

AJUSTES DE JUICIO: IMPACTO DE APOYOS REDUCTORES DE SESGOS EN LA
PRECISIÓN Y EN LOS TIEMPOS DE PROCESAMIENTO DEL PRONÓSTICO

Propuesta de Investigación

LINA MARCELA MERCHÁN LEÓN

JAIME ANDRÉS VARGAS CRUZ

DIRECTOR

JORGE ANDRÉS ALVARADO VALENCIA

M.S Analytics, Master en Educación

PONTIFICIA UNIVERSIDAD JAVERIANA

FACULTAD DE INGENIERÍA

BOGOTÁ

2014

Contenido

| | |
|--|----|
| Glosario | 4 |
| Resumen Ejecutivo..... | 7 |
| Introducción | 8 |
| Antecedentes | 10 |
| Pronósticos | 10 |
| Cualidades Deseables de los Pronósticos..... | 10 |
| Mecanismos para la Mejora de Pronósticos | 11 |
| Entrenamiento en Pronósticos | 11 |
| Interfaz y Sesgo..... | 12 |
| El Efecto Bouba-Kiki | 12 |
| Formulación del Problema | 14 |
| Pregunta Principal | 14 |
| Preguntas Secundarias..... | 14 |
| Objetivos | 15 |
| Objetivo General | 15 |
| Objetivos Específicos | 15 |
| Marco Teórico..... | 16 |
| Pronósticos | 16 |
| Precisión..... | 16 |
| Métricas del Error..... | 16 |
| Sesgos y Tipos de Sesgos..... | 18 |
| Entrenamiento | 19 |
| Métodos..... | 22 |
| Generalidades | 22 |
| Medición del Error | 22 |
| Diseño de Experimentos..... | 22 |
| Tamaño de Muestra..... | 23 |
| Hipótesis..... | 24 |
| Poblaciones (Participantes) | 25 |
| Instrumento..... | 26 |
| Análisis Estadístico | 28 |

| | |
|---|----|
| Resultados | 31 |
| 0. Consideraciones Iniciales | 31 |
| 1. Variable Respuesta Precisión | 31 |
| 1.1 Entrenamiento | 33 |
| 1.2 Separación Por Clásico | 36 |
| 1.3 Ruido | 37 |
| 1.4 Combinación de Entrenamiento y Ruido..... | 38 |
| 1.5 Tendencia..... | 40 |
| 1.6 Persona..... | 41 |
| 1.7 Combinación de Tendencia y Ruido..... | 42 |
| 2. Variable Respuesta Tiempo..... | 43 |
| 2.1 Orden | 45 |
| 2.2 Persona..... | 46 |
| 3. Relación Intervariables..... | 46 |
| 3.1 Factual | 46 |
| 3.2 Discusión | 47 |
| Conclusiones | 48 |
| Recomendaciones..... | 50 |
| Limitaciones del Experimento..... | 50 |
| Futuras Investigaciones | 50 |
| Referencias Bibliográficas | 52 |
| Apéndices | 58 |
| Apéndice A..... | 59 |
| Apéndice B:..... | 60 |
| Apéndice C:..... | 64 |
| Apéndice D:..... | 67 |
| Apéndice E:..... | 68 |
| Apéndice F: | 69 |
| Apéndice G:..... | 70 |
| Apéndice H:..... | 71 |

Glosario

A continuación se definen una serie de términos y conceptos utilizados a lo largo de la propuesta, ello a fin de facilitar la comprensión del documento.

- Boubá: Figura con rasgos redondeados ([Ramachandran & Hubbard, 2001](#)). En lo referente a este trabajo, este término se usará para aquellos gráficos de dispersión que unen sus puntos por una línea suavizada. Un ejemplo se puede evidenciar en la figura 1.

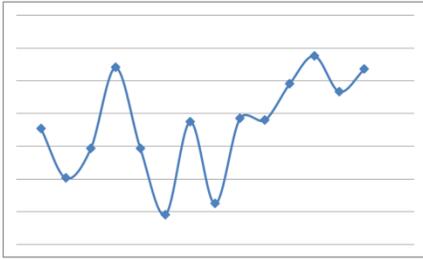


Figura 1. Gráfico de dispersión tipo Boubá

- Diseño de Experimentos: Un diseño experimental es esencialmente una estrategia de la experimentación industrial donde se puede cambiar simultáneamente un número de factores en un proceso para estudiar su efecto en las salidas del proceso o sistema ([Antony, 2003](#)).
- Efecto: Cambio en la respuesta promedio en las combinaciones de dos factores o entre dos condiciones experimentales ([Mason, Gunts & Hess, 2003](#)).
- Factor: Variables experimentales controlables que pueden influenciar los valores observados en las variables de respuesta ([Mason, Gunts & Hess, 2003](#)).
- Grupo Control: Es el grupo que no recibe alguno de los tratamientos que se están estudiando. Este grupo se compara con los grupos que reciben algún tratamiento para determinar si el nuevo tratamiento tiene efectos significativos.
- Interacción: Existencia de un efecto de factor conjunto en el cual el efecto de cada factor depende de los niveles de los otros factores ([Mason, Gunts & Hess, 2003](#)).
- Kiki: Figura con picos o ángulos cerrados. ([Ramachandran & Hubbard, 2001](#)). En lo referente a este trabajo, este término se usará para aquellos gráficos de dispersión que unen sus puntos por una línea no suavizada. Un ejemplo se puede evidenciar en la figura 2.

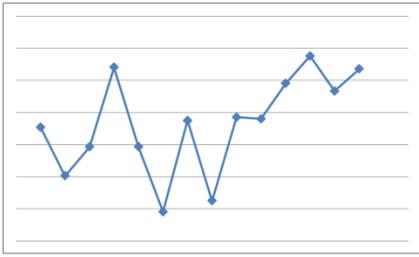


Figura 2. Gráfico de dispersión tipo Kiki

- Muestra: Grupo de observaciones que pertenecen a una población o proceso ([Mason, Gunts & Hess, 2003](#)).
- Nivel: Valor específico de un factor ([Mason, Gunts & Hess, 2003](#)).
- Observación: La recolección de información en un experimento o los valores obtenidos durante este. ([Mason, Gunts & Hess, 2003](#)).
- Población: Una población estadística consiste de todos los posibles miembros o unidades que poseen una o más características comunes al estar sometidas a unas condiciones experimentales u observacionales ([Mason, Gunts & Hess, 2003](#)).
- Población con entrenamiento tipo doble (Tradicional y Enfocado): Población en la cual sus individuos han pasado por una serie de entrenamientos que les ha permitido conocer sobre análisis de series de tiempo, modelos de pronóstico de la demanda, sesgos más comunes y como evitar que estos conlleven a errar a la hora de ajustar un pronóstico.
- Población con entrenamiento tipo enfocado: Población en la cual sus individuos han pasado por un entrenamiento en el cual se les ha enseñado cuales son los sesgos más comunes y que conllevan a errar a la hora de ajustar un pronóstico.
- Población con entrenamiento tipo tradicional: Población en la cual sus individuos han pasado por un entrenamiento en análisis de series de tiempo y en modelos de pronóstico de la demanda.
- Población sin entrenamiento: Población en la cual sus individuos no tienen ningún tipo de entrenamiento en el área de pronósticos.
- Réplica: Repetición de un experimento o una porción de él, bajo 2 o más grupos de condiciones ([Mason, Gunts & Hess, 2003](#)).

- Respuesta: El *outcome* de un experimento ([Mason, Gunts & Hess, 2003](#)).
- Tratamiento: Combinación de uno o más niveles de factores.
- Variable de respuesta: Resultado de un experimento ([Mason, Gunts & Hess, 2003](#)).

Resumen Ejecutivo

En la búsqueda de pronósticos más precisos se hace necesario ajustar los pronósticos de la demanda emitidos por un software estadístico por medio de ajustes de juicio, a fin de agregar conocimiento del producto al pronóstico, logrando así pronósticos que hagan la cadena de suministros más eficiente y de esa manera agregar valor a la operación. Sin embargo, en algunos casos la eficiencia de los ajustes de juicio se puede ver afectada por sesgos humanos.

Por lo tanto la presente investigación tuvo como finalidad determinar los efectos en la precisión y en los tiempos de procesamiento de un pronóstico ajustado por juicio, cuando este se realiza bajo la presencia de apoyos reductores de sesgos de tipo consciente e inconsciente específicos.

Respecto a los apoyos reductores de sesgos de tipo consciente, fueron comparadas cuatro poblaciones con diferentes niveles de entrenamiento: Tradicional, Enfocado, Doble (Tradicional + Enfocado) y Sin Entrenamiento, dónde se encontró que poblaciones con componentes de entrenamiento tradicional fueron más imprecisas respecto de aquellas que no lo recibieron. Así mismo también se encontró que en ausencia de información privilegiada, con base en la cual el pronosticador pueda ajustar, la precisión se perjudica en mayor medida por optar por grandes ajustes, en cambio, en presencia de pequeños ajustes se obtienen resultados similares a no ajustar.

La estrategia de tipo no consciente, buscó reducir el impacto del ruido sobre la calidad del ajuste, comparando el efecto de la forma visual del histórico mediante el uso de gráficas tipo Bouba (redondeadas) y tipo Kiki (puntiagudas). Se encontró que estos apoyos reductores de sesgos de naturaleza no consciente no generaron efectos estadísticamente significativos en la precisión y en el tiempo de procesamiento del ajuste de juicio.

Introducción

Planear y administrar adecuadamente la cadena de suministros es un objetivo principal en las compañías a nivel mundial. La planeación y pronóstico de la demanda es un punto clave en la administración de la cadena, ya que -en el marco de buenas prácticas del manejo de cadenas- es con base en los pronósticos que se da inicio a toda la operación. Debido a que los pronósticos suministran los datos de entrada para planear y controlar diversas áreas empresariales, tales como logística, producción, marketing y finanzas, así como para programar capacidades, recursos financieros y no financieros, el impacto del nivel de precisión del pronóstico repercute a gran escala en todos los eslabones de la cadena y etapas de operación ([Ballou, 2004](#)).

Existe una relación inversa entre el aumento de la precisión del pronóstico de la demanda y la disminución de costos en la empresa. Ello debido a que la varianza del error del pronóstico repercute de forma directa y creciente en la varianza asociada a los niveles de inventarios presentes en cada uno de los eslabones de la cadena, originando niveles de inventario diferentes al requerido ([Syntetos, Boylan & Disney, 2009](#)). Poseer mayores niveles de inventario, o niveles de inventario negativo (faltantes), conllevan a que se incurra en mayores costos de operación.

Esto se hace evidente en casos prácticos como el ocurrido a Nike en el 2001, cuando debido a malos pronósticos, tuvo pérdidas asociadas a altos inventarios y faltantes en varios de sus productos, lo cual llevo a una fuerte baja en el precio de sus acciones ([Worthen, 2003](#)). En sectores industriales particulares como el químico, se encontró que realizar cambios al método de pronóstico aumenta la probabilidad de conseguir ahorros significativos en los costos del área de inventarios ([Usuga & Alvarado, 2014](#)).

Los casos anteriormente descritos, así como la importancia de incorporar buenas prácticas en la administración de la cadena hacen evidente que pronosticar de manera eficiente (entendiéndose por eficiente, emitir pronósticos precisos dado una restricción de tiempo) se convierte en un paso

ineludible y diferenciador en las compañías. Por lo anterior, se hace necesario el desarrollo de una investigación encaminada a lograr pronósticos más precisos, disminuyendo con ello los costos asociados a la mala calidad de éstos, en los cuales incurren las compañías.

En el desarrollo de tal investigación, uno de los mecanismos de mejora es la determinación de métodos que permitan superar los sesgos e ineficiencias asociadas a los ajustes de juicio realizados a los pronósticos, con el ánimo de mejorar la precisión. Diferentes tipos de sesgos han sido documentados, dentro de los cuales se encuentran anclaje, aplicación de pequeños ajustes, dirección del ajuste/sobre-optimismo y modelamiento del ruido. ([Fildes, Goodwin, Lawrence & Nikolopoulos, 2009](#)).

A fin de superar los sesgos existentes en el proceso de pronosticar y con base en investigaciones previas, -que evidencian mejora en la precisión del pronóstico gracias a la implementación de técnicas de retroalimentación basadas en brindar a los pronosticadores información de sus propias prácticas- ([Legerstee & Franses, 2014](#)), éste proyecto de grado de carácter investigativo se enfoca en determinar la factibilidad de intervenciones al proceso de modelamiento de los pronósticos, tales como el nivel de entrenamiento y forma visual del histórico. Queriendo con ello contribuir a la solución de la problemática global asociada a la carencia de precisión de los pronósticos, dada una restricción de tiempo de procesamiento presente en el modelamiento y la existencia de sesgos.

Antecedentes

Pronósticos

Debido a que pronosticar es una herramienta fundamental en las empresas, a lo largo de las últimas décadas la academia ha dedicado una amplia investigación al tema ([Lawrence, Goodwin, O'Connor & Onkal, 2006](#); [Syntetos, Boylan & Disney, 2009](#)). El objetivo de pronosticar es estimar el comportamiento de una variable en el futuro, es por ello que la precisión asociada a dichas predicciones cobra gran importancia. A fin de garantizar un alto nivel de precisión, se han diseñado varias maneras de medir y comparar la calidad de los diferentes métodos usados para construir pronósticos.

Cualidades Deseables de los Pronósticos

Alta precisión y bajo tiempo de procesamiento del pronóstico son cualidades deseables al momento de pronosticar, por lo cual se han denominado dichas cualidades las variables respuesta del experimento.

Una de las cualidades más relevantes para cuantificar la calidad de un pronóstico es la precisión, es por ello que se han diseñado varias medidas de desempeño a fin de medirla ([Hyndman & Koehler, 2006](#)). Otra cualidad de gran importancia es el tiempo de procesamiento, el cual se define como el tiempo requerido para realizar una actividad, y en este caso particular, el tiempo utilizado para formular el pronóstico correspondiente. Este tiempo puede llegar a ser un inconveniente cuando se deben pronosticar amplios portafolios de SKU -Stock Keeping Units- ([Hyndman, 2014](#)), por lo tanto, buscando reducir esta deficiencia, se han realizado diferentes aproximaciones al área. Sin embargo, usualmente estos caminos se orientan a reducir el tiempo por medio de procedimientos sin intervención humana, o con la menor posible ([Gomez & Maravall, 2011](#); [Hyndman & Khandakar, 2007](#); [Hyndman, Koehler, Snyder & Grose, 2002](#)).

