

# UNA MODELIZACIÓN PARA LOS ACCIDENTES DE TRABAJO EN ESPAÑA Y ANDALUCÍA

Por

**Mónica Ortega Moreno**

Profesora Estadística. Departamento Economía General y Estadística

## RESUMEN

El aumento de la siniestralidad laboral en los últimos años ha coincidido, de una parte, con un período de crecimiento económico y, de otra de una serie de factores entre los que se encuentra el cambio normativo producido en el año 1995, tras la aprobación de la Ley 31/1995 de Prevención de Riesgos Laborales.

En este trabajo se analiza el comportamiento de las series anuales: accidentes de trabajo en España (1977 – 2002) y accidentes de trabajo en Andalucía (1985 – 2002), se realizan previsiones futuras y se discute sobre la influencia de la Ley 31/1995 de Prevención de Riesgos Laborales. Y además, se observan simultáneamente un grupo de variables, con el objeto de analizar posibles relaciones y su evolución conjunta.

## 1. INTRODUCCIÓN

El estudio de la siniestralidad laboral implica la necesidad de desarrollar un sistema metodológico para la modelización de los accidentes de trabajo.

Usualmente los datos necesarios en los estudios sobre accidentes de trabajo son recogidos en forma de variable de recuento, definida como el número de sucesos o eventos que ocurren en un período de tiempo definido. Las variables de recuento son tratadas a menudo como variables continuas y, en consecuencia, el efecto de un conjunto de variables explicativas sobre dichas variables se analiza mediante un Modelo de Regresión Lineal (MRL).

El tratamiento de variables de recuento mediante el MRL, a pesar de ser una práctica extendida, es problemática por diferentes motivos, entre ellos: el incumplimiento de los supuestos de normalidad y homocedasticidad, las predicciones fuera de rango de los posibles valores de un recuento, así como la ausencia de linealidad (Hair et al., 2000). Por ello, el uso del MRL para modelar variables de respuesta de recuento suele proporcionar estimaciones sesgadas, ineficientes e inconsistentes.

Una alternativa es el análisis de la serie temporal, que consiste en un conjunto de técnicas estadísticas que permiten, además de estudiar y modelizar el comportamiento de un fenómeno que evoluciona a lo largo del tiempo, realizar previsiones de los valores que se alcanzarán en el futuro. Nuestro objetivo es construir el modelo que mejor se ajuste a la variable sobre siniestralidad laboral de interés y que permita hacer predicciones de dicha variable. Los valores de una serie temporal van ligados a instantes de tiempo, de manera que el análisis de una serie implica el manejo conjunto de dos variables; la variable en estudio propiamente dicha y la variable tiempo.

En el presente trabajo se presentan, en primer lugar, las variables objeto de estudio, fuentes consultadas, incidencias detectadas y periodo de tiempo objeto de observación. Para continuar con el análisis de las series anuales: accidentes de trabajo en España (1977 – 2002) y accidentes de trabajo en Andalucía (1985 – 2002), realizar predicciones futuras y discutir sobre la influencia de la Ley 31/1995 de Prevención de Riesgos Laborales.

## **2. MÉTODO**

El estudio llevado a cabo partió de la elaboración de una base de datos dinámica sobre la siniestralidad laboral en España y Andalucía. Los datos referidos a España fueron extraídos a partir de las consultas realizadas en los diferentes anuarios y publicaciones monográficas del Ministerio de Trabajo y Asuntos Sociales, así como del Instituto Nacional de Estadística; mientras que los de Andalucía se completaron con los anuarios publicados por el Instituto de Estadística de Andalucía. Dicha base de datos ha dado lugar a un conjunto de series temporales, colección de observaciones realizadas de forma secuencial en el tiempo, que se han organizado en función de los programas que manipularan los datos y recoge, además de las fuentes originales, las incidencias encontradas a la hora de su elaboración. Para concretar, hemos de especificar que el

tratamiento de los datos se ha realizado con el programa estadístico informático SPSS (*Statistical Product and Service Solutions*) (Pérez, 2001).

El período de tiempo objeto de observación ha sido condicionado por la disponibilidad de los datos. Para España se han conseguido datos a partir del año 1977 hasta el año 2002, mientras que en Andalucía la información es obtenida a partir del año 1985 hasta 2002.

