El método Holt Winters

La plantilla Excel utiliza el método "Holt Winters" para hacer un pronóstico del comportamiento de una serie temporal a partir de los datos obtenidos anteriormente. En este caso el pronóstico se hace para el año siguiente a partir del último dato obtenido. El método se basa en un algoritmo iterativo que a cada tiempo (mes o semana) realiza un pronóstico sobre el comportamiento de la serie en base a promedios debidamente ponderados de los datos anteriores. El peso que se le da a los datos decrece de manera exponencial hacia atrás en el tiempo. En este sentido, es similar a los métodos de alisamiento exponencial pero mejorado, ya que incorpora información sobre la tendencia de la serie. El algoritmo tiene tres parámetros, cada uno de ellos asociado a diferentes componentes de la serie. El valor de estos parámetros se ajusta, comparando la serie real con la pronosticada para ese mismo lapso. Una vez realizados los ajustes, se procede a hacer el pronóstico para el período en donde no hay datos.

Las tres componentes consideradas en el método son el valor medio, la tendencia y la estacionalidad. Al contemplar componentes estacionales, el método se adapta muy bien para el pronóstico de series asociadas enfermedades con fuerte carácter estacional.

Cada una de estas componentes está asociada a un parámetro, generalmente llamados α , β , γ . Los valores de α , β , γ pueden estar fijados por el usuario o escogerse de manera que minimicen el error cuadrático medio comparando el comportamiento de la serie real y la serie pronosticada en la zona en la que se superponen. Para usar este método es necesario estimar condiciones iniciales. Una descripción detallada del método puede encontrarse en *Holt Winters forecasting: some practical issues*. C. Chatfield & M. Yar, *The statistician* (1988) 37, pp 129-140.

La plantilla Excel – Generalidades

La plantilla Excel "Pronóstico Holt Winters" calcula la proyección de una serie con datos mensuales o semanales con estacionalidad anual. El usuario **solamente** debe ingresar los datos de la serie temporal en la columna C, modificando adecuadamente las columnas A y B que son sólo etiquetas para el gráfico.

Dado un juego de valores para los parámetros α , β , γ la plantilla calcula, **de manera automática**, las tres componentes necesarias para el algoritmo y las muestra en las columnas D, E y F, asociadas a media, tendencia y estacionalidad respectivamente que el método H-W necesita para calcular el pronóstico (columna G).

Para que este pronóstico sea el mejor posible, los valores de α , β , γ deberían ser los óptimos (en el sentido de hacer que la diferencia entre la proyección y los datos sea la menor posible). Para obtener esos valores, se debe entonces minimizar el error cuadrático medio, lo cual puede hacerse usando el complemento **Solver** de Excel (ver apéndice). La plantilla tiene una implementación automática del Solver al clicar el botón optimizar. Esta acción explora automáticamente todas las combinaciones de parámetros α , β , γ , elige la que minimiza el error cuadrático medio y los devuelve en los casilleros correspondientes. Las celdas K6 a K8 contienen los parámetros del método optimizados por **Solver**. El error cuadrático medio está calculado en la celda K10. Con los valores de la tabla y las columnas D, E, y F la plantilla calcula la proyección y la escribe en la columna G. Todo esto ocurre automáticamente luego de haber clicado el botón de optimizar.

La proyección en columna G empieza un año más tarde que los primeros datos, ya que usa el primer año como **condiciones iniciales** para el algoritmo, y se extiende un año más allá de los últimos datos. Este último año es la proyección realizada. En las columnas H e I se calculan los extremos superior e inferior del intervalo de predicción. En el gráfico aparecen en azul los datos reales, en rojo punteado la proyección y en morado y verde los errores que limitan la región de confiabilidad.

A la derecha del gráfico aparecen algunos números relacionados con la proyección, que pueden resultar útiles. En las celdas Y13 e Y15 (en rosa) aparecen el máximo número de casos y el total

anual de casos para el año de la predicción. En la celda Y10 (amarillo) aparece el error relativo promedio de la proyección (para los años anteriores a la predicción). Por ejemplo, si indica 0.2 quiere decir que la diferencia entre la proyección y los datos es de 20%, en promedio. El valor de la celda Y2 es aproximadamente el cuadrado de este valor. En la celda Y6 se tiene la diferencia promedio entre el intervalo superior de confianza, y la predicción. En la celda Y18 (verde) aparece la autocorrelación de la diferencia entre proyección y datos (respecto a su utilidad, ver la sección Recomendaciones).

La plantilla Excel – Uso

El uso de la plantilla Excel es simple.

El usuario sólo debe completar la columna C con los datos de la serie y modificar las columnas A (año) y B (mes) a los solos fines de modificar las etiquetas del gráfico. El contenido de las columnas A y B no afecta al pronóstico. **Es imperativo no borrar filas**, solo modificar su contenido, ya que si se borrarán filas se modificaría la automatización de la plantilla. **Solamente pueden modificarse las columnas A, B y C. Y si hiciera falta borrar solo “borrar el contenido” pero no borrar celdas.**

Si bien la plantilla comienza desde 2004, esto puede modificarse para que se ajuste a los datos con los que se cuenta. Siempre debe comenzarse a ingresar los datos en la columna C desde la fila 2 (la fila 1 está reservada para etiquetas) Luego de completar la columna C, se cliquee el botón **optimizar** y en la columna G aparecerá la serie pronosticada, incluyendo la proyección a un año a partir del último dato. En el gráfico aparecerán las curvas de las series temporales según lo descripto anteriormente.

Apéndice

La plantilla incluye un macro para invocar al **Solver** y minimizar el error cuadrático medio (MSE). En consecuencia es necesario habilitar los macros e instalar el complemento Solver. Este ultimo se habilita desde **Complementos de Excel**. La manera de hacerlo depende de la versión de Excel que se está usando. Una vez habilitado, queda incorporado a Excel y no hay que volver a realizar esta operación en futuras utilizaciones de la plantilla.

Los macros, en cambio, sí deben ser habilitados cada vez que se abre la plantilla (por seguridad suelen estar deshabilitados). Esto es en general muy simple. En Excel 2007 aparece una barra entre la barra de herramientas habitual y la plantilla, en la que esta la posibilidad de habilitar los macros.

Las columnas ocultas entre Y y AV contienen las fórmulas para el cálculo de los errores. Se ocultaron para evitar confusiones, pero pueden volver a mostrarse marcando las columnas Y y AV, y en formato de columna elegir la opción “mostrar columnas”.

Importante: Podría ocurrir que debido a diferencias en versiones de Excel, el macro no funcione correctamente. En ese caso, se puede ejecutar el **Solver** manualmente. Para eso hay que buscar el **Solver** en la barra superior, en general en la pestaña “Herramientas” en versiones 2000 y 2003 y en “Datos” en versiones superiores. Una vez encontrado hay que cliquee **Solver** y pedir que lo ejecute. Luego aceptar el resultado propuesto por **Solver**

Recomendaciones

Como se explica más arriba, la predicción que realiza la plantilla no depende solo de los datos, sino también de los valores iniciales de varios parámetros. Por ello, variando estos pueden obtenerse diferentes predicciones. Sin embargo, no todas son “equivalentes”. Hay varios requerimientos que la

predicción debería tratar de cumplir. Algunos de ellos son necesarios para que los intervalos de confianza calculados sean correctos. Si estos requerimientos no se cumplen, los intervalos de confianza dejan de ser confiables (paradójicamente), y conviene no utilizarlos.

Un requerimiento importante es que el MRSE (celda Y10) sea aproximadamente menor que 0.4. Otro es que α y β sean bastante menores que 1. En particular, es recomendable que α no exceda 0.6 (aproximadamente). De no ser así, el alisamiento exponencial resulta demasiado brusco y la proyección para cada tiempo sólo utilizará el dato del tiempo inmediatamente anterior.

También es importante chequear que r1 (celda Y18, la autocorrelación de los errores) sea pequeño, en módulo. Sugerimos que su valor esté en el rango -0.35 a 0.35. De no ser así, la predicción puede usarse, pero no los intervalos de confianza.

Por razones obvias, conviene también que los intervalos de confianza no sean demasiado amplios. Numéricamente esto puede chequearse con la celda Y6 (su valor debería ser menor que 1).

Hay varias formas de obtener diferentes predicciones. Algunas de ellas son:

- Ajustar la tendencia inicial (celda O8) y luego optimizar. El valor por default es 1, pero puede cambiarse a 0. Esto puede ser útil cuando los dos primeros años de los datos (que se usan para calcular las variables iniciales) presentan (a ojo) una tendencia diferente al resto.
- Ajustar el peso del intervalo de confianza (celda O6) y luego optimizar. El valor por default es 0. Si es mayor, el algoritmo tratará de minimizar la diferencia entre datos y proyección, y a la vez también de conseguir intervalos de confianza pequeños. Si el peso se fija en 1, solo minimizará el intervalo de confianza. En general hay que comenzar con 0 y, si hiciera falta cambiarlo, no conviene que este peso sea mayor que 0,1.
Hay que tener cuidado porque incrementar el peso suele dar predicciones con r1 demasiado grande, con lo cual, como se explica más arriba las bandas de confianza son más pequeñas pero dejan de ser confiables.
- Poner diferentes cantidades de años anteriores al año que se quiera predecir. Por ejemplo, no es claro que para predecir 2011 sea necesario usar los datos desde 2000. Quizás con 2007-2010 sea suficiente. Si se hace esto, recordar no borrar celdas, simplemente “borrar el contenido” de la columna C.
- Otra posibilidad es, después de haber optimizado, tocar un poco los parámetros α , β , y γ “a mano”. La idea aquí es que a veces ocurre que si bien los parámetros que devuelve el Solver son los óptimos, lo son por muy poco. O sea que cambiando un poco algún parámetro puede por ejemplo mejorarse el r1 o los intervalos de confianza, a expensas de apenas un poco más de error medio en la proyección.

Por último, hacemos notar que en general el modelo de Holt-Winters se “comporta” mejor para datos mensuales que para datos semanales. Así que en caso de no poder resolver los problemas de predicción para datos semanales, puede ser una buena idea convertir estos en mensuales (o cuatrisesmanales) y ver cómo dan las predicciones.

Ante cualquier duda contactarse con Marcelo Kuperman (mkuperman@gmail.com) o Sebastián Risau (srisaug@gmail.com).