ya planteada, para sí obtener fuerzas de concordancias buenas o muy buenas, que puedan ser representativas.

## Bibliografía

- Cohen, J. (1968), “Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit”, *Psychological Bulletin* **70**, 213–220.
- Craig, S., Graesser, A., Sullins, J. y Gholson, B. (2004), “Affect and learning: an exploratory look into the role of affect in learning with autotutor”, *Journal of Educational Media* **29**.
- D’Mello, S., Craig, S. y Graesser, A. (2006), “Multi-method assessment of affective experience and expression during deep learning, running head: Multi-method assessment of affect”, *University of Memphis* pp. 1–42.
- Gameloft (2011), “Brain challenge”, <http://www.gameloft.com>.
- McDaniel, B., D’Mello, S., King, B., Chipman, P., Tapp, K. y Graesser, A. (2007), “Facial features for affective state detection in learning environments”.
- Rodrigo, M., Baker, R., Jadud, M., Amarra, A., Dy, T., Espejo-Lahoz, M., Lim, S., Pascua, S., Sugay, J. y Tabanao, E. (2009), “Affective and behavioral predictors of novice programmer achievement”, *ITiCSE’09*.

# Desarrollo de un módulo de pruebas de Estadística No Paramétrica con Java y R

Francisco Javier Landa-Torres<sup>a</sup>, Sergio Hernández-González<sup>b</sup>, Genaro Rebolledo-Méndez<sup>c</sup>, Nery Sofía Huerta-Pacheco<sup>d</sup>  
*Facultad de Estadística e Informática. Universidad Veracruzana*

## 1. Introducción

El proceso estadístico para el análisis de datos puede resultar complejo cuando se realiza de forma manual y más cuando es necesario analizar una gran cantidad de información, en donde el control y la organización de la información se denota demasiado compleja, y esto aparenta una imposibilidad en el procesamiento y el análisis de los datos. Actualmente existe un software que tiene la capacidad de analizar y procesar funciones estadísticas contando con entradas para el análisis de una importante cantidad de datos y mostrando los resultados en un lapso de tiempo corto gracias a las características de una computadora, teniendo confiabilidad y rapidez en sus tareas estadísticas y ganándose un brillante prestigio, sin embargo algunos de ellos pueden estar disponibles con la adquisición de licencias costosas provocando estancamiento o límites para su uso.

R, se considera un lenguaje de programación estadístico descendiente de S en su sintaxis, que a su vez la base de su código está en C, C++ y Fortran. Después de su lanzamiento obtuvo tanto éxito que su utilidad no fue orientada solamente al campo estadístico, sino que se expandió a la economía, medicina, geografía y por supuesto la informática, por mencionar algunos más. Este lenguaje está posicionado en el lugar 27 dentro del ranking de programación en noviembre del 2011 que publica The TIOBE Software mensualmente, mientras

---

<sup>a</sup>fco.j.landa@gmail.com

<sup>b</sup>sehernandez@uv.mx

<sup>c</sup>grebolledo@uv.mx

<sup>d</sup>sofia.huerta.p@gmail.com

que otros lenguajes de la misma naturaleza como MATLAB y SAS se encuentran en los lugares 15 y 35 respectivamente. El lenguaje y el ambiente de desarrollo de R es de licencia GNU, es decir que es código abierto y software libre; sin embargo, el lado negativo que tiene este lenguaje estadístico es que su mayor funcionalidad se encuentra en modo de consola, pudiendo ejecutarse todas sus instrucciones ingresando comandos, presentando al usuario una interfaz poco amigable para su uso, por lo cual la consulta de fuentes de sintaxis puede llegar a ser una pérdida de tiempo; ver Barrios-Zamudio (2010) y Vienna University of Economics and Business (2011).

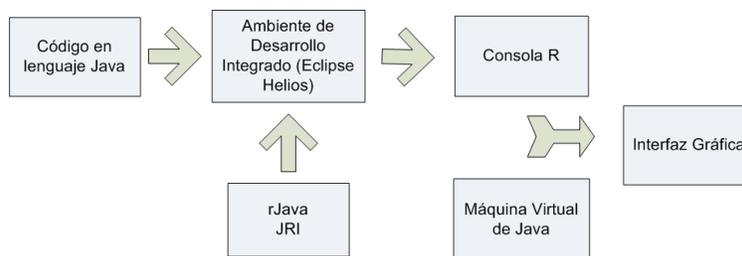
El objetivo de este artículo es desarrollar una interfaz gráfica aprovechando los beneficios de R, implementando una sencilla prueba de Estadística No Paramétrica en combinación con la tecnología de Java ORACLE (2010), el cual es un lenguaje de programación de código abierto que cuenta con librerías para desarrollar interfaces gráficas amigables para el usuario, la finalidad del programa es utilizar los beneficios del procesamiento de funciones estadísticas que ofrece R siendo así manipulados por elementos gráficos creadas con Java en complementación con una herramienta de comunicación rJava, logrando el intercambio de instrucciones entre los dos lenguajes mencionados. La importancia en el enfoque de desarrollo a Estadística No Paramétrica se debe a que existen pocos sistemas con la capacidad de realizar funciones de este tipo, al menos no la totalidad, provocando que se dificulte su análisis y tal vez descartando una importante base de datos con información que ofrezcan resultados importantes.

## 2. Metodología

rJava es una herramienta que es posible utilizarse con Java, la cuál es una interfaz que habilita un puente de comunicación entre Java y R, creando objetos y llamando métodos adecuados para su comunicación vía JNI (forma de comunicación que se realiza con base en el lenguaje C). Existen dos vías de comunicación entre los dos lenguajes mencionados, la primera es con JRI (Java R Interface) que es la librería que logra la comunicación de Java a R, mientras que rJava lo hace en sentido contrario, pero las dos interfaces se encuentran incluidas en el proyecto que lleva por nombre rJava y este paquete es utilizado actualmente en los programas iPlots, JavaGD, JGR y JavaStat. Esta herramienta está disponible tanto para Linux, como Mac y Windows, teniendo la capacidad de poderse desarrollar aplicaciones

para estas tres plataformas; ver Xue y Harner (2001) y RForge.net (2011).

El funcionamiento para este prototipo está basado en el código Java, el cuál se encuentra codificado dentro de Eclipse Helios<sup>1</sup> (The Eclipse Foundation (2011)) que tendrá como función absorber el funcionamiento de las librerías que proporciona rJava para que se logre el intercambio de mensajes entre los dos lenguajes con código Java. Una vez unido lo anterior, Eclipse es encargado de crear la liga funcional entre los dos software que logra el funcionamiento de ambos lenguajes: la consola de R y la Máquina Virtual de Java quienes serán los encargados de procesar los mensajes que pueden transferirse entre ellos. Logrando esto, es posible la traducción entre los dos motores de procesamiento y generar, a través de la Máquina Virtual de Java, una serie de interfaces gráficas que muestren al usuario los resultados que procesó la consola R. En la Figura 1 se muestra el esquema de funcionamiento descrita anteriormente; ver Ceballos (2008) y Ceballos (2010).



**Figura 1:** Proceso de desarrollo del módulo en Java y R

Una vez descrito el proceso de comunicación, se muestra su implementación con el desarrollo de una técnica de Estadística No Paramétrica que es la prueba del signo para confirmar su utilidad.

El prototipo se desarrolló creando una ventana simple que traspase los comandos mostrados de la consola de R a la ventana creada por Java. La prueba del signo se codificó en un archivo script de R. La prueba de los signos se utiliza para hacer pruebas de hipótesis acerca de la mediana de una población. La prueba estadística está basada en la distribución Binomial con probabilidad de éxito  $p=1/2$ , puesto que la probabilidad de que un dato sea mayor o menor que la mediana es  $1/2$ . Para calcularla se determinan las diferencias de los

<sup>1</sup>Ambiente de Desarrollo Integrado (siglas en Inglés IDE) que permite interpretar y compilar el código Java.

datos con respecto al valor dado de la mediana y se cuentan los signos positivos y negativos. Si la hipótesis alterna es “menor que” y el número de diferencias positivas es mayor que el número de diferencias negativas entonces “valor-p” =  $P_2$ , en caso contrario “valor-p” =  $P_1$ . Cuando la hipótesis alterna es de dos lados y el número de diferencias positivas son mayores que el número de diferencias negativas entonces el “valor-p” =  $2P_2$ , si hay menor número de diferencias positivas entonces “valor-p” =  $2P_1$ , y si hay igual número de diferencias positivas y negativas entonces, “valor-p” = 1. Si  $n > 20$  se puede usar aproximación Normal a una Binomial con  $p = q = 0.5$ , para calcular los “valores-p”. Es decir:  $Z \frac{X - .5n}{.5\sqrt{n}}$ .

Dentro de la prueba realizada se siguió el patrón manual para la resolución de la prueba, con la finalidad de que se mostraran los procesos ejecutados para el análisis desde el ingreso de los datos a una matriz (se pueden incrementar o disminuir tanto las columnas como las filas a placer del usuario). Por consiguiente se realiza el proceso de comparación de la información con el valor de la mediana deseada y automáticamente se crea una matriz nueva con las diferenciaciones y las asignaciones de signos correspondientes a cada caso. Asimismo, se realiza un conteo de signos positivos y negativos, y finalmente se ejecuta una suma de probabilidades (valor-p), el cual nos dará razón de aceptación o rechazo de hipótesis. (Ver Figura 2.)

El resultado de la compilación es una ventana de Java que muestra la información de inicio de la consola de R como bienvenida, pero además muestra los datos que fueron analizados por la prueba del signo importados desde el script, además en la pantalla se incluyen los resultados de la prueba.

### 3. Conclusiones

La fusión de Java y R como lenguajes de programación podrían ser una muy buena herramienta para la creación de un software libre complementando rJava que demuestra ser la herramienta de programación ideal que permite la conexión entre los dos lenguajes interesantes dentro de su propia área, impulsando uno al crecimiento del otro, donde su utilización sea simple y reduzca los riesgos de errores, además de ser utilizable; aunque la desventaja de ésta es que su manual de uso es un tanto complejo y su lectura podría retrasar el proceso de desarrollo de un sistema con base en estas tecnologías.

El desarrollo de este módulo puede dejar ver una arquitectura en conjunto con los



**Figura 2:** Pantallas de Proceso

dos lenguajes mencionados, haciendo una fusión bastante atractiva para su desarrollo. La descripción de la arquitectura consiste en tener como motor de funcionamiento la consola de R que se encarga de la realización de las funciones estadísticas; la interfaz que logra interpretar y presentar la información al usuario sería la tarea que desempeñaría Java y en conjunto construir una herramienta que también pueda ser interesante para la muestra de información de series de tiempo con una representación en movimiento, pues R permite enviar datos en bruto de una gráfica y Java interpretarla para crear una gráfica dinámica y obtener mejores detalles para su interpretación.

El desarrollo de una propuesta de software con estas tecnologías permite la creación de un paquete estadístico orientado a las funciones no paramétricas, debido a que existe poco software que implemente la mayoría de éstas, teniendo a veces que desartar en el análisis

o solamente buscar el software apropiado que realice dicha función. El desarrollo de este software podría ejecutarse en los sistemas operativos más populares actualmente, también ser un sistema completamente gráfico que facilite el ingreso de variables, la selección de parámetros y arrojar la gráfica de forma interesante; pero la característica más importante es la depuración de información basura en memoria, la cual son datos que utiliza el programa y las almacena para realizar una función, pero cuando el mismo programa identifique que la información ya no es necesaria, el Recolector de Basura, se encarga de limpiar esa información innecesaria agilizando así el rendimiento y acelerando el procesamiento de los datos.

El desarrollo de este sistema se tiene contemplado también para su incorporación como herramienta de aprendizaje de estadística en la carrera de Ciencias y Técnicas Estadísticas en la Universidad Veracruzana, ya que se fomenta la utilización de R para el análisis de datos, pero la población que ocupe el sistema propuesto se tendrá como base para el perfeccionamiento de este sistema y posteriormente liberarlo de manera pública.

## Bibliografía

- Barrios-Zamudio, E. (2010), “R: Un lenguaje para análisis de datos y graficación”.
- Ceballos, F. (2008), *Java 2: Interfaces gráficas y aplicaciones para Internet*, Editorial RA-MA, Madrid, España.
- Ceballos, F. (2010), *Java 2: Curso de programación*, Editorial RA-MA, Madrid, España.
- ORACLE (2010), “Java 2 SDK, Standard Edition Documentation”, <http://docs.oracle.com/javase/1.4.2/docs/>.
- RForge.net (2011), “rJava, low-level R to Java interface”, <http://www.rforge.net/rJava>.
- The Eclipse Foundation (2011), “Eclipse Helios is here!”, <http://www.eclipse.org/helios/>.
- Vienna University of Economics and Business (2011), “The R project for statistical computing”, <http://www.r-project.org>.
- Xue, H. y Harner, E. (2001), JavaStat: A distributed Statistical Computing Enviroment, en “2nd International Workshop on Distributed Statistical Computing”.

# Modelo para estimar el riesgo de inundaciones en la cuenca Grijalva Usumacinta

Ana Laura Reyes Olvera<sup>a</sup>  
*Colegio de Postgraduados*

Eduardo Gutiérrez González  
*Instituto Politécnico Nacional*

## 1. Introducción

Tabasco es uno de los estados de la República Mexicana que presenta con mayor frecuencia inundaciones en los municipios de la cuenca Grijalva Usumacinta. Por esta razón en el presente trabajo se propone un modelo basado en los índices de riesgo de inundaciones.

Existen diferentes formas de estudiar las inundaciones, Hunt (2002) analiza las inundaciones como un desastre natural en el que se debe identificar los peligros y la vulnerabilidad para poder superarlos. El CENAPRED (Centro Nacional de Prevención de Desastres) por el contrario no maneja los desastres como naturales sino como producto de las condiciones de vulnerabilidad y exposición de las comunidades. Pero en general todas las formas de estudiar las inundaciones llevan al riesgo de inundaciones. El riesgo abarca dos aspectos: Peligro y Vulnerabilidad (Mileti, 1999). En el peligro se deben obtener las probabilidades de que ocurra un fenómeno que pueda dañar lo expuesto (Ordaz, 1996). La vulnerabilidad está compuesta de la exposición y susceptibilidad (daño potencial y pérdida) (Begum *et al.* 2007), por tales razones Blaikie *et al.* (1994) analiza las condiciones socio-económicas que aumentan el grado de vulnerabilidad.

Las inundaciones se presentan cuando se registran valores extremos, que pueden estudiarse de diferentes formas. Por ejemplo, Young (2002) combina los métodos estadísticos y

---

<sup>a</sup>reyes.ana@colpos.mx

determinísticos para analizar datos históricos y predecir los niveles de inundación y de los ríos, mientras que los investigadores Bayliss & Reed (2001) discuten diferentes formas de estudiar las inundaciones históricas y revisan métodos para incorporar los datos históricos al análisis de frecuencias de inundaciones. En el trabajo se utiliza la teoría de valores extremos presentada en 1928 por Fisher & Tippett quienes derivan la forma límite de la distribución del valor máximo de una m. a., encontrando tres posibles distribuciones (DVE): Gumbel, Weibull y Fréchet. A partir de esto se desarrollan aportaciones de interés tales como las de Gumbel (1934), quien presenta el primer libro de importancia (Gumbel, 1958). En 1943 Gnedenko presenta una prueba rigurosa del Teorema de DVE propuestos por Fisher & Tippett. Otra aportación la dio Von Mises (1936) y Jenkinson (1955) que propusieron una distribución que combina las tres familias de DVE, la distribución de valores extremos generalizada (DVEG). De 1990-2011 el desarrollo estadístico ha sido sustancial en el área, por mencionar algunos, Tawn (1992), Rosbjerg & Madsen (1996), Coles & Dixon (1999) y Coles (2001).

## 2. Marco teórico

La definición de riesgo (R):  $R = C \times V \times P$ , en donde  $C$ – valor de los bienes expuestos,  $V$ – vulnerabilidad y  $P$ – peligro, fue propuesta por Ordaz (1996).

**Peligro.** La probabilidad de ocurrencia de eventos extremos calculada bajo la teoría de valores extremos será utilizada como el peligro.

**Periodos de retorno.** Si  $P(X_i > u) = p$ , su periodo de retorno es  $p^{-1}$ . Entonces para el evento  $\{X > x\}$  el periodo de retorno es:

$$\tau_x = \frac{1}{1 - F(x)}.$$

**Vulnerabilidad.** La vulnerabilidad  $V$  es una medida de qué tan susceptible es un bien expuesto a la ocurrencia de un fenómeno perturbador.

**Costo.** El costo  $C$  mide la cuantía de lo que es susceptible de afectarse durante la ocurrencia de un fenómeno perturbador.

Obtenida la medida del riesgo se recomienda el uso de mapas de riesgo para ubicar las zonas por categorías y desarrollar un plan de manejo (Tawatchi & Mohammed, 2005).

### 3. Método propuesto

Se calculan los tres factores del riesgo de inundaciones y se trazan los mapas de riesgo.

**Peligro.** Se propone una prueba Bootstrap para determinar el ajuste de las observaciones de los valores extremos a una distribución DVE y DVEG.

Sea  $x_1, x_2, \dots, x_k$  valores de alguna variable aleatoria  $X$ , para este caso particular corresponden a un periodo de un año de  $n$  observaciones formando los vectores  $\mathbf{x}^{(1)} = (x_1^1, \dots, x_n^1) \dots \mathbf{x}^{(k)} = (x_1^k, \dots, x_n^k)$ . Calculando los máximos por vector se aplica a éstos la prueba Bootstrap propuesta y se elige el mejor modelo y con éste se calculan sus respectivos periodos de retorno.

**Vulnerabilidad.** Para la vulnerabilidad se usan dos variables: los servicios que tiene la comunidad (energía eléctrica, agua potable y drenaje) y sus recursos hidrológicos (ríos). Se construye un índice de servicios y uno de ríos, mediante la teoría de componentes principales del análisis multivariado.

**Costo.** Se construye un indicador de Costo con dos factores: los bienes con los que cuenta la población y el número de habitantes por comunidad. El número de habitantes se considera porque “A mayor población, mayor riqueza”. Para complementar se construye un índice de bienes con las variables: ocupantes que disponen de televisión, computadora, refrigerador, lavadora y automóvil.

### 4. Resultados

Los datos de precipitación fueron proporcionados por CONAGUA. Para obtener la base de datos por municipio se delimitó una región, para esto se consideró la información de las subcuencas que forman parte de cada municipio y se tomó en cuenta aquellas que la aportación de su caudal influyó en el municipio. A las series de cada municipio se les aplicó el método propuesto, resultando un mejor ajuste con una distribución Gumbel  $F(x) = \exp \left\{ - \exp \left[ - \left( \frac{x-\mu}{\sigma} \right) \right] \right\}$ ,  $x \in \mathbb{R}$  y VEG  $F(x) = \exp \left\{ - \left[ 1 + \xi \left( \frac{x-\mu}{\sigma} \right) \right]^{-\frac{1}{\xi} +} \right\}$ . En los cuadros 1 y 2 se muestran las estimaciones del municipio de Balancán. En el cuadro 3 se tienen los resultados del índice de Riesgo de Inundación Propuesto (IRP) con un nivel de retorno (nr) de 150mm y en la figura 1 el mapa de riesgo.

## 5. Conclusiones

El mapa de riesgo que se obtuvo fue contrastado con los resultados históricos de las inundaciones en los 17 municipios de Tabasco, lo obtenido resultó acorde a los registros. En el cuadro 4 se muestra el comparativo del índice de riesgo propuesto (IRP) y los registros históricos (RH). Se puede concluir que los factores que fueron incluidos en la construcción de los índices fueron correctos, así como el uso de los componentes principales para obtener las ponderaciones adecuadas para cada uno de los factores considerados.

## Bibliografía

- [1] Bayliss, A. C. y Reed, D. W. (2001). The use of historical data in flood frequency estimation. Walingford: Report to MAFF. Center for Ecology and Hidrology.
- [2] Begum, S., Merz, B., Thiekem, A. H. y Gocht, M. . (2007). Flood risk Management in Europe. Springer , 231–251.
- [3] Blaikie P., Wisner B., Davis I. y Cannon T. (1994). At Risk. Natural Hazards, people´s vulnerability and disasters. New York: Routledge, London.
- [4] Centro Nacional de Prevención de Desastres. (2006). Guía Básica para la Elaboración de Atlas Estatales y Municipales de Peligros y Riesgos. México.
- [5] Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values. London: Springer-Verlag.
- [6] Coles, S. y Dixon, M. (1999). Likelihood-Based Inference for Extreme Values Models Extremes. Journal of the Royal Statistical Society, 2(1), 5–23.
- [7] Fisher, R. A. y Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest memeber of a sample. Proceedings of the Cambridge Philosophical Society, 24, 180–190.
- [8] Gnedenko, B. (1943). Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire. The Annals of Mathematics, 44(3), 423–453.

- [9] Gumbel, E. J. (1934). Les moments des distribution finales de la première et de la dernière valeur. *Comptes Rendus de l'Académie des Sciences*, 198, 141–143.
- [10] Gumbel, E. J. (1958). *Statistics of extremes*. New York: Columbia Univ. Press.
- [11] Hunt, J. C. (2002). Floods in a Changing Climate: A Review. *Philosophical Transaction: Mathematical, Physical and Engineering Sciences*, 1531–1543.
- [12] Jenkinson, A. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quart. J. Roy. Meteor. Soc.*, 81, 158–171.
- [13] Mileti, D. E. (1999). *Disasters by design. A reassessment of natural hazard in the United States*. Washington, DC: Joseph Henry Press.
- [14] Ordaz, M. (1996). Algunos Conceptos del Análisis de Riesgo. *CENAPRED*, 14, 6–12.
- [15] Rosbjerg, D. y Madsen, H. (1996). The role of regional information in estimation of extreme point rainfall. *Atmos. Res.*, 42, 113–122.
- [16] Tawatchi, T. y Mohammed, F. K. (2005). Flood Hazard and risk analysis in the south-west region of Bangladesh. *Hydrological Processes*, 19, 2055–2069.
- [17] Tawn, J. A. (1992). Estimating Probabilities of Extreme Sea-Levels. *Journal of the Royal Statistical Society*, 41(1), 77–93.
- [18] Von Mises, R. (1936). La distribution de la plus grande de  $n$  valeurs. *Revue Mathématique de l'Union Interbalkanique (Athens)*, 1, 141–160.
- [19] Young, P. C. (2002). Advances in real time flood forecasting. *Philosophical Transaction: Mathematical, Physical and Engineering Sciences*, 1433–1450.

Estimación de parámetros, Balancán					
Distribución	Escala	Forma	Localidad	Log-veros	Criterio de Akaike
Gumbel	31.992	.	117.654	-236.491	476.534
Veg	32.384	0.121	136.028	-235.987	475.432

Cuadro 1: Estimación de los parámetros de la DVEG.

Intervalos de Confianza al 95 %			Periodo de Retorno, Balancán		
	VEG	Gumbel	mm	VEG	Gumbel
Escala	(21.002, 34.678)	(27.172, 31.564)	100	1.18	1.17
Forma	(-0.092, 0.351)	.	120	1.69	1.59
Localidad	(108.315, 126.274)	(110.456, 128.139)	150	3.42	3.19

Cuadro 2: Intervalos de confianza y Periodos de Retorno.

<i>i</i>	Municipios	C	V	P	IR
1	Balancán	2.514	0.221	0.285	0.158
2	Cárdenas	2.955	0.556	0.290	0.476
3	Centla	2.624	0.664	0.374	0.652
4	Centro	3.403	0.523	0.345	0.614
5	Comalcalco	2.866	0.366	0.241	0.253
6	Cunduacán	2.757	0.632	0.316	0.550
7	Emiliano Zapata	2.571	0.065	0.481	0.080
8	Huimanguillo	2.770	0.533	0.718	1.000
9	Jalapa	2.577	0.405	0.555	0.580
10	Jalpa de Méndez	2.734	0.373	0.421	0.429
11	Jonuta	2.313	0.138	0.322	0.103
12	Macuspana	2.806	0.564	0.297	0.470
13	Nacajuca	2.901	0.166	0.425	0.205
14	Paraíso	2.820	0.467	0.241	0.317
15	Tacotalpa	2.333	0.220	0.682	0.350
16	Teapa	2.564	0.124	0.680	0.216
17	Tenosique	2.596	0.218	0.481	0.273

Cuadro 3: Índice de Riesgo por municipio con un nr de 150mm

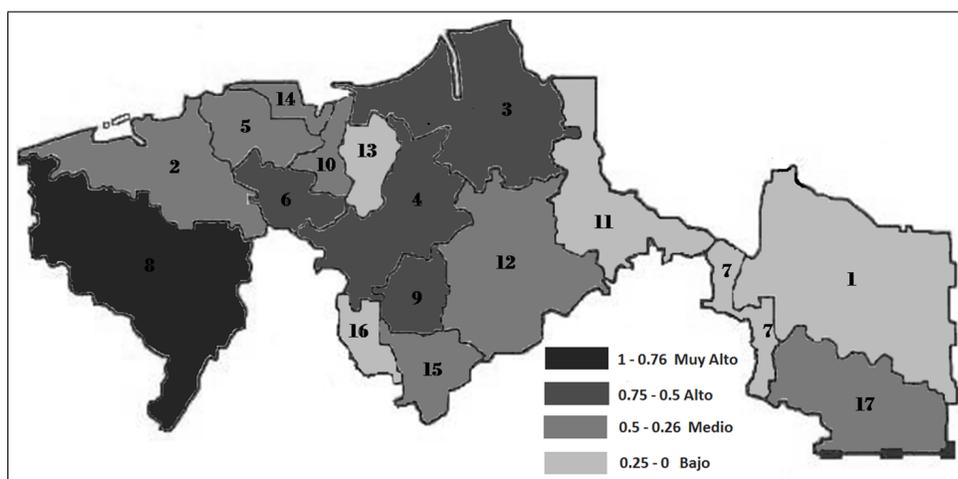


Figura 1: Índice de Riesgo con un nr de 150mm

<b>Mun.</b>	<b>Balancán</b>	<b>Cárdenas</b>	<b>Centla</b>	<b>Centro</b>	<b>Comalcalco</b>	<b>Cunduacán</b>
IH	Bajo	Alto	Alto	Muy Alto	Alto	Alto
IRP	Bajo	Medio	Alto	Alto	Medio	Alto
<b>Mun.</b>	<b>E. Zapata</b>	<b>Huimang.</b>	<b>Jalapa</b>	<b>Jalpa</b>	<b>Jonuta</b>	<b>Macuspana</b>
IH	Bajo	Alto	Alto	Medio	Bajo	Medio
IRP	Bajo	Muy alto	Alto	Medio	Bajo	Medio
<b>Mun.</b>	<b>Nacajuca</b>	<b>Paraíso</b>	<b>Tacotalpa</b>	<b>Teapa</b>	<b>Tenosique</b>	
IH	Medio	Medio	Bajo	Bajo	Medio	
IRP	Bajo	Medio	Medio	Bajo	Medio	

Cuadro 4: Comparativo de inundaciones históricas e IRP por municipio (Mun.)



# Elección de variables meteorológicas relacionadas con ozono en la estación Plateros

Sara Rodríguez R.<sup>a</sup>, Hortensia Reyes C.<sup>b</sup>  
*Facultad de Ciencias Físico-Matemáticas, BUAP*

Humberto Vaquera H.<sup>c</sup>, Paulino Pérez R.<sup>d</sup>  
*Colegio de Postgraduados*

## 1. Introducción

En la actualidad, el desarrollo y progreso tecnológico ha originado diversas formas de contaminación, las cuales alteran el equilibrio físico y mental del ser humano. Debido a esto, la actual contaminación se convierte en un problema más crítico que en épocas pasadas. El viento, las variaciones de temperatura, la cantidad de radiación solar y la lluvia son los principales factores meteorológicos que influyen y determinan la distribución, dispersión y concentración espacial de los contaminantes. En particular la contaminación del aire, que es la presencia de contaminantes tóxicos, representa uno de los principales problemas ambientales.

El ozono es una sustancia venenosa que se encuentran en el aire, que daña el medio ambiente y la salud de los seres humanos en altas concentraciones, en particular agrava enfermedades respiratorias, afecta funciones cerebrales, produce irritación en los ojos y causa problemas de conducta del ser humano.

En México varios científicos de diferentes disciplinas, están investigando conexiones importantes entre el ozono, cambio climático y calidad del aire, para crear nuevos compuestos químicos que no afecten el medio ambiente.

Para este trabajo se hace la suposición de que las concentraciones de ozono siguen una distribución de Valores Extremos Generalizada, además de ingresar un modelo de regresión

---

<sup>a</sup>rguez.sara@gmail.com

<sup>b</sup>hreyes@fcfm.buap.mx

<sup>c</sup>hvaquerah@gmail.com

<sup>d</sup>perpdgo@gmail.com

lineal que incluye covariables ambientales y químicas, de esta manera se verifica si dichas covariables influyen en las concentraciones de ozono (Rodríguez et al., 2009). Sin embargo para diferentes combinaciones de covariables se obtienen diferentes adecuaciones de los datos al modelo teórico, por lo que en este trabajo nos enfocamos a realizar una selección del mejor subconjunto de covariables que influye en la tendencia de las concentraciones de ozono.

## 2. Distribución de Valores Extremos

Sean  $y_1, \dots, y_n$  realizaciones independientes de una variable aleatoria  $Y$  que tiene densidad de probabilidad  $VEG(y; \mu, \sigma, \varepsilon)$ .

$$VEG_{\mu, \sigma, \varepsilon}(z) = \exp\{-[1 + \varepsilon(\frac{z - \mu}{\sigma})]^{-\frac{1}{\varepsilon}}\} \quad (1)$$

donde  $1 + \varepsilon(\frac{z - \mu}{\sigma}) > 0$

$$VEG_{\mu, \sigma, 0}(z) = \exp(-\exp^{-\frac{z - \mu}{\sigma}}) \quad (2)$$

donde  $\sigma = g(X^t \beta) = e^{X^t \beta}$ . Esta información son máximos de un periodo de la variable de interés para un tiempo fijo. Esta idea fue propuesta por Cox y Chu (1996), donde introduce un modelo de Regresión en la Distribución Pareto Generalizada (PG), donde  $X^t$  es el vector de covariables y  $\underline{\beta}$  es el vector de parámetros desconocidos, así  $\sigma = \sigma(\underline{x}^t \underline{\beta})$ . En el caso de VEG, como  $\sigma > 0$ , es natural suponer que  $\sigma(\underline{x}^t \underline{\beta}) = \exp(\underline{x}^t \underline{\beta})$ . En esta expresión se introducen los máximos de las covariables para el periodo de tiempo que se tome.

La función de log-verosimilitud  $\log L(Y; \mu, \underline{\beta}, \varepsilon)$ , se maximiza de manera numérica empleando el algoritmo de Nelder-Mead en el programa R (R Core Development Team, 2010), encontrando los estimadores de los parámetros  $\mu, \underline{\beta}$  y  $\varepsilon$ . Para el caso particular de tener sólo la covariable tiempo para una muestra del modelo anterior, se tendría que  $\mu$  es el parámetro de localización,  $\beta_1$  es el parámetro que se está asociando a la tendencia y  $t_i$  representa al  $i$ -ésimo periodo de tiempo. Es decir,  $\log L(X; \mu, \exp(\beta_0 + \beta_1 t_i), \varepsilon)$ , entonces si  $\beta_1 = 0$  se tiene que no existe tendencia, si  $\beta_1 > 0$  implica que hay tendencia creciente y por último,  $\beta_1 < 0$  habrá tendencia decreciente; para mayor información ver Reyes, et al. (2009).

Cuando se tiene una muestra de  $n$  observaciones y  $k$  covariables meteorológicas, sólo se analizará si influye o no en la tendencia. Se prueba la hipótesis,  $H_0 : \beta_j = 0$  vs  $H_a : \beta_j \neq 0$

con todos los coeficientes de las  $k$  variables. Se rechaza  $H_0$  si  $|\frac{\widehat{\beta}_j}{\sqrt{v_{jj}}}| > z_{\alpha/2}$  donde  $\alpha$  es el nivel de confianza y  $v_{jj}$  es el  $j$ -ésimo término de la diagonal de la matriz  $(E(-\frac{\partial^2 \log L}{\partial \beta_i \partial \beta_j} |_{\beta_j = \widehat{\beta}_j}))^{-1}$ . Para esto, las varianzas se obtienen usando la matriz de información de Fisher observada.

Se muestra un procedimiento para la selección de modelos propuesto en Coles (2004). Se supone un modelo  $\mathcal{M}_1$  con vector paramétrico  $\theta$  y que el modelo  $\mathcal{M}_0$  es un subconjunto del modelo  $\mathcal{M}_1$  ( $\mathcal{M}_0 \subset \mathcal{M}_1$ ) obtenido al restringir los  $k$  componentes de  $\theta$ , digamos cero. Así,  $\theta$  puede ser particionado como  $\theta_0 = (\theta^{(1)}, \theta^{(2)})$  bajo la restricción de que el  $k$ -dimensional sub-vector  $\theta^{(1)} = 0$ . Estableciendo  $\ell_0(\mathcal{M}_0)$  y  $\ell_1(\mathcal{M}_1)$  los máximos del logaritmo de la verosimilitud para el modelo  $\mathcal{M}_0$  y  $\mathcal{M}_1$  respectivamente, una prueba de la validez del modelo de relación de  $\mathcal{M}_0$  con  $\mathcal{M}_1$  a un nivel de significancia  $\alpha$ , es la siguiente:

$$H_0 : \theta_0 = \theta^{(1)} \quad (3)$$

$$H_a : \theta_0 \neq \theta^{(1)} \quad (4)$$

Se rechaza la hipótesis nula si  $D = 2\{\ell_1(\mathcal{M}_1) - \ell_0(\mathcal{M}_0)\} > c_\alpha$ , donde  $c_\alpha$  es el cuantil  $(1 - \alpha)$  de la distribución  $\chi_k^2$ .

### 3. Metodología

Asumiendo las condiciones de regularidad para la distribución VEG, se usan datos de las concentraciones de ozono de 1990 a 2009 de la estación Plateros, la cual es parte del SIMAT (2010) de la ciudad de México. Posteriormente se realizan bloques de 72 horas para disminuir la correlación entre los datos. Las covariables ambientales y químicas que se usan son: Ozono ( $O_3$ ), Temperatura ( $tmp$ ), Velocidad del Viento ( $wsp$ ), Humedad Relativa ( $rh$ ), Monóxido de Carbono ( $co$ ), Dióxido de Nitrógeno ( $no2$ ) y Dióxido de Azufre ( $so2$ ).

La estimación de parámetros se hace usando máxima verosimilitud por medio del algoritmo escrito en el programa R. El modelo que incluye todas las covariables es el siguiente VEG( $\widehat{\mu}, \widehat{\sigma}, \widehat{\varepsilon}$ ) donde  $\widehat{\sigma} = \exp(\widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2 + \widehat{\beta}_3 x_3 + \widehat{\beta}_4 x_4 + \widehat{\beta}_5 x_5 + \widehat{\beta}_6 x_6 + \widehat{\beta}_7 x_7)$ . Aquí  $x_1 = t$ ,  $x_2 = tmp$ ,  $x_3 = rh$ ,  $x_4 = wsp$ ,  $x_5 = co$ ,  $x_6 = no2$  y  $x_7 = so2$ .

Los modelos que se comparan son los siguientes. Representamos al modelo completo por  $\mathcal{M}_1$  con ocho variables, y lo comparamos con cada uno de los siguientes modelos:  $\mathcal{M}_A$  con  $t$ ,

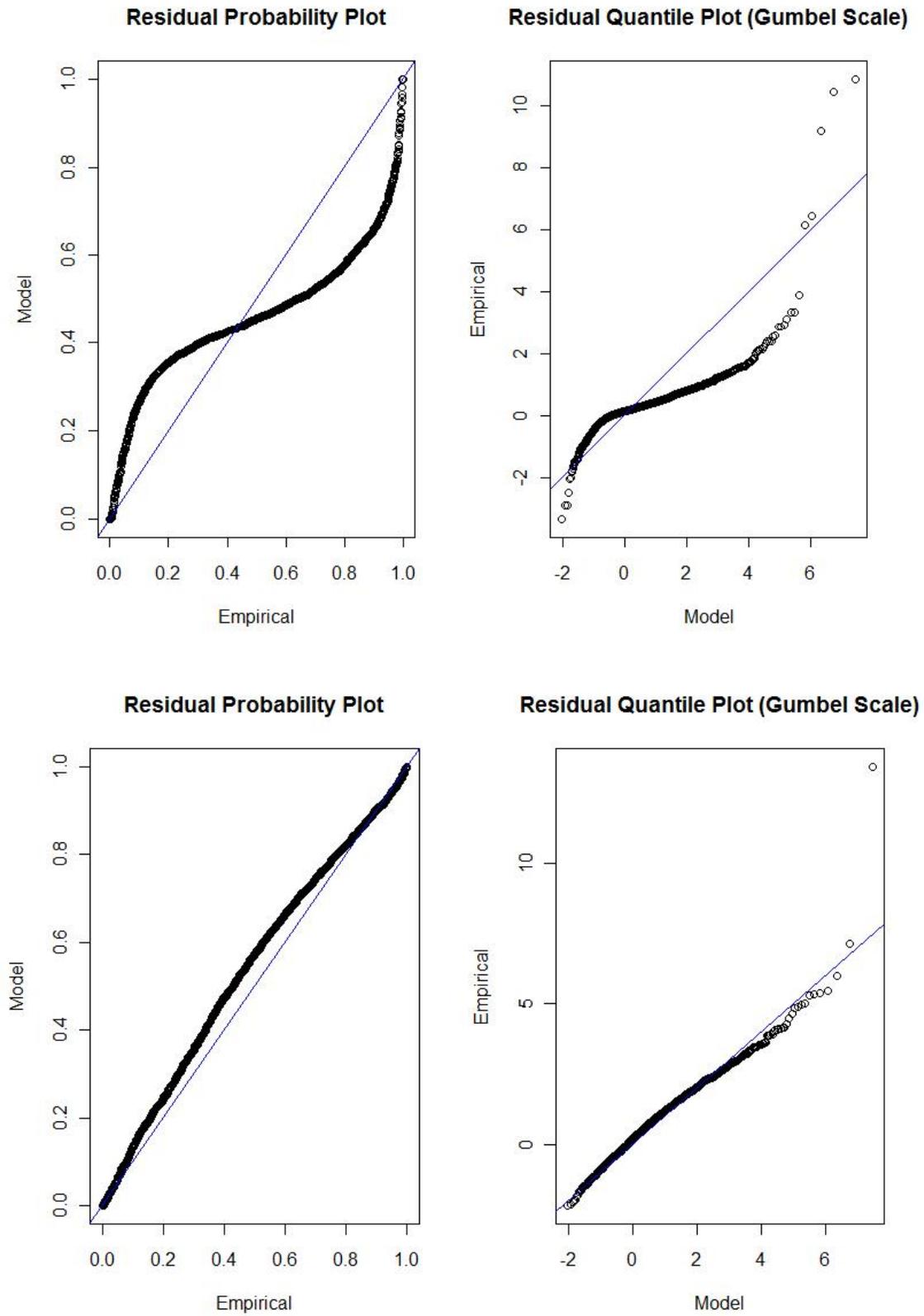


Figura 1: Gráfica Cuantil para  $\mathcal{M}_1$  y  $\mathcal{M}_B$

$\mathcal{M}_B$  con  $t, tmp$ ,  $\mathcal{M}_C$  con  $t, tmp, co$ ,  $\mathcal{M}_D$  contiene  $t, tmp, co, wsp$ ,  $\mathcal{M}_E$  con  $t, tmp, co, wsp, no2$ ,  $\mathcal{M}_F$  con  $t, tmp, co, wsp, no2, so2$ .

En la Figura 1 mostramos la gráfica de cuantiles para el modelo  $\mathcal{M}_1$  y del modelo que tiene la desviación más grande en valor absoluto. En el Cuadro 1 presentamos los resultados obtenidos

Modelo	log-verosimilitud	Devianza	G.l.
$\mathcal{M}_1$	1690.488		
$\mathcal{M}_A$	2825.267	-2269.558	6
$\mathcal{M}_B$	2826.945	-2272.914	5
$\mathcal{M}_C$	2634.511	-1888.046	4
$\mathcal{M}_D$	2509.235	-1637.494	3
$\mathcal{M}_E$	1696.392	-11.808	2
$\mathcal{M}_F$	2507.322	-1516.874	1

**Cuadro 1:** Función Deviance para la estación Plateros

## 4. Conclusiones

- En este trabajo presentamos una metodología para la selección de los modelos con diferentes número de covariables y elegimos cuál es más probable de explicar la variación de las concentraciones de ozono para la estación Plateros.
- Al realizar la comparación del modelo que tiene las siete variables con los modelos que contienen menor número de variables, se obtuvo que cualquier modelo reducido es mejor a excepción del modelo  $\mathcal{M}_E$ . Se puede deber al hecho de que el modelo que contiene las siete covariables no se adecuaba a la Distribución de Valores Extremos Generalizada.

## Bibliografía

- [1] Coles, S. (2004). *An introduction to Statistical Modeling of Extreme Values*. London: Springer-Verlag.
- [2] Cox, W., y Chu, S. (1996). Assessment of interannual ozone variation in urban areas from a climatological perspective. *Atmospheric Environment*, 30:2615–2625.

- 
- [3] R Core Development Team. (2010). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.
- [4] Reyes, H., Vaquera, H. y Villaseñor, J. (2009). Estimation of trends in high ozone levels using the quantiles of the distribution GEV. <http://www.interscience.wiley.com/>: *Environmetric*.
- [5] Rodríguez, S., Reyes, H., Vaquera, H. y Linares, G. (2009). Modelación estadística en contaminación del aire para dos estaciones vecinas del estado de México. *Aportaciones y Aplicaciones de la Probabilidad y la Estadística*, Benemérita Universidad Autónoma de Puebla 3(1):43–53.
- [6] SIMAT. (2010). Sistema de Monitoreo Atmosférico de la Ciudad de México. <http://www.sma.df.gob.mx/simat2/>.

# Generalización de la condición de la matriz $\mathbf{Z}$ para la caracterización del BLUP de $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$

Fernando Velasco Luna<sup>a</sup>, Mario Miguel Ojeda Ramírez<sup>b</sup>

*Facultad de Estadística e Informática. Universidad Veracruzana*

## 1. Introducción

La teoría de estimación en áreas pequeñas trata la estimación de características poblacionales para el área pequeña, cuando el tamaño de la muestra es pequeño, tal como la media poblacional  $\bar{Y} = T/N$  Valliant et al. (2000), donde  $T$  denota el total poblacional y  $N$  denota el número de unidades en la población. Uno de los enfoques de inferencia en la teoría de estimación en áreas pequeñas es el basado en el Modelo Lineal Mixto. En este enfoque se considera el modelo lineal jerárquico  $\mathbf{Y}_j = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{u}_j + \mathbf{e}_j$ , donde  $\mathbf{Y}_j$ ,  $\mathbf{X}_j$  y  $\mathbf{Z}_j$  denotan el vector respuesta, y las matrices de diseño (variables explicatorias a nivel 1 y a nivel 2), respectivamente, en la  $j$ -ésima unidad de nivel 2 (área pequeña), la cual cuenta con  $N_j$  unidades,  $\boldsymbol{\beta}$  denota el vector de parámetros fijos, y  $\mathbf{u}_j$  y  $\mathbf{e}_j$  los efectos aleatorios de nivel 1 y nivel 2, respectivamente; sea  $\mathbf{s}_j$  la muestra de  $n_j$  unidades en la  $j$ -ésima área pequeña, la cual cuenta con  $N_j$  unidades en la población,  $\mathbf{r}_j$  denotando las unidades en la  $j$ -ésima área que no están en  $\mathbf{s}_j$  y  $r_j = N_j - n_j$  el número de unidades no muestreadas. Una vez que la muestra  $\mathbf{s}_j$  ha sido obtenida se tiene la descomposición del modelo para la parte observada, que está dado por:

$$\mathbf{Y}_{j\mathbf{s}} = \mathbf{X}_{j\mathbf{s}}\boldsymbol{\beta} + \mathbf{Z}_{j\mathbf{s}}\mathbf{u}_j + \mathbf{e}_{j\mathbf{s}} \quad (1)$$

y el modelo para la parte no observada, que está dado por:

$$\mathbf{Y}_{j\mathbf{r}} = \mathbf{X}_{j\mathbf{r}}\boldsymbol{\beta} + \mathbf{Z}_{j\mathbf{r}}\mathbf{u}_j + \mathbf{e}_{j\mathbf{r}}. \quad (2)$$

---

<sup>a</sup>fvelasco@uv.mx

<sup>b</sup>mojeda@uv.mx

Entre todos los predictores, el mejor predictor lineal insesgado (*BLUP*) de la media poblacional finita  $\bar{Y}_j = N_j^{-1} \sum_{i=1}^{N_j} Y_{ij}$  en la  $j$ -ésima área pequeña está dado por  $f_j \bar{Y}_{js} + (1 - f_j) \left[ \bar{\mathbf{X}}_{jr} \hat{\boldsymbol{\beta}}_s + \bar{\mathbf{Z}}_{jr} \mathbf{G} \mathbf{Z}_{js}^t \mathbf{V}_{jss}^{-1} \left( \mathbf{Y}_{js} - \mathbf{X}_{js} \hat{\boldsymbol{\beta}}_s \right) \right]$ , donde  $f_j = n_j/N_j$ , y  $\bar{\mathbf{X}}_{jr}$  y  $\bar{\mathbf{Z}}_{jr}$  son los vectores de medias para las  $r_j$  unidades no muestreadas en la  $j$ -ésima unidad de nivel 2. La media de la población finita se puede descomponer en la media obtenida de la muestra  $\bar{Y}_{js}$  más la media de las unidades no muestreadas  $\bar{Y}_{jr}$ . Para la parte no muestreada se debe de tener una estimación de la media poblacional  $\mu_j$  de la  $j$ -ésima área pequeña, la cual es un efecto mixto. Dos de los modelos más usados en este enfoque son el modelo intercepto aleatorio y el modelo de pendientes aleatorias. En una serie de trabajos Velasco y Ojeda (2010a, 2010b, 2011) desarrollan la caracterización del *BLUP* del efecto mixto  $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$ , del *BLUP* de la media poblacional  $\mu_j$  correspondiente a la  $j$ -ésima unidad de nivel 2 y del *BLUP* de la media poblacional finita  $\bar{Y}_j$  correspondiente a la  $j$ -ésima unidad de nivel 2, en términos de los operadores proyector, ortogonal  $\mathbf{P}_Z = \mathbf{Z}(\mathbf{Z}^t\mathbf{Z})^{-1}\mathbf{Z}^t$  y oblicuo  $\mathbf{P}_{\mathbf{XV}} = \mathbf{X}(\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{V}^{-1}$ , definidos sobre los subespacios generados por las matrices de diseño que intervienen en el modelo intercepto aleatorio. Para lograr la caracterización de los predictores la matriz de diseño  $\mathbf{Z}_j$  del modelo bajo estudio debe cumplir la condición  $n_j\mathbf{P}_{\mathbf{Z}_j} = \mathbf{Z}_j\mathbf{Z}_j^t$ . El modelo de pendientes aleatorias no cumple la condición  $n_j\mathbf{P}_{\mathbf{Z}_j} = \mathbf{Z}_j\mathbf{Z}_j^t$  por lo no se puede llevar a cabo la caracterización de los predictores mencionados en términos de los operadores proyector. En este trabajo se presenta la generalización de la condición que debe de cumplir la matriz de diseño  $\mathbf{Z}_j$  que interviene en un caso particular del modelo de pendientes aleatorias para poder llevar a cabo la caracterización de predictores en términos de los operadores proyector  $\mathbf{P}_Z$  y  $\mathbf{P}_{\mathbf{XV}}$ .

## 2. Modelo de pendientes aleatorias

El modelo intercepto aleatorio está dado por

$$Y_{ij} = \mu + u_j + e_{ij}, \quad i = 1, \dots, n_j, \quad j = 1, \dots, J. \quad (3)$$

donde  $\mu$  es un parámetro fijo;  $u_j$  es el efecto aleatorio;  $u_j$  y  $e_{ij}$  son independientes, con  $u_j \sim N(0, \sigma_{u0}^2)$  y  $e_{ij} \sim N(0, \sigma_e^2)$ .

En el modelo intercepto aleatorio con variables explicatorias a nivel 1, sólo el intercepto

se supone aleatorio. En ocasiones la relación entre cada una de las variables explicatorias y la variable respuesta puede ser diferente en las unidades de nivel 2. Lo anterior da surgimiento al modelo de pendientes aleatorias. Para el caso de una variable explicatoria a nivel 1 lo anterior se expresa en el siguiente modelo:

$$Y_{ij} = \beta_{00} + \beta_{10}X_{ij} + u_{0j} + u_{1j}X_{ij} + e_{ij}, \quad i = 1, \dots, n_j; \quad j = 1, \dots, J,$$

$$E(e_{ij}) = 0, \quad Var(e_{ij}) = \sigma_e^2, \quad e_{ij} \sim N(0, \sigma_e^2), \quad (4)$$

$$E(u_{0j}) = 0, \quad Var(u_{0j}) = \sigma_{u_0}^2, \quad E(u_{1j}) = 0, \quad Var(u_{1j}) = \sigma_{u_1}^2 \text{ y } Cov(u_{0j}, u_{1j}) = \sigma_{u_{01}},$$

el cual se denomina modelo de pendientes aleatorias con una variable explicatoria a nivel 1.

Los modelos anteriores son casos especiales del modelo lineal mixto dado por:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e},$$

$$\mathbf{u} \sim N(\mathbf{0}, \sigma_u^2 \mathbf{I}_q), \quad \mathbf{e} \sim N(\mathbf{0}, \sigma_e^2 \mathbf{I}_n), \quad Cov(\mathbf{e}, \mathbf{u}^t) = \mathbf{0}, \quad (5)$$

donde  $\mathbf{Y} \in \mathbb{R}^n$ ,  $\mathbf{X}$  y  $\mathbf{Z}$  son matrices de orden  $n \times p$  y  $n \times q$ , respectivamente y  $\boldsymbol{\beta} \in \mathbb{R}^p$ . En este caso la matriz de varianzas y covarianzas de  $\mathbf{Y}$  está dada por  $\mathbf{V} = \sigma_u^2 \mathbf{Z}\mathbf{Z}^t + \sigma_e^2 \mathbf{I}_n$ .

Henderson (1975) obtiene el mejor estimador lineal insesgado de  $\boldsymbol{\beta}$  y el *BLUP* de  $\mathbf{u}$ , que están dados por  $\hat{\boldsymbol{\beta}} = (\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{V}^{-1} \mathbf{Y}$  y  $\hat{\mathbf{u}} = \mathbf{G}\mathbf{Z}^t \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}})$ , respectivamente.

Además obtiene el *BLUP* del efecto mixto  $\mathbf{k}^t \boldsymbol{\beta} + \mathbf{m}^t \mathbf{u}$  que está dado por  $\mathbf{k}^t \hat{\boldsymbol{\beta}} + \mathbf{m}^t \hat{\mathbf{u}}$ .

### 3. Caracterización de predictores

En esta sección se presentan los resultados de la caracterización de predictores en términos de los operadores proyector  $\mathbf{P}_Z$ ,  $\mathbf{P}_{XV}$  y  $\mathbf{Q}_{XV} = \mathbf{I} - \mathbf{P}_{XV}$ , bajo la condición  $n_j \mathbf{P}_{Z_j} = \mathbf{Z}_j \mathbf{Z}_j^t$ .

#### 3.1. Caracterización del *BLUP* del efecto mixto $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$

Velasco y Ojeda (2010b) desarrollan la caracterización del *BLUP* del efecto mixto  $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$  en términos de los operadores  $\mathbf{P}_{XV}$  y  $\mathbf{P}_Z$ .

**Teorema 3.1.** *Bajo el modelo (5), si se cumple la condición  $n_j\mathbf{P}_{\mathbf{Z}_j} = \mathbf{Z}_j\mathbf{Z}_j^t$ , entonces el BLUP del efecto mixto  $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$  se expresa en términos de los operadores, proyector oblicuo  $\mathbf{P}_{\mathbf{XV}}$  sobre  $S(\mathbf{X})$  y proyector ortogonal  $\mathbf{P}_{\mathbf{Z}}$  sobre  $S(\mathbf{Z})$  por:*

$$[\mathbf{P}_{\mathbf{XV}} + \mathbf{P}_{\mathbf{Z}}\mathbf{B}\mathbf{Q}_{\mathbf{XV}}]\mathbf{Y}, \quad (6)$$

donde  $\mathbf{B} = \bigoplus_{j=1}^J (b_j\mathbf{I}_{n_j})$  y  $b_j = n_j\sigma_{u_0}^2 / (n_j\sigma_{u_0}^2 + \sigma_e^2)$ .

*Demostración.* Para una demostración ver Velasco y Ojeda (2010b).

### 3.2. Caracterización del BLUP de la media poblacional $\mu_j$

Una vez que la muestra  $\mathbf{s}$ , de tamaño  $n$ , ha sido obtenida el vector  $\mathbf{Y}$ , las matrices  $\mathbf{X}$  y  $\mathbf{V}$ , los operadores  $\mathbf{Q}_{\mathbf{XV}}$  y  $\mathbf{P}_{\mathbf{Z}}$ , la estimación del parámetro  $\boldsymbol{\beta}$  y la predicción del efecto aleatorio  $u_j$  se denotarán por medio de  $\mathbf{Y}_s$ ,  $\mathbf{X}_s$ ,  $\mathbf{V}_s$ ,  $\mathbf{Q}_{\mathbf{X}_s\mathbf{V}_s}$ ,  $\mathbf{P}_{\mathbf{Z}_s}$ ,  $\hat{\boldsymbol{\beta}}_s$  y  $\hat{u}_{js}$ , respectivamente.

