

ANÁLISIS TEMPORAL DE NACIMIENTOS EN BURGOS

PREMIOS A LA REALIZACIÓN DE TRABAJOS RELACIONADOS CON LA
ESTADÍSTICA 2012

ÍNDICE:

1- INTRODUCCIÓN	Pág. 2
2- ESTUDIO GRÁFICO	Pág. 3
3- ANÁLISIS DE LAS COMPONENTES	Pág. 4
4- CONCLUSIONES	Pág. 10
5- BIBLIOGRAFÍA	Pág. 10

1- INTRODUCCIÓN

DEFINICIÓN:

Una serie temporal se define como una colección de observaciones de una variable recogidas en sucesivos instantes de tiempo.

OBJETIVOS:

En el análisis estadístico de una serie temporal se pueden considerar varios objetivos:

- **Descripción:** En este sentido se sacarán conclusiones a cerca de su tendencia, de la influencia de algún factor (estaciones...) y si existen observaciones extrañas o discordantes.
- **Predicción:** Se pretende estimar valores futuros de la variable, que puede servir para una mejor planificación de recursos

Mientras estudiábamos el tema, nos dimos cuenta de la enorme cantidad de aplicaciones en numerosos campos que tiene el análisis estadístico de estas series: economía, demografía, medioambiente....Por esta razón, cuando decidimos presentarnos al concurso convocado por la Consejería de Hacienda, los alumnos de 1º de Bachillerato quisimos poner en práctica los conocimientos adquiridos y hacer un análisis del número de nacimientos que se han producido en Burgos en los últimos años.

Los datos fueron tomados de la página oficial del Instituto Nacional de Estadística (INE). Aún no están publicados los datos correspondientes al año 2011.

La herramienta utilizada para realizar todos los cálculos y los gráficos de este trabajo ha sido el programa Microsoft Excel.

DATOS: Nacimientos producidos en Burgos

	2005	2006	2007	2008	2009	2010
ENERO	245	240	259	271	277	284
FEBRERO	210	213	227	284	252	209
MARZO	253	269	268	265	266	299
ABRIL	286	269	264	304	268	282
MAYO	268	267	303	352	275	294
JUNIO	265	271	281	299	295	284
JULIO	248	275	309	305	312	291
AGOSTO	255	294	294	308	309	280
SEPTIEMBRE	253	268	271	304	266	288
OCTUBRE	277	287	285	286	299	274
NOVIEMBRE	253	242	247	261	244	264
DICIEMBRE	244	228	242	286	284	311

Tabla Nº 1

2- ESTUDIO GRÁFICO

El primer paso en el estudio de una serie cronológica es el gráfico temporal, que nos da una primera idea del comportamiento de los datos.

Representación gráfica del número de nacimientos en Burgos (2005-2010)

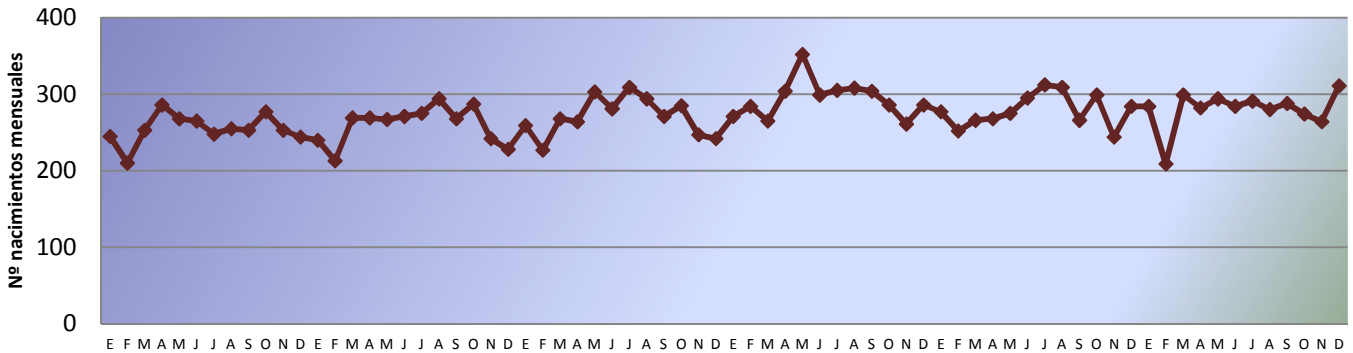


Fig. Nº 1

3- ANÁLISIS DE LAS COMPONENTES

El segundo paso en el estudio de una serie temporal es el análisis de cada una de las componentes que forman la serie cronológica.

