



COLEGIO DE GRADUADOS EN CIENCIAS ECONOMICAS

128 años al servicio de la profesión (1891-2019)

XXXI JORNADAS DE ACTUACIÓN JUDICIAL



"SEMINARIO DE INTEGRACIÓN Y APLICACIÓN
CONTADOR PÚBLICO NACIONAL
MINERÍA DE DATOS Y CONTABILIDAD"

AUTOR: PATRICIO AGUSTÍN IGLESIAS
E-mail: patricio.iglesias@yahoo.com.ar

21 y 22 de agosto de 2019
Ciudad Autónoma de Buenos Aires

Índice

Introducción	3
Capítulo I: El rol del Contador Público y las nuevas tecnologías de la información y las comunicaciones.....	4
Capítulo II: Estadística, informática y minería de datos. Aplicaciones.....	7
Capítulo III: Auditoría, estadística e informática. La ley de Benford.....	14
Conclusiones	19
Bibliografía	22
Anexos	24
Materias Relacionadas: Álgebra, Análisis Matemático, Historia Económica y Social General, Historia Económica y Social Argentina, Estadística, Teoría Contable, Sistemas Contables, Contabilidad Patrimonial, Cálculo Financiero, Auditoría.	

Introducción

Hoy la minería de datos está en boca de todos, si bien este campo comenzó a tomar forma en la década de 1980. Ya no es más un tema que atrae a científicos o ingenieros. Periodistas, empresarios, funcionarios públicos, sindicalistas, jueces y cuantas personas uno pueda imaginar están interesadas en las aplicaciones que puede tener en sus respectivos ámbitos. Aprendizaje automático, ciencia de datos, inteligencia artificial o datos masivos (*big data*) son términos de uso cada día más popular, aunque no por ello menos vagos.

Esta nueva disciplina, que conjuga conocimientos de estadística y computación con los propios del campo de aplicación, no deja de resultar de interés para los profesionales de las ciencias económicas. No por nada, una de sus primeras aplicaciones estuvo relacionada con la detección de fraudes, y se usa extensivamente tanto en campañas publicitarias como en el armado de indicadores de incobrabilidad.

Este trabajo se propondrá, en un primer capítulo, analizar el nuevo rol del Contador Público en medio de la transformación tecnológica de las últimas décadas, realizando un repaso de los lineamientos a nivel nacional, regional e internacional. El segundo capítulo tendrá como objetivo definir la minería de datos no sólo por intensidad, sino también por extensión, estudiando las subdisciplinas y problemáticas que suelen asociarse a la misma, dando un breve marco histórico que incluya una referencia a raíces más profundas vinculadas a la estadística (en particular inferencial) y la computación. El tercero tendrá como propósito estudiar la denominada ley de Newcomb-Benford, cuyas perspectivas han dado un salto cualitativo con la incorporación del nuevo instrumental informático, realizando una pequeña aplicación con datos de *warrants* de soja y arroz de Argentina. Por último se brindará un espacio de conclusiones.

Capítulo I: El rol del Contador Público y las nuevas tecnologías de la información y las comunicaciones

Generalidades

Las nuevas Tecnologías de la Información y las Comunicaciones (TIC) han representado un reto para los profesionales de las ciencias económicas, los cuales han sido planteados en el ámbito universitario, legal, gremial e internacional. Esto es consecuencia lógica de que, más allá de las enormes disidencias que caracterizan a la contabilidad, existe un amplio consenso sobre la idea de que la información es el núcleo de la labor profesional, por lo que todos los cambios en los sistemas de información implican una necesidad por parte del Contador Público de actualizarse a fin de cumplir de forma eficaz y eficiente su rol dentro de las organizaciones.

Esta ola de transformaciones ya fue percibida en una publicación de la *American Accounting Association* de 1966, en la cual se advirtió que la tecnología del computador y el diseño de sistemas, junto con las técