

REVISTA ESTADISTICA

Vol. 63 - NRO. 181 – 2011

CONTENIDO – CONTENTS

ANALISIS DE LA EVOLUCION DE LAS TÉCNICAS DE SERIES TIEMPO. UN ENFOQUE UNIFICADO

JUAN CARLOS ABRIL

*Consejo Nacional de Investigaciones, Científicas y Técnicas (CONICET),
Facultad de Ciencias Económicas, Universidad Nacional de Tucumán*

jabril@herrera.unt.edu.ar

RESUMEN

Los sucesos variables en el tiempo reciben el nombre genérico de series de tiempo o series cronológicas. Es extremadamente difícil presentar una descripción breve del campo de las series de tiempo. La dificultad se basa en el hecho de que la materia es por sí misma muy compleja, siendo una rama de la estadística pero con su metodología y su propio vocabulario peculiar. La idea básica de una *serie de tiempo* es muy simple, consiste en el registro de cualquier cantidad fluctuante medida en diferentes puntos del tiempo. La característica común de todos los registros que pertenecen al dominio de las “series de tiempo” es que ellos están influenciados, aunque sea parcialmente, por fuentes de *variación aleatoria*. Entonces, si deseamos explicar la estructura de las fluctuaciones en una serie de tiempo debemos recurrir a lo que llamamos el *estudio de las series de tiempo*. Hay dos aspectos en el estudio de las series de tiempo: el análisis y el modelado. El objetivo del análisis es resumir las propiedades de una serie y remarcar sus características salientes. La principal razón para modelar una serie de tiempo es para permitir la predicción de sus valores futuros.

En este trabajo se presenta una breve introducción al estudio de las series de tiempo seguido de un conjunto de ejemplos que ocurren en áreas tales como medicina, astronomía, economía, etc. Luego se introduce una pequeña historia del desarrollo de esta parte de la estadística, comenzando en 1664 con los trabajos de Sir Isaac Newton, hasta llegar al gran desarrollo experimentado en los últimos sesenta años.

Posteriormente se compara el enfoque de espacio de estado (EE) para el análisis de las series de tiempo con el enfoque ARIMA de Box-Jenkins (BJ). Luego de definir lo que se entiende y cómo trabajan los enfoques de EE y BJ, nos retrotraemos a los orígenes históricos del análisis moderno y aplicado de las series de tiempo conocido como suavizado exponencial que nació en los ‘50, y

tratamos de mostrar que ambos sistemas pueden ser considerados como que evolucionaron naturalmente desde esos orígenes. A continuación se realiza una comparación amplia de los méritos relativos de los dos sistemas, concluyendo en favor del enfoque de EE. Después de esto, se describen brevemente algunos trabajos recientes en donde se aplica el enfoque de EE. Finalmente se realizan algunas consideraciones sobre el uso potencial de los métodos de EE en el trabajo sobre series de tiempo en las estadísticas oficiales y se dan algunas líneas nuevas de desarrollo del área.

Palabras clave

ARIMA; Box-Jenkins; espacio de estado; estadísticas oficiales; series de tiempo; suavizado exponencial.

ABSTRACT

Time-varying events receive the generic name of time series. It is extremely difficult to present a brief description of the field of time series. The difficulty lies in the fact that the subject is in itself complex, being a branch of statistics but with its methodology and its own peculiar vocabulary. The basic idea of a time series is very simple, consisting of the registration of any fluctuating quantity measured at different points in time. The common feature of all records that belong to the domain of "time series" is that they are influenced, even partially, by sources of random variation. So if we wish to explain the structure of the fluctuations in a time series we use what we call the study of time series. There are two aspects of the study of time series: analysis and modeling. The objective of the analysis is to summarize the properties of a series and highlight its salient features. The main reason for a time series model is to allow the prediction of their future values.

In this paper we present a brief introduction to the study of time series followed by a set of examples that occur in areas such as medicine, astronomy, economics, etc. Then we present a short history of the development of this part of statistics, beginning in 1664 with the works of Sir Isaac Newton, until the great development experienced over the last sixty years.

Later we compares the state-space approach (SS) for the analysis of time series with the approach of Box-Jenkins ARIMA (BJ). After defining what is meant and how they work both SS and BJ approaches, we go back to the historical origins of modern and applied time series known as exponential smoothing which was born in the 50s, and try to show that both systems can be considered that they naturally evolved from those origins. Following that we give a comprehensive comparison of the relative merits of the two systems, concluding in favor of the SS approach. After that we briefly describe some recent work where the SS approach is applied. Finally, some considerations are made on the potential use of SS methods in the time series work of official statistics and provides some new lines for the development of the area.

Key words

ARIMA; Box-Jenkins; exponential smoothing; official statistics; state space; time series.

**DEMOGRAPHIC STUDY OF CUBAN BLUE LODGE
MASONS:
A TECHNICAL DISCUSSION**

JORGE LUIS ROMEU
Syracuse University, NY 13224
romeu@cortland.edu

GUSTAVO PARDO VALDÉS
Gran Logia de Cuba, La Habana, Cuba

ABSTRACT

Cuban Blue (or Symbolic) Lodge masons constitute one of the oldest, largest and geographically most widely spread organizations of Cuban civil society. We examine the evolution of such masons, using annual membership data from the confederation of all Blue Lodges (or Grand Lodge). Defining the equivalent of population at risk, a new Masonic indicator is derived. Using such indicator and Immigration and Naturalization Service data, we estimate the number of Cuban Blue Lodge masons that left Cuba in the wake of Castro's revolution. Using United Nations rates, we estimate deceased masons. Using annual membership trends at different epochs, we estimate the number of Communist Party and administration officers that may have joined the Cuban Blue Lodge masons. We thus describe the evolution of the members of the Grand Lodge, through the second half of the XX Century, and through the seven epochs of Cuban sociopolitical developments, that we have identified.

Key Words

Demographic study; statistical models; time series; Freemasons.

RESUMEN

Los masones cubanos de las Logias Azules (o Simbólicas) constituyen una de las organizaciones más antiguas, geográficamente más expandidas y más numerosas de la sociedad civil cubana. Examinamos la evolución de estos masones, utilizando datos anuales de membresía en la confederación de todas las Logias Azules (o Gran Logia). Definimos el equivalente de población en riesgo, como un nuevo indicador de estos masones. Utilizando este indicador y datos del INS (Servicio de Inmigración y Naturalización de EEUU), estimamos el número de masones de Logias Azules que emigraron tras la revolución de Fidel Castro. Utilizando tasas de Naciones Unidas, estimamos el número de masones fallecidos. Utilizando tasas de crecimiento de la membresía anual, en distintas épocas, estimamos el número de miembros provenientes del Partido Comunista y del gobierno cubano, que ingresaron en sus filas después de 1992. Así, describimos la evolución de los miembros de la Gran Logia, o confederación de Logias Azules, durante la segunda mitad del siglo XX, y a través de siete épocas que hemos identificado en estos años, y que describen los acontecimientos socio-políticos ocurridos en ella.

Palabras Clave

Estudio demográfico; modelos estadísticos; series de tiempo; masonería.

STOPPING TIMES DETERMINATION FOR LINEAR DEGRADATION PROCESSES

SILVIA RODRIGUEZ

Department of Statistics, Universidad Autónoma de Aguascalientes, Mexico

srodrign@correo.uaa.mx

ENRIQUE R. VILLA

Department of Probability and Statistics, CIMAT, Mexico

villadi@ciamat.mx

ABSTRACT

Reliability studies of high-reliability products require long testing periods to obtain failures. For highly reliable products with a degradation characteristic related to the failure process, it is advisable to use degradation tests instead of life tests. We consider the situation where the reliability study is conducted to estimate a quantile t_p of the failure time distribution. We propose a criterion to determine the stopping time for linear degradation processes with random coefficients. The stopping rule depends on planning values for the model parameters

Key words

Linear mixed effects model; failure time distribution; degradation model; planning tests; Fisher information matrix.

RESUMEN

Los estudios de confiabilidad de productos de alta confiabilidad requieren largos períodos de prueba para tener fallas. Para productos altamente confiables con una característica de degradación relacionada con el proceso de fallas, es recomendable usar pruebas de degradación en lugar de pruebas de vida. Consideramos la situación donde el estudio de confiabilidad se conduce para estimar un cuantil t_p de la distribución del tiempo a la falla. Proponemos un criterio para determinar el tiempo de paro para procesos de degradación lineal con coeficientes aleatorios. La regla de paro depende de los valores de planeación de los parámetros del modelo.

Palabras clave

Modelo lineal de efectos mixtos; distribución de tiempos de falla; modelo de degradación; pruebas de planeación; matriz de información de Fisher.

STATISTICAL METHODS TO QUANTIFY THE EFFECT OF MITE PARASITISM ON THE PROBABILITY OF DEATH IN HONEY BEE COLONIES

CHELSEA DEROCHE

Department of Experimental Statistics, Louisiana State University
cderoc2@lsu.edu

JOSÉ D. VILLA

USDA-ARS Honey Bee Breeding, Genetics and Physiology Laboratory
Jose.Villa@ars.usda.gov

LUIS A. ESCOBAR

Department of Experimental Statistics, Louisiana State University
luis@lsu.edu

ABSTRACT

Varroa destructor is a mite parasite of European honey bees, *Apis mellifera*, that weakens the population, can lead to the death of an entire honey bee colony, and is believed to be the parasite with the most economic impact on beekeeping. The purpose of this study was to estimate the probability of death for a honey bee colony as a function of the mite-infestation level present in the colony and to explore the influence of other variables (such as genetic origin of the colony and season of the year) on the relationship. Preliminary analyses showed that there was an association between season and mite infestation that needed to be considered in later analyses. Two analytical approaches were considered to account for the lack of deaths in colonies from two genetic origins which led to divergence of the maximum likelihood method when including origin as one of the variables in the logistic regression model. In the first approach, we used Firth's penalized likelihood method which has the double effect of correcting the bias of maximum likelihood (ML) estimates and providing estimates of the parameters. The second approach consists of forcing a death at the largest mite infestation for each of the two genetic origins without deaths. This approach, in general, would tend to provide slightly larger colony-death probability estimates. Because there were multiple observations on the same colony over a period of time, the data are longitudinal and the observations may not be independent. For this reason, we used a Generalized Estimating Equations (GEE) approach, which considers the dependency among the observations and compared it with the simple logistic regression that ignores the dependency. The GEE analysis showed increasing odds of death with increasing mite infestation and found no influence of season or genetic origin on the relationship. The results of the analysis using simple logistic regression are similar to those obtained using the more complex GEE analysis, suggesting that, for the data set considered, the longitudinal observations can be treated as statistically independent.

Key words

Generalized estimating equations; logistic regression; *Varroa destructor*.