COMPONENTES

Cada uno de los valores que toma la variable considerada es fruto que la unión de múltiples factores que actúan sobre el fenómeno estudiado.

Tendencia secular: Es una variación lenta, regular y de larga duración que refleja la evolución de la serie a largo plazo. Corresponde con la dirección predominante de la serie observada en un espacio de tiempo suficientemente amplio.

Variaciones estacionales: Son oscilaciones de mayor o menor regularidad de periodo un año (a veces menos) motivadas por ciertos acontecimientos periódicos (vacaciones, navidades...).

Variaciones cíclicas: Son oscilaciones que abarcan un intervalo de tiempo superior a un año.

Variaciones accidentales: Son variaciones fortuitas e irregulares cuya causa es un acontecimiento impredecible (huelgas, catástrofes naturales...)

Vamos a considerar la hipótesis de que estas cuatro fuerzas se unen por productos para dar cada uno de los valores de la variable. $X_i = T \times E \times C \times A$

En nuestro caso nos centraremos en el estudio de la tendencia secular y las variaciones estacionales.

3.1 TENDENCIA SECULAR

3.3.1. Método de las medias móviles.

Consiste en hallar la tendencia secular de la serie diluyendo la importancia individual de cada observación. Esto se consigue promediando cada dato mediante una media aritmética con los datos inmediatamente anteriores y posteriores. Cada valor X_t vendrá sustituido por una media. En el caso de que el número de elementos tomados sea par es necesario volver a calcular otra nueva serie de medias móviles de tamaño 2 sobre la primera con el objeto de centrar los datos en los momentos originales.

Ejemplo:

Para hallar las medias móviles de tamaño 5 sustituimos cada observación por el siguiente valor:

$$y_t = \frac{x_{t-2} + x_{t-1} + x_t + x_{t+1} + x_{t+2}}{5}$$

Obsérvese que siempre quedarán sin determinar algunos valores de la media al principio y al final de la serie

MEDIAS MÓVILES DE TAMAÑO 12						
	2005	2006	2007	2008	2009	2010
Enero		256,04	267,75	285,08	282,21	279,63
Febrero		258,79	269,17	285,50	282,54	277,54
Marzo		261,04	269,29	287,46	281,00	277,25
Abril		262,08	269,33	288,88	279,96	277,13
Mayo		262,04	269,46	289,50	279,79	276,92
Junio		260,92	270,25	291,92	279,00	278,88
Julio	254,54	261,04	271,33	294,00	279,21	
Agosto	254,46	262,42	274,21	292,92	277,71	
Septiembre	255,25	262,96	276,46	291,63	277,29	
Octubre	255,21	262,71	278,00	290,17	279,25	
Noviembre	254,46	264,00	281,71	285,46	280,63	
Diciembre	254,67	265,92	284,50	282,08	280,96	

Tabla Nº 2

En el siguiente gráfico podemos ver juntas la serie original y la serie formada por las medias móviles. Observamos como la oscilación de la segunda se suaviza respecto de la de la primera.