Las etapas llevadas a cabo para construir el modelo que mejor se ajuste a la serie son:

1. **Identificación.** Determinación del posible modelo al que puede ajustarse la serie, seleccionamos entre el modelo de suavizado exponencial, el modelo autorregresivo integrado de media móvil (ARIMA) univariante o los modelos ARIMA (o modelos de función de transferencia) multivariantes para series temporales,
2. **Estimación** de los parámetros del modelo previamente seleccionado.
3. **Validación.** Si un ajuste es válido los parámetros deben ser significativos y el Test de Ljung-Box para los residuos nos lleva a no rechazar la hipótesis nula de que los residuos son ruido blanco. La etapa de validación puede conducir de nuevo a la primera si el modelo no es válido. Puede ocurrir también que haya más de un modelo que se ajuste a la serie, en tal caso seleccionamos el modelo de acuerdo con las medidas de bondad de ajuste que proporciona el SPSS.
4. **Predicciones:** Cuando el modelo estimado supera la fase de verificación se puede utilizar para realizar predicciones del comportamiento futuro de la serie. Además es posible facilitar un intervalo de confianza para dichas predicciones, denotaremos LCS al extremo superior de dicho intervalo y LCI al extremo inferior.  
Para cada modelo, el programa SPSS facilita las predicciones comenzando después del último valor no perdido del rango del período de estimación solicitado y finalizando en el último período para el que hay disponibles valores no perdidos de todos los predictores o en la fecha de finalización del período de predicción solicitado, lo que ocurra antes.  
Para generar predicciones mediante modelos con variables independientes, el conjunto de datos activo debe contener valores de estas variables para todos los casos del período de predicción.

- **Serie de accidentes de trabajo en España (1977-2002)**

El modelo estimado para la serie anual que recoge los accidentes de trabajo en España entre 1977 y 2002 es de la forma:

$$F_t = 0'944E_{t-1} + 0'066F_{t-1}$$

donde F denota el pronóstico y E la variable en estudio.

El modelo es válido, dado que los parámetros son significativos y los residuos ruido blanco (test de Ljung-Box). Destacar una alta estimación de la proporción de la variación total de la serie que es explicada por el modelo ( $R^2=0'939$ ) y modelo con menor raíz del error cuadrático medio.

La Figura 1 muestra la serie observada (línea roja) y el ajuste de la misma (línea azul fina) en el período 1977 – 2002, junto con una previsión (línea azul gruesa) y un intervalo de confianza para dicha previsión (líneas discontinuas) en el período comprendido entre 2003 y 2007. Las previsiones indicadas se detallan en la Tabla 1.

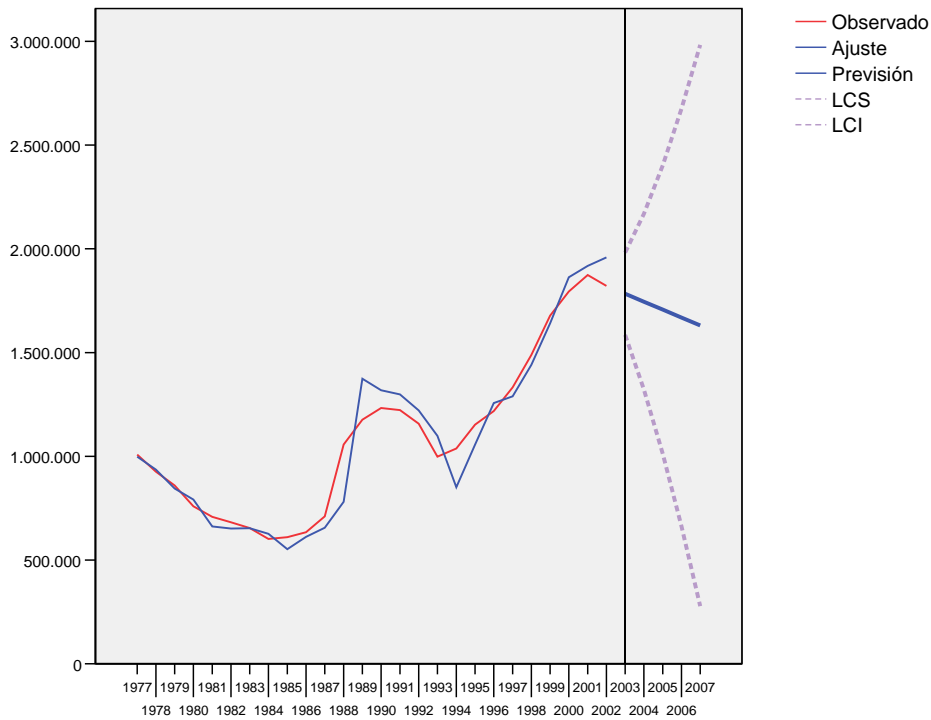


Figura 1

	2003	2004	2005	2006	2007
Previsión	1783391	1745199	1707007	1668815	1630623
LCS	1981525	2168641	2401867	2674941	2983446
LCI	1585258	1321758	1012148	662689	277800

Tabla 1

- **Serie de accidentes de trabajo en Andalucía (1985-2002)**

En el caso de la serie anual de accidentes de trabajo en Andalucía en el período comprendido entre 1985 y 2002 la ecuación que se ajusta es de la forma:

$$F_t = 0'988E_{t-1} + 0'022F_{t-1}$$

donde F denota el pronóstico y E la variable en estudio.

En dicho modelo los parámetros significativos y los residuos ruido blanco (test de Ljung-Box), por lo que el modelo es válido. Además, destacar la alta estimación de la proporción de la variación total de la serie que es explicada por el modelo ( $R^2=0'952$ ) y que se trata del modelo con menor raíz del error cuadrático medio.

La Figura 2 muestra la serie objeto de estudio (línea roja) junto con el ajuste de la misma (línea azul fina) en el período 1977 – 2002; además de una previsión (línea azul gruesa) y un intervalo de confianza (líneas discontinuas) para el período comprendido entre 2003 y 2007. Las previsiones indicadas se detallan en la Tabla 2.

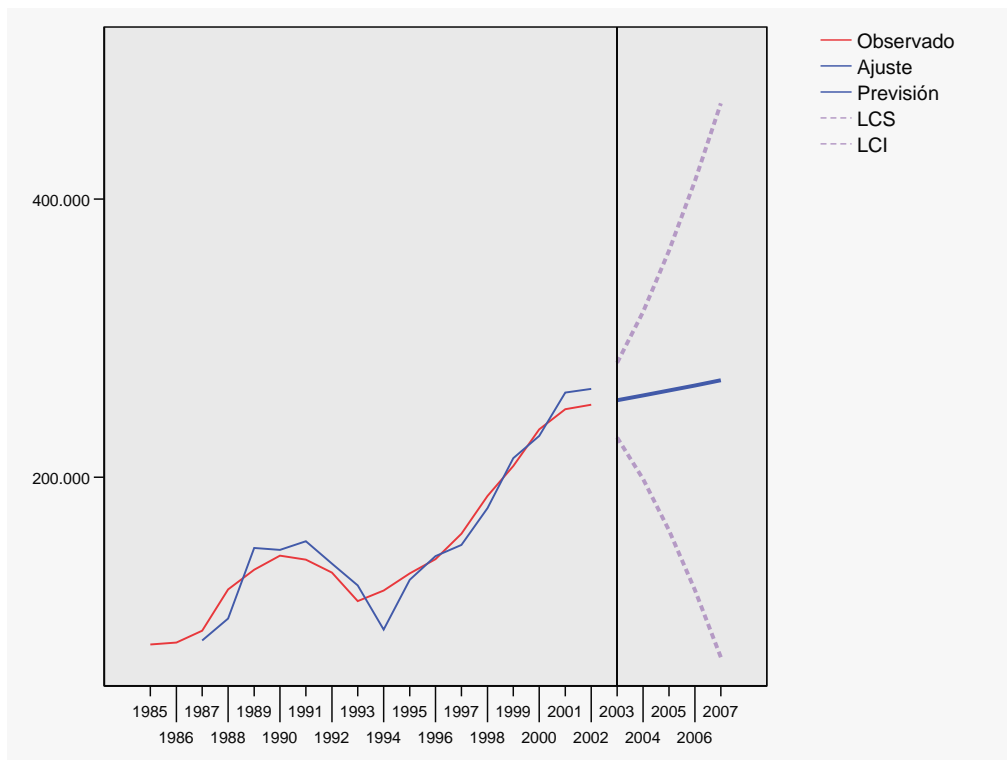


Figura 2

	2003	2004	2005	2006	2007
Previsión	255529	258995	262460	265925	269391
LCS	280493	314278	354584	400470	451297
LCI	230566	203711	170336	131381	87485

Tabla 2

### 3. INFLUENCIA DE LA LEY 31/1995 DE PREVENCIÓN DE RIESGOS LABORALES.

Los datos disponibles para estudio de la serie accidentes de trabajo en España y Andalucía hasta 1995 son escasos, por lo que se ha aplicado técnicas multivariantes de análisis de series temporales con el objeto de realizar un mejor ajuste. Para ello, se ha analizado posibles relaciones entre las variables: población activa, población ocupada, tasa de paro, índice de incidencia y producto interior bruto (tasa de variación interanual)- y las series accidentes de trabajo en España y Andalucía. La Tabla 3 muestra las correlaciones significativas, siendo en ambos casos la variable que muestra mayor correlación de Pearson la población ocupada.