**Teorema 3.2.**  $\mathbf{T}_{js}$  dada por  $\mathbf{T}_{js} = \mathbf{X}_j(\mathbf{X}_s^t\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s^t\mathbf{V}_s^{-1}$  define una transformación lineal de  $\mathbb{R}^n$  a  $\mathbb{R}^{N_j}$ .

*Demostración.* Sean  $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \in \mathbb{R}^n$  y  $c \in \mathbb{R}$ ,

$$\begin{aligned} \mathbf{T}_{js}(c\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2) &= \mathbf{X}_j(\mathbf{X}_s^t\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s^t\mathbf{V}_s^{-1}(c\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2) \\ &= \mathbf{X}_j(\mathbf{X}_s^t\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s^t\mathbf{V}_s^{-1}(c\boldsymbol{\alpha}_1) + \mathbf{X}_j(\mathbf{X}_s^t\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s^t\mathbf{V}_s^{-1}(\boldsymbol{\alpha}_2) \\ &= c\mathbf{T}_{js}(\boldsymbol{\alpha}_1) + \mathbf{T}_{js}(\boldsymbol{\alpha}_2). \end{aligned}$$

La media poblacional  $\mu_j$  bajo el modelo (2) con  $\mathbf{Z}_j = \mathbf{1}_{n_j}$  está dada por  $\bar{X}_{jr}^t\boldsymbol{\beta} + u_j$ , que es un efecto mixto.

**Teorema 3.3.** *Bajo el modelo (5) si  $n_j\mathbf{P}_{\mathbf{Z}_{js}} = \mathbf{Z}_{js}\mathbf{Z}_{js}^t$ , entonces el BLUP de  $\mu_j$  se expresa en términos de los operadores proyector  $\mathbf{Q}_{\mathbf{X}_s\mathbf{V}_s}$  y  $\mathbf{P}_{\mathbf{Z}_s}$ , y de la transformación lineal  $\mathbf{T}_{js}$  por:*

$$\left[ \frac{\mathbf{1}_{N_j}^{*jr^t}}{r_j}\mathbf{T}_{js} + \frac{\mathbf{1}_n^{*jst}}{n_j}(\mathbf{P}_{\mathbf{Z}_s}\mathbf{B}_s\mathbf{Q}_{\mathbf{X}_s\mathbf{V}_s}) \right] \mathbf{Y}_s,$$

donde  $\mathbf{1}_n^{*jst}$  es un vector de 0's en  $\mathbb{R}^n$  con un 1 en las posiciones correspondientes a las unidades que pertenecen a la  $j$ -ésima área pequeña,  $\mathbf{B}_s = \bigoplus_{j=1}^J (b_j\mathbf{I}_{n_j})$  y  $b_j = n_j\sigma_{u_0}^2 / (n_j\sigma_{u_0}^2 + \sigma_e^2)$ .

*Demostración.* Para una demostración ver Velasco y Ojeda (2010a).

### 3.3. Caracterización del *BLUP* de la media poblacional finita $\bar{Y}_j$

$\mathbf{1}_{N_j}^{*jr}$  denota un vector de 0's en  $\mathbb{R}^{N_j}$  con un 1 en las posiciones correspondientes a las unidades de nivel 1, que pertenecen a la  $j$ -ésima área pequeña, que no están en la muestra.

**Teorema 3.4.** *Bajo el modelo (5), con  $\mathbf{Z}_j = \mathbf{1}_{n_j}$ , el *BLUP* de la media poblacional finita  $\bar{Y}_j$  está dado por:*

$$f_j \bar{Y}_{js} + (1 - f_j) \left[ \bar{\mathbf{X}}_{jr} \hat{\boldsymbol{\beta}}_s + \hat{u}_{js} \right]. \quad (7)$$

*Demostración.* Para una demostración ver Velasco y Ojeda (2011).

**Teorema 3.5.** *Bajo el modelo (5), con  $\mathbf{Z}_j = \mathbf{1}_{n_j}$ , si  $n_j \mathbf{P}_{\mathbf{Z}_{js}} = \mathbf{Z}_{js} \mathbf{Z}_{js}^t$ ,  $n_j \in \mathbb{R}$ , entonces el *BLUP* de la media poblacional finita  $\bar{Y}_j$ , se expresa en términos de los operadores proyector  $\mathbf{Q}_{\mathbf{X}_s \mathbf{V}_s}$  y  $\mathbf{P}_{\mathbf{Z}_s}$ , y de la transformación lineal  $\mathbf{T}_{js}$  por medio de:*

$$\left( \frac{\mathbf{1}_n^{*jst}}{N_j} + \frac{r_j}{N_j} \left[ \frac{\mathbf{1}_{N_j}^{*jr}}{r_j} \mathbf{T}_{js} + \frac{\mathbf{1}_n^{*jst}}{n_j} (\mathbf{P}_{\mathbf{Z}_s} \mathbf{B}_s \mathbf{Q}_{\mathbf{X}_s \mathbf{V}_s}) \right] \right) \mathbf{Y}_s,$$

donde  $\mathbf{B}_s = \bigoplus_{j=1}^J (b_j \mathbf{I}_{n_j})$  y  $b_j = n_j \sigma_u^2 / (n_j \sigma_u^2 + \sigma_e^2)$ .

*Demostración.* Para una demostración ver Velasco y Ojeda (2011).

## 4. Generalización de la condición $n_j \mathbf{P}_{\mathbf{Z}_j} = \mathbf{Z}_j \mathbf{Z}_j^t$

En este trabajo se considera el caso que en el modelo de pendientes aleatorias el intercepto es fijo. El modelo considerado está dado por:

$$Y_{ij} = \beta_0 + u_j + \beta_1 x_{ij} + u_{1j} x_{ij} + e_{ij}, \quad i = 1, \dots, n_j, \quad j = 1, \dots, J. \quad (8)$$

De las secciones anteriores para poder llevar a cabo la caracterización tanto del *BLUP* del efecto mixto, del *BLUP* de la media poblacional y del *BLUP* de la media poblacional finita es necesario que se cumpla la condición  $n_j \mathbf{P}_{\mathbf{Z}_j} = \mathbf{Z}_j \mathbf{Z}_j^t$ , donde  $n_j$  es el número de unidades de nivel 1 en la  $j$ -ésima unidad de nivel 2. Definiendo  $\boldsymbol{\beta} = (\beta_0 \ \beta_1)^t$ ,  $\mathbf{u}_j = u_{1j}$  y  $\mathbf{X}_j^* = \begin{pmatrix} \mathbf{1}_{n_j} & \mathbf{X}_j \end{pmatrix}$  donde  $\mathbf{X}_j$  es la matriz de variables explicatorias a nivel 1, el modelo para la  $j$ -ésima unidad de nivel 2 está dado por  $\mathbf{Y}_j = \mathbf{X}_j^* \boldsymbol{\beta} + \mathbf{X}_j \mathbf{u}_j + \mathbf{e}_j$ .

En este caso la matriz de diseño  $\mathbf{Z}_j$  está dada por la matriz  $\mathbf{X}_j$ . Ahora bien el modelo (8) no cumple la condición  $n_j \mathbf{P}_{\mathbf{Z}_j} = \mathbf{Z}_j \mathbf{Z}_j^t$  por lo cual no se puede llevar la caracterización de los predictores en términos de los operadores proyector bajo la condición  $n_j \mathbf{P}_{\mathbf{Z}_j} = \mathbf{Z}_j \mathbf{Z}_j^t$ . Desarrollando el producto  $\mathbf{Z}_j^t \mathbf{Z}_j$  se obtiene:

$$\mathbf{Z}_j^t \mathbf{Z}_j = \mathbf{X}_j^t \mathbf{X}_j = \sum_{i=1}^{n_j} X_{ij}^2.$$

Definiendo  $c_j = \sum_{i=1}^{n_j} X_{ij}^2$  tenemos que el modelo (8) cumple la condición  $c_j \mathbf{P}_{\mathbf{Z}_j} = \mathbf{Z}_j \mathbf{Z}_j^t$ .

## 5. Conclusiones

En este trabajo se llevó a cabo la generalización de la condición  $n_j \mathbf{P}_{\mathbf{Z}_j} = \mathbf{Z}_j \mathbf{Z}_j^t$  que debe cumplir la matriz  $\mathbf{Z}_j$ . Se espera que la condición  $c_j \mathbf{P}_{\mathbf{Z}_j} = \mathbf{Z}_j \mathbf{Z}_j^t$  permita obtener la caracterización de los predictores en términos de los operadores proyector en el modelo de pendientes aleatorias tal como sucede en el modelo intercepto aleatorio bajo la condición  $n_j \mathbf{P}_{\mathbf{Z}_j} = \mathbf{Z}_j \mathbf{Z}_j^t$ .

## Bibliografía

- Henderson, C. (1975), “Best Linear Unbiased Estimation and Prediction Under a Selection Model”, *Biometrics* **31**, 423–447.
- Valliant, R., Dorfman, A. H. y Royall, R. M. (2000), *Finite Population Sampling and Inference: A Prediction Approach*, John Wiley, New York.
- Velasco, L. y Ojeda, R. (2010a), “Caracterización del *BLUP* de la media poblacional en el modelo lineal general mixto”, *Memorias del XXIV Foro Nacional de Estadística* (INEGI: México), 81–87.
- Velasco, L. y Ojeda, R. (2010b), *Caracterización del BLUP del efecto mixto  $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$* , *Aportaciones y Aplicaciones de la Probabilidad y la Estadística*, Editores: Taponar S.F.S., Cruz S.H.A., Reyes C.H. y Zacarías F.J.D. edn, Benemérita Universidad Autónoma de Puebla, Puebla, México.

Velasco, L. y Ojeda, R. (2011), “Caracterización del *BLUP* de la media poblacional finita  $\bar{Y}_j$  en predicción en áreas pequeñas (Small Area Estimation)”, *Memorias del XXV Foro Nacional de Estadística* (INEGI: México), 105–111.



## Sección II

### Resúmenes cortos



## Conferencias Magistrales

### **Time Series Methods for describing structure and uncertainty in small angle X-Ray scattering experiments**

F. Jay Breidt

Colorado State University

Small-angle X-ray scattering (SAXS) is a technique for obtaining low-resolution structural information about biological macromolecules, by exposing a dilute solution to a high-intensity X-ray beam and capturing the resulting scattering pattern on a two-dimensional detector. The two-dimensional pattern is reduced to a one-dimensional curve through radial averaging; that is, by averaging across annuli on the detector plane. Subsequent analysis of structure relies on these one-dimensional data. This talk reviews the technique of SAXS and investigates autocorrelation structure in the detector plane and in the radial averages. Across a range of experimental conditions and molecular types, spatial autocorrelation in the detector plane is present and is well-described by a stationary kernel convolution model. The corresponding autocorrelation structure for the radial averages is non-stationary. Implications of the autocorrelation structure for inference about macromolecular structure are discussed. In particular, changepoint detection methods from time series analysis are applied to the radial averages to obtain new and improved estimates and confidence intervals for the radius of gyration, a key structural parameter.

This is joint work with Cody Alsaker, Andreea Erciulescu, Julien Neveu and Mark van der Woerd.

## Bondad de Ajuste: cómputo intensivo

Federico O'Reilly Togno

IIMAS, UNAM

En este trabajo se revisará como ha ido cambiando la manera de enfrentar el problema clásico de bondad de ajuste, con el incremento cualitativo y cuantitativo del apoyo computacional. Se presentarán algunas ideas que en su momento fueron muy valiosas, cuyo propósito fue el reducir el cómputo debido a que era un recurso muy “caro”. Se comentarán algunas propuestas que implican un procedimiento de aleatorización exógena, que ya era reconocido, que pueden disminuir la potencia de las pruebas tradicionales, además de que violan el principio de verosimilitud de Fisher. Entre ellas se comentará la propuesta de una sustitución aleatoria (Durbin, 1961, *Biometrika*, **48**, pp 41-55), y el método que utiliza la mitad de la muestra, seleccionada al azar, para la bondad de ajuste y la otra mitad para estimar parámetros (Stephens, 1978, *J. Roy Stat Soc, B*, **40** pp 64-70). Se comentará también el enorme esfuerzo por encontrar las distribuciones asintóticas de las estadísticas basadas en la distribución empírica (estadísticas EDF), las tablas que se fueron generando; y en varios casos, la obtención de correcciones para  $n$  finita, ya que el uso de distribuciones asintóticas produce necesariamente un resultado aproximado. Se revisa finalmente el uso de la simulación (con cierto grado de sofisticación) para no requerir tablas ni correcciones, pretendiendo aproximar, tanto como el número de simulaciones utilizadas lo permita, los valores de significancia ( $p$ -valores o  $p$ -values) de las estadísticas EDF. Se comentará asimismo el uso del remuestreo paramétrico (de cuando produce simulaciones “exactas”), y la obtención de  $p$ -valores condicionales. Además, se comentará la generación de muestras condicionalmente independientes (Monte Carlo condicional). En este último enfoque se comentarán los trabajos de Engen y Lillegård, (1997, *Biometrika* **84**, pp 235-240), el de Lindqvist y Taraldsen, (2005, *Biometrika*, **92**, pp 451-464), el de O'Reilly y Gracia Medrano, (2006, *Comms in Stats, Theory and Methods*, Vol **35**, pp 541-549), el de González-Barrios et.al, (2006, *Metrika* , **64**, pp 77-94, y 2010, *Metrika* , Vol **72**, pp 369-383), así como los de Lockhart et.al, (2007, *Biometrika*, Vol **94**, pp 992-998, y 2009, *Journal of Statistical Theory and Practice*, Vol **3**, pp 543-554).

## Clasificación difusa mediante el enfoque de suavizamiento hiperbólico

Javier Trejos Zelaya  
Universidad de Costa Rica

El método de clasificación o agrupamiento con suavizamiento hiperbólico es una nueva estrategia general para resolver problemas de clasificación automática; corresponde a hacer una sucesión de clasificaciones difusas. Analizamos estos hechos y presentamos un nuevo algoritmo para clasificación difusa. El enfoque tiene tres etapas principales: primero un relajamiento de la asignación al centro de clase más cercano, y luego un suavizamiento de las funciones máximo y la norma Euclidiana. Esto lleva a un problema de optimización continua que puede ser abordado con iteraciones de Newton–Raphson, cuya solución son los centroides de las clases. Luego, la asignación a las clases se hace para cada valor del paso de relajamiento de acuerdo con una regla simple, lo que lleva esencialmente a una clasificación difusa. Los resultados computacionales obtenidos para resolver problemas de prueba de la literatura muestran la eficiencia y las potencialidades del método propuesto. Mostraremos la posibilidad de obtener soluciones duras (no difusas) del problema particular de suma de cuadrados en clasificación mediante una estrategia difusa. Creemos que el mismo principio se podría utilizar en otros problemas similares.

Este es un trabajo conjunto con Adilson E. Xavier (Universidad Federal de Río de Janeiro), Eduardo Piza (Universidad de Costa Rica) y Alex Murillo (Universidad de Costa Rica).

## Modelación de valores extremos en medio ambiente

José A. Villaseñor Alva  
Programa de Estadística, Colegio de Postgraduados

El estudio del comportamiento estocástico del máximo  $M$  de un número aleatorio  $N$  de observaciones es de gran interés con el propósito de predecir la ocurrencia de eventos ambientales extremos que afectan a los habitantes de una población. En esta plática se discutirá la modelación probabilística conjunta de las variables aleatorias  $M$  y  $N$ . Se discutirán los dos

casos posibles: observaciones independientes y observaciones dependientes. En el caso de las observaciones independientes se incluye una aplicación de los resultados a un estudio de datos diarios de precipitación en la región centro del estado de Tabasco. En el caso de las observaciones dependientes, se incluyen aplicaciones al estudio de máximos diarios de niveles de ozono de las ciudades de México y Guadalajara, las cuales presentan altos niveles de ozono observados.

## Contribuciones Libres

### Estimación del valor en riesgo en la bolsa mexicana de valores usando modelos de heterocedasticidad condicional y teoría de valores extremos

Alejandro Iván Aguirre Salado, Humberto Vaquera Huerta y Martha Elva Ramírez Guzmán  
ISEI, Colegio de Postgraduados

José René Valdez Lazalde, Ciencias Forestales, Colegio de Postgraduados

Carlos Arturo Aguirre Salado, Facultad de Ingeniería, Universidad Autónoma de San Luis  
Potosí

Se propone una metodología para la estimación del valor en riesgo (VaR) para el índice de precios y cotizaciones de la Bolsa Mexicana de Valores mediante el uso combinado de modelos autoregresivos y de medias móviles (ARMA), modelos autoregresivos generalizados de heterocedasticidad condicional (GARCH) y la teoría de valores extremos. Los modelos ARMA se usaron para obtener residuales no correlacionados, que sirvieron de base para el análisis de valores extremos. Los modelos GARCH, al incluir en el modelo las volatilidades pasadas, son particularmente útiles en periodos de inestabilidad como de calma. De manera paralela se usó el método *riskmetrics*, a manera de comparación para las metodologías propuestas. Se concluye que el método propuesto presentó un mejor desempeño con respecto al método de *riskmetrics*, evaluadas al mismo nivel de error y confiabilidad en el ajuste de kupiec.

### **Aplicaciones de modelos lineales dinámicos**

Víctor Aguirre Torres, Depto. de Estadística, ITAM  
Gabriel Cervantes, ITESM

Se muestran los conceptos asociados a los Modelos Lineales Dinámicos para series de tiempo, los cuales tienen como punto de partida el enfoque Bayesiano para permitir que los parámetros del modelo evolucionen en el tiempo. Esto hace que sea una herramienta de pronóstico flexible, oportuna y capaz de incorporar fácilmente las previsiones del usuario. Se presentan unos ejemplos y finalmente se hace un análisis comparativo breve con otros enfoques de pronóstico.

### **Análisis bayesiano y modelos demográficos**

Alejandro Alegría Hernández, UMIE, Facultad Latinoamericana de Ciencias Sociales

El propósito del presente trabajo es aplicar los métodos bayesianos para estudiar el comportamiento de variables demográficas, tales como tasas de mortalidad, natalidad, etc. y realizar los pronósticos correspondientes. La metodología bayesiana permite tratar el caso de observaciones faltantes sin mayor problema, lo cual es importante cuando se trabaja con estadísticas vitales de países que tienen registros incompletos o poco confiables. También es posible considerar variabilidad no constante para diferentes grupos de edad. En esta plática se presentan algunos modelos usados en demografía y el análisis bayesiano de los mismos.

### **Calidad de plantas forestales en el parque nacional Gral. Juan Álvarez, en el estado de Guerrero**

Lorena Alonso Guzmán, Unidad Académica de Ingeniería, UAGro  
Alfredo Méndez Bahena, Unidad Académica de Ciencias Químicas Biológicas, UAGro  
Juan Carlos Medina Martínez, Unidad Académica de Ingeniería, UAGro

Los Estudios de Impacto Ambiental suelen identificar áreas que deberán ser preservadas y otras que por su sensibilidad o riesgo, deben tener algún manejo específico; como así también,

algunas en las que se deben aplicar medidas de mitigación o corrección. Este proyecto tiene por objetivo conocer la calidad morfológica de la especie arbórea del Parque Nacional Gral. Juan Álvarez, en el Estado de Guerrero. El área contaba con la cubierta vegetal de la formación boscosa para denominarse parque descrita para la zona, hoy presenta una cobertura vegetal moderada a pobre, y por la cantidad, altura y diversidad no constituye un bosque, sin embargo, se sugiere pensar en alguna medida compensatoria por pérdida de cubierta vegetal, considerando para esta acción, por medio de un análisis de estadístico multivariado, determinar que especies arbóreas, con el fin de evitar su deterioro y modificar políticas, planes y estrategias de carácter económico social.

### **Óptimos forestales y captura de carbono: caso Balancán, Tabasco**

Plácido Salomón Álvarez López, DICEA, Universidad Autónoma Chapingo

Víctor Rafael Valdovinos Chávez, Universidad Autónoma Chapingo

En la presente investigación se determinó la relación de cinco tipos de óptimos forestales: dos técnicos (máxima producción total y máxima producción media), dos económicos (Hotelling y Faustman) y un ambiental (Hartman), para tres modelos de crecimiento: polinomial, Schumacher y Chapman-Richards. Los datos referentes a la plantación fueron proporcionados por la empresa “PROPLANSE”, para dos especies tropicales: *Eucalyptus grandis* y *Gmelina arborea*. Al comparar el óptimo ambiental con los económicos, se encontró que al incorporar aspectos ambientales el turno registró un incremento de 2.5 hasta 9.9 meses para *E. grandis*, y de 1 hasta 23.4 meses para *G. arborea*. En términos generales, el óptimo de Hartman es la opción más adecuada para un aprovechamiento forestal sustentable, debido a que incorpora los aspectos económicos y, los servicios ambientales y recreativos de la plantación.

## Explotación de registros administrativos sobre el medio ambiente

Elizabeth Amézquita Muñoz

INEGI/DGGMA/DGARNMA/Dirección de Estadísticas Ambientales/  
Subdirección de Explotación de Registros Administrativos

Es conocido que las unidades del sector público y privado, dentro de sus actividades de gestión y fiscalización interna, llevan registros con el fin de tener elementos para sus procesos de evaluación y contabilidad. Sin embargo, hasta ahora, México no cuenta con un sistema de registros administrativos ambientales organizado, actualizado y articulado, por lo que se ha identificado como una fuente alterna de información, que bajo ciertos parámetros puede considerarse como información estadística. Proyecto de Explotación de Registros Administrativos sobre el Medio Ambiente Establecer, en conjunto con las Unidades de Estado, un Sistema Nacional de Registros Administrativos sobre el Medio Ambiente, como instrumento estándar para el aprovechamiento de registros administrativos y la generación de estadísticas, básicas y derivadas, que describan el estado y las tendencias del medio ambiente en nuestro país.

Relevancia estadística de los Registros Administrativos. En el desarrollo de los Registros Administrativos es preciso tener en cuenta:

- La existencia de Marco Jurídico o Legislación como sustento a la Unidad de Estado para generar el RA.
- Cobertura temática y geográfica.
- Proceso de captación de datos, aplicación de nuevas tecnologías, su procesamiento e integración de resultados.

---

## **Modelos Lineales Jerárquicos para el análisis del logro en matemáticas en estudiantes de bachillerato de la UAGro.**

Francisco Julián Ariza Hernández, Flaviano Godínez Jaimes, Ramón Reyes-Carreto y  
Enrique C. Crisóstomo-Santos

Unidad Académica de Matemáticas, Universidad Autónoma de Guerrero

En el presente trabajo se usan modelos lineales jerárquicos para conocer que variables influyen en el logro alcanzado por los alumnos en las asignaturas de matemáticas de nivel preparatoria de la Universidad Autónoma de Guerrero. El modelo propuesto contiene tres niveles, en el primer nivel las unidades de estudio son los alumnos, el segundo a nivel de escuelas y el tercer a nivel de las regiones del estado. Los datos se obtuvieron usando un muestreo complejo, donde los estratos son las regiones y los conglomerados las escuelas y grupos. Se seleccionaron 15 escuelas distribuidas en siete regiones, con un tamaño de muestra de 2496 alumnos. El logro alcanzado en matemáticas y los factores se registran haciendo uso de un cuestionario en el que se propusieron 13 variables, las cuales incluyen factores socioeconómicos del alumno, su situación escolar y el nivel académico de sus padres.

## **Análisis de tendencia en los extremos de variables meteorológicas, en el Estado de Durango**

Ma. Esperanza Blanco Ramírez, Colegio de Postgraduados

Humberto Vaquera Huerta, ISEI, Colegio de Postgraduados

Michel Rosengaus, Servicio Meteorológico Nacional

José René Valdez Lazalde, Ciencias Forestales, Colegio de Postgraduados

Se describirá una metodología para analizar tendencias en series de tiempo de variables climáticas extremas (precipitación máxima, temperatura mínima, temperatura máxima), en el estado de Durango. Esto, con el fin de acumular evidencia sobre los efectos del cambio climático en eventos extremos. Se utilizará la distribución de valores extremos generalizada considerando no estacionariedad en las series. Para estimar tendencias nos basamos en un modelo de regresión (Vector Generalized Linear Models), para probar la hipótesis de estacionariedad a partir de la prueba de razón de verosimilitudes generalizada. La base de datos

utilizada es la Maya 1.0 proporcionada por Servicio Meteorológico Nacional, y en particular utilizamos los máximos anuales (1961- 2000). Los resultados se presentan en un mapa de tendencias de eventos extremos para el estado de Durango.

### **Nuevas medidas para evaluar la capacidad predictiva en selección genómica**

Juan Burgueño y José Crossa  
CRIL, CIMMYT Int

En selección genómica se usa información genética de los individuos para predecir el fenotipo de los mismos con lo que se ahorra mucho tiempo y dinero. Una selección efectiva supone una predicción precisa. Existen diferentes metodologías y modelos para obtener la predicción, por lo que es necesario evaluar el poder predictivo de éstas, para lo cual se han utilizado diversas medidas, coeficiente de correlación y error cuadrático medio. Se proponen nuevas medidas para ser utilizadas en el proceso de comparar métodos y modelos de predicción en el contexto de la selección genómica, las cuales son más fáciles de interpretar que las usadas habitualmente, son más sensibles para diferenciar entre modelos y están relacionadas directamente con el objetivo del mejorador que va a hacer uso de las predicciones.

### **Series de Tiempo con puntos de cambio y observaciones censuradas con segmentos AR**

René Castro Montoya<sup>1</sup>, Gabriel A. Rodríguez Yam<sup>2</sup> y José Vidal Jiménez Ramírez<sup>1</sup>

<sup>1</sup>Fac. de Ciencias Físico-Matemáticas, Universidad Autónoma de Sinaloa, <sup>2</sup>Universidad Autónoma Chapingo

En este trabajo se propone un modelo bayesiano para series de tiempo con un número desconocido de puntos de cambio y observaciones censuradas, donde cada segmento es un proceso autoregresivo de orden variable. Se consideran distribuciones iniciales conjugadas para las medias y las varianzas en cada segmento. Para los coeficientes autoregresivos se condiciona para que la serie sea estacionaria en los segmentos. Para estimar el número y las localizaciones de los puntos de cambio se implementa el algoritmo de cadenas de Markov

Monte Carlo con saltos reversibles desarrollado por (green95), éste algoritmo consiste en crear una cadena de Markov irreducible y aperiódica que alterna saltos entre varios modelos con espacios de parámetros de diferente dimensión, que cumpla la condición de probabilidad de equilibrio, asegurando la convergencia a la distribución final. Los valores censurados se simulan de una distribución normal multivariada dada la parte observada ( Jung et al., 2005). Para ilustrar el algoritmo se analiza un conjunto de datos simulados de series de tiempo con dos puntos de cambio y distintos porcentajes de observaciones censuradas (10 %, 40 %).

### **Determinación del parámetro de suavizamiento de regresión por splines penalizados en muestreo de poblaciones finitas**

Luis Fernando Contreras Cruz, ISEI, Colegio de Postgraduados

José Elías Rodríguez Muñoz, Depto. de Matemáticas, Universidad de Guanajuato

Javier Suárez Espinosa, ISEI, Colegio de Postgraduados

Proponemos una forma de obtener el parámetro de suavizamiento en splines penalizados en el contexto de muestreo de poblaciones finitas. Dicho parámetro se obtiene fijando el número equivalente de parámetros o grados de libertad y resolviendo una igualdad para el parámetro de suavizamiento basado en la información de una muestra. Se hizo un estudio de simulación para ver el comportamiento empírico de la determinación del parámetro de suavizamiento. Además, con una población real se observó el efecto que tiene la determinación del parámetro de suavizamiento propuesto en la estimación del total poblacional. Para la población real, hay evidencia empírica de que el parámetro de suavizamiento propuesto produce un error relativo de estimación similar al que se obtendría determinando dicho parámetro con toda la información auxiliar de la población.

## **Un ejemplo de aplicación estadística sobre evaluación de las matemáticas en preparatorias de la BUAP**

Diana Jazmín Cruz Armenta, Hortensia Reyes Cervantes, Francisco Tajonar Sanabria, y  
Hugo Cruz Suárez

Facultad de Ciencias Físico Matemáticas/Estadística, BUAP

En este trabajo se presentan los resultados de una encuesta que se realizó sobre algunos de los factores que consideramos influyen en el gusto por las matemáticas en alumnos de las preparatorias de la BUAP mediante la aplicación de una encuesta, de la cual nos interesó saber los hábitos de estudio y factores sociales que afectan en interesarles o no las materias de matemáticas. Algunas de las variables que consideramos fueron: horas de estudio, confianza con el profesor, promedio en matemáticas, si les gusta el método de enseñanza de su profesor, etc. Por medio de un modelo de regresión logística se analizaron los resultados y se encontró que se han perdido hábitos de estudio e interés por las matemáticas. De aquí observamos que de querer hacer algo, por estas nuevas generaciones se debe poner más cuidado en la motivación, difusión por parte de los maestros que imparten esas materias.

## **Análisis estadístico descriptivo sobre las bases de datos de monitoreo atmosférico en las 4 estaciones meteorológicas de Puebla (REMA) durante el año 2003**

Juan Antonio Cruz Juárez, Hortensia Reyes Cervantes y Hugo Cruz Suárez

Facultad de Ciencias Físico Matemáticas/Estadística, BUAP

Se trabajó con las bases del REMA (que después de realizar una “limpieza” de la información se eliminaron algunos años) realizando métodos de análisis descriptivos y no paramétricos para verificar los supuestos de normalidad e independencia, así como la homogeneidad de varianzas para modelar los datos de ozono (O<sub>3</sub>) a través del método de regresión lineal múltiple. Se presentan los resultados usando el paquete estadístico “Minitab”.

---

## **Utilización de algunos modelos estadísticos y de la teoría de juegos como apoyo a la gobernanza ambiental**

Luis Cruz Kuri<sup>1</sup>, Agustín Jaime García Banda<sup>1</sup>, Juan Ruiz Ramírez<sup>2</sup>, Ismael Sosa Galindo<sup>1</sup>  
y Bertha Alicia Arce Castro<sup>1</sup>

<sup>1</sup>Fac. de Ciencias Administrativas y Sociales, <sup>2</sup>Fac. de Economía, Universidad Veracruzana

Gran parte de la reglamentación ambiental en las sociedades complejas ha tenido que ver con un enfoque autoritario y de control. Si se cuenta con suficientes recursos para monitorear y hacer que se tenga cumplimiento, tales enfoques son efectivos; pero cuando los gobiernos no tienen la voluntad política o los recursos para proteger a las áreas supuestamente “protegidas”, cuando ocurren grandes daños ambientales en puntos difíciles de detectar y cuando la necesidad es la de promover la innovación en los comportamientos o en las tecnologías en lugar de prohibir o requerir las familiares, los enfoques autoritarios o de control son menos efectivos. En el presente trabajo se abordan conceptos específicos sobre algunos modelos estadísticos, así como los de cooperación vía la teoría de juegos, y se menciona su aplicabilidad al tema de la gobernanza ambiental.

## **Análisis de la resistencia del concreto suministrado para la construcción de la torre de Ingeniería**

Alfredo Cuevas Sandoval<sup>1</sup>, Flaviano Godínez Jaimes<sup>2</sup>, Sulpicio Sánchez Tizapa<sup>1</sup>, y Raziel Barragán Trinidad<sup>1</sup>

<sup>1</sup>Unidad Académica de Ingeniería, <sup>2</sup>Unidad Académica de Matemáticas, Universidad Autónoma de Guerrero

El concreto hidráulico tiene variaciones en su calidad debido a los diversos factores que la componen, como son los agregados usados, la dosificación, el proceso de fabricación y procedimientos en obra. Una característica importante del concreto hidráulico es la resistencia a compresión y durabilidad, que son consecuencia del control realizado sobre los factores mencionados. Es imprescindible tener un control de calidad en una obra de infraestructura como lo es la Torre de la Unidad Académica de Ingeniería de la UAGro. Por esta razón se

realizaron muestreos de concreto durante el tiempo de ejecución de la obra a las unidades de suministro, con el fin de verificar el cumplimiento de los criterios de las normas mexicanas para campo y laboratorio. El concreto suministrado fue del tipo industrializado listo para usarse. En este trabajo se hace un análisis estadístico descriptivo.

### **Estudio de la industria ladrillera artesanal del municipio de Chilpancingo, Guerrero**

Alfredo Cuevas Sandoval<sup>1</sup>, Sulpicio Sánchez Tizapa<sup>1</sup>, Flaviano Godínez Jaimes<sup>2</sup>, y Raziel Barragán Trinidad<sup>1</sup>

<sup>1</sup>Unidad Académica de Ingeniería, <sup>2</sup>Unidad Académica de Matemáticas, Universidad Autónoma de Guerrero

El ladrillo o tabique es un elemento básico en la conformación de mampostería utilizado para la construcción de vivienda en el mundo y en particular en México y Latinoamérica. La industria en la región de Chilpancingo está conformada por una gran variedad de fábricas de tipo artesanal. El bajo nivel de desarrollo tecnológico en los procesos de extracción y de producción enmarca a esta actividad como un trabajo de subsistencia más que una actividad industrial, representada en altos costos de producción. Debido a los vaivenes económicos en el país y a una competencia de fuerte impacto por la producción de piezas de concreto, este ha sido uno de los sectores más afectados, generando el cierre de varias ladrilleras. El presente estudio técnico actualiza el estado del sector, evalúa sus estadísticas, define algunos parámetros que lo afectan y en consecuencia permite establecer algunas posibles estrategias a implementar con el fin de lograr una recuperación del mismo.

### **Verificación del supuesto de homogeneidad de varianzas bajo censura**

Violeta De La Huerta Contreras, y Humberto Vaquera Huerta  
ISEI, Colegio de Postgraduados

Existen muchas pruebas para verificar el supuesto de homocedasticidad para comparar dos poblaciones, sin embargo se tiene una limitante en el ámbito práctico cuando se tienen observaciones censuradas. El objetivo de este trabajo es investigar el comportamiento de la prueba

de Bartlett bajo censura por la derecha tipo II. Se utiliza el método de máxima verosimilitud (ML) para estimar las varianzas, se investiga también el comportamiento asintótico de este estadístico. Se realizó también un estudio de simulación para obtener la distribución del estadístico de Bartlett bajo censura para tamaños de grupos tanto iguales como diferentes y para distintos niveles de censura. Esta metodología es útil ya que conforme aumenta la censura en las muestras, la media y varianza de la distribución del estadístico de Bartlett aumenta considerablemente. Tablas de valores de la distribución fueron generadas así como el tamaño, potencia y robustez de la prueba.

### **Aplicaciones de la herramienta estadística conocida como la huella ecológica en aspectos de sustentabilidad**

Agustín Jaime García Banda<sup>1</sup>, Luis Cruz-kuri<sup>1</sup>, Ismael Sosa Galindo<sup>1</sup>, Juan Ruiz Ramirez<sup>2</sup>  
y Bertha Alicia Arce Castro<sup>1</sup>,

<sup>1</sup>Fac. de Ciencias Administrativas y Sociales, <sup>2</sup>Fac. de Economía, Universidad Veracruzana

La huella ecológica, propuesta por William Rees y más tarde desarrollada por Mathis Wackernagel entre los años de 1992 y 1994, es una medida de la demanda humana sobre los ecosistemas terrestres, la cual permite calcular la presión humana que se está ejerciendo sobre el planeta a nivel global, sobre una nación en particular, o sobre una región más pequeña de un país, como lo puede ser una entidad federativa o una ciudad. Año tras año este número se vuelve a calcular (con un retraso de tres años, debido al tiempo que le toma a las naciones Unidas el recolectar y publicar todas las estadísticas que la fundamentan). En el presente trabajo se mencionan varias maneras de obtener la huella ecológica desde un enfoque estadístico, buscando una que permita realizar comparaciones válidas.

---

## **Análisis, usando estadística descriptiva, de la obesidad en jóvenes de la UTM entre 18-23 años**

Johnatan Adonai García López y Marisol López Cerino  
Instituto de Física y Matemáticas, Universidad Tecnológica de la Mixteca

La obesidad es una enfermedad que conlleva a múltiples factores que dan origen a diversas enfermedades. En el caso de México, las medidas que se utilizan para detectar posible padecimiento de esta enfermedad, para hombres es una circunferencia abdominal mayor a 90 cm, mientras que para las mujeres, mayor a 80 cm. Se estudiará y analizará una muestra tomada de 200 estudiantes de la UTM, de los cuales, la mitad serán hombres y la otra mitad mujeres, jóvenes entre 18 y 23 años. El interés es dar a conocer una de las muchas aplicaciones de la Estadística Descriptiva en el análisis de problemas cotidianos que se presentan en la vida real. Se resalta el uso de comparaciones con la distribución normal y el acoplamiento entre la teoría y las aplicaciones.

## **Una comparación de la metodología Box-Jenkins y el análisis espectral singular en una serie de tiempo con datos faltantes**

Flaviano Godínez Jaimes, Francisco Julián Ariza Hernández, Ramón Reyes Carreto y  
Rigoberto Menor Hernández

Unidad Académica de Matemáticas, Universidad Autónoma de Guerrero

Algunas series de tiempo presentan observaciones faltantes. Un método simple para imputar datos es usar la media o la moda. En este trabajo se comparan dos metodologías para imputar datos en una serie de tiempo estacional que presenta tendencia y heterogeneidad de varianzas con dos bloques de datos perdidos. La primera es la metodología Box y Jenkins que consiste en ajustar un número de modelos ARIMA igual al número de bloques de datos ausentes en la serie de tiempo. La segunda es el Análisis Espectral Singular que se basa en el análisis de componentes principales en el dominio del tiempo para extraer los componentes de la serie de tiempo tales como periodicidad y tendencia para el llenado de datos perdidos. Para comparar ambas metodologías se usa la media del error cuadrado de predicción.

## **Una medición de la calidad de la consulta externa de hospitales públicos de Chilpancingo**

Flaviano Godínez Jaimes, Ramón Reyes Carreto, Francisco Julián Ariza Hernández

Unidad Académica de Matemáticas, Universidad Autónoma de Guerrero

Rodrigo Pérez Cabañas

Unidad Académica de Enfermería, Universidad Autónoma de Guerrero

Elia Barrera Rodríguez

Hospital de la madre y el niño Guerrerense

Los hospitales públicos tienen una demanda que en muchos casos rebasa su capacidad de infraestructura y de personal. Aún en esas condiciones los usuarios esperan servicios de salud eficiente, oportunos y de calidad. Una forma de medir la calidad de los servicios sanitarios es mediante la percepción de sus usuarios. Este trabajo presenta los resultados de un estudio transversal descriptivo para evaluar la satisfacción de los usuarios de consulta externa de cuatro de los principales hospitales públicos de Chilpancingo.

## **Utilización del pez sapo en nutrición animal como ensilaje en la cuenca baja del Balsas**

Flaviano Godínez Jaimes y Moisés Cipriano Salazar

Unidad Académica de Matemáticas, UAGro

Blas Cruz Lagunas, Luis Miguel Camacho Díaz y José Alonso Galeana

Unidad Académica de Medicina Veterinaria y Zootecnia, UAGro

El pez sapo o diablo es una especie invasora que ha reducido la pesca en la tierra de caliente de Guerrero. Esto se debe a que destruyen los implementos de pesca y reducen la población de tilapia. Cuando es capturado solo es desechado contaminando el medio ambiente. Para reducir su impacto se busca aprovechar la especie para elaborar alimento para ganado en forma de ensilado. Este alimento debe cubrir requisitos como un pH aceptable que garantice su calidad. En este trabajo se presenta un estudio para determinar que factores reducen el pH de ensilaje de pez sapo. Se hizo un experimento factorial con factores fijos donde la variable

respuesta es el pH y los factores son Melaza con tres niveles y Modalidad de inclusión del pez diablo con cuatro niveles. Se encontró que hay efecto de ambos factores y de la interacción y se identificaron los mejores tratamientos.

### **Pruebas de hipótesis para modelos no lineales no anidados**

Roxana Góngora Hernández, Área de Matemáticas, Instituto Politécnico Nacional  
Jorge Domínguez Domínguez, CIMAT A.C.

En muchas situaciones de la realidad se está interesado en saber el comportamiento de un fenómeno, por tanto la construcción de modelos estadísticos es utilizado surgiendo en ocasiones varios modelos alternativos, entonces surge el problema de seleccionar el modelo más adecuado del conjunto que se tenga. Un método usado para seleccionar entre los modelos es a través de pruebas de hipótesis. Las pruebas de hipótesis implican generalmente modelos anidados, en la cual el modelo que representa la hipótesis nula es un caso especial de un modelo más general que representa la hipótesis alternativa. Para este modelo, siempre se puede probar la hipótesis nula mediante pruebas de las restricciones que se imponen a la alternativa. Pero a menudo surgen modelos los cuales no están anidados. Esto significa que ninguno de los modelos puede ser escrito como un caso especial del otro sin imposición de restricciones en ambos modelos. En tal caso, no podemos simplemente probar uno de los modelos contra el otro, al imponer restricciones en uno de ellos. Existe una extensa literatura sobre las pruebas de hipótesis no anidadas. Esta proporciona un gran número de maneras de probar la especificación de los modelos estadísticos cuando una o más alternativas no anidadas existen. En esta presentación se presentará algunas pruebas de hipótesis no anidadas, principalmente en el contexto de los modelos de regresión, para la seleccionar un modelo dentro de un conjunto de modelos alternativos. Las pruebas estudiadas fueron, la prueba de Davidson y MacKinnon, la prueba de Hoel y los métodos de discriminación presentados en el libro “Model Discrimination for Nonlinear Regression Models” de Borowiak Dale S. (1989), además se realizó una revisión de la prueba del criterio de Información de Akaike.

Para ilustrar la aplicación de estas pruebas y formas de discriminación se utilizaron dos conjuntos de datos, un conjunto de datos médicos, en específico datos de mediciones

realizados a fetos en desarrollo y un conjunto de datos de bioquímica, donde se midió la producción de Nisina en un cultivo de leche

### **Comparación de métodos de reducción de dimensión para predecir sobrevivencia**

Elizabeth González Estrada, ISEI, Colegio de Postgraduados

Javier Rojo, Rice University, USA

En esta presentación se discuten los resultados de un estudio de simulación en el que se comparan procedimientos de dos etapas para predecir el tiempo de sobrevivencia a enfermedades usando datos de microarreglos, en los cuales se tienen mediciones sobre niveles de expresión genética de un número de genes mucho mayor que el número de pacientes bajo estudio. En la primera etapa de cada procedimiento se reduce la dimensión del conjunto de datos usando matrices aleatorias de Achlioptas y con entradas iid normal estándar, y el método de componentes principales. En la segunda etapa, las variables del conjunto de datos reducido son introducidas como covariables en el modelo de riesgos proporcionales de Cox para predecir el tiempo de sobrevivencia. Los criterios de comparación usados son el error cuadrado medio de los pesos asignados a las covariables, el error cuadrado medio de la función de sobrevivencia estimada usando el promedio de las covariables y el error cuadrado medio de la función de sobrevivencia estimada usando las covariables correspondientes a los individuos.

### **Una nota sobre un converso en teoría de la medida**

José María González-Barrios Murguía, IIMAS, UNAM

En este trabajo se analiza un converso natural en la construcción de una gran familia de medidas en el sigma-álgebra de Borel en los reales, conocida como la familia de medidas de Lebesgue-Stieltjes. En particular se estudia el caso de este converso en dimensiones mayores o iguales a dos.

### **Modelación de valores máximos de ozono**

María Guzmán Martínez, Colegio de Postgraduados  
José Aurelio Villaseñor Alva, ISEI, Colegio de Postgraduados  
Javier Suarez Espinoza, ISEI, Colegio de Postgraduados  
Enrique Raúl Villa Diharce, CIMAT A.C.  
Elizabeth González Estrada, ISEI, Colegio de Postgraduados

El cambio climático se ha convertido en un tema importante en los últimos años. Algunas Ciudades más desarrolladas, tienen sus propios sistemas de monitoreo para evaluar la calidad del aire. Tal es el caso de la zona metropolitana de Guadalajara, México, que tiene su red de monitoreo atmosférico. En este trabajo se analizaron los máximos diarios del ozono, procedentes de la estación de monitoreo Centro, ubicada en la zona centro de Guadalajara, México. La base de datos que se analizó contiene las observaciones del periodo 1997 -2008. El objetivo de este trabajo es comparar el ajuste del modelo de probabilidad conjunta propuesto por Villaseñor y González (2010) y el de la distribución Pareto Generalizada (DPG). Con base en pruebas estadísticas realizadas, se obtuvo que los dos modelos propuestos proporcionan un buen ajuste; no obstante el modelo de probabilidad conjunta ofrece un análisis estadístico más completo para este tipo de datos, dado que se considera una variable adicional, la cual es útil para propósitos de predicción.

### **Meta-análisis usando modelos jerárquicos bayesianos**

Juan Diego Hernández Jarquín, COLPOS, CIMMYT - CRIL  
Sergio Pérez Elizalde, COLPOS, CIMMYT - CRIL  
José Luis Crossa Hiriart, CIMMYT - CRIL

En el análisis de la interacción de tablas de doble entrada ha sido difícil estimar los parámetros de interacción, además en el contexto de los experimentos de plant breeding existe la necesidad de incorporar información disponible para el análisis de la interacción GxE y de los efectos lineales. Se propone el modelo Jerárquico Bayesiano para la realización del meta análisis, usando el muestreador de Gibbs se ahorra tiempo computacional y se hace estable

el algoritmo. Datos de experimentos de maíz en multi-ambientes (CIMMYT) de dos años consecutivos (2009 y 2010). La distribución von Mises-Fisher fue usada como a priori para el estudio de la interacción GE, y la incorporación de información se utilizó para estimar los parámetros lineales y bilineales. Los resultados muestran que usando el modelo propuesto se obtiene mejores estimaciones lo cual se refleja en un encogimiento (shrinkage).

### **Pronóstico de la demanda de energía eléctrica mediante series de tiempo**

José Del Carmen Jiménez Hernández, Felipe De J. Trujillo Romero, y Williams Gómez  
López

Instituto de Física y Matemáticas, Universidad Tecnológica de la Mixteca

El pronosticar algún fenómeno repercute tanto en la toma de decisiones de una empresa como en la planificación de recursos para una producción mayor y más eficiente, además, el conocer el suceso que va a pasar en el futuro permite tomar ciertas medidas preventivas. El objetivo principal de este trabajo es realizar la predicción para un conjunto de datos, los cuales corresponden a la demanda máxima mensual de energía eléctrica en una subestación de distribución proporcionados por la Comisión Federal de Electricidad. Dicha predicción se realiza mediante la metodología de Box and Jenkins de series de tiempo, además de realizar la comparación de dichas predicciones con las realizadas mediante redes neuronales artificiales usando “backpropagation” como el algoritmo de aprendizaje de la red neuronal.

### **Probabilidad de ruina en el modelo clásico de Cramér-Lundberg**

José Del Carmen Jiménez Hernández y Alma Delia Maldonado Santiago  
Instituto de Física y Matemáticas, Universidad Tecnológica de la Mixteca

Ekaterina Todorova Kolkovska

CIMAT, A.C.

Se presentan aproximaciones de la probabilidad de ruina en una compañía de seguros empleando el modelo clásico de Cramer-Lundberg cuando las distribuciones del tamaño de los

reclamos es de cola ligera, en particular, mezcla de exponenciales y gamma, además, se realizan simulaciones de las trayectorias del proceso de riesgo para estas distribuciones con diferentes capitales iniciales.

### **Cálculo de formaciones geológicas para la perforación de un pozo petrolero por correlación**

Marcos Andrés Jiménez Moreno<sup>1</sup>, Trinidad Cruz Sánchez<sup>1</sup>, Luis Alfonso Ruiz Ochoa<sup>2</sup>,  
Marlen Marín Solís<sup>2</sup>, y Nancy Elena Hernández Morales<sup>1</sup>

<sup>1</sup>Universidad Tecnológica de Tabasco, <sup>2</sup>Universidad Juárez Autónoma de Tabasco

Uno de los aspectos de mayor trascendencia en el diseño de la perforación de pozos es determinar la mayor cantidad de características técnicas, por ejemplo: formaciones a encontrar, estabilidad, etc., así como también, posibles problemas que se podrían presentar durante la perforación e intervención del mismo, dicha actividad es denominada correlación de pozos petroleros. Utilizando 37 estados mecánicos de pozos reales en su mayoría aun en etapa de producción junto con sus coordenadas UTM se obtuvo un plano de distribución espacial. Posteriormente se utilizó el Origen 6.0 para analizar la información, logrando seleccionar los siguientes 6 pozos: 163, 123, 125, 141, 145, 161. Con AutoCAD se obtuvieron planos en 3D, logrando así corroborar la calidad y cantidad de información y el perfil de las formaciones geológicas a perforar.

### **Aplicaciones de la estadística en la hidrología superficial**

Marcos Andrés Jiménez Moreno<sup>1</sup>, Trinidad Cruz Sánchez<sup>1</sup>, Luis Alfonso Ruiz Ochoa<sup>2</sup>,  
Marlén Marín Solís<sup>2</sup>  
y Gabriel Ramón Hernández<sup>1</sup>,

<sup>1</sup>Universidad Tecnológica de Tabasco, <sup>2</sup>Universidad Juárez Autónoma de Tabasco

La erosión de cuencas de drenaje y la degradación es uno de los problemas medioambientales más complejos a nivel mundial, dado su impacto en el diseño y planeación de las obras hidráulicas, logrando así afectar su funcionamiento, eficiencia y periodo de vida, éstas

están relacionadas con eventos hidrológicos futuros. Los sistemas hidrológicos son afectados algunas veces por eventos extremos, tales como tormentas severas, crecientes y sequías. Las diversas aplicaciones de la estadística la hacen una herramienta imprescindible para la determinación de volumen de acarreo de sólidos, calendarios de riego, área de precipitación, etc. Utilizando ecuaciones y métodos como por ejemplo: método de Thorntwaite, Bleney-Criddle, Polígono de Thiessen, Isoyetas, curvas de intensidad-duración-periodo de retorno, Distribuciones Pearson, Gumbel, Normal y General de Valores Extremos, por mencionar.

### **Ecuaciones estructurales en la identificación de actitudes hacia la conservación del medio ambiente**

Bertha Laura Letipichia  
ISEI, Colegio de Postgraduados

El objetivo de este trabajo es identificar la actitud de los estudiantes del Colegio de Postgraduados hacia la conservación del medio ambiente, como son sus hábitos personales, la inclusión de medidas de conservación en sus proyectos de investigación y la influencia de la formación recibida por la Institución en dicho tema. En inicio, se comparan los métodos de estimación de Mínimos Cuadrados Ponderados y de Máxima verosimilitud para observar diferencias en el tratamiento de variables ordinales y continuas. Empleando el Modelo de Ecuaciones Estructurales se analiza el instrumento “Conservación del Medio Ambiente”. Se proponen las fases: implementación del instrumento, prueba piloto y propuesta de ajustes para ampliar la muestra. Este proyecto es contribución a la Línea Prioritaria de Investigación 14 del Colegio de Postgraduados: Educación, Desarrollo Humano y Gestión del Conocimiento.

## **Selección de variables para el modelo lasso bayesiano de emisiones CO<sub>2</sub>**

Gladys Linares Fleites, Departamento de Investigación en Ciencias Agrícolas,  
Instituto de Ciencias - BUAP

Arturo Abraham López Pérez, Facultad de Ciencias Físico Matemáticas BUAP

El tratamiento bayesiano de la regresión lasso corresponde a la moda posterior cuando una distribución inicial independiente doble-exponencial es asignada a los coeficientes de regresión. Sin embargo, generalmente no considera la selección de variables. La incertidumbre inherente a la especificación de un modelo de regresión puede ser incorporada en el análisis Bayesiano de la regresión lasso a través de la especificación de una distribución inicial sobre el espacio de modelos. En este trabajo se recopila el análisis del modelo de regresión lasso bayesiano considerando su incertidumbre y se presenta un ejemplo de aplicación sobre emisiones de CO<sub>2</sub>, aproximado con el paquete `blasso`, disponible en R.

## **Estimación de la densidad bivariada bajo el enfoque bayesiano no paramétrico**

Gladys Linares Fleites, Miguel Ángel Valera Pérez y Maribel Castillo Morales  
Instituto de Ciencias, Benemérita Universidad Autónoma de Puebla

La estadística bayesiana no paramétrica brinda nuevas posibilidades a los investigadores de las Ciencias Ambientales, permitiendo la estimación bivariada de la densidad del Carbono según regímenes de humedad y temperatura de los suelos. El Carbono Orgánico del Suelo (COS) es uno de los componentes fundamentales en el estudio del Secuestro de Carbono, una de las formas de mitigación del Cambio Climático. Es conocido que parámetros climáticos tales como la temperatura y la precipitación se encuentran fuertemente ligados al contenido del COS. Utilizando el paquete `DPpackage` en R se desarrolló la inferencia bayesiana vía simulación de la distribución posterior bivariada del carbono con la temperatura y con la precipitación, considerando mezclas del Proceso de Dirichlet para la distribución inicial y se obtuvieron gráficos e información predictiva de gran utilidad para la toma de decisiones sobre el almacenamiento de carbono en suelos de la región Teziutlán, Puebla, México.