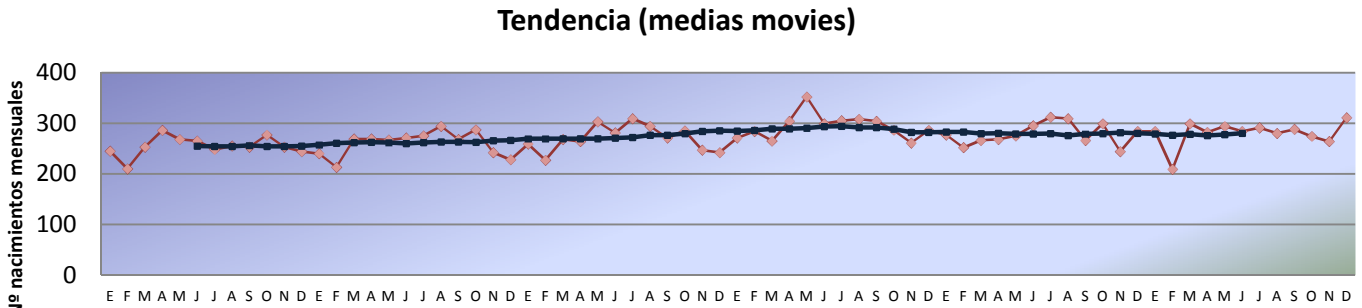


Fig. Nº 2

3.1.2.Método analítico.

Este método consiste en sustituir la serie por una recta de la forma $X=bt+a$ que mejor se ajuste a los puntos. El método más general para encontrar esta recta es el de mínimos cuadrados. Los coeficientes a y b son la solución del siguiente sistema:

$$\left. \begin{aligned} \sum_i X_i &= an + b \sum_i t_i \\ \sum_i t_i \cdot X_i &= a \cdot \sum_i t_i + b \cdot \sum_i t_i^2 \end{aligned} \right\}$$

t_i = momentos temporales
 X_i = observaciones en cada instante.

En el caso que nos ocupa trabajaremos con datos anuales. De esta forma la influencia de las variaciones estacionales es prácticamente nula. Además utilizaremos en este caso datos de 10 años.

Para facilitarnos la resolución del sistema de ecuaciones podemos hacer el cambio de variable

$$t'_i=t_i-2000$$

t_i	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
X_i	2671	2797	2818	2968	3057	3123	3250	3525	3311	3360
t'_i	1	2	3	4	5	6	7	8	9	10
$t'_i \cdot X_i$	2671	5594	8454	11872	15285	18738	22750	28200	29799	33600
$t_i'^2$	1	4	9	16	25	36	49	64	81	100

Tabla Nº 3

<p>SUMAS:</p> $n = 10$ $\sum_i X_i = 30.880;$ $\sum_i t'_i = 55;$ $\sum_i t'_i \cdot X_i = 176.963;$ $\sum_i t_i'^2 = 385$	<p>SISTEMA:</p> $\left. \begin{aligned} 30880 &= 10a + 55b \\ 176963 &= 55a + 385b \end{aligned} \right\}$ <p>RESULTADO</p> $a=2.613,13 \quad b=86,34$ <p>RECTA</p> $X=86,34+t+2.613,13$
---	---

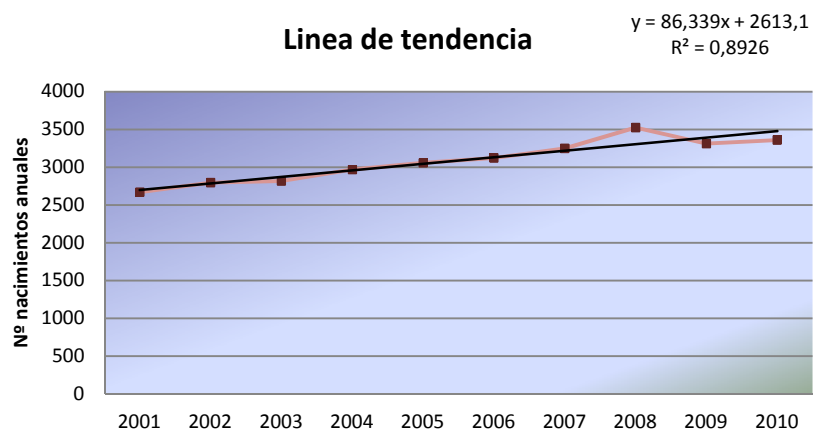


Fig. Nº 3

PREDICCIÓN:

La importancia de hallar la recta que se ajuste a los datos consiste en permitir la extrapolación; es decir, permite predecir valores futuros de la variable. Evidentemente, cuanto más próximo sea el futuro que quiere predecirse, mayor fiabilidad tendrán los valores predichos por lo que se desaconseja extrapolar a momentos demasiado alejados del actual. De esta forma podemos hacer una previsión de recursos, costes.... Para hacer las predicciones solo hay que sustituir t por el valor del cual queremos hacer la predicción.