Mecanismos para la Mejora de Pronósticos

Paralelamente, se han buscado diferentes maneras de mejorar la precisión de los pronósticos; y para ello han propuesto diferentes soluciones, de las cuales ésta investigación se concentrará en: Entrenamiento del pronosticador y mejor diseño de la forma visual para representar los datos históricos.

Los ajustes de juicio tienen como objetivo combinar procedimientos estadísticos con información no contenida en los históricos, información de la cual tiene conocimiento la persona encargada de ajustar el pronóstico. Sin embargo, los pronósticos con ajustes de juicio tienen algunos inconvenientes que disminuyen la efectividad de este método. Uno de estos se denomina “sesgo”, y en él se incluyen aquellos errores del proceso mental en los que incurre el pronosticador. La literatura ha estudiado una gran variedad de sesgos tales como el anclaje, el uso de pequeños ajustes, el sobre-optimismo, la dirección del ajuste y la modelación de ruido ([Eroglu & Croxton, 2010](#); [Fildes et al. 2009](#)). De todos los anteriores la modelación del ruido se ha convertido en uno de los focos de la investigación, dado que se ha llegado a conclusiones disímiles respecto al impacto de éste en la calidad del pronóstico ([Lawrence & Makridakis, 1989](#); [Mosteller et al., 1981](#); [Fildes et al., 2009](#); [Goodwin & Wright 1993](#)); sin embargo, poco se ha investigado en opciones que minimicen el impacto de éste sesgo.

Entrenamiento en Pronósticos

Con miras a mejorar la capacidad de realizar predicciones, la academia ha buscado brindar a los futuros pronosticadores herramientas que les permitan sortear estos inconvenientes de manera adecuada. A partir de esto, se ha brindado un entrenamiento en universidades enfocado al conocimiento de los diferentes métodos de pronóstico y de los componentes de la demanda tales como demanda promedio, tendencia, estacionalidad y variación aleatoria.

La existencia de evidencia contradictoria respecto a la eficiencia de este entrenamiento en la capacidad de realizar mejores pronósticos ([Thomson, Pollock, Henriksen & Macaulay, 2004](#); [Yates, McDaniel, & Brown, 1991](#); [Onkal & Muradoglu, 1994](#)), ha conllevado a que algunos académicos consideren un entrenamiento alternativo, que no se fundamente en información teórica sobre series de tiempo, sino que se enfoque en brindar a los sujetos información acerca de prácticas deseadas y no deseadas al momento de pronosticar, a fin de mejorar la exactitud del pronóstico ([Sanders & Ritzman, 1992](#)).

Interfaz y Sesgo

A fin de superar los sesgos asociados a los ajustes de juicio, se han desarrollado investigaciones que estudian los cambios asociados a la presentación de la información. Gran cuantía de estas comparaciones provienen de la dicotomía tablas vs gráficas ([Desanctis & Jarvenpaa, 1989](#); [Harvey & Bolger, 1996](#)), concluyendo que cambios estéticos influyen o disminuyen la aparición de diferentes de sesgos.

El Efecto Bouba-Kiki

Por otro lado, desde el campo de la psicología se ha estudiado el impacto del uso de figuras puntiagudas -Kiki- y figuras redondeadas -Bouba- en la percepción del sujeto, adoptándose el nombre de efecto “Bouba and Kiki” para referirse a tal impacto. Los primeros estudios en el tema evidenciaron la capacidad cerebral para relacionar formas de naturaleza Bouba y Kiki con sonidos ([Köhler, 1929, 1947](#); [Ramachandran, & Hubbard, 2001](#); [Davis, 1961](#); [Nielsen, 2011](#)), dicha relación se observó no solo en poblaciones adultas, sino también en niños de temprana edad que aún no sabían escribir ([Maurer, Pathman, & Mondloch, 2006](#)).

Posteriormente se ha buscado trasladar las relaciones inconscientes entre Bouba-Kiki y sonidos al plano de la psicología y el comportamiento humano. Estudios realizados por Larson, Aronoff, Sarinopoulos y Zhu ([2009](#)) apoyados en imágenes de encefalogramas, concluyeron la

existencia de diferentes reacciones del cerebro ante círculos y triángulos posicionados hacia arriba o hacia abajo, encontrando principalmente que los triángulos apuntando hacia abajo, activan las zonas cerebrales relacionadas con la sensación de peligro en los seres humanos. Larson, Aronoff y Steuer ([2012](#)) mediante un procedimiento conocido como IAT (Implicit Association Test) - el cual mide relaciones inconscientes entre diferentes conceptos- concluyeron que se asociaban con mayor inmediatez los conceptos de no placentero con triángulos hacia abajo que con círculos o triángulos apuntados hacia arriba, adicionalmente, se encontró una relación de los círculos con placentero.

Lo anterior resulta importante para lograr mejorar la precisión por medio de brindar ayudas conscientes (entrenamiento) e inconscientes (forma visual del histórico) a los pronosticadores, a fin de que reduzcan el impacto de sus sesgos y puedan ajustar adecuadamente; ocasionando pronósticos más precisos.

Formulación del Problema

Pregunta Principal

¿Al realizar ajustes de juicio, apoyos reductores de sesgos de tipo consciente como el entrenamiento y de tipo inconsciente como la forma visual del histórico, afectan las variables respuesta: precisión y/o tiempo de procesamiento del pronóstico?

Preguntas Secundarias

¿El entrenamiento enfocado conlleva a hacer ajustes más precisos?

¿El nivel de entrenamiento afecta el tiempo de procesamiento del pronóstico?

¿Las gráficas tipo Bouba, cuando representan la forma visual del histórico, permiten realizar pronósticos más precisos que las tipo Kiki?

Objetivos

Objetivo General

Determinar el impacto de apoyos reductores de sesgos de tipo consciente como el entrenamiento y de tipo inconsciente como la forma visual del histórico, en las variables respuesta: precisión y tiempo de procesamiento del pronóstico, al realizar ajustes de juicio.

Objetivos Específicos

- Determinar los efectos de los entrenamientos (tradicional, enfocado y doble) en la precisión de un pronóstico.
- Determinar el impacto de los entrenamientos (tradicional, enfocado y doble) en el tiempo de procesamiento del pronóstico.
- Comparar los efectos de la forma visual del histórico (gráficas tipo Bouba y tipo Kiki) en la precisión de los pronósticos.

Marco Teórico

Pronósticos

Se define pronóstico como un dato de entrada por medio del cual se hace una estimación o conjetura de un evento futuro ([Armstrong, 1985](#); [Oxford English Dictionary, 2009](#)). Dado que “Un pronóstico perfecto es imposible, ya que en un ambiente de negocios hay demasiados factores que no se pueden pronosticar con certeza” ([Chase, Jacobs & Aquilano, 2009, p.468](#)) es importante revisar continuamente el modelo o la metodología de pronosticar, a fin de mejorarla ([Chase et al., 2009](#)).

Precisión

Una cualidad fundamental para tal fin es la precisión, entendiéndose ésta como el grado de cercanía entre el valor pronosticado y el valor real ([Joint Committee for Guides in Metrology \[JCGM 1\], 2008](#)). El nivel de precisión de un pronóstico aumenta cuando el valor absoluto de la diferencia entre el valor real y el pronosticado se reduce, ello debido a que existe una relación inversa entre la magnitud del error y el nivel de precisión asociado al pronóstico. En otras palabras la precisión del pronóstico consiste en minimizar la magnitud del error sistemático ([JCGM, 2008](#)).

Métricas del Error

Medir el grado de precisión de un pronóstico implica hacer una medición del error; es por ello que es significativo seleccionar la medida de error apropiada. En la práctica usualmente se han utilizado inapropiadas medidas de error, conllevando a que se concluya erróneamente acerca del real desempeño del pronóstico ([Davydenko & Fildes, 2013](#)). A fin de que una medida del error sea adecuada, esta debe satisfacer las siguientes propiedades: la primera hace referencia a ser independiente de la escala, y la segunda, respecto a su capacidad de no ser sensible ante la presencia de valores atípicos ([Davydenko & Fildes, 2013](#)). En el

[A](#) se detalla información de algunas de las medidas de error más usadas.

Para el desarrollo del presente experimento se usará *Median Absolute Percentage Error* (MdAPE) y *Mean Absolute Percentage error* (MAPE) como medida del error, con la salvedad de que Y_t no denota la observación real en el tipo t , sino la señal de la observación real en el tiempo t . En la Tabla 1 se amplía la información de éstas medidas del error.

Tabla 1.

Medidas del Error

| MdAPE | | |
|--------------|--|---|
| Expresión | Mediana $\left \frac{(Y_t - F_t)}{Y_t} \right $ | Dónde, Yt: Denota la señal de la observación en el tiempo t Ft: Denota el pronóstico de Yt. |
| Ventajas | Independiente de la escala Simple de aplicar Robusta frente a valores extremos | |
| Desventajas | Indeterminada cuando el dato real asume el valor de cero Distribución poco simétrica cuando el comportamiento real de los datos asume valores cercanos a cero | |
| MAPE | | |
| Expresión | Media Aritmética $\left \frac{(Y_t - F_t)}{Y_t} \right $ | Dónde, Yt: Denota la señal de la observación en el tiempo t Ft: Denota el pronóstico de Yt. |
| Ventajas | Independiente de la escala Usada ampliamente en la industria Simple de aplicar Indeterminada cuando el dato real asume el valor de cero | |
| Desventajas | Poco robusta frente a valores extremos Distribución poco simétrica cuando el comportamiento real de los datos asume valores cercanos a cero | |

Nota. Adaptado de Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4), 679-688.

Seleccionar una medida de precisión apropiada es tan solo uno de los factores relevantes a fin de maximizar la precisión del pronóstico. La precisión también se ve afectada debido a que en la práctica los pronosticadores realizan ajustes de juicio y al ajustar se incurre en un tiempo de procesamiento - espacio de tiempo que se requiere para realizar un proceso – el cual puede

aumentar significativamente al impartir un ajuste, disminuyendo con ello la utilidad del pronóstico, puesto que los pronósticos deben ser eficientes en términos de calidad o precisión y tiempo oportuno de entrega. Así mismo los ajustes presentan sesgos alterando de manera directa la calidad del pronóstico.

Sesgos y Tipos de Sesgos

Evans (citado por [Eroglu & Croxton, 2010](#)) define sesgo como la desviación o error sistemático que se tiene de un estándar, la cual puede ser controlada. A continuación se evidencian cuatro tipos de sesgos asociados a los ajustes de juicio, los cuales son una mala práctica a la hora de realizar ajustes de juicio.

El primer sesgo, que se analizará en esta investigación, se denomina dirección del ajuste/sobre-optimismo, cuyo impacto en la precisión del pronóstico varía dependiendo de la dirección del ajuste. Por lo general se presenta más inexactitud cuando los ajustes son positivos y tienden a sobreestimar el pronóstico final ya que se hacen bajo un sesgo de naturaleza optimista en el cual los pronosticadores usan indebidamente un optimismo injustificado. Por el contrario, impartir ajustes negativos generalmente es más efectivo ya que estos tienden a reducir el sesgo medio del pronóstico, al parecer porque una vez el pronosticador ha decidido hacer un ajuste negativo, es realista acerca de los niveles probables de la demanda ([Fildes et al., 2009](#)).

El segundo sesgo asociado a los ajustes de juicio, se denomina tamaño del ajuste; estudios anteriores demostraron que grandes ajustes tienden en promedio a mejorar la precisión, mientras que pequeños ajustes usualmente la dañan ([Fildes et al., 2009](#)).

El tercer sesgo hace referencia al modelamiento del ruido. El ruido, o error experimental, está compuesto por el error aleatorio, el cual recoge las diferencias existentes entre diferentes observaciones que no se pueden explicar, ya sea por su naturaleza o por la alta complejidad asociada a evaluar la variable ([Antony, 2003](#)). Estas variaciones no controlables pueden conllevar a modelaciones inadecuadas de la serie de tiempo, ya que los pronosticadores pueden confundir el ruido con la señal ([Harvey,](#)

[1995](#)) y tal como dicen Goodwin y Wright ([1993](#)), este es uno de los factores que pueden conllevar a una mayor complejidad de la serie.

El cuarto y último sesgo hace referencia al anclaje, el cual nos dice que las personas son propensas a darle una ponderación adicional a un dato particular, lo cual puede llevar a que si este es un valor atípico o no representativo, se perjudique la precisión del pronóstico.

Entrenamiento

Buscando realizar pronósticos de calidad, la industria junto a la academia han buscado entrenar y capacitar a la gente en esta área. El diccionario de Oxford ([2009](#)) define entrenamiento como el proceso de adquirir unas habilidades requeridas para realizar un trabajo.

La academia contemporánea brinda a los pronosticadores un entrenamiento aquí denominado como tradicional; el cual se expone en el Figura 3.

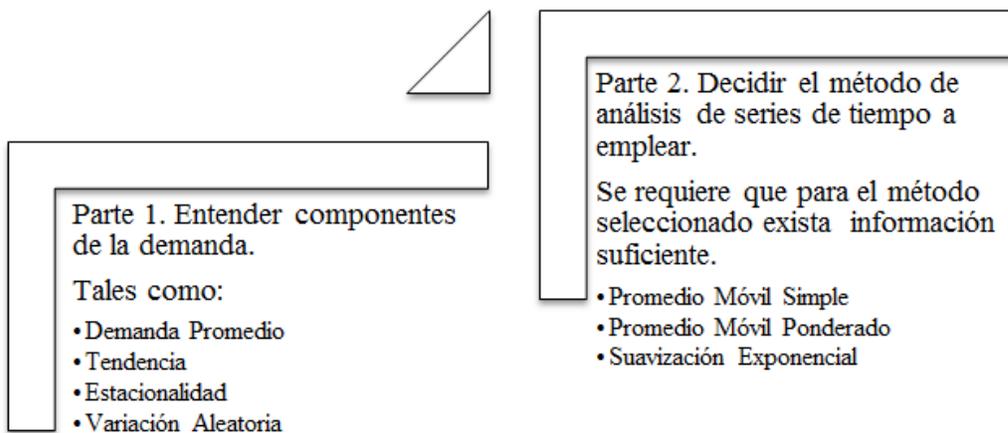


Figura 3. Entrenamiento Tradicional

Nota. Adaptado de [Chase, R. B., Jacobs, F. R., Aquilano, N. J. \(2009\)](#). *Administración de operaciones: Producción y cadena de suministros*. México, D.F.: McGraw-Hill.