---

## Aplicación de tratamientos de amplio espectro para una población mixta

Marisol López Cerino y José Del Carmen Jiménez Hernández

Instituto de Física y Matemáticas, Universidad Tecnológica de la Mixteca

Las migraciones están estrechamente relacionadas con el desarrollo de las enfermedades del tipo infeccioso, entre las que se encuentran las parasitosis intestinal (huéspedes invasores). El inmigrante es considerado como un foco potencial para el desarrollo de las mismas, ya sea por las infecciones que trae consigo desde su lugar de origen o por las condiciones de vida en la ciudad receptora que puede favorecer el desarrollo de dichas infecciones. Se realizó un estudio para determinar la prevalencia de los huéspedes invasores en una muestra de inmigrantes del estado de Oaxaca. Se entrevistaron 30 personas y se le administró un tratamiento de amplio espectro. El estudio fue de tipo caso-control encontrando que el parásito intestinal más común es *Entamoeba Histolytica* que se encuentra entre las diez infecciones más comunes reportadas por el IMSS.

## Aplicación del Elastic Net LASSO y modelos relacionados en selección asistida por marcadores *moleculares*

Marco Antonio López Cruz y Paulino Pérez Rodríguez

ISEI, Colegio de Postgraduados

José Crossa

CRIL, CIMMYT Int

Gustavo De Los Campos

University of Alabama at Birmingham, Section on Statistical Genetics

El Elastic Net Bayesiano (BEN), es un método de regresión que utiliza una mezcla de las penalizaciones L1 y L2 (Qing y Lin, 2010; Kyung et al 2010). Se ha demostrado que este modelo puede ser usado cuando el tamaño de muestra es mucho menor que el número de predictores ( $n \ll p$ ). En este trabajo se muestra como utilizar este modelo para incluir de forma conjunta marcadores moleculares (MM) y pedigree, ampliamente utilizados en genética cuantitativa en la llamada selección asistida por MM. El poder predictivo del BEN se compara

con el de otros modelos: LASSO Bayesiano, Regresión Ridge Bayesiana, etc., por medio de validación cruzada usando datos reales de rendimiento de cultivares trigo e índice de masa corporal de ratones. Los resultados muestran que el BEN tiene un poder predictivo igual o superior que el resto de los modelos.

### **Distribución geográfica de la mortalidad por tumor maligno de la mama y cervicouterino en el Estado de México: 2005-2009**

Miguel Ángel López Díaz, Enrique Castañeda Alvarado y Fernando Orozco Zitli  
Facultad de Ciencias, UAEMex

En estudios epidemiológicos el empleo de mapas para representar la distribución geográfica de enfermedades o decesos permite generar hipótesis acerca de las posibles causas en la variación del riesgo relativo así como la identificación de clústers. Los enfoques tradicionales para la construcción de mapas de mortalidad se basan en la categorización de la distribución de las tasas de mortalidad estandarizadas. Sin embargo, estos enfoques no permiten estimar a la distribución de probabilidad subyacente. Recientemente, los modelos de mezclas finitas ofrecen una herramienta, simple y efectiva, para determinar clústers. En este trabajo se presentan dos métodos estadísticos Bayesianos que permiten construir mapas de mortalidad considerando la dependencia espacial. Los métodos son comparados a través del análisis de la mortalidad por cáncer de mama y cervicouterino en el Estado de México entre los años 2005 y 2009. Al final de la plática también se presentan los clústers de mortalidad más elevada empleando técnicas como el Templado Simulado

## **Una aplicación de los Vectores Autorregresivos para estudiar los factores determinantes de los salarios en la maquila y en la manufactura**

Julio López Gallardo

Facultad de Economía-UNAM

Armando Sánchez Vargas

Instituto de Investigaciones Económicas-UNAM

Alberto Contreras Cristán y Miguel Ángel Chong Rodríguez

IIMAS-UNAM

Los Vectores autorregresivos (VAR) introducidos por Sims (1980), son una extensión multi-dimensional de los modelos Autorregresivos unidimensionales y son utilizados para capturar la evolución y la interdependencia entre múltiples series de tiempo. Estos modelos pueden tener sentido y ser útil para hacer interpretaciones aún si las componentes del vector no son estacionarias de segundo orden, sino que tengan una tendencia estocástica común. Una herramienta para examinar si las componentes del VAR tienen una tendencia común es la prueba de cointegración de Johansen (1991). En este trabajo estimamos un modelo VAR para estudiar el comportamiento de los salarios de dos sectores de la economía en México (el manufacturero y la industria de la maquila) y encontramos que los salarios en los dos sectores están determinados por un conjunto similar de variables y están relacionados funcionalmente.

### **Análisis de máximos para datos espaciales de lluvias**

Ehyter Martín y Joaquín Ortega

Depto. de Probabilidad y Estadística, CIMAT A.C.

En este estudio se analizan los datos máximos de lluvias correspondientes al Estado de Guanajuato utilizando la teoría de valores extremos y los modelos espaciales para extremos multivariados propuestos por Smith y Schlather en sus trabajos de 1990 y 2002, respectivamente. El análisis se hizo considerando valores máximos mensuales y ajustando los modelos de Smith y Schlather a los valores máximos correspondientes a cada mes del año, con fines

comparativos. En total se tienen 24 modelos ajustados. En esta plática se presentarán algunos de estos modelos, así como las interpretaciones y debilidades correspondientes a cada uno.

### **Comparación de técnicas de minería de datos y componentes principales: El modelo de árboles de decisión**

Juan Carlos Medina Martínez, Lorena Alonso Guzmán y León Julio Cortez Organista  
Unidad Académica de Ingeniería, Universidad Autónoma de Guerrero

En esta investigación se presenta un comparativo de técnicas de manejo de datos aplicadas a la búsqueda de patrones para determinar el patrón espacial del maguey papalote, se destaca el uso de minería de datos como técnica de análisis de datos, ya que permite con la construcción de redes de interacción biótica representar una herramienta potente que permita predecir factores para la producción y declinamiento del maguey papalote; otra técnica es el Análisis de Componentes Principales, la cuál es una técnica en estadística multivariada muy utilizada para reducir variables sin la pérdida de información; para el desarrollo de estas técnicas se utilizan los resultados de la investigación “Desarrollo de un Sistema de Inventario y Monitoreo de Maguey Papalote (*Agave cupreata* Trel et Berger)”, (SIMMP), para conocer patrones ambientales que nos indiquen la etapa a labrar de la planta y con esto a conservar el maguey papalote en el Estado de Guerrero.

### **Estimación de una tabla de mortalidad regionalizada**

Miguel Medina Pinto y Jorge A. Argáez Sosa  
Facultad de Matemáticas, Universidad Autónoma de Yucatán

Las tablas de mortalidad son la herramienta principal para el cálculo del costo de un seguro de vida. En México se encuentran en vigor las Tablas CNSF 2000-I y 2000-G. Sin embargo, una pregunta es: ¿Por qué debe usarse la misma tabla para calcular el costo de un seguro de vida en cualquier parte del país, cuando las condiciones de vida no son homogéneas? En este trabajo se estudia el comportamiento de la tasa de mortalidad para la zona de la península

de Yucatán. Asimismo, se presenta una manera para estimar la tabla de vida correspondiente utilizando para ello el enfoque de estadística Bayesiana. Finalmente, se discute cómo, usando la forma de modelar que se propone, se pueden obtener tablas de mortalidad para diferentes escenarios que reflejen comportamientos diferentes de la tasa de mortalidad.

### **Estimación en áreas pequeñas, caso de pobreza**

Ignacio Méndez Ramírez, IIMAS, UNAM

Hortensia Moreno Macías, Economía/Área de modelación y sistemas, UAM-I

Se plantea una forma de efectuar estimaciones en “áreas pequeñas”. La metodología propuesta se considera útil, cuando se tenga una muestra probabilística planeada para estimar variables de interés en toda una población. Se consideran dominios o subconjuntos que se forman al interior de la población con el uso de conglomerados. El objetivo es estimar promedios de las variables de interés en esos dominios de la población. Lo usual es que al considerar la muestra inicial, los tamaños de muestra que ocurren en los dominios son pequeños o nulos, las “áreas pequeñas”. Además se cuenta con datos censales de variables auxiliares para las áreas pequeñas y para la población. En la propuesta de investigación, como un ejemplo, se utiliza la estimación de dos variables de pobreza en cada uno de los 210 municipios de Veracruz. Los resultados son muy satisfactorios, tener estimadores de promedios con intervalos de confianza al 95%, suficientemente pequeños, para todos los municipios. Este problema se trabajó mediante un contrato con el Consejo Nacional de Evaluación de la Política Social, CONEVAL, y con la participación de El Colegio de México.

### **Familias exponenciales y distribuciones iniciales**

Manuel Mendoza Ramírez, Depto. de Estadística, ITAM

Eduardo Gutiérrez Peña, Depto. Probabilidad y Estadística, IIMAS-UNAM

Las familias exponenciales de distribuciones han sido ampliamente estudiadas en la literatura pero siguen ofreciendo un campo fértil para la investigación. En este trabajo se examina este

tipo de modelos desde la perspectiva de optimización. Se muestra como, cuando se considera una medida de información específica, estos modelos pueden considerarse una solución apropiada para describir fenómenos aleatorios. En una dirección relacionada, también se muestra que, con un argumento similar, se justifica el empleo de distribuciones conjugadas para el análisis Bayesiano y como se pueden inducir distribuciones iniciales no informativas.

### **Evaluación de la interacción gen-ambiente para estudios familiares con fenotipos cuantitativos: Una revisión y extensión a estudios longitudinales**

Hortensia Moreno Macías, Economía/Área de modelación y sistemas, UAM-I

Isabelle Romieu, International Agency for Research on Cancer

Stephanie London, Epidemiology Branch, National Institute of Environmental Health Sciences, NIH

Nan Laid, Department of Biostatistics, Harvard School of Public Health

Los estudios longitudinales son una herramienta importante para analizar fenotipos que cambian en el tiempo dependiendo de características individuales y exposiciones ambientales. En poblaciones latinas es importante tomar en cuenta la heterogeneidad genética debido a la mezcla de poblaciones. En este trabajo se propone una metodología cuyo objetivo es proporcionar estimadores insesgados de los efectos marginales de factores genéticos y ambientales así como de su interacción a través de la combinación de las ventajas de los estudios longitudinales y los diseños basados en familias. Usando simulaciones, el método es aplicado a un estudio de cohorte donde la respuesta es la función respiratoria de niños asmáticos mexicanos expuestos a altos niveles de ozono y que aleatoriamente recibieron un placebo o un complemento antioxidante.

## **Paquete R para estimación de densidades y clasificación**

Luis E. Nieto Barajas y Ernesto Barrios

Depto de Estadística, ITAM

Igor Pruenster

University of Turin

En esta plática se presenta la implementación en R de un modelo para estimación de densidades. El modelo se basa en la inferencia Bayesianas no paramétrica y en particular en mezclas de procesos de incrementos independientes normalizados. Se dará énfasis en el uso del paquete y la interpretación de resultados.

## **Una cota para el sesgo relativo del efecto del diseño**

Alberto Manuel Padilla Terán, Banco de México

El estimador del efecto del diseño, Kish (1965), se emplea comúnmente en los diseños muestrales complejos para el cálculo del tamaño de muestra, así como para la construcción de intervalos de confianza, entre otros usos. En este trabajo se muestra que dicho estimador es sesgado y se exhiben, tanto el sesgo del estimador, como una cota para el sesgo relativo. Se ilustrará con algunos ejemplos la magnitud del sesgo y del sesgo relativo.

## **Análisis de sendero y componentes principales aplicado en un experimento con chile serrano**

Emilio Padrón Corral<sup>1</sup>, Ignacio Méndez Ramírez<sup>2</sup>, Armando Muñoz Urbina<sup>1</sup>, Haydée De La Garza Rodríguez<sup>1</sup>, y Manuel Antonio Torres Gomar<sup>1</sup>

<sup>1</sup>Facultad de Ciencias Físico-Matemáticas, Universidad Autónoma de Coahuila

<sup>2</sup>IIMAS, UNAM

Cinco genotipos de chile serrano (*Capsicum annuum* L.) y sus 10 cruzas posible fueron evaluadas en el campo experimental del sur de Tamaulipas (CESTAM) del Instituto Nacional

de Investigaciones Agrícolas y Pecuarias (INIFAP), ubicado en la Estación Cuauhtémoc, Tamaulipas, México, durante el ciclo invierno-primavera del 2006-2007, Las variables evaluadas fueron: rendimiento, número de frutos por planta, peso individual de fruto, altura de la planta, días a inicio de floración, días a inicio de cosecha, longitud de fruto y diámetro de fruto. El objetivo del presente trabajo fue: Efectuar un análisis de coeficientes de sendero para estudiar las relaciones de causa y efecto entre los componentes del rendimiento, así como efectuar un análisis de componentes principales para estudiar los patrones de interrelación entre las variables evaluadas.

### **Pruebas de bondad de ajuste bajo el esquema de muestras en conjuntos ordenados**

Francisco Javier Pérez Arredondo, Departamento de Economía y Finanzas, Universidad de Guanajuato

Adolfo José Quiroz Salazar, Universidad Simón Bolívar (Venezuela)

Se presentan resultados teóricos y empíricos de cuantiles y potencia de cuatro pruebas clásicas de bondad de ajuste en el caso de que las observaciones provengan de muestreo en conjuntos ordenados (ranked set sampling) en poblaciones normales. Se comparan con los resultados de las mismas pruebas bajo muestreo aleatorio simple.

### **Regresión espacial de precipitaciones extremas en el estado de Tabasco**

Rebeca Alejandra Pérez Figueroa y Sergio Pérez Elizalde

ISEI, Colegio de Postgraduados

En México se ha creado una malla a lo largo de todo el territorio, creando con esto una serie de nodos los cuales contienen información meteorológica del país. En este trabajo se propone un modelo espacial para valores extremos empleando un proceso espacial latente de variación de parámetros empleando información de precipitaciones máximas mensuales en el área del estado de Tabasco, estado que se ha visto sumamente afectado por las inundaciones extraordinarias en años pasados.

## **Métodos estocásticos para determinar la región de rechazo de pruebas clínicas de no inferioridad**

Iván Pérez Laguna, ISEI, Colegio de Postgraduados

Gabriel A. Rodríguez Yam, Universidad Autónoma Chapingo

Javier Suarez Espinoza, ISEI, Colegio de Postgraduados

Las pruebas exactas de no inferioridad están diseñadas para averiguar si un tratamiento experimental no es “muy inferior” a un control dentro de un margen de no inferioridad. Dicho margen es usualmente pequeño, conocido y fijo. Excepto cuando la región de rechazo cumple con la condición de convexidad de Barnard, la construcción de la región de rechazo de estas pruebas es computacionalmente intensivo. La condición de convexidad de Barnard es a veces, sólo verificable numéricamente, limitando así el uso de algunos estadísticos de prueba propuestos en la literatura. En este trabajo se implementará un método de optimización estocástico para encontrar la región de rechazo de estas pruebas, que como se notará, no exige que la región de rechazo cumpla la condición de convexidad de Barnard.

## **Un modelo probabilístico para el metro de la ciudad de México**

Blanca Rosa Pérez Salvador y Patricia Saavedra Barrera

Depto. de Probabilidad y Estadística, UAM-I

Estudiar al Metro de la ciudad de México utilizando un modelo probabilístico, representa un reto por la complejidad de los elementos que lo conforman. Por un lado, están los trenes que se mueven a lo largo de los rieles, deteniéndose en cada andén para dejar y recoger pasajeros. Por otro lado se tiene a los usuarios que llegan a los andenes para abordar los trenes desde el exterior, o de otras líneas cuando la estación objetivo es de correspondencia. En este trabajo se revisará un modelo probabilístico mediante el cual se describen los movimientos de los trenes y de los usuarios, así como la relación entre estos dos movimientos. Este modelo sirve de base para la elaboración futura de un micro simulador.

## **Uso de transformaciones para probar interacción en experimentos factoriales con variable respuesta Poisson**

Gustavo Ramírez-Valverde, ISEI, Colegio de Postgraduados

Benito Ramírez-Valverde, Colegio de Postgraduados

Experimentos con variable respuesta Poisson son frecuentes. Siendo la variable respuesta Poisson, entonces se violan los supuestos del ANOVA. Para solucionar el problema se usan transformaciones. Otra alternativa es el modelo lineal generalizado con distribución Poisson y función liga logarítmica, el cual cumple con el supuesto distribucional. Sin embargo, el análisis se basa en máxima verosimilitud y en pruebas asintóticas. En los diseños experimentales típicamente se tienen muestras pequeñas, por lo que es necesario explorar su comportamiento con muestras pequeñas. En este trabajo, con un estudio de simulación se compara la potencia y el tamaño de la prueba para probar interacción en un factorial con dos factores utilizando: a) ANOVA b) ANOVA con transformaciones, y c) Modelo Lineal generalizado con función de liga logarítmica.

## **Modelo de ecuaciones estructurales para determinar las causas que afectan el rendimiento académico en estudiantes de bachillerato**

Ramón Reyes Carreto, Francisco Julián Ariza Hernández y Flaviano Godínez Jaimes

Unidad Académica de Matemáticas, Universidad Autónoma de Guerrero

El mejoramiento de la calidad académica usa como estrategia la evaluación del rendimiento del alumno. Diversos estudios enfatizan que el rendimiento está determinado por factores personales y contextuales. Este trabajo analiza factores socioeconómicos, institucionales, motivacionales, aptitudinales y rendimiento académico de alumnos de bachillerato para identificar relaciones causales. Se utilizan indicadores de trayectoria escolar del bachillerato e información del CENEVAL. Se aplica un modelo de ecuaciones estructurales para mejor comprensión de la complejidad de las relaciones causales entre los factores. Se concluye que el factor aptitudinal influye significativamente sobre el factor rendimiento; el factor motivacional influye sobre el factor aptitudinal; y el factor motivacional también tiene efecto significativo sobre el rendimiento.

## **Coefficiente de determinación para la regresión polinomial local en el muestreo de poblaciones finitas**

José Elías Rodríguez Muñoz, Depto. de Matemáticas, Universidad de Guanajuato

Luis Alejandro Escobar López, Universidad Autónoma de Aguascalientes

Para valorar si un modelo estimado es adecuado, se requiere de una medida de la bondad de ajuste. Una medida de esta bondad de ajuste para los modelos de regresión está dada por el coeficiente de determinación. Para la regresión polinomial local existe en la literatura y en un contexto general una propuesta para estimar dicho coeficiente de determinación. En este trabajo proponemos una forma de estimar el coeficiente de determinación para la regresión polinomial local en el contexto de muestreo de poblaciones finitas. La propuesta contempla la estimación del coeficiente a nivel global y local. Además las características de este estimador se muestran a través de algunos experimentos de simulación sobre poblaciones artificiales y reales.

## **Análisis del robo de vehículos: Una aplicación de las series de tiempo**

Alejandro Román Vásquez, Blanca Rosa Pérez Salvador y Carlos Omar Jiménez Palacios

Departamento de Matemáticas, UAM-I

En la actualidad el robo de vehículos es un fenómeno social que además de las consecuencias que por sí mismo conlleva el delito, afecta a la sociedad por el encarecimiento de los costos de los seguros. Este delito lesiona agresivamente al sector asegurador, por lo que su análisis es de vital importancia. En este trabajo se presenta, utilizando un modelo ARIMA, el análisis de la serie mensual del robo de vehículos asegurados en México en el periodo de enero 2000 a diciembre 2010. Este análisis se realiza con dos propósitos: medir la calidad de la serie a través de la detección de datos aberrantes y pronosticar la serie a corto y mediano plazo. La intención es que los resultados obtenidos sean empleados para complementar las técnicas actuales en el cálculo de primas de riesgo.

## **Técnicas de interpolación espacial aplicadas mediante un sistema de información geográfica para el manejo de los recursos hídricos superficiales**

Luis Alfonso Ruiz Ochoa<sup>1</sup>, Javier Martin Heredia<sup>2</sup>, José Alberto Martin Ruiz<sup>3</sup>, Trinidad Cruz Sánchez<sup>1</sup> y

Marcos Andrés Jiménez Moreno<sup>1</sup>

<sup>1</sup>Universidad Tecnológica de Tabasco, <sup>2</sup>Colegio de Bachilleres de Tabasco, <sup>3</sup>UNAM

Una mejor interpretación de los datos y el análisis de modelos son cualidades de los sistemas de información geográfica. Se creó un sistema que facilitara la consulta, el análisis y el despliegue de gráficos para una mejor interpretación de la información, como herramienta de apoyo para el aprovechamiento de los recursos hídricos superficiales que convergen en el área de influencia del distrito Zanapa-Tonalá, en Tabasco. Para ello se utilizó el software GvSig, a partir del cual se generaron capas de información tanto geográfica como hidrológica del distrito, el sistema desarrollado muestra información clasificada en una geodatabase para fines de planeación y de predicción, con las cuales se aplicaron técnicas de interpolación espacial que permitieron obtener mapas de distribución de variables hidrológicas, estas permitirán conocer los principales mecanismos de transferencia de agua.

## **Eficiencia relativa y post-análisis de los diseños experimentales en caña de azúcar**

Juan Ruiz Ramírez, Fac. de Economía, Universidad Veracruzana

Luis Cruz Kuri, Fac. de Ciencias Admvas. y Sociales, Universidad Veracruzana

Bertha Alicia Arce Castro, Fac. de Ciencias Admvas. y Sociales, Universidad Veracruzana

Agustín Jaime García Banda, Universidad Veracruzana

Ismael Sosa Galindo, Fac. de Ciencias Admvas. y Sociales, Universidad Veracruzana

En la experimentación agrícola, comúnmente se utiliza el diseño de bloques al azar (DBA) y no se calcula la eficiencia relativa (ER) y por consecuencia, no se realiza el post-análisis con el diseño completamente al azar (DCA) cuando es requerido. Es por ello que se plantea el siguiente objetivo: “Calcular la ER del DBA vs. el DCA al azar y realizar el post-análisis

estadístico”. Para ello se analizaron 12 experimentos de fertilización en caña de azúcar con el SAS y se calculó la ER con el software “Eficiencia Relativa”. Se obtuvo que el 42% de ellos aplicaron el DBA de manera incorrecta, por lo que se tuvieron que post-analizar con el DCA, obteniéndose valores de probabilidad (p-values) y CV, menores o iguales a los del DBA. Se concluye que es importante calcular la ER y así obtener resultados de investigación confiables.

### **Pronósticos de poblaciones con series de tiempo multivariadas: el caso de la ZMCM contrastado con datos del censo 2010**

José Eliud Silva Urrutia, Actuaría, UNAM

Manuel Ordorica, El Colegio de México

Con base a metodologías de series de tiempo multivariadas, se muestra cómo es factible realizar pronósticos de población y se evidencia cómo estos son satisfactorios de acuerdo con datos censales observados para la Zona Metropolitana de la Ciudad de México, tomando en consideración los resultados del Censo de Población y Vivienda 2010. Se enuncia de manera general la metodología empleada para realizar los pronósticos, partiendo de técnicas de desagregación, ajustes y la elaboración de los pronósticos restringidos. También se mencionan pruebas de compatibilidad estadísticas, como herramienta objetiva en la formulación de planes de crecimiento demográfico. Se considera que tener herramientas metodológicas capaces de proyectar de manera eficiente, resultan indispensables para la elaboración de planes que tengan por objeto, entre otras cosas, sociedades en armonía con el medio ambiente.

## **Muestreo de respuesta aleatorizadas con probabilidades desiguales: el estimador de Rao-Hartley-Cochran**

Víctor H. Soberanis Cruz, y Jaime D. Cuevas Domínguez  
Universidad de Quintana Roo

La técnica de Respuestas Aleatorizadas (RR) ha sido introducida para reducir el riesgo de evasión o no respuesta de preguntas sensitivas en encuestas con muestreo para estimar el total de individuos con alguna característica sensitiva. Reconociendo que las estrategias de alta eficiencia en la estimación para estos casos requieren de variables auxiliares fuertes, en este trabajo se propone el estimador de Rao-Hartley-Cochran (RHC), tanto para el Modelo de Warner, como para el Modelo de Greenberg. Asimismo, se muestra, mediante simulación, que el estimador RHC que del Modelo de Greenberg resulta más eficiente que el del Modelo de Warner.

## **Aspectos estadísticos relacionados con el cálculo de la huella ecológica en la ciudad de Xalapa, Veracruz**

Ismael Sosa Galindo, Bertha Alicia Arce Castro,  
Agustín Jaime García Banda y Luis Cruz Kuri  
Fac. de Ciencias Administrativas y Sociales, Universidad Veracruzana  
Juan Ruiz Ramírez  
Fac. de Economía, Universidad Veracruzana

La Huella Ecológica del Agua constituye una metodología que se ocupa en varios países, tales como EE.UU. e Inglaterra. En este trabajo se describe una adaptación de esos resultados a la ciudad de Xalapa, Veracruz, para estimar la cantidad total de agua que una vivienda o una persona utilizan. Para tales propósitos se recaba la información pertinente a las viviendas o a los individuos acerca del uso del agua y de los hábitos de sus habitantes. Se considera no tan solo el agua utilizada en la vivienda, sino también el agua para producir los alimentos que se consumen, productos que se compran, y el agua ahorrada para reciclar. Constituye un seguimiento de una encuesta reciente en Xalapa, donde a los hogares seleccionados se les

---

aplicó un cuestionario sobre cantidades utilizadas de agua tanto para ingestión individual y aseo personal como para actividades relacionadas con el consumo de alimentos, entre otros.

### **Desempeño escolar y “bullying” en escuelas de educación media superior**

Belem Trejo Valdivia y Airaín Montoya Rodríguez

Centro de Investigación en Evaluación de Programas y Bioestadística, Instituto Nacional de Salud Pública

El bajo rendimiento escolar es un problema en México y se requiere identificar los factores que lo determinan. Se ha reportado un aumento en la frecuencia y gravedad del bullying (manifestación de violencia) así como su íntima relación con el desempeño escolar. No hay estudios en México sobre esta asociación. Se analizó la información de la 2ª Encuesta Nacional sobre Exclusión, Intolerancia y Violencia en las Escuelas de Educación Media Superior, para identificar y evaluar los principales patrones de asociación entre el desempeño escolar y el bullying, considerando factores individuales/familiares/escolares. El enfoque fue por ecuaciones estructurales para determinar la relevancia de asociaciones directas e indirectas entre el promedio escolar y bullying, junto con sexo, edad, nivel socioeconómico, autoestima y percepción del ambiente escolar.

### **Tipologías en los hábitos de estudio del estudiante universitario, construidas en base a técnicas de análisis multivariante**

Victoria Velásquez de la Cruz, Claudio Rafael Castro López y Ma. de los Ángeles Silva Mar  
Universidad Veracruzana

Un hábito es cualquier acto adquirido por la experiencia, una práctica o costumbre que se realiza con frecuencia. El Proyecto institucional “Hábitos de Estudio del Estudiante Universitario” que se lleva a cabo en la región Poza Rica-Tuxpan de la Universidad Veracruzana, tiene la finalidad estudiar comportamientos en los estudiantes frente a sus hábitos de estudio. Se realizó una encuesta a 2,450 estudiantes de diversas carreras para conocer estos comportamientos. En este trabajo se presentan resultados de tipologías en los hábitos de estudio creadas a partir de un Análisis de Clases Latentes, de igual forma, aplicamos la técnica de

Análisis de Regresión Logística para obtener un modelo que determine las variables relevantes de la pertenencia de los individuos de cada tipología creada.

### **Identificabilidad y comparación de métodos de estimación para mezclas Weibull univariadas**

Olga Vladimirovna Panteleeva y Humberto Vaquera Huerta

ISEI, Colegio de Postgraduados

Eduardo Gutiérrez González, UPIICSA-IPN

Los modelos de mezclas finitas de distribuciones Weibull son estudiados mostrando sus propiedades generales. Se prueba la identificabilidad del modelo y se comparan las estimaciones de los parámetros para el caso de dos componentes. Los estimadores de los parámetros se buscan por Máxima Verosimilitud calculados bajo diferentes algoritmos: EM, Fisher *scoring*, *backfitting*, optimización de los vecinos más cercanos y un último algoritmo de recorridos aleatorios a través de simulación Monte Carlo. Por medio de simulación se calcula y compara el sesgo y el error cuadrado medio de los estimadores por tres métodos.

### **Un modelo de regresión logística ordinal para el análisis de deslizamiento de laderas**

Martín Zúñiga Gutiérrez

Unidad Académica de Ingeniería, Universidad Autónoma de Guerrero

Francisco Julián Ariza Hernández, Ramón Reyes Carreto y Flaviano Godínez Jaimes

Unidad Académica de Matemáticas, Universidad Autónoma de Guerrero

Los problemas de deslizamientos en laderas sin duda constituyen uno de los desastres geológicos más destructivos a nivel mundial, causando pérdidas económicas de varios miles de millones de pesos y también lamentables pérdidas humanas. En este trabajo de investigación se usa un modelo de regresión logística ordinal para el análisis de la amenaza de riesgo por deslizamiento de laderas en la Cd. de Chilpancingo, Guerrero. Se considera el tipo de

amenaza como variable respuesta con cinco categorías: muy baja, baja, moderada, alta y muy alta. Las covariables que se integran al modelo representan factores topográficos, históricos, geotécnicos, geomorfológicos y ambientales, todos éstos considerados por Protección Civil del Estado de Guerrero.



# Trabajos del Concurso Nacional de Estadísticos Jóvenes (patrocinado por SAS)

## Contaminación del aire

Israel Emmanuel Ambriz Lobato

FES-Acatlán, UNAM

*Mención Honorífica Licenciatura*

Es de suma importancia generar una estrategia para reducir los niveles de contaminación del aire, grave problema que no solo afecta a las grandes ciudades del mundo, ya que esta depende de muchos factores que se pueden presentar en distintas regiones. Para poder lograrlo es necesario conocer de antemano el comportamiento histórico de las variables contaminantes y cómo interactúan con el entorno de estudio. En este trabajo se proponen inicialmente métodos de imputación múltiple en caso de que exista una cantidad considerable de valores perdidos en las variables, lo cual no es raro en variables que representen contaminantes o condiciones atmosféricas. Con la finalidad de encontrar patrones de comportamiento conjunto entre las variables atmosféricas y contaminantes, se hace uso de medidas de concordancia (discordancia) y dependencia, siendo éstas de gran utilidad al momento de fomentar acciones sociales para reducir los niveles de contaminación. Al llevar a cabo acciones sociales, es fundamental un monitoreo constante de los niveles contaminantes para precisar si dichas acciones han tenido impacto o no, para lo anterior, se hizo uso de ANOVA no paramétrico analizando si existía variación en la escala, forma y localización de las variables en el tiempo de estudio. Al analizar los resultados anteriores se determina una estación de monitoreo como crítica para la salud y se exaltan datos alarmantes en las demás estaciones. Conocer el comportamiento futuro de las variables ayuda a prever y controlar el impacto de éstas en la población. Para finalizar se propone un pronóstico a un año del Ozono, variable que

presentó niveles alarmantes durante todo el periodo de estudio. El desarrollo del trabajo se llevó a cabo con procedimientos (PROC) de SAS tales como, PROC MI, PROC MEANS, PROC TRANSPOSE, PROC FREQ, PROC CORR y PROC NPAR1WAY, con ayuda de SAS 9.2 y SAS Enterprise Guide 4.2.

### **Estadística y Medio Ambiente**

Ángel Manuel Godoy Aguilar

IIMAS, UNAM

*2º Lugar Posgrado*

En este trabajo, presentado en el Concurso Nacional de Estadísticos Jóvenes en la Categoría de Posgrado, se aborda el análisis de un conjunto de datos proporcionados por el Sistema de Monitoreo Atmosférico de la Ciudad de México, referentes a la presencia de contaminantes en la atmósfera de ésta ciudad. Dichos contaminantes son consecuencia de actividades cotidianas de los habitantes, condiciones geográficas y climatológicas de la región, afectando de manera directa la salud de las personas y el ecosistema. Con la finalidad de comprender el impacto ambiental de éstos contaminantes, se presenta evidencia obtenida mediante métodos estadísticos; para identificar zonas de riesgo, relación entre contaminantes y variables atmosféricas, cambio en los niveles registrados y predicciones para algunas variables de interés.

### **La calidad del aire de la Ciudad de México**

Carlos González Munguía y Lany Muñoz Mota

Facultad de Ciencias, UNAM

*2º Lugar Licenciatura*

Se analizó la base de datos del SIMAT en el periodo del 2000 al 2009 de las estaciones atmosféricas Tlalnepantla, Merced, San Agustín, Pedregal y UAM Iztapalapa. Ya que se tenía más del 10% de datos faltantes se realizó imputación de cada dato faltante con el valor obtenido de realizar una regresión polinomial por estación meteorológica y después se hizo el

análisis estadístico descriptivo e inferencial para examinar los datos asociados en el contexto de medio ambiente y contaminación atmosférica. Se crearon modelos estadísticos y realizaron las pruebas necesarias para justificar cada uno de los siguientes aspectos: a) Descripción de las diversas variables en el estudio. b) La relación existente entre las variables atmosféricas y contaminantes. c) Cambios en los niveles de los contaminantes en el periodo de estudio. d) Pronóstico de las variables contaminantes. e) Identificación de las estaciones de monitoreo críticas, según normas ambientales vigentes.

**Estudio de los indicadores de la calidad del aire en la zona metropolitana del Valle de México de enero del 2000 a diciembre de 2009**

Fernando Daniel Pérez Arriaga y Oscar Ortega Ibáñez

IIMAS-UNAM

*Mención Honorífica Posgrado*

En este trabajo se presenta un análisis estadístico del comportamiento de los distintos contaminantes y variables atmosféricas de la zona metropolitana del valle de México monitoreados en cinco de los 34 sitios del SIMAT (Tlalnepantla, San Agustín, Merced, Pedregal y UAM-Iztapalapa). Esto tiene como finalidad evaluar las políticas y acciones emitidas por los distintos programas para la mejora de la calidad del aire (PROAIRE II) analizando tendencias, estableciendo relaciones entre los contaminantes y las variables meteorológicas y determinando estaciones de monitoreo críticas. Con base en los resultados obtenidos concluimos que los efectos de dichas políticas son positivos ya que las emisiones de CO, NO<sub>2</sub>, NO<sub>x</sub> y O<sub>3</sub> han disminuido considerablemente. Sin embargo, el análisis del clima no basta para determinar los niveles de los contaminantes en las distintas zonas de monitoreo. Aunado al decremento en los niveles de estos contaminantes, el número de ocasiones en las que los contaminantes exceden los límites establecidos por las Normas Oficiales Mexicanas de Salud Ambiental es cada vez menor. Pese a esto, se han establecido zonas de monitoreo críticas: San Agustín y Tlalnepantla, pues afectan de forma directa la concentración de algunos contaminantes en otros sectores debido a las características geográficas de la ZMVM.

## **Análisis estadístico de la calidad del aire**

María José Urcelay Novelo e Ignacio Puch Puc

Facultad de Matemáticas, Universidad Autónoma de Yucatán

*Mención Honorífica Licenciatura*

Los datos analizados en este trabajo provienen de mediciones de ciertos contaminantes en la atmósfera en diferentes áreas de la ciudad de México, recolectados durante los últimos 10 años, de manera diaria y a cada hora. La primera parte de este trabajo consta de un análisis exploratorio de dichos contaminantes y algunas variables atmosféricas que se miden también en las estaciones meteorológicas, usando histogramas y medidas de dispersión y centralización, con el objetivo de observar el comportamiento de estas variables. Luego se usaron modelos de regresión lineal múltiple, para probar la existencia de asociación entre cada contaminante y las variables atmosféricas, y en caso afirmativo medir el grado de asociación. Además se compararon el comportamiento de los contaminantes en el periodo en el que no se tenían implementados programas ambientales en la zona contra el periodo en el que ya se tenían programas. También se realizaron pronósticos puntuales y por intervalos, a corto plazo para el ozono y las partículas menores a 10 micrómetros para cada estación de monitoreo, usando análisis Bayesiano con distribución a priori no informativa. Finalmente se identificaron estaciones de monitoreo críticas.