En nuestro caso la previsión de nacimientos para el año 2011 ($t'=11$) es de 3563 nacimientos.

COEFICIENTE DE DETERMINACION:

Una vez calculada la recta que mejor se ajusta a los puntos se utiliza el coeficiente de determinación como medida de la bondad del ajuste realizado.

$$R^2 = \frac{\sum_i (\hat{X}_i - \bar{X})^2}{\sum_i (X_i - \bar{X})^2}$$

Este coeficiente mide la proporción de la variabilidad total de la variable dependiente (en este caso X) respecto a su media que es explicada por el modelo de regresión.

Se verifica que $0 \leq R^2 \leq 1$. Cuanto más cerca se encuentre su valor de 1, mejor será el ajuste realizado

Coeficiente de determinación:

$$R^2=0,8926$$

3.2. VARIACIONES ESTACIONALES**Análisis de variaciones estacionales**

Tiene por objeto determinar oscilaciones de periodo relativamente corto (generalmente un año). De todos los métodos existentes utilizaremos uno basado en la hipótesis multiplicativa, es decir las cuatro componentes se unen en forma de producto.

$$Y=T*E*C*A$$

El método consiste en lo siguiente:

- Se halla la tendencia mediante el método de las medias móviles tomando como tamaño un año (tamaño 12). *Ver tabla N° 2*
- Se elimina de la tendencia dividiendo cada dato de la serie original entre los valores hallados de la tendencia en cada instante de tiempo. *Ver tabla N° 4*

	2005	2006	2007	2008	2009	2010
Enero		0,937	0,967	0,951	0,982	1,016
Febrero		0,823	0,843	0,995	0,892	0,753
Marzo		1,030	0,995	0,922	0,947	1,078
Abril		1,026	0,980	1,052	0,957	1,018
Mayo		1,019	1,124	1,216	0,983	1,062
Junio		1,039	1,040	1,024	1,057	1,018
Julio	0,974	1,053	1,139	1,037	1,117	
Agosto	1,002	1,120	1,072	1,051	1,113	
Septiembre	0,991	1,019	0,980	1,042	0,959	
Octubre	1,085	1,092	1,025	0,986	1,071	
Noviembre	0,994	0,917	0,877	0,914	0,869	
Diciembre	0,958	0,857	0,851	1,014	1,011	

Tabla Nº 4

c) Se eliminan las variaciones cíclicas y accidentales hallando las medias aritméticas de los valores obtenidos en cada periodo. (En nuestro caso meses.) Ver Tabla Nº 5

MEDIAS MENSUALES	
Enero	0,970
Febrero	0,861
Marzo	0,995
Abril	1,007
Mayo	1,081
Junio	1,036
Julio	1,064
Agosto	1,072
Septiembre	0,998
Octubre	1,052
Noviembre	0,914
Diciembre	0,938

Tabla Nº 5

INDICES DE VARIACION ESTACIONAL	
Enero	97,1
Febrero	86,2
Marzo	99,5
Abril	100,8
Mayo	108,2
Junio	103,7
Julio	106,5
Agosto	107,3
Septiembre	99,9
Octubre	105,3
Noviembre	91,5
Diciembre	93,9

Tabla Nº 6

d) Sobre estos últimos valores se calculan los índices de variación estacional. Éstos consisten en la variación porcentual de los distintos momentos del año, respecto a la media anual, es decir, es la fuerza de la componente estacional una vez eliminadas las otras tres. Ver Tabla Nº 6

$$IVE = \frac{\text{media mensual}}{\text{media anual}} \times 100$$

Si eliminásemos por cociente la influencia de las estaciones, obtendremos la serie "desestacionalizada".

e) Por último, para comprender mejor el significado de los índices de variación estacional, los representaremos en un gráfico de coordenadas.