Análisis previos han demostrado que los pronósticos y los ajustes de juicio pueden ser sesgados, ineficientes y se pueden ver afectados por la dirección y amplitud del ajuste, por lo que se considera pertinente explotar más tales características a fin de lograr pronósticos más precisos ([Fildes et al., 2009](#)). Para ello se propone un tipo de entrenamiento denominado enfocado, el cual se orienta a entrenar a los pronosticadores (sujetos de prueba) en evitar que cometan errores

típicos que empeoran la exactitud del pronóstico. En la Figura 4 se exponen los errores típicos a impactar con el entrenamiento propuesto. Con ello se lograría que el tiempo del pronosticador se emplee en lo que realmente agrega valor al pronóstico, como lo es identificar y tener en cuenta en la predicción dada variables externas tales como conocimiento de mercado, a fin de que influyan satisfactoriamente en la exactitud del pronóstico.

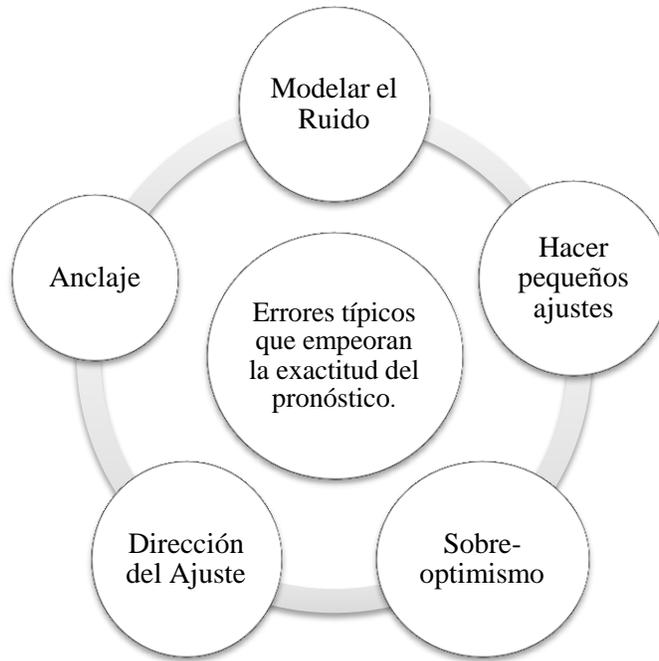


Figura 4. Objetivo del Entrenamiento Enfocado.

Nota. Adaptado de [Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. \(2009\)](#). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, 25(1), 3-23.

Adicional a lo anterior, existen diversas maneras de abordar los diferentes sesgos e intentar mitigar su impacto en la precisión del pronóstico, una de estas es mediante la modificación de la forma visual del histórico ([Desanctis & Jarvenpaa, 1989](#); [Harvey & Bolger, 1996](#)). Similarmente, en psicología se describe un fenómeno conocido como “Bouba and Kiki”, este fenómeno describe la existencia de asociaciones inconscientes de conceptos en el ser humano, en la Tabla 2 se exponen detalles acerca de tal fenómeno.

Tabla 2:

Revisión de Bouba – Kiki.

| Cita (Año) | Conclusión |
|--|---|
| Nielsen (2009) | El ser humano, sin importar su cultura o su educación, relaciona determinados sonidos con formas en particular. La palabra Bouba se relacionaba con formas redondeadas y la palabra Kiki con puntiagudas. |
| Larson, Aronoff, Sarinopoulos y Zhu (2009) | Se comprobó que las figuras de tipo Bouba and Kiki también se relacionaban inconscientemente con otros procesos mentales. Figuras Bouba se relaciona con el bienestar. Figuras Kiki con agresividad |

Por lo cual, la forma visual del histórico, podría tener un efecto sobre la manera en la cual se realiza el pronóstico, y finalmente en la precisión de éste. Particularmente se espera que la representación tipo Bouba de la forma visual del histórico de datos evite la modelación del ruido que se presenta en las gráficas tipo Kiki, puesto que la representación tipo Bouba se parece más a un modelamiento ya que se suavizan los puntos, conllevando a que sean más agradables para el usuario. Las gráficas Kiki son la forma tradicional en que se presenta el histórico de los datos.

Métodos

Generalidades

El experimento consistió en analizar las variables respuesta: precisión y tiempo de procesamiento del pronóstico en cuatro poblaciones diferentes, la precisión fue medida a través del Mean Absolut Percentage Error (MAPE) en lo que respecta al análisis de entrada (Pruebas de normalidad, homogeneidad e independencia) y a través del Median Absolute Percentage Error (MdAPE) en lo que respecta a las pruebas de Kruskal Wallis y de comparaciones múltiples de Dunn (La [Tabla 1](#) en Marco Teórico, amplia información de ambas medidas). Por su parte la variable tiempo de procesamiento del pronóstico fue medida en los segundos empleados en realizar el ajuste de juicio.

Medición del Error

El error fue medido individualmente a cada uno de los ajustes de juicio realizados por los sujetos a través del Absolut Percentage Error $APE = \left| \frac{(Señal_t - Ajuste\ de\ Juicio_t)}{Señal_t} \right|$, y el error total asociado al individuo a través del MdAPE o MAPE. En el cálculo del error se empleó la señal como referente de medición y no el dato real, debido a que se consideró que en el largo plazo en términos de precisión es más sano identificar adecuadamente la señal y pronosticar con base en ella, que con base en el dato real, pues éste último se ve afectado por el ruido.

Diseño de Experimentos

Se realizó un diseño experimental $4*3*2*2$ con orden como factor de bloqueo y persona como factor de bloqueo anidado al factor entrenamiento, para evaluar el impacto de diferentes niveles de entrenamiento, ruido, tendencia y forma visual del histórico en la precisión y el tiempo de procesamiento de un pronóstico ajustado por juicio. Los niveles de los factores pueden ser observados en la Figura 5. Factores y Niveles. Se estudiaron 48 combinaciones de niveles de factores o tratamientos.

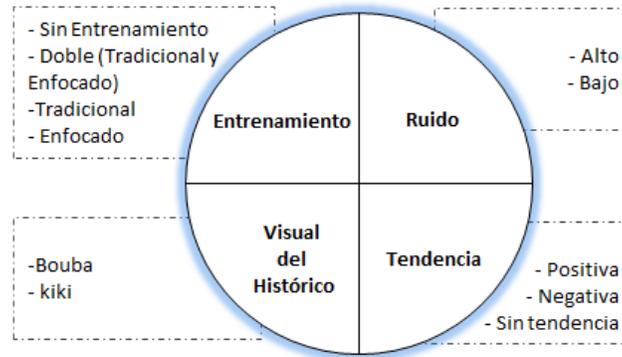


Figura 5. Factores y Niveles

Buscando reducir los efectos relacionados al aprendizaje y/o cansancio de los sujetos experimentales, se manejó un diseño de bloques, de tal manera que las 4 primeras observaciones de los sujetos pertenecieron a un bloque, las 4 siguientes a otro y las 4 finales a un último bloque, por lo cual se tuvo un total de tres bloques.

Respecto a bloquear por persona, estudios previos encontraron poca variación en la calidad de los pronósticos entre grupos de personas homogéneas, razón por la cual no se consideró al individuo un aspecto central de ésta investigación, sin embargo, a fin de aumentar la calidad del experimento y disminuir la variabilidad atribuida al error del mismo, se introdujo a persona como un factor de bloqueo anidado al factor entrenamiento, ya que así fue posible contemplar aspectos intrínsecos a cada individuo que ayudaron a explicar de mejor forma el modelo.

Tamaño de Muestra

El tamaño de muestra del experimento se calculó a partir del número de réplicas requeridas para cada tratamiento y fue calculado para la variable respuesta tiempo de procesamiento y precisión, sin embargo, dado que se usaron los mismos valores para los parámetros, el cálculo fue muy similar y arrojó los mismos valores de tamaño de muestra. Debido a que no existen fórmulas de tamaño de réplicas para más de dos factores, de las dos fórmulas mostradas en la tabla 3 se empleó la que arrojó el mayor número de réplicas luego de tomar la relación más restrictiva de a y b .

Tabla 3.

Tamaño de Muestra

| Fórmulas | Parametros: (Dean & Voss, 1999) |
|--------------------------------------|---|
| $r = 2b\sigma^2\phi^2/(a\Delta_B^2)$ | r: Número de ajustes de juicio por tipo de entrenamiento. a: Número de niveles correspondiente al factor que más niveles tuvo. b: Número de niveles correspondiente al factor que menos niveles tuvo. Δ : Hizo referencia a la diferencia mínima que se quiso detectar. Sus unidades fueron las de la variable respuesta (MAPE o Segundos empleados en ajustar) |
| $r = 2a\sigma^2\phi^2/(b\Delta_A^2)$ | ΔA : Para el primer factor . ΔB : Para el segundo σ^2 : Valor que se esperaba de varianza para la variable de respuesta. Sus ϕ^2 : Valor adimensional que dependió de la potencia y de la significancia que se buscó. |

El valor de Δ/σ fue manejado como una relación entre lo mínimo que se quiso detectar y la varianza de los datos, al cual se le asignó un valor de 0.3, debido a que esta proporción se consideró un nivel adecuado para detectar cambios significativos y no conllevar a tamaños de muestra poco prácticos. De igual manera, se tomaron valores de 4 para a, de 2 para b, un α del 5% y una potencia $(1-\beta)$ del 90%, lo cual conllevó a que r asumiera un valor mínimo de 178 ajustes de pronóstico. Debido a que fueron 4 niveles de entrenamiento, se tuvo un mínimo tamaño de muestra de 712 ajustes de pronóstico, de tal manera que cada uno de estos ajustes de pronóstico fue la unidad experimental. Tomando en cuenta que se requería por lo menos un tamaño de muestra de este valor y que se propusieron 12 tomas de datos por persona, fueron requeridas por lo menos 60 personas para mantener la significancia y potencia deseada.

Hipótesis

Las hipótesis a comprobar junto a su justificación se encuentran en la Tabla 4.

Tabla 4

Hipótesis y Justificación

| Hipótesis | Justificación |
|--|---|
| H ₁ : Los pronósticos emitidos con base en gráficas de características tipo Bouba son más precisos que los realizados en gráficas tipo Kiki. | Debido a los efectos psicológicos asociados a Bouba y Kiki, y a que al estar el histórico suavizado se le pondrá menor atención al ruido inherente a los datos, se considerará que la suavización conllevará a una mejor precisión. |
| H ₂ La precisión de los ajustes realizados por personas sin entrenamiento es igual a la de los realizados por aquellos con entrenamiento tradicional, pero inferiores a las de aquellos con entrenamiento Enfocado o doble. | Los sesgos son en parte culpables de la falta de precisión en los ajustes de juicio, si el entrenamiento tiene efecto, esto podría tener un efecto en la reducción de la imprecisión. |
| H ₃ : Los tiempos de procesamiento de los ajustes realizados por aquellos sin entrenamiento van a ser diferentes a los realizados por aquellos con algún tipo de entrenamiento. | Debido a la existencia de dos posibles efectos, el del experto que sabe donde mirar y por lo tanto es más veloz, y el del inexperto que al no saber, puede ignorar información y por lo tanto reducir el tiempo, se considerará que los tiempos serán diferentes, más no cual será mayor. |

Poblaciones (Participantes)

Los cuatro grupos contemplados para realizar el experimento fueron estudiantes Javerianos de Ingeniería Industrial y se diferenciaron principalmente por el tipo de entrenamiento en pronósticos bajo el cual fueron instruidos. A fin de determinar si los conocimientos teóricos en el área que poseía el individuo eran los requeridos por cada tipo de entrenamiento, se evaluaron los prerrequisitos de las clases que estaban cursando. La tabla 5 describe cada una de las poblaciones.

Tabla 5. *Poblaciones*

| | Entrenamiento Tradicional | Sin Entrenamiento | Entrenamiento Enfocado | Entrenamiento Doble |
|-------------------|---|---|---|--|
| Individuos | Estudiantes que en el momento de las pruebas se encontraron durante su curso de producción y ya habían visto el tema de pronósticos en clase. | Estudiantes que no habían cursado producción ni logística, debido a que en estos cursos se ve con cierta profundidad los temas de pronósticos. Se buscó que fueran estudiantes con cierto grado de avance de la carrera, para que al ser comparados con aquellos de conocimiento tradicional, no existieran tantas diferencias conceptuales ni de pensamiento no atribuibles a la asignatura de producción. | Estudiantes con características similares al grupo sin entrenamiento. | Estudiantes con el mismo perfil de la población con entrenamiento tradicional. |

| | | | | |
|----------------------|--|------------------------------|--|--|
| Entrenamiento | Corto entrenamiento en las temáticas definidas como “clásicas” (Series de Tiempo). | No se les dio entrenamiento. | En prácticas recomendadas por la literatura para realizar mejores ajustes de juicio. | En prácticas recomendadas por la literatura para realizar mejores ajustes de juicio y en las temáticas definidas como “clásicas” (Series de Tiempo). |
| Material | El apéndice B evidencia el entregable dado. | No aplica | El apéndice C evidencia el entregable dado. | A ésta población le fue dado los entregables del apéndice B y C. |

A fin de estimular la participación de cada integrante se tuvo un plan de incentivos, de manera que aquellas personas que se sometieron a las pruebas aumentaran su desempeño en el experimento. La recompensa consistió en dar a cada individuo que participo en el experimento una décima a la nota de un porcentaje previamente definido por el maestro de cada grupo, y dar una décima adicional a todas las personas cuya precisión de los ajustes de juicio realizados se encontrara por encima de la mediana de su grupo experimental. Esta recompensa varió un poco en parte del grupo que recibió entrenamiento tradicional, donde la bonificación tan solo se dio a aquellos que obtuvieron el mejor resultado.