<b>CORRELACIONES DE PEARSON</b>	<b>Accidentes de Trabajo España</b>	<b>CORRELACIONES DE PEARSON</b>	<b>Accidentes de Trabajo Andalucía</b>
<b>Población Ocupada</b>	0'967	<b>Población Ocupada</b>	0'991
<b>Población Activa</b>	0'908	<b>Población Activa</b>	0'915
<b>Índice de Incidencia</b>	0'881	<b>Tasa de Paro</b>	-0'748

Tabla 3

Tras introducir la variable explicativa población ocupada en el modelo el análisis de las series objeto de estudio se detalla a continuación:

- **Serie de accidentes de trabajo en España (1977-1995)**

Los datos disponibles dan lugar al modelo:

$$F_t = E_{t-1} + 0'299(I_t - I_{t-1})$$

donde F denota el pronóstico, E la variable en estudio y I es una variable independiente, en este caso la población ocupada en España.

Dado que los parámetros del modelo son significativos y los residuos ruido blanco (test de Ljung-Box) el modelo es validado. Destacar una alta estimación de la proporción de

la variación total de la serie que es explicada por el modelo ( $R^2=0'927$ ) y modelo con menor raíz del error cuadrático medio.

La Figura 3 muestra la serie observada (línea roja) para accidentes de trabajo en España entre 1977 y 2002, junto con las previsiones (línea azul) y un intervalo de confianza al 95% (líneas discontinuas) a partir de 1995 hasta 2002 con el ajuste realizado.

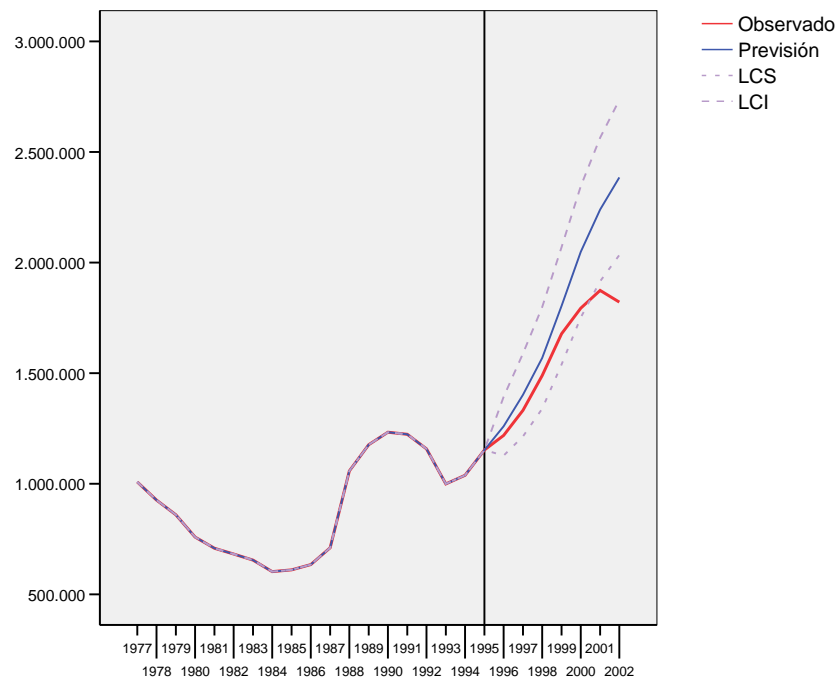


Figura 3

Observemos que las previsiones para los accidentes de trabajo antes de la Ley 31/1995 de Prevención de Riesgos Laborales eran superiores a las que se han dado tras dicha ley. Sin embargo, los datos observados se encuentran dentro del intervalo de confianza al 95% proporcionado por el ajuste hasta el año 2002. Por consiguiente, todo lo que se ha venido haciendo desde 1995 en materia de prevención parece no haber afectado hasta la reducción observada en 2002.

- **Serie de accidentes de trabajo en Andalucía (1985-1995)**

En el caso de Andalucía los datos disponibles se ajustan al modelo:

$$F_t = E_{t-1} + 0'189(I_t - I_{t-1})$$

donde F denota el pronóstico, E la variable en estudio y I es una variable independiente, en este caso la población ocupada en Andalucía.

El modelo es validado, dado que los parámetros son significativos y los residuos ruido blanco (test de Ljung-Box). Destacar alta estimación de la proporción de la variación

total de la serie que es explicada por el modelo ( $R^2=0.932$ ) y modelo con menor raíz del error cuadrático medio.

La Figura 4 muestra la serie observada (línea roja) para accidentes de trabajo en Andalucía entre 1985 y 2002, previsiones (línea azul) e intervalo de confianza al 95% (línea discontinua) a partir de de 1995 hasta 2002 con el ajuste realizado.

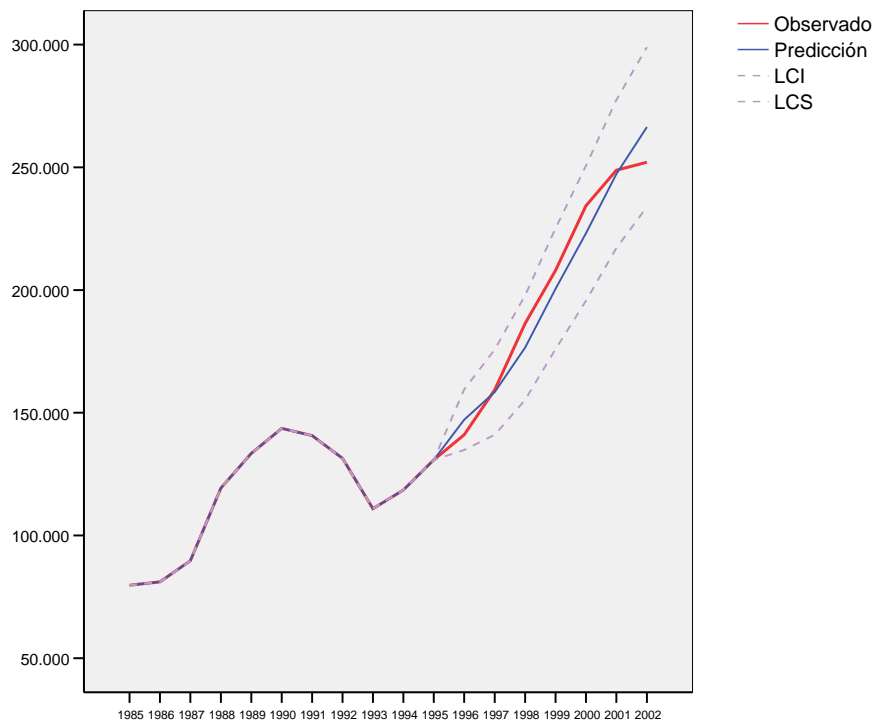


Figura 4

Observando la serie objeto de estudio junto con las previsiones realizadas a partir de 1995 podemos afirmar que en el caso de Andalucía la Ley 31/1995 de Prevención de Riesgos Laborales no ha afectado a los accidentes de trabajo.

#### 4. DISCUSIÓN DE RESULTADOS

El análisis realizado de las series accidentes de trabajo en España y Andalucía en los últimos años nos permite señalar una escasa o nula influencia del cambio normativo producido en el año 1995, tras la aprobación de la Ley de Prevención de Riesgos Laborales.

Sin embargo, téngase en cuenta que el crecimiento económico puede ser un factor explicativo del aumento de la siniestralidad producido en los últimos años. Es por ello necesario estudiar las relaciones existentes entre las variables que conectan, como

pueden ser la carga de trabajo, las condiciones del riesgo, la tasa de temporalidad, la inmigración, el tamaño de las empresas ... Y, tras recoger la influencia que ejercen dichas variables en los accidentes de trabajo tener en cuenta que nuestra mayor siniestralidad puede deberse a la influencia de otra serie de perturbaciones aleatorias no observadas, como podrían ser las deficientes puesta en marcha de las acciones preventivas, o de formación.

## **REFERENCIAS**

HAIR, J.F., ANDERSON, R.E., TATHAM, R.L., BLACK, W.C. (2000): Análisis Multivariante. Prentice Hall. Madrid.

LÉVY MANGIN, J.P., VARELA MALLOU, J. (2003): Análisis Multivariable para las Ciencias Sociales. Prentice Hall.

PÉREZ, C. (2001): Técnicas estadísticas con SPSS. Prentice Hall. Madrid.