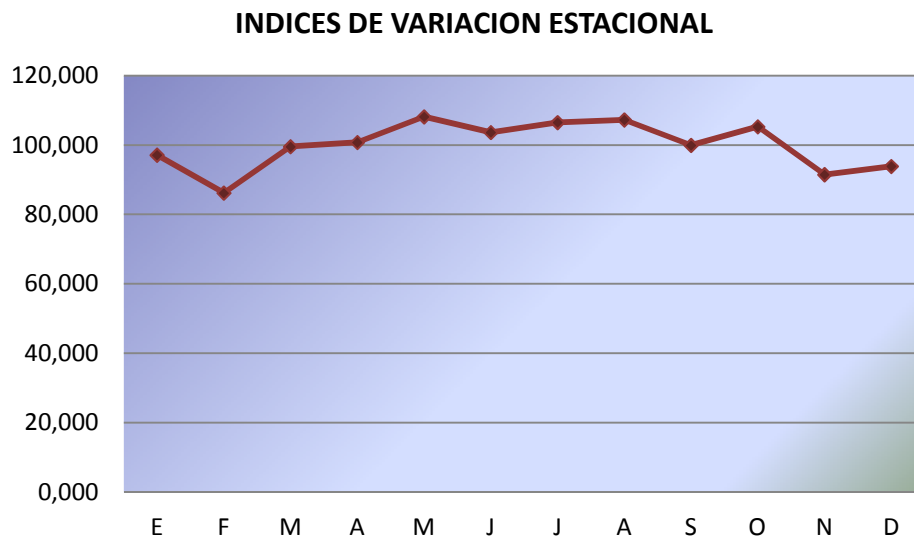


Fig. Nº 4

4- CONCLUSIONES

Del análisis de la tendencia deducimos que, efectivamente ha habido un incremento del número de nacimientos en Burgos durante los años 2001 a 2010. Ese incremento lo podemos apreciar tanto en el gráfico de las medias móviles (*Fig. Nº 2*) como en la recta de tendencia (*Fig. Nº 3*) ya que su pendiente es positiva. Además podemos señalar, observando la propia ecuación de la recta que ese crecimiento es, en promedio, a razón de 86 nacimientos por año.

En cuanto a la predicción realizada del número de nacimientos esperado para el año 2011, diremos que al estar el coeficiente de determinación R^2 bastante próximo a 1, nos permite asegurar que el ajuste lineal es correcto y que, por lo tanto, la predicción realizada del número de nacimientos para el año 2011 es bastante fiable. Esto no significa que el valor pueda diferir del que finalmente publique el INE. Si existiese una diferencia grande, ésta podría deberse a variaciones accidentales, posibles en cualquier serie temporal: época de crisis, reducción de las ayudas estatales, retorno a su país de la población emigrante...

Finalmente, del estudio de la estacionalidad se deduce, tanto en los resultados obtenidos de la tabla Nº 6 como en su representación gráfica (*Fig. Nº 4*), que existe una serie de meses consecutivos, desde abril hasta agosto (y excepcionalmente octubre), en los que el índice de natalidad se sitúa por encima de la media anual, mientras que hay otra serie de meses consecutivos, de noviembre a febrero, en los cuales este índice está por debajo de la media anual. Septiembre y marzo están en la media. Es decir, parece, efectivamente que en los meses más calurosos se producen más nacimientos que en los meses más fríos.

5- BIBLIOGRAFÍA

- CASA ARUTA E., *200 problemas de estadística descriptiva*, Vicens-Vives, Barcelona 1986
- LÓBEZ URQUIA J., *Estadística*, Boixareu Editores, Barcelona 1975
- www.ine.es