RESUMEN

El acaro *Varroa destructor* es un parásito de las abejas europeas (*Apis mellifera*) que reduce el tamaño de la población de abejas, puede llevar a la muerte de colonias y es considerado como el parásito de mayor impacto económico sobre la apicultura. El objetivo de este estudio fue estimar la probabilidad de muerte de las colonias en función del nivel de infestación presente en la colonia e investigar la influencia de otras variables sobre esta relación (como el origen genético de la colonia y la estación del año). Análisis preliminares mostraron una asociación entre la estación del año y la infestación, requiriendo ser considerada en los análisis posteriores. Debido a la ausencia de muertes en colonias de dos orígenes genéticos lo cual produjo divergencia en el modelo de estimación usando máxima verosimilitud cuando se incluyó la variable origen genético en la regresión logística, se utilizaron dos metodologías alternativas. La primera, fue el uso del método penalizado de Firth el cual produce el doble efecto de corregir el sesgo de los estimadores de máxima verosimilitud y el producir estimadores de los parámetros. La segunda metodología, consistió en considerar como muertas aquellas colonias con el nivel máximo de infestación observado en los dos orígenes sin mortalidad. Este método tiende a producir probabilidades de muerte más altas. Debido al uso de observaciones múltiples sobre el tiempo para cada colonia, las observaciones son longitudinales y dependientes. Por esto, se utilizó el método de ecuaciones de estimación generalizadas (GEE), el cual incorpora la dependencia entre observaciones y permite comparaciones con funciones de regresión logística que ignoran la dependencia. El método de las GEE indica probabilidades de muerte más altas con aumentos en el nivel de infestación, pero los efectos de origen genético y de estación del año no fueron estadísticamente importantes. Los resultados con regresión logística fueron similares a los obtenidos con el método de GEE, sugiriendo, que para este grupo de datos, las observaciones longitudinales pueden ser consideradas como independientes.

Palabras clave

Ecuaciones generalizadas de estimación; regresión logística; *Varroa destructor*.

RESEÑA: EXACTITUD DE LA INFERENCIA EN POBLACIONES FINITAS, escrito y editado por Mariano Ruiz Espejo, Madrid, 2011, 169 páginas.

Breve descripción: este libro es un curso de introducción a los tipos de muestreo y a la inferencia bajo el diseño muestral en poblaciones finitas. Lectores potenciales son estudiantes, profesores o profesionales que, conociendo los fundamentos de la teoría de la probabilidad, estén interesados en introducirse en este campo. Así, el libro comprende los contenidos básicos de cualquier asignatura de introducción al muestreo en poblaciones finitas para titulaciones de las ramas de ciencias. Incluye todos los diseños muestrales simples habituales y algunos diseños más complejos como el bietápico. Además de describir los distintos tipos de muestreo esenciales, cubre adecuadamente la inferencia básica bajo cada uno de estos diseños.

Comentarios: a pesar de ser un manual relativamente conciso, introduce de forma apropiada la teoría básica del muestreo sin omisiones importantes. Por ejemplo, se incluyen los estimadores indirectos basados en el uso de información auxiliar. Además, mientras que muchos manuales pasan muy por encima los diseños con reposición, aquí se estudian de forma más extensiva al ser incluidos como un caso particular en la definición general de diseño muestral.

Esta concisión, sin dejar de lado temas importantes, ayuda ciertamente al aprendizaje de la materia en un relativamente corto periodo de tiempo. Además, un estudiante puede encontrar no pocos ejercicios para comprender mejor los conceptos y familiarizarse en el uso de las herramientas de cálculo necesarias. Un profesor o investigador agradecerá la formalidad matemática con la que está expuesta la materia, y un profesional podrá aprender las técnicas básicas del muestreo sin excesivo esfuerzo.

Cabe resaltar también el énfasis que realiza el autor, véase e.g. el título del libro, en la idea de que la inferencia bajo poblaciones finitas es completamente *exacta*, no asume ninguna hipótesis distribucional ni modelo alguno sobre las variables que se miden. Se buscan estimadores con buenas propiedades en promedio respecto de todas las posibles muestras que se pueden extraer de la población finita mediante el diseño muestral seleccionado. Por tanto, es una teoría para los que renuncien a asumir hipótesis que en la práctica nunca son contrastables de manera exacta.

Como conclusión, encuentro el libro muy adecuado para aprender en un breve plazo de tiempo, sin excesivo esfuerzo y en castellano, los fundamentos de muestreo y la inferencia bajo el diseño en poblaciones finitas.

Isabel Molina
Departamento de Estadística
Universidad Carlos III de Madrid