Instrumento

Cada persona desarrolló el ejercicio en 14 demandas diferentes, las cuales incluyeron 2 de práctica y 12 que entraron a hacer parte de los datos. Cada demanda tuvo una gráfica con 14 datos históricos y con un pronóstico del mes 15, el cual fue realizado por el modelizador de pronósticos del software SPSS. Finalmente, con base en estos datos el sujeto se encargó de ajustar por juicio la demanda del mes 15.

Buscando medir el efecto de la forma visual del histórico y, por lo tanto, el efecto de un apoyo reductor de sesgo inconsciente, cada individuo realizó la mitad de sus pronósticos con base en gráficas que tuvieron unión suavizada entre sus puntos (tipo Bouba), y la otra mitad, con graficas que tuvieron sus puntos unidos por líneas rectas (tipo Kiki).

Las 14 gráficas tuvieron características diversas y previamente definidas, en términos de tendencia, demanda promedio y ruido a fin de darle mayor validez a los resultados del experimento en diferentes contextos de series de tiempo. Respecto a tendencia se manejaron tres niveles (Tendencia Positiva, Tendencia Negativa y Sin Tendencia), en demanda promedio se tuvieron diferentes valores los cuales fueron seleccionados de manera aleatoria, y en relación al ruido se diferenciaron dos niveles (Ruido Alto y Ruido Bajo). La tabla 6 evidencia el procedimiento incurrido a fin de generar cada una de las gráficas, debido a que los datos históricos no fueron reales y por el contrario fueron generados por los investigadores.

Tabla 6.**Generador de Gráficas**

| Explicación | | Tipo de Tendencia | | |
|---|--|---|--|---|
| | | Positiva | Negativa | Sin Tendencia |
| Ecuación a través de la cual fueron generados los datos históricos de cada una de las gráficas. | | $Y=a*x +b+e$ | | |
| a | Este término de la ecuación representa la pendiente de la gráfica. | Variable aleatoria uniforme que tomó valores entre 1 y 10 | Variable aleatoria uniforme que tomó valores entre 1 y 10 | Constante y asumió el valor 0. |
| b | Este término representa la parte estable de la gráfica. | Multiplicación de “a” por una variable aleatoria uniforme que tomó valores entre 1 y 10 | | Variable aleatoria uniforme que tomó valores entre 1 y 10 |
| x | Este representa el periodo actual | Tomó valores iguales al periodo que se generaba | El primer periodo tomó como valor 15, el siguiente 14, hasta que el último tomó 1. | Tomó valores iguales al periodo que se generaba |
| e | Representa el ruido blanco de la demanda. | Variable aleatoria generada mediante una distribución normal de media cero y desviación estándar igual al promedio de los periodos antes de agregar el ruido, multiplicado por un porcentaje, del 10% para el caso de ruido bajo y del 50% para ruido alto. | | |

La figura 4 evidencia el número de gráficas existentes en cada combinación de visual del histórico, ruido y tendencia.

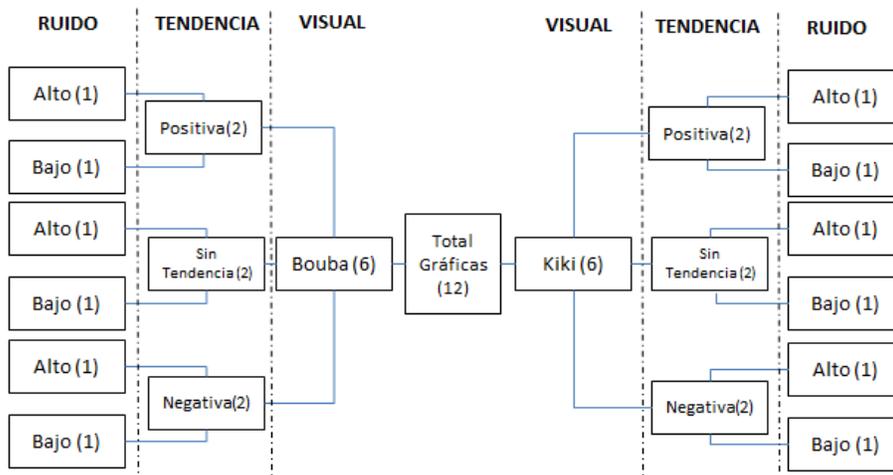


Figura 4. Distribución de Gráficas

Adicionalmente el orden en el cual las personas realizaron los diferentes pronósticos fue aleatorizado para cada individuo, buscando bloquear el efecto de variables ajenas al interés del experimento.

El proceso de ajuste se realizó de manera individual y la información pertinente se recolectó mediante un macro realizado en Excel -Visual Basic (Ver Apéndice H. Instrumento en digital). Buscando disminuir distracciones o sesgos por el uso de colores se procedió a trabajar el histórico de los datos con un color neutro tal como el azul.

Análisis Estadístico

Se realizó un análisis de entrada a cada una de las dos variables respuesta: tiempo y precisión. Con una confianza del 95% se rechazó en ambas variables la hipótesis nula de que la variable siguió una distribución normal así como la hipótesis nula de igualdad de varianzas en las poblaciones, sin embargo, por medio de una prueba de rachas el análisis de entrada demostró que tanto la variable respuesta tiempo como precisión aludieron independencia entre sus datos. Las rachas fueron realizadas por persona y no por ajuste de juicio -obteniendo un total de 83- debido a que se buscó comprobar independencia estadística entre las personas y no entre los ajustes de

juicio hechos por una misma persona, pues para mitigar la dependencia intra-persona el orden de las gráficas fue aleatorizado y el efecto del aprendizaje fue bloqueado en el diseño experimental. Las Tablas 7 y 8 evidencian el valor p, estadístico y conclusión a la que se llegó en cada una de las pruebas para precisión y tiempo de procesamiento, respectivamente.

Tabla 7

Análisis de Entrada Precisión

| Prueba | Test | Valor P | Estadístico | Conclusión |
|---------------|--------------------|---------|-------------|------------|
| Normalidad | Kolgomorov Smirnov | <0.01 | 0.214 | Rechaza |
| | Ryan Joiner | <0.01 | 0.746 | Rechaza |
| | Anderson-Darling | <0.005 | 118.41 | Rechaza |
| Homogeneidad | Levene | 0.001 | 1.46 | Rechaza |
| | Barlett | <0.001 | 11.47 | Rechaza |
| Independencia | Prueba de Rachas | 0.784 | | No Rechaza |

Tabla 8

Análisis de entrada de Tiempo de Procesameinto del Pronóstico

| Prueba | Test | Valor P | Estadístico | Conclusión |
|---------------|--------------------|---------|-------------|------------|
| Normalidad | Kolgomorov Smirnov | <0.01 | 0.248 | Rechaza |
| | Ryan Joiner | <0.01 | 0.746 | Rechaza |
| | Anderson-Darling | <0.005 | 118.41 | Rechaza |
| Homogeneidad | Levene | <0.001 | 1.51 | Rechaza |
| | Barletts | <0.001 | 9.15 | Rechaza |
| Independencia | Prueba de Rachas | 0.753 | | No Rechaza |

Debido a que los datos no se ajustaron a una distribución normal, ni tampoco presentaron varianzas iguales, los resultados se analizaron a través de las siguientes dos pruebas no paramétricas:

- Kruskal-Wallis: Se empleó para identificar si las fuentes de variación afectaron la variable respuesta.
- Prueba de Comparaciones Múltiples Dunn: Se empleó para identificar las diferencias existentes entre los niveles o tratamientos que pertenecen a una misma fuente de variación.

Posteriormente se procedió a utilizar la Prueba Chi Cuadrado, con la cual se buscó identificar la causa de las diferencias existentes entre los niveles o tratamientos que pertenecen a una misma fuente de variación, mediante la elaboración de variables categóricas que contemplaron ciertos comportamientos o sesgos. Finalmente, y buscando tener un criterio base acerca de las interacciones que pudieron haber afectado las variables respuesta, se realizaron previamente pruebas paramétricas de análisis de varianza (ANOVA).

Resultados

0. Consideraciones Iniciales

En el estudio se analizaron 996 muestras de ajustes de juicio de pronósticos, superando el valor de 712 ajustes de juicio mínimo requerido, conllevando a que se trabajaran con 83 personas, las cuales se detallan por su tipo de entrenamiento en la tabla 9. Se realizaron 5 pruebas de independencia Chi cuadrado, las cuales se trabajaron con una significancia del $\alpha= 0.022085$, a fin de tener una confianza total de la prueba del 90%. Por su parte en la prueba de comparaciones múltiples de Dunn se trabajó con la significancia requerida (dado el número de fuentes de variación analizadas para cada factor) a fin de tener una confianza total en la prueba del 80%.

Tabla 9

Detalle tamaño de muestra

| | Entrenamiento Tradicional | Sin Entrenamiento | Entrenamiento Enfocado | Entrenamiento Doble |
|------------------------------|----------------------------------|--------------------------|-------------------------------|----------------------------|
| No. Individuos | 16 | 27 | 21 | 19 |
| No. Ajustes de Juicio | 192 | 324 | 252 | 228 |

En ésta sección se evidencian los resultados obtenidos en tres grandes capítulos: 1. Variable Respuesta Precisión, 2. Variable Respuesta Tiempo de Procesamiento y 3. Relación Intervariables. Al interior de cada capítulo se realizó un análisis por fuente de variación, y para las fuentes de variación más importantes se dio una división entre los resultados del efecto de la fuente de variación y las posibles explicaciones de dicho resultado

1. Variable Respuesta Precisión

Para el análisis de precisión, se aplicó un ANOVA y se encontró un bajo valor del R cuadrado, por lo que se presume la existencia de variables no contempladas en el experimento que ayudan a explicar mejor la precisión de cada ajuste, lo cual es lógico cuando se estudia el comportamiento de seres humanos. La tabla con el ANOVA se evidencia en el apéndice D.

A fin de determinar aquello que afecta la precisión de los pronósticos ajustados por juicio, se analizaron 11 posibles fuentes de variación por medio de una prueba Kruskal Wallis, en la cual se empleó una significancia del $\alpha = 0,0095316$ para tener una confianza total en la prueba del 90% dado que ésta prueba pierde confianza proporcionalmente con la cantidad de fuentes de variación analizadas. Las fuentes de variación analizadas fueron: los factores principales, los factores de bloqueo, combinaciones de factores principales que se consideraron relevantes y la fuente de variación denominada separación por clásico, la cual se profundizó más adelante. La tabla 10 evidencia el estadístico, grados de libertad, valor p y conclusión a la que llegó la prueba en cada una de las fuentes de variación propuestas.

Tabla 10

Prueba Kruskal Wallis (Precisión)

| Fuente de Variación | Estadístico | Grados de Libertad | Valor P | Conclusión |
|---------------------------|-------------|--------------------|---------|---------------|
| Entrenamiento | 12.47 | 3 | 0.006 | Rechaza |
| Visual | 2.73 | 1 | 0.066 | No se rechaza |
| Entrenamiento*Visual | 17.22 | 7 | 0.016 | No se rechaza |
| Tendencia | 212.48 | 2 | <0.001 | Rechaza |
| Ruido | 3.52 | 1 | <0.001 | Rechaza |
| Orden | 0.53 | 2 | 0.767 | No se rechaza |
| Persona | 174.84 | 82 | <0.001 | Rechaza |
| Entrenamiento * Tendencia | 227.91 | 11 | <0.001 | Rechaza |
| Entrenamiento*Ruido | 190.34 | 7 | <0.001 | Rechaza |
| Separación por Clasico | 9.82 | 1 | 0.002 | Rechaza |
| Tamaño del Ajuste | 240.34 | 2 | <0.001 | Rechaza |

En lo que respecta al factor visual, la combinación de entrenamiento*visual y al orden, la prueba demostró con valores p superiores a la significancia ($\alpha = 0,0095316$) evidencia suficiente para no rechazar la hipótesis nula de que las medianas de los grupos no son diferentes entre ellas, y con ello afirmar que dichas fuentes de variación no generaron un efecto significativo en la variable dependiente precisión. Sin embargo, dado que los valores p del factor visual y combinación de entrenamiento*visual fueron cercanos a la significancia se consideró que aunque estadísticamente no afecten la variable, es sano no descartar fuertemente la hipótesis de que el factor visual del histórico de los datos afecta la precisión.

Por otro lado, las fuentes de variación entrenamiento, tendencia, ruido, persona, combinación de entrenamiento y tendencia, combinación de entrenamiento y ruido, combinación de tendencia y ruido, separación por clásico y tamaño del ajuste, demostraron con valores p inferiores a la significancia ($\alpha= 0,0095316$) que afectaron la variable respuesta precisión, lo que sugiere que al menos una de las medianas de las k poblaciones fue diferente. A continuación se hace un análisis detallado de cada una de estas fuentes de variación.

1.1 Efectos del Entrenamiento en la Precisión

El tipo de entrenamiento fue una de la variables de mayor interés para el experimento, ya que con él se buscó afectar la inferencia de los diferentes sesgos, y con ello cambiar la precisión.

1.1.a Resultados: Efectos del Entrenamiento en la Precisión

Se determinó que el entrenamiento afectó la precisión del pronóstico, de tal manera que las poblaciones sin entrenamiento o que recibieron entrenamiento Enfocado generaron los pronósticos más precisos. La población con entrenamiento Doble conllevó a los ajustes de juicio más imprecisos, originando pronósticos menos precisos respecto de las poblaciones sin entrenamiento y con entrenamiento Enfocado. Por su parte la población con entrenamiento tradicional fue estadísticamente igual de precisa a todas las otras poblaciones de entrenamientos. La tabla 11 evidencia los grupos formados por el test de Dunn y sus medianas.

Tabla 11

Test de Comparaciones Múltiple de Dunn (Precisión – Fuente de Variación Entrenamiento).

| Grupo | Niveles | Mediana |
|--------------|-------------------|----------------|
| Grupo A | Enfocado | 0.08333 |
| | Sin Entrenamiento | 0.09073 |
| Grupo A y B | Tradicional | 0.11287 |
| Grupo B | Doble | 0.11875 |

1.1.b Discusión e implicaciones del efecto del entrenamiento en precisión

Se supuso que una de las causas de que el entrenamiento afectara la precisión del pronóstico, fue el apetito o aversión que tuvo cada población de entrenamiento por ajustar el pronóstico. Por lo cual posteriormente se procedió a realizar un análisis considerando si estos habían ajustado o no. La tabla 12 evidenció que la población con entrenamiento enfocado fue la que más presento aversión por ajustar (el resultado observado del grupo se incrementó un poco más de seis desviaciones estándar del resultado esperado), por su parte la población con entrenamiento doble fue la que menos aversión por ajustar presento, un poco menos de tres desviaciones estándar en relación a lo esperado. A causa de lo anterior se infirió que el hecho de que la población con entrenamiento doble haya ajustado en mayor medida que el resto de las poblaciones fue determinante en hacerlos la población más imprecisa, y por el contrario la buena práctica de no impartir ajustes cuando éste no se basa en un argumento contundente influyó en que la población con entrenamiento enfocado fuera la más precisa.

Tabla 12

Residuales Estandarizados (Entrenamiento Vs ¿Apetito - Aversión?)

| <i>¿Ajustar?</i> | <i>Entrenamiento</i> | | | |
|-----------------------------|----------------------|-----------------|--------------------------|--------------------|
| | <i>Doble</i> | <i>Enfocado</i> | <i>Sin Entrenamiento</i> | <i>Tradicional</i> |
| <i>Apetito por Ajustar</i> | 1.435 | -2.616 | 1.27 | -0.216 |
| <i>Aversión por Ajustar</i> | -3.608 | 6.579 | -3.193 | 0.544 |

La población sin entrenamiento no se abstuvo de ajustar, sin embargo cuando decidieron ajustar, incurrieron en pequeños ajustes, y es por esta razón que aun cuando incurrieron en la mala práctica de ajustar por ajustar, obtuvieron pronósticos más precisos respecto a la población con entrenamiento Doble. La tabla 13 evidencia el apetito que tuvo cada una de las poblaciones de entrenamiento por incurrir en ajustes grandes, ajustes pequeños y no ajustar.

Tabla 13
Residuales Estandarizados (Tamaño del Ajuste)

| Tamaño del Ajuste | Entrenamiento | | | |
|-------------------|---------------|----------|-------------------|-------------|
| | Doble | Enfocado | Sin Entrenamiento | Tradicional |
| Ajuste Grande | 2.502 | -2.713 | -0.646 | 1.222 |
| No ajustar | -3.608 | 6.579 | -3.193 | 0.544 |
| Ajuste Pequeño | -0.593 | -0.918 | 2.571 | -1.641 |

Al analizar la causa de porque incurrir en pequeños ajustes dañó en menor medida la precisión del pronóstico respecto de incurrir en ajustes grandes, aun cuando los dos tipos de ajustes carecieron de información privilegiada, se encontró por medio de un Test de comparaciones múltiples de Dunn (ver Tabla 14) que el efecto generado en la precisión al no ajustar e incurrir en ajustes pequeños fue estadísticamente igual, y que por el contrario incurrir en grandes ajustes en ausencia de información privilegiada disminuyó en mayor medida la calidad del pronóstico. Se presume que el efecto de los ajustes pequeños en la precisión del pronóstico, pudo ser resultado de un alto componente aleatorio no justificado, y que a causa de ello tal como pudieron acercarse un poco más a la señal, pudieron también alejarse, pero debido a que un ajuste pequeño, por su misma magnitud habrá tenido un efecto pequeño, dicho efecto pudo no haberse detectado dadas las significancias usadas en la prueba.

Tabla 14

Test de Comparaciones Múltiples de Dunn (Precisión – Fuente de Variación: Tamaño del Ajuste).

| Grupo | Niveles | Mediana |
|---------|----------------|---------|
| Grupo A | Ajuste Pequeño | 0,03972 |
| | No Ajustar | 0,05787 |
| Grupo B | Ajuste Grande | 0,24861 |

Nótese que a pesar de que la literatura converge a que en términos de precisión, incurrir en grandes ajustes es mejor que hacer pequeños ajustes, en éste experimento se demostró que en un escenario en el cual la población no cuenta con información privilegiada, incurrir en pequeños

ajustes es menos perjudicial para la precisión que incurrir en grandes ajustes y muy similar a no ajustar.

1.2 Efectos del Entrenamiento Clásico en la Precisión

Debido a que la población con entrenamiento tradicional fue estadísticamente igual de precisa al resto de las poblaciones, se quisieron evaluar las diferencias existentes entre los ajustes de juicio realizados por personas que recibieron componentes de entrenamiento clásico tales como las poblaciones con entrenamiento tradicional y entrenamiento doble, y los ajustes realizados por individuos que no recibieron ningún tipo de entrenamiento clásico, lo cual incluyó a las poblaciones de entrenamiento enfocado y sin entrenamiento.

1.2.a Resultados: Efectos del Entrenamiento Clásico en la Precisión

La prueba de Kruskal Wallis identificó que ésta fuente de variación afectó de manera significativa la precisión del ajuste de juicio. Mediante una diferencia de medianas se evidenció que los ajustes de juicio realizados por personas que no recibieron entrenamiento clásico fueron 26% más precisos que los realizados por personas que si recibieron entrenamiento clásico, la tabla 15 evidencia la mediana de cada población y la diferencia encontrada.

Tabla 15

Análisis de Medianas – Fuente de Variación Separación por Clásico

| Análisis de Medianas | | |
|---|-------------|-------------|
| Precisión-Fuente de Variación Separación por Clásico | | |
| Nivel | Sin Clásico | Con Clásico |
| Mediana | 0.08708 | 0.11750 |
| Diferencia | 26% | |

1.2.b Discusión e implicaciones de los Efectos del Entrenamiento Clásico sobre la Precisión

Los resultados de la tabla 15 indujeron a sospechar que componentes de entrenamiento clásico conllevaron a disminuciones de la precisión del pronóstico, lo cual pudo ocurrir debido a que

éstas instruyen a las personas en cómo hacer un pronóstico y no en como ajustarlo. En la práctica se tienen potentes herramientas computacionales que emiten buenos pronósticos y lo que se espera del individuo es que ajuste dicho valor basado en variables que la máquina no contempla y no en que compita con ella, que fue tal vez en lo que cayó la población con entrenamiento clásico.

1.3 Ruido en Datos Históricos y su Efecto en la Precisión

El ruido, como variable inherente a los datos no depende del investigador, pero conocer su impacto en la precisión permitió conjeturar la importancia de reducir su impacto al realizar un ajuste de juicio.

1.3.a Resultados: Efecto del Ruido en la Precisión

Se identificó mediante una diferencia de medianas que los ajustes de juicio realizados en gráficas con características de ruidos bajos fueron 77% más precisos que los realizados en gráficas con características de ruidos altos. La tabla 16 evidencia la mediana de cada grupo y la diferencia encontrada.

Tabla 16.

Análisis de Mediana – Fuente de Variación Ruido

| Análisis de Medianas | | |
|--|---------|---------|
| Precisión-Fuente de Variación Ruido | | |
| Nivel | Alto | Bajo |
| Mediana | 0.24331 | 0.05699 |
| Diferencia | 77% | |

1.3.b Discusión e Implicaciones de los Efectos del Ruido sobre la Precisión

Éste resultado sugirió que no sólo características intrínsecas al individuo tales como el entrenamiento incidieron en la precisión del pronóstico, sino que también factores ajenos al individuo y de naturaleza única del comportamiento de los datos, tales como la presencia de ruidos altos o bajos generó un efecto el resultado.

1.4 Efectos de la Interacción Entrenamiento-Ruido sobre la Precisión

Ahora bien, se hizo pertinente determinar si el tipo de entrenamiento repercutió en la interpretación y manejo del ruido asociado a los datos, lo cual permitió conocer si el entrenamiento tuvo un efecto sobre la manera en la cual se interpretaba el ruido.

1.4.a Resultados: Interacción Entrenamiento-Ruido y su efecto en la Precisión.

La tabla 17 evidencia los valores que tomaron las medianas de las distintas combinaciones de entrenamiento y ruido.

Tabla 17

*Test de Comparaciones Múltiple de Dunn (Precisión – Fuente de Variación Entrenamiento*Ruido).*

| Niveles | Mediana |
|------------------------|----------------|
| Enfocado*Bajo | 0,04935 |
| Sin Entrenamiento*Bajo | 0,05471 |
| Tradicional*Bajo | 0,06022 |
| Doble*Bajo | 0,06982 |
| Sin Entrenamiento*Alto | 0,20969 |
| Enfocado*Alto | 0,22111 |
| Tradicional*Alto | 0,23363 |
| Doble*Alto | 0,3405 |

En presencia de gráficas con ruidos altos el entrenamiento no generó ningún tipo de efecto ya que todas las poblaciones de entrenamientos (enfocado, doble, tradicional y sin entrenamiento) fueron estadísticamente igual de precisas. Por medio de una prueba Chi Cuadrada se concluyó que en presencia de ruidos altos no existió una correlación entre el tipo de entrenamiento y el hecho de que el ajuste de juicio haya mejorado el pronóstico dado por SPSS (ver apéndice E),

ello sugirió que en presencia de ruidos altos todos los entrenamientos incurrieron en la mala práctica de modelar el ruido y no identificaron adecuadamente la señal de la gráfica.

Por el contrario, en presencia de ruidos bajos la población sin entrenamiento fue más precisa que la población con entrenamiento doble y las poblaciones con entrenamiento tradicional y enfocado fueron estadísticamente igual de precisas a todas las poblaciones. Por medio de una prueba Chi Cuadrada se concluyó que en gráficas con características de ruidos bajos existió asociación entre el tipo de entrenamiento recibido y el hecho de que el ajuste haya mejorado la precisión del pronóstico dado por SPSS (calidad del ajuste de juicio impartido). En lo que respecta a la población sin entrenamiento, se identificó que la frecuencia observada de la cantidad de ajustes de juicio que mejoraron la precisión del pronóstico fue 1.6 desviaciones estándar superior a la esperada, por lo cual se presume que en presencia de ruidos bajos la población sin entrenamiento modeló menos el ruido de la gráfica e identificó en mayor medida la señal con relación a las demás poblaciones de entrenamiento. La Tabla 18 evidencia los residuos.

Tabla 18.

*Residuales Estandarizados Entrenamiento Vs Calidad de Ajuste (**Ruido Bajo)*

| <i>Calidad del Ajuste</i> | <i>Entrenamiento</i> | | | |
|----------------------------|----------------------|-----------------|--------------------------|--------------------|
| | <i>Doble</i> | <i>Enfocado</i> | <i>Sin Entrenamiento</i> | <i>Tradicional</i> |
| <i>No Mejoro Precisión</i> | 1.3409 | -0.2522 | -1.2105 | 0.3629 |
| <i>Si Mejoro Precisión</i> | -1.847 | 0.3474 | 1.6674 | -0.4998 |

Nota: Ésta tabla sólo contempla los ajustes de juicio realizados en ruido bajo

1.4.b Discusión e implicaciones la Interacción Entrenamiento-Ruido y su efecto en la Precisión.

El ruido afectó la precisión de los ajustes, cuando éste era alto ninguno de los entrenamientos tuvo un efecto estadísticamente significativo, lo cual nos sugiere que la misma búsqueda de patrones se encuentra tan interiorizada en nuestras mentes que se requiere algo más fuerte para lograr abandonarlo. Por otro lado, en presencia de ruido bajo podemos evidenciar que Doble y Tradicional (este último sin ser estadísticamente significativo), fueron más propensos a ajustes

que perjudicaron la precisión, lo cual indica que una vez el efecto del ruido va a un segundo plano, lo aprendido en los entrenamientos toma fuerza.

1.5 Tendencias en Datos Históricos y sus Efectos sobre la Precisión

Al igual que el ruido, la tendencia es otra característica intrínseca al comportamiento de la serie de tiempo en estudio, y por lo tanto, conocer si esta tuvo un efecto sobre la precisión y la magnitud de este, permitió dilucidar la importancia de entrenamientos u otros tipos de ayudas que reduzcan su efecto sobre la precisión de ajustes de juicio.

1.5.a Resultados: Tendencia y su Efecto en la Precisión

Se encontró que la tendencia intrínseca a la gráfica que representa el histórico de datos afectó la calidad del ajuste y con ello, la precisión del pronóstico. Ajustes de juicio realizados en gráficas con tendencias negativas fueron menos precisos que aquellas realizadas en gráficas sin tendencia y con tendencia positiva, siendo estas dos últimas no estadísticamente diferentes. Lo cual concuerda con la literatura, la cual converge a que existen mayores dificultades cuando el ajuste se hace sobre pendientes negativas. La tabla 19 evidencia los grupos que se formaron mediante el uso del test de comparaciones múltiple de Dunn.

Tabla 19

Test de Comparaciones Múltiples de Dunn Precisión – Fuente de Variación Tendencia

| Grupo | Niveles | Mediana |
|--------------|----------------|----------------|
| Grupo A | Sin tendencia | 0.06167 |
| | Positiva | 0.06466 |
| Grupo B | Negativa | 0.39119 |

A fin de identificar el causal de las diferencias de precisión en escenarios de tendencias positivas, negativas y sin tendencia, se aplicó una prueba de independencia Chi Cuadrada entre la dirección del ajuste y el tipo de tendencia asociado a la gráfica concluyendo la prueba que dichas variables están correlacionadas. La Tabla 20 evidencia los residuos estándar.

Tabla 20

Residuales Estandarizados (Dirección del Ajuste Vs Tendencia)

| Tendencia | Dirección del Ajuste | | |
|------------------|-----------------------------|--------|-----------|
| | Abajo | Arriba | No Ajusto |
| Negativa | -0,616 | 1,132 | -0,792 |
| Positiva | 2,117 | -1,48 | -1,386 |
| Sin Tendencia | -1,501 | 0,348 | 2,178 |

1.5.b Discusión e Implicaciones de los Efectos de la Tendencia en la Precisión

El análisis del error estandarizado evidenció que en gráficas con tendencias positivas la cantidad de ajustes observados en dirección “abajo” fueron mayores un poco más de dos desviaciones estándar respecto de los esperados. Lo anterior sugirió que bajo tendencias positivas los individuos cayeron en "*Trend Damping*", que consiste en intentar estabilizar una pendiente sustentado en una visión conservadora, y contrario a lo que se esperaba no incurrieron en la mala práctica de sobre optimismo. Así mismo, se identificó que la no existencia de una tendencia en la gráfica, conllevó a que los individuos se inclinaron dos desviaciones estándar más de lo esperado por el no ajuste de los datos, originando ello que fueran más precisos pues no cayeron en la mala práctica de ajustar por ajustar.

1.6 La Singularidad del individuo y su efecto en la Precisión

La prueba de Kruskal- Wallis concluyó que las personas generaron un efecto en la precisión del pronóstico, por lo cual se sospecha fuertemente que particularidades relacionadas a cada individuo que no fueron contempladas en el experimento influyeron en la calidad del ajuste de juicio. No obstante, no se incurrió en un análisis más profundo, puesto que no es de interés en ésta investigación conocer la particularidad de que individuo fue mejor que otro.

1.7 Efectos de la Interacción Tendencia-Ruido en la Precisión

Analizar el efecto de las interacciones entre tendencia y ruido, dos variables intrínsecas a los datos históricos, permitieron dilucidar la forma en que estas dos variables afectan la percepción de la información de los sujetos.

1.7.a Resultados: Efecto en la Precisión de la Interacción Tendencia-Ruido

Al analizar dicha combinación de factores se identificó que en presencia de tendencias negativas el nivel de ruido afectó en menor medida la precisión del ajuste respecto de la tendencia, por el contrario cuando se ajustaron pronósticos bajo graficas con tendencias positivas y graficas sin tendencia, el nivel de ruido asociado a las gráficas primo en mayor medida que la tendencia en la manera como el individuo ajustó por juicio el pronóstico. La tabla 21 evidencia las medianas de cada combinación de tendencia y ruido y los grupos estadísticos que encontró la prueba de Dunn.

Tabla 21

Medianas (Tendencia Vs Ruido)

| Grupo | Niveles | Mediana |
|----------------|--------------------------------|----------------|
| Grupo A | Ruido Bajo- Tendencia Positiva | 0,02638 |
| | Ruido Bajo-Sin Tendencia | 0,04042 |
| | Ruido Alto-Sin Tendencia | 0,10375 |
| Grupo B | Ruido Alto-Tendencia Positiva | 0,11111 |
| | Ruido Bajo-Tendencia Negativa | 0,14273 |
| Grupo C | Ruido Alto-Tendencia Negativa | 0,64286 |

1.7.b Discusión e Implicaciones del Efecto de la Interacción Tendencia-Ruido en la Precisión

Lo anterior permitió conocer que en diferentes combinaciones de tendencia y ruido, la precisión se afectó de diversas maneras, lo cual sugirió que los diferentes sesgos pueden verse alterados por más de una una característica de las series de tiempo a la vez.

1.8 Recapitulación de Precisión

La tabla 22 sintetiza los hallazgos de ésta investigación en lo que a precisión respecta. Se evidencia que la precisión se vio afectada tanto por factores propios a la persona, como aquellos que dependen de la naturaleza de los datos.

Tabla 22

Tabla Resumen Variable Respuesta Precisión

| Fuente de Variación | Hallazgo |
|-------------------------------|--|
| Entrenamiento | Poblaciones sin entrenamiento o que recibieron entrenamiento enfocado generaron los pronósticos más precisos, ello a causa de que el grupo de entrenamiento enfocado fue averso a ajustar. La población con entrenamiento tuvo <u>apetito por ajustar lo que originó que fuera la más imprecisa de todas.</u> |
| Separación por Clásico | Ajustes de juicio realizados por personas que no recibieron entrenamiento clásico fueron más precisos que los realizados por personas que si recibieron entrenamiento clásico, debido a que técnicas clásicas instruyen a las personas en cómo hacer un pronóstico y no en como ajustarlo. |
| Visual | No Afectó, sin embargo dado que los valores p fueron cercanos a la significancia no se descarta la hipótesis de que pueda afectar. |
| Tendencia | Ajustes de juicio realizados en gráficas con tendencias negativas fueron menos precisos que aquellos realizados en gráficas sin tendencia y con tendencia positiva. Debido a que bajo tendencias positivas los individuos no incurrieron en la mala práctica de sobre optimismo, y ante la no existencia de una tendencia en la gráfica, los individuos se inclinaron por el no ajuste de los datos. |
| Ruido | Ajustes de juicio realizados en gráficas con características de ruidos bajos fueron más precisos que los realizados en gráficas con características de ruidos altos. |
| Entrenamiento*Ruido | En presencia de gráficas con ruidos altos el entrenamiento no generó ningún tipo de efecto, debido a que todos los entrenamientos tipos de entrenamiento incurrieron en la mala práctica de modelar el ruido y no identificar adecuadamente la señal de la gráfica. Por el contrario, en presencia de ruidos bajos la población sin entrenamiento fue más precisa que la población con entrenamiento doble . |

2. Variable Respuesta Tiempo

Los resultados del ANOVA sugirieron la inexistencia de interacciones entre factores que afecten el tiempo empleado en realizar los ajustes de juicio a los pronósticos. Adicionalmente, al igual que con la precisión, el bajo valor del R cuadrado obtenido presume la existencia de variables no

contempladas en el experimento que pudieron ayudar a explicar mejor el tiempo de procesamiento de cada ajuste, lo cual fue de esperarse cuando las variables dependen de seres humanos y de su manera de pensar. La tabla ANOVA se evidencia en el apéndice F.

A fin de conocer los factores que afectaron el tiempo de procesamiento del ajuste se analizaron 7 posibles fuentes de variación por medio de la prueba Kruskal Wallis – en la cual se empleó una significancia del $\alpha=0,015$ para tener una confianza total en la prueba del 90% ya que ésta prueba pierde confianza a medida de que aumenta la cantidad de fuentes de variación analizadas- . Se analizaron como fuente de variación: los factores principales. Los factores de bloqueo y la combinación de Entrenamiento con Visual, que a pesar de no mostrarse como significativa en el ANOVA, por ser la combinación de los factores centrales de ésta investigación fue analizada. La tabla 23 evidencia el estadístico, grados de libertad, valor p, y conclusión a la que llegó la prueba en cada una de las fuentes de variación propuestas.

Tabla 23

Pruebas Kruskal Wallis (Tiempo)

| Fuente de Variación | Tiempo | | | Conclusión |
|----------------------|-------------|--------------------|---------|---------------|
| | Estadístico | Grados de Libertad | Valor P | |
| Entrenamiento | 4.17 | 3 | 0.243 | No se rechaza |
| Tendencia | 3.83 | 2 | 0.147 | No se rechaza |
| Ruido | 2.78 | 1 | 0.095 | No se rechaza |
| Visual | 3.31 | 1 | 0.069 | No se rechaza |
| Orden | 82.61 | 2 | <0.001 | Rechaza |
| Persona | 536.53 | 82 | <0.001 | Rechaza |
| Visual*Entrenamiento | 9.03 | 7 | 0.251 | No se rechaza |

La prueba de Kruskal Wallis evidenció que los factores de bloqueo orden y persona afectaron el tiempo de procesamiento del ajuste de juicio. A pesar de que el visual del histórico de los datos tiene un valor p por encima de la significancia, no es un valor muy alto, lo cual, al igual que en el caso de precisión no permite descartarlo decisivamente como un factor que podría afectar los tiempos de ajuste. Debido a que ninguno de los factores principales ni sus interacciones fueron estadísticamente significativas se evidenció que el tiempo es una variable con una fuerte

tendencia a mantenerse inalterada. Sin embargo, los tiempos de procesamientos se vieron afectados por factores que no fueron centro de estudio de esta investigación tales como orden y persona. A continuación se evidencia el análisis de estas fuentes de variación que resultaron significativas.

2.1 Efectos del Orden en el Tiempo de Procesamiento del Ajuste

2.1.a Resultados: Orden en el Tiempo de Procesamiento del Ajuste

El orden en el cual fueron realizados los ajustes de juicio afectó el tiempo incurrido en realizar el ajuste de juicio, el tiempo se redujo a medida que los sujetos de prueba se acercaron más al final de la prueba. La tabla 24 evidencia la mediana de cada grupo en donde al pasar de los primeros 4 ajustes, a los próximos 4, el tiempo se redujo en un 33%, y al pasar de este grupo, a los 4 últimos, se tuvo una reducción del 19% del tiempo. Con lo cual se observó una reducción total de más del 45% en el tiempo empleado por ajuste desde el inicio la prueba hasta su finalización.

Tabla 24

Test de Comparaciones Múltiple de Dunn Tiempo – Fuente de Variación Orden

| | Grupo A | Grupo B | Grupo C |
|--------------------------|----------------|----------------|----------------|
| Niveles | Primeras | Medio | Últimas |
| Mediana del Grupo | 55 | 37 | 30 |
| Diferencia | 33% | 19% | 45% |

2.1.a Discusión de los Efectos del Orden en el Tiempo de Procesamiento del Ajuste

Esta reducción en el tiempo se conjetura que pudo ser resultado de un aprendizaje ocurrido durante la prueba y /o del efecto de un desgaste o cansancio.

2.2 La Singularidad del individuo y su efecto en el Tiempo de Procesamiento

Al igual que con precisión, no se ahondó en el análisis hecho al factor persona, debido a que no fue de interés para la investigación conocer que individuo realizó ajustes de juicio en menor tiempo respecto de otros. Se determinó que existen factores intrínsecos a cada persona, que afectaron el tiempo empleado en la prueba.

2.3 Recapitulación de Tiempo

Finalmente se pudo apreciar que el tiempo es una variable robusta, debido a que frente a fuentes de variación que afectaron a la precisión, varió muy poco.

3. Relación Intervariables

Se realizó una correlación de Pearson para observar si mayores tiempos de procesamiento conllevaron a mejores pronósticos o viceversa. Debido a que se conoce que la aversión por ajustar mejora la precisión, se excluyeron aquellos datos donde la diferencia entre el ajuste de juicio y el pronóstico dado por SPSS fue cero, es decir cuando no se ajustó.

3.1a Resultados sobre la Relación Intervariables

Al realizar la prueba correspondiente se encontró que el valor de la correlación tuvo un valor muy cercano a 0, lo cual sugirió fuertemente que los valores que tomó una variable fueron independientes a los valores que tomó la otra variable, aunque esta conclusión puede verse afectada por el promedio de efectos, por tal motivo se realizó la prueba separando por tipo de entrenamiento. Los resultados pueden apreciarse en la tabla 25.

Tabla 25

Correlación precisión tiempo

| Correlaciones | | |
|-------------------------------|--------------------|----------------|
| Entrenamiento | Correlación | Valor P |
| Sin dividir por Entrenamiento | 0,037 | 0,245 |
| Enfocado | -0,203 | 0,006 |
| Sin Entrenamiento | 0,29 | <0,001 |
| Tradicional | 0,0042 | 0,598 |
| Doble | -0,117 | 0,084 |

Con una confianza del 95% dos de las correlaciones por tipo de entrenamiento fueron significativas, las cuales incluyen a enfocado y a sin entrenamiento.

3.1b Discusión acerca de la Relación Intervernables

Las correlaciones donde existió un valor significativo, correspondieron a aquellos grupos donde no se tuvo acceso al entrenamiento clásico, lo cual parece aportar a la teoría expuesta previamente de que el entrenamiento clásico tuvo un efecto diferenciador entre los grupos. Así mismo se evidenció otro evento de gran interés y fue que en las dos correlaciones existentes los signos fueron diferentes, lo cual sugirió que cuando alguien con entrenamiento Enfocado se demoró más tiempo, tendió a tener mejores pronósticos, en cambio alguien que no tuvo ninguno de los entrenamientos al demorarse más, tendió a tener pronósticos más imprecisos.

El anterior análisis debe tomarse con cautela, ya que como correlación no implica causalidad, no se pudo determinar cuál fue la variable que afectó a la otra, o si hubo una tercera variable en juego.

Conclusiones

1. Los grupos que recibieron entrenamiento clásico, fueron más imprecisos respecto de aquellos que no lo recibieron. Una posible causa de ello, es que los primeros buscaron competir contra el pronóstico realizado por la máquina.
2. En el escenario en el cual no se posee información privilegiada con base en la cual el pronosticador pueda ajustar, en términos de precisión es menos perjudicial tener apetito por no ajustar u optar por impartir ajustes pequeños, que incurrir en grandes ajustes.
3. Apoyos reductores de sesgos de naturaleza inconsciente “forma visual del histórico” no generan ningún tipo de efecto en la precisión y en el tiempo de procesamiento del ajuste de juicio. Sin embargo no se descarta la hipótesis de que éste apoyo reductor de sesgos genere algún efecto, pues se cree que la diferencia visual entre las gráficas tipo Bouba y tipo Kiki pudo ser más notoria, por lo que se sugiere se hagan nuevas investigaciones en el tema.
4. Incurrir en mayores tiempos de procesamiento conlleva a pronósticos más precisos cuando los ajustes de juicio son impartidos por individuos instruidos en entrenamiento enfocado, quizá porque al demorarse más el individuo analiza mejor la gráfica y evita caer en malas prácticas al momento de ajustar; por el contrario cuando el pronosticador no ha sido instruido en ningún tipo de entrenamiento demorarse más disminuye la precisión del pronóstico.
5. El tiempo incurrido en ajustar el pronóstico disminuye a medida que transcurre la prueba, se conjetura puede ser resultado de un aprendizaje ocurrido durante la prueba y /o del efecto de un desgaste o cansancio.
6. No sólo características intrínsecas al individuo tales como el entrenamiento incidieron en la precisión del pronóstico, factores ajenos al individuo y de naturaleza única del comportamiento de los datos tales como el ruido y la tendencia generaron un efecto en la precisión del pronóstico.

7. Ajustes de juicio realizados en gráficas con características de ruidos bajos originaron pronósticos más precisos que los realizados en gráficas con características de ruidos altos, debido a que en presencia de ruidos altos los pronosticadores incurrieron en la mala práctica de modelar el ruido y no identificar adecuadamente la señal de la gráfica.
8. Ajustes de juicio impartidos en graficas con tendencias negativas son considerablemente más imprecisos que los realizados en gráficas con tendencias positivas y graficas sin tendencia. Se sugiere entrenar mejor a los pronosticadores en escenarios de tendencias negativas a fin de que sean más cuidadosos al impartir el ajuste y con ello más precisos.
9. Particularidades relacionadas a cada individuo influyen en la calidad del pronóstico y tiempo incurrido en ajustar por juicio.

Recomendaciones

Se divide éste capítulo en dos secciones, una hace referencia a las limitaciones del estudio, las cuales deben ser tenidas en cuenta al momento de analizar los resultados y evaluar las conclusiones, la otra contempla lo que a juicio de los investigadores y basados en los resultados hallados en el experimento, podrían ser futuras investigaciones.

Limitaciones del Experimento

Históricos de Datos no Reales: Los históricos de datos que se contemplaron en éste experimento no fueron reales y por el contrario fueron generados de manera controlada, por lo que las conclusiones expuestas en la investigación podrían no ser completamente aplicables al sector real.

Limitación Entrenamiento Clásico: En ésta investigación las gráficas bajo las cuales los individuos ajustaron por juicio no presentaron características de estacionalidad, lo cual pudo afectar el desempeño de las poblaciones con entrenamiento clásico, pues buena parte de su formación se enfoca en el análisis de estacionalidades. Así mismo, se cree que entrenar a las poblaciones con entrenamiento clásico en técnicas de pronósticos tales como regresiones lineales habría aumentado su desempeño, pues a causa del comportamiento del histórico de los datos presentados en éste experimento, ajustar con base en técnicas de regresión lineal simple habría originado un ajuste de buena calidad.

Futuras Investigaciones

1. **Visual del Histórico de los Datos (Bouba – Kiki):** Se recomienda realizar futuras investigaciones en las cuales la diferencia entre gráficas Bouba y gráficas Kiki sea más agresiva, y con ello determinar si el visual del histórico de los datos afecta de manera contundente la precisión del pronóstico. De igual manera, un mayor tamaño de muestra podría facilitar la detección.

2. **Recompensa por Participación en el Experimento:** A fin de garantizar que todos los participantes tengan la misma motivación por participar y tener un buen desempeño en el experimento, se recomienda dar recompensas uniformes a los participantes. Si bien en el presente experimento se intentó dar a todos los participantes la misma recompensa en términos cuantitativos (1 ó 2 décimas dependiendo del desempeño), el hecho de que las materias a las cuales iba a ser ponderada la recompensa tuvieran niveles de dificultad diferentes pudo generar discrepancia en la percepción que los individuos tuvieron de la recompensa.
3. **Tiempos de Procesamiento:** Realizar nuevos experimentos donde se mida la precisión de los pronósticos en condiciones de diferentes tiempos máximos, medir el efecto de recomendar a los pronosticadores reducir el tiempo de procesamiento del ajuste o manejar recompensas por reducir los tiempos incurridos en ajustar, podrían conllevar a conclusiones interesantes.
4. **Bitácora de la Toma de Datos:** Incluir en la toma de datos una bitácora en la cual se describa el procedimiento usado por cada individuo al realizar el pronóstico, permitiría agrupar en categorías los diferentes métodos usados y de este modo concluir con mayor profundidad.
5. **Diversidad de Gráficas:** Realizar estudios similares en los cuales las gráficas que los pronosticadores ajustan por juicio contemplen una mayor variedad de características respecto de estacionalidad, tendencia, ruido, demanda promedio, puntos atípicos entre otras, permitiría evaluar el efecto de los entrenamientos en un escenario más amplio y cercano a la realidad.

Referencias Bibliográficas

- Antony, J. (2003). *Design of experiments for engineers and scientists*. Butterworth-Heinemann.
- Ballou, R. H. (2004). *Logística: Administración de la cadena de suministro*. Pearson Educación.
- BIPM, I., IFCC, I., ISO, I., & IUPAP, O. (2008). The international vocabulary of metrology—basic and general concepts and associated terms (VIM), JCGM 200: 2008.
- Chase, R. B., Jacobs, F. R., Aquilano, N. J. (2009). *Administración de operaciones: Producción y cadena de suministros*. México, D.F. : McGraw-Hill.
- Davis, R. (1961). The fitness of names to drawings. A cross-cultural study in tanganyika. *British Journal of Psychology*, 52(3), 259-268.
- Davydenko, A., & Fildes, R. (2013). Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts. *International Journal of Forecasting*, 29(3), 510-522.
- Dean, A., & Voss, D. (1999). *Design and Analysis of Experiments*, Springer, New York
- Desanctis, G., & Jarvenpaa, S. L. (1989). Graphical presentation of accounting data for financial forecasting: An experimental investigation. *Accounting, Organizations and Society*, 14(5), 509-525.
- Eroglu, C., & Croxton, K. L. (2010). Biases in judgmental adjustments of statistical forecasts: The role of individual differences. *International Journal of Forecasting*, 26(1), 116-133.

Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, 25(1), 3-23.

Gomez, V., & Maravall, A. (2001). Automatic modeling methods for univariate series. *A course in time series analysis*, 171-201.

Goodwin, P., & Wright, G. (1993). Improving judgmental time series forecasting: A review of the guidance provided by research. *International Journal of Forecasting*, 9(2), 147-161.

Harvey, N. (1995). Why are judgements less consistent in less predictable task situations? *Organizational Behavior and Human Decision Processes*, 63, 247– 263.

Harvey, N., & Bolger, F. (1996). Graphs versus tables: Effects of data presentation format on judgemental forecasting. *International Journal of Forecasting*, 12(1), 119-137.

Hyndman, R. (2014, February). Automatic time series forecasting. New Trends on Intelligent Systems and Soft Computing 2014. *Lecture Conducted from University of Granada, Spain*

Hyndman, R. J., & Khandakar, Y. (2007). *Automatic time series for forecasting: the forecast package for R* (No. 6/07). Monash University, Department of Econometrics and Business Statistics.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4), 679-688

Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3), 439-454.

Köhler, W. (1929). Gestalt psychology. *New York: Liveright Publishing Corporation*.

Köhler, W. (1947). Gestalt psychology (2nd edn.). *New York: Liveright Publishing Corporation*

Larson, C. L., Aronoff, J., & Steuer, E. L. (2012). Simple geometric shapes are implicitly associated with affective value. *Motivation and Emotion*, 36(3), 404-413.

Larson, C. L., Aronoff, J., Sarinopoulos, I. C., & Zhu, D. C. (2009). Recognizing threat: A simple geometric shape activates neural circuitry for threat detection. *Journal of cognitive neuroscience*, 21(8), 1523-1535.

Lawrence, M., & Makridakis, S. (1989). Factors affecting judgmental forecasts and confidence intervals. *Organizational Behavior and Human Decision Processes*, 43, 172– 187.

Lawrence, M., Goodwin, P., O'Connor, M., & Önköl, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting*, 22(3), 493-518.

Legerstee, R., & Franses, P. H. (2014). Do experts' SKU forecasts improve after feedback?. *Journal of Forecasting*, 33(1), 69-79.

Mason, R. L., Gunst, R. F., & Hess, J. L. (2003). *Statistical design and analysis of experiments: with applications to engineering and science* (Vol. 474). John Wiley & Sons.

Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of Boubas: Sound–shape correspondences in toddlers and adults. *Developmental science*, 9(3), 316-322

Mosteller, F., Siegel, A., Trapido, E., & Youtz, C. (1981). Eye fitting straight lines. *American Statistician*, 35(3), 150.

Nielsen, A. K. S. (2011). *Sound symbolism and the Bouba-Kiki effect: uniting function and mechanism in the search for language universals* (Doctoral dissertation, Lethbridge, Alta.: University of Lethbridge, Dept. of Psychology, c2011)

O'nnkal, D., & Muradoglu, G. (1994). Evaluating probabilistic forecasts of stock prices in a developing stock market. *European Journal of Operational Research*, 74, 350–358.

Oxford Dictionary (2009), "Oxford Dictionary," *Oxford University Press: UK*.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia--a window into perception, thought and language. *Journal of consciousness studies*, 8(12), 3-34.

Sanders, N. R., & Ritzman, L. P. (1992). The need for contextual and technical knowledge in judgmental forecasting. *Journal of Behavioral Decision Making*, 5(1), 39-52.

Sanders, N. R., & Ritzman, L. P. (2001). Judgmental adjustment of statistical forecasts. In J. S. Armstrong (Ed.), *Principles of forecasting* (pp. 405–416). Boston: Kluwer Academic Publishers [The process of using a forecasting support system]

Syntetos, A. A., Boylan, J. E., & Disney, S. M. (2009). Forecasting for inventory planning: a 50-year review. *Journal of the Operational Research Society*, S149-S160

Thomson, M. E., Pollock, A. C., Henriksen, K. B., & Macaulay, A. (2004). The influence of the forecast horizon on judgmental probability forecasts of exchange rate movements. *The European Journal of Finance*, 10(4), 290– 307.

Usuga Espinal, S. T., & Alvarado, J. A. (2014). Metodología para la elaboración de pronósticos en la unidad de negocio de agroquímicos de una multinacional química integrando juicio humano con computador

Worthen, D. (2003). Future results not guaranteed: contrary to what vendors tell you, computer systems alone are incapable of producing accurate forecasts. *CIO*, *16*(19), 1.

Yates, J. F., McDaniel, L. S., & Brown, E. S. (1991). Probabilistic forecasts of stock prices and earnings: The hazards of nascent expertise. *Organizational Behavior and Human Decision Processes*, *49*, 60–79.

Apéndices

Apéndice A

Medidas del Error.

| MEDIDA | | CARACTERÍSTICAS |
|--|---|--|
| | Dónde, | $e_t = Y_t - F_t$ |
| RMSE (Root Mean Square error) | $\sqrt{\text{mean}((e_t)^2)}$ | Se encuentra en la misma escala de los datos. Importancia teórica relevante en modelos estadísticos. Más sensible ante valores atípicos que otras medidas como |
| MSE (Mean Square Error) | $\text{mean}((e_t)^2)$ | No se encuentra en la misma escala de los datos. Importancia teórica relevante en modelos estadísticos. Más sensible ante valores atípicos que otras medidas como Se encuentra en la misma escala de los datos. |
| MAE (Mean Absolute Error) | $\text{mean}(e_t)$ | Se encuentra en la misma escala de los datos. |
| MdAE (Median Absolute Error) | $\text{median}(e_t)$ | Se encuentra en la misma escala de los datos. |
| Medidas Basadas en Errores Porcentuales | | |
| MAPE (Mean Absolute Percentage Error) | $\text{mean}\left(\left \frac{100 e_t}{Y_t}\right \right)$ | Es infinita o indefinida cuando el dato real asume el valor de Muy usada en el sector real. Es independiente de la escala. |
| MdAPE (Median Absolute Percentage Error) | $\text{median}\left(\left \frac{100 e_t}{Y_t}\right \right)$ | Es independiente de la escala. |
| RMSPE (Root Mean Square Percentage Error) | $\sqrt{\text{mean}\left(\left(\frac{100 e_t}{Y_t}\right)^2\right)}$ | Es independiente de la escala. |
| RMdSPE(Root Median Square Percentage Error) | $\sqrt{\text{median}\left(\left(\frac{100 e_t}{Y_t}\right)^2\right)}$ | Es independiente de la escala. |

Nota. Adaptado de Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4), 679-688.

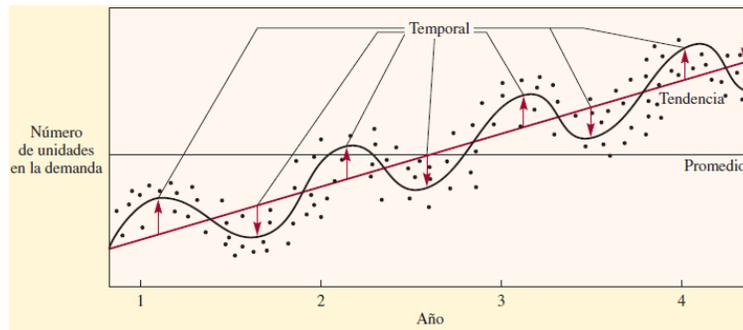
Apéndice B:

SÍNTESIS ENTRENAMIENTO TRADICIONAL

¿Qué es un Pronóstico?

- ✓ Estimar o conjeturar un evento futuro. (1)
- ✓ Es un dato de entrada que permite planear el futuro. (2)

Componentes de la Demanda (3)



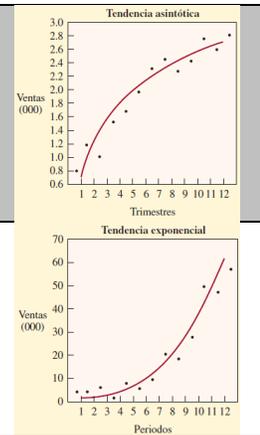
- ✓ Demanda Promedio: Valor que representa de la mejor manera a todos los datos del conjunto de periodos analizados.
- ✓ Tendencia: Indica un patrón de comportamiento de los datos que permanece en una ventana de tiempo, usualmente es el punto de inicio para pronosticar, e intervienen en ella elementos estacionales, y eventos esperados que puedan influir en el pronóstico final.

Tipos de Tendencias (4)

| | | |
|-------------------|---|---|
| Lineal | Es una relación continua y directa. | <p>Tendencia lineal</p> <p>Ventas (000)</p> <p>Trimestres</p> |
| Curva en S | Tendencia que presenta puntos en los cuales la velocidad de crecimiento cambia de rápido a lento y viceversa. | <p>Tendencia de curva S</p> <p>Ventas (000)</p> <p>Periodos</p> |

Asintótica Se tiene un crecimiento alto al inicio y posteriormente se da una reducción en la tasa de crecimiento.

Exponencial Tendencia que sugiere un crecimiento infinito.



✓ Estacionalidad: Representa las variaciones de ascenso o descenso por encima de la media que se repiten a lo largo del tiempo.

✓ Variación Aleatoria: Las variaciones aleatorias son causadas por eventos fortuitos, los cuales resultan luego de restar todas las causas conocidas de la demanda (promedio, tendencias, estacionalidad y auto correlación), esa fracción que queda sin explicar se supone es aleatoria.

✓ Auto Correlación: Hace referencia a la persistencia de la ocurrencia, un ejemplo puede ser, si una línea es larga en un momento dado, tiempo después es de esperar que la línea siga siendo larga.

Método de Análisis de Series de Tiempo (5)

Los análisis de series de tiempo se basan en la premisa de emplear datos históricos para pronosticar. Los métodos más usados son:

✓ Promedio Móvil Simple: Se calcula sumando puntos de datos, y dividiendo entre el número total de datos utilizados; conllevando a que cada punto tenga el mismo peso porcentual. La base del promedio móvil influye en las características del promedio, ya que cuanto más larga sea, más se uniformaran los elementos aleatorios.

$$F_{t+1} = \frac{D_t + D_{t-1} + \dots + D_{t-N+1}}{N}$$

✓ Promedio Móvil Ponderado: Es en esencia similar al promedio móvil simple, pero la diferencia radica en que el ponderado permite asignar ponderaciones diferentes a cada punto de datos. Una desventaja de usar promedios ya sea simple o móvil es que si existen tendencias, éstos las retrasan.

$$F_{t+1} = W_1 D_t + W_2 D_{t-1} + \dots + W_N D_{t-N+1}$$

$$\sum_{i=1}^N W_i = 1$$

✓ Suavización Exponencial: Se basa en ponderar a los puntos más recientes con un peso mayor e ir reduciendo su peso exponencialmente en un factor de $(1-\alpha)$ a medida que se hacen más antiguos los datos, dónde α es una constante de suavización la cual establece la velocidad de reacción a las diferencias entre los datos pronosticados y los datos reales. A fin de tener en cuenta la tendencia de la serie en el modelo de pronósticos, se puede hacer una corrección al método de suavización exponencial agregando un ajuste de tendencias, para lo cual es necesario adicionar una constante de suavización delta δ , la cual tiene por objeto reducir el impacto del error asociado a la tendencia.

SUAVIZACIÓN EXPONENCIAL

$$F_{t+1} = F_t + \alpha (D_t - F_t)$$

F_{t+1} = Pronóstico suavizado exponencialmente para el periodo t + 1

Simple

F_t = Pronóstico suavizado exponencialmente para el periodo t

D_t = Demanda real para el periodo t

α = Constante de suavización

$$FIT_t = F_t + T_t$$

$$F_t = FIT_{t-1} + \alpha (D_{t-1} - FIT_{t-1})$$

$$T_t = T_{t-1} + \delta (F_t - FIT_{t-1})$$

F_t = Pronóstico suavizado exponencialmente para el periodo t

T_t = Tendencia suavizada exponencialmente para el periodo t

Doble

FIT_t = Pronóstico incluida la tendencia para el periodo t

FIT_{t-1}

= Pronóstico incluida la tendencia hecha para el periodo anterior

D_{t-1} = Demanda real para el periodo

α = Constante de suavización

δ = Constante de suavización

BIBLIOGRAFÍA

- (1)Oxford Dictionary (2009), “Oxford Dictionary,” *Oxford University Press: UK*.
- (2)Sanders, N. R., & Ritzman, L. P. (2001). Judgmental adjustment of statistical forecasts. In J. S. Armstrong (Ed.), *Principles of forecasting* (pp. 405–416). Boston: Kluwer Academic Publishers [The process of using a forecasting support system]
- (3)Chase, R. B., Jacobs, F. R., Aquilano, N. J. (2009). *Administración de operaciones: Producción y cadena de suministros*. México, D.F. : McGraw-Hill. 469.
- (4)Chase, R. B., Jacobs, F. R., Aquilano, N. J. (2009). *Administración de operaciones: Producción y cadena de suministros*. México, D.F. : McGraw-Hill. 472.
- (5)Chase, R. B., Jacobs, F. R., Aquilano, N. J. (2009). *Administración de operaciones: Producción y cadena de suministros*. México, D.F. : McGraw-Hill. 472.

Apéndice C:

SÍNTESIS ENTRENAMIENTO ENFOCADO

¿Qué es un Pronóstico?

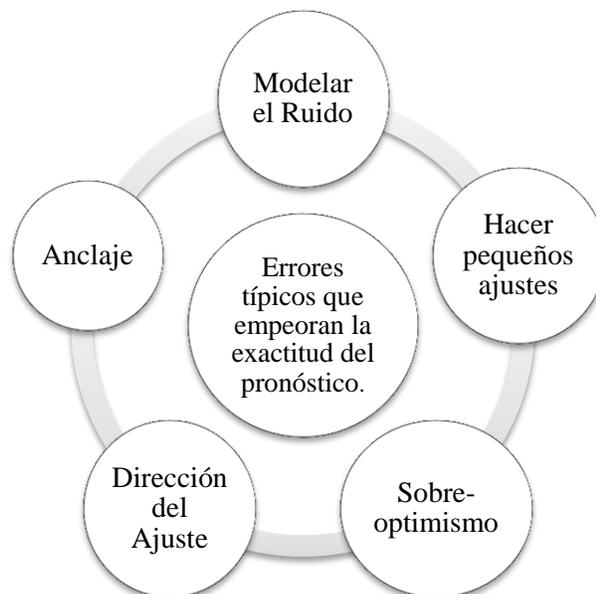
- ✓ Estimar o conjeturar un evento futuro. (1)
- ✓ Es un dato de entrada que permite planear el futuro. (2)

¿Qué es un ajuste de pronóstico?

- ✓ Los ajustes de juicio tienen como objetivo combinar procedimientos estadísticos con información no contenida en los históricos, información de la cual tiene conocimiento la persona encargada de ajustar el pronóstico.
- ✓ Los pronósticos con ajustes de juicio tienen algunos inconvenientes que disminuyen la efectividad de este método. Uno de estos se denomina “sesgo”.

¿Qué es un Sesgo?

- ✓ Los sesgos son aquellos errores del proceso mental en los que incurre el pronosticador. Un sesgo se puede ver como una desviación sistemática de un pensamiento racional. A continuación se presentan algunos sesgos que impactan en la precisión de un pronóstico (3)(4)



- ✓ **Hacer pequeños ajustes:** Los ajustes pequeños por lo general no son muy efectivos, dado que cuando una persona los realiza, los realiza no porque tenga seguridad de que un cambio se deba hacer, sino para cumplir su necesidad de modificar el pronóstico. (4)

- ✓ **Modelación de Ruido:** El ruido, o error experimental, está compuesto por el error aleatorio, el cual recoge las diferencias existentes entre diferentes observaciones que no se pueden explicar, ya sea por su naturaleza o por la alta complejidad asociada a evaluar la variable (5). Estas variaciones no controlables pueden conllevar a modelaciones inadecuadas de la serie de tiempo, ya que los pronosticadores pueden confundir el ruido con la señal (6) y tal como dicen Goodwin y Wright (7), este es uno de los factores que pueden conllevar a una mayor complejidad de la serie.

- ✓ **Sesgo de dirección del ajuste y sobre optimismo:** su impacto en la precisión del pronóstico varía dependiendo de la dirección del ajuste. Por lo general se presenta más inexactitud cuando los ajustes son positivos, tienden a sobreestimar el pronóstico final ya que se hacen bajo un sesgo de naturaleza optimista en el cual los pronosticadores usan indebidamente un optimismo injustificado. Por el contrario, impartir ajustes negativos generalmente es más efectivo ya que estos tienden a reducir el sesgo medio del pronóstico, al parecer porque una vez el pronosticador ha decidido hacer un ajuste negativo, es realista acerca de los niveles probables de la demanda. (4)

- ✓ **Anclaje:** el cual nos dice que las personas son propensas a darle una ponderación adicional a un dato particular, lo cual puede llevar a que si este es un valor atípico o no representativo, se perjudique la precisión del pronóstico. Por lo cual se debe evitar caer en esto.

BIBLIOGRAFÍA:

- (1) Oxford Dictionary (2009), “Oxford Dictionary,” Oxford University Press: UK.
- (2) Sanders, N. R., & Ritzman, L. P. (2001). Judgmental adjustment of statistical forecasts. In J. S. Armstrong (Ed.), *Principles of forecasting* (pp. 405–416). Boston: Kluwer Academic Publishers [The process of using a forecasting support system]
- (3) Eroglu, C., & Croxton, K. L. (2010). Biases in judgmental adjustments of statistical forecasts: The role of individual differences. *International Journal of Forecasting*, 26(1), 116-133.
- (4) Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, 25(1), 3-23.
- (5) Antony, Jiju. (2003). *Design of experiments for engineers and scientists*. Butterworth-Heinemann.
- (6) Harvey, N. (1995). Why are judgements less consistent in less predictable task situations? *Organizational Behavior and Human Decision Processes*, 63, 247– 263.
- (7) Goodwin, P., & Wright, G. (1993). Improving judgmental time series forecasting: A review of the guidance provided by research. *International Journal of Forecasting*, 9(2), 147-161.

Apéndice D:**ANOVA variable Precisión**

| Fuente de Variación | Grados de Libertad | Suma de Cuadrados | Cuadrado Medio | F | P |
|----------------------------|---------------------------|--------------------------|-----------------------|----------|----------|
| Visual | 1 | 0,039 | 0,039 | 0,31 | 0,578 |
| Tendencia | 2 | 34,8403 | 17,4201 | 138,6 | 0 |
| Visual*Tendencia | 2 | 0,4321 | 0,216 | 1,72 | 0,18 |
| Entrenamiento | 3 | 2,0812 | 0,6937 | 5,52 | 0% |
| Visual*Entrenamiento | 3 | 0,1033 | 0,0344 | 0,27 | 84% |
| Ruido | 1 | 16,3014 | 16,3014 | 129,7 | 0% |
| Visual*Ruido | 1 | 0,0279 | 0,0279 | 0,22 | 64% |
| Tendencia*Entrenamiento | 6 | 1,1854 | 0,1976 | 1,57 | 15% |
| Entrenamiento*Ruido | 3 | 0,1953 | 0,0651 | 0,52 | 67% |
| Tendencia*Ruido | 2 | 14,3364 | 7,1682 | 57,03 | 0% |
| Orden | 2 | 0,0324 | 0,0162 | 0,13 | 0,879 |
| Persona(Entrenamiento) | 79 | 44,3982 | 0,562 | 4,47 | 0 |
| Error | 890 | 111,8588 | 0,1257 | | |
| Total | 995 | | | | |

Apéndice E:**Residuales estandarizados: Entrenamiento vs Calidad de Ajuste**

| | NO | SI | Todos |
|--------------------------|-----------|-----------|--------------|
| Doble | 78 | 31 | 109 |
| | 74,5 | 34,5 | 109 |
| | 0,4055 | -0,59588 | * |
| Enfocado | 61 | 28 | 89 |
| | 60,83 | 28,17 | 89 |
| | 0,02176 | -0,03198 | * |
| Sin Entrenamiento | 100 | 52 | 152 |
| | 103,89 | 48,11 | 152 |
| | 0,38164 | 0,56082 | * |
| Tradicional | 59 | 27 | 86 |
| | 58,78 | 27,22 | 86 |
| | 0,02872 | -0,0422 | * |
| Todos | 298 | 138 | 436 |
| | 298 | 138 | 436 |
| | * | * | * |

Apéndice F:**ANOVA Tiempo**

| Fuente de Variación | Grados de Libertad | Suma de Cuadrados | Cuadrado Medio | F | P |
|----------------------------|---------------------------|--------------------------|-----------------------|----------|----------|
| Visual | 1 | 419 | 419 | 0,08 | 0,778 |
| Tendencia | 2 | 51959 | 25980 | 4,95 | 0,007 |
| Visual*Tendencia | 2 | 5113 | 2557 | 0,49 | 0,615 |
| Entrenamiento | 3 | 61528 | 20509 | 3,9 | 0,009 |
| Visual*Entrenamiento | 3 | 25400 | 8467 | 1,61 | 0,185 |
| Ruido | 1 | 1320 | 1320 | 0,25 | 0,616 |
| Visual*Ruido | 1 | 5042 | 5042 | 0,96 | 0,327 |
| Tendencia*Entrenamiento | 6 | 46172 | 7695 | 1,47 | 0,187 |
| Entrenamiento*Ruido | 3 | 338 | 113 | 0,02 | 0,996 |
| Tendencia*Ruido | 2 | 5050 | 2525 | 0,48 | 0,618 |
| Orden | 2 | 398574 | 199287 | 37,94 | 0 |
| Persona(Entrenamiento) | 79 | 2059025 | 26064 | 4,96 | 0 |
| Error | 890 | 4674917 | 5253 | | |
| Total | 995 | | | | |

Apéndice G:**Generador de Gráficas**

| | | Positiva | Negativa | Sin Tendencia |
|-----------------|--|---|--|---|
| Ecuación | Ecuación de una recta con un término adicional. | $Y=a*X +b+r$ | | |
| a | Este término de la ecuación representa la pendiente de la gráfica. | Variable aleatoria uniforme que tomó valores entre 1 y 10 | Variable aleatoria uniforme que tomó valores entre 1 y 10 | Constante y asumió el valor 0. |
| b | Este término representa la parte estable de la gráfica. | Multiplicación de “a” por una variable aleatoria uniforme que tomó valores entre 1 y 10 | | Variable aleatoria uniforme que tomó valores entre 1 y 10 |
| X | Este representa el periodo actual | Tomó valores iguales al periodo que se generaba | El primer periodo tomó como valor 15, el siguiente 14, hasta que el último tomó 1. | Tomó valores iguales al periodo que se generaba |
| r | | Representa el ruido blanco de la demanda. Variable aleatoria generada mediante una distribución normal de media cero y desviación estándar igual al promedio de los periodos antes de agregar el ruido, multiplicado por un porcentaje, del 10% para el caso de ruido bajo y del 50% para ruido alto. | | |

Apéndice H:

Instrumento en Digital

Se adjunta en digital el instrumento a través del cual fueron tomados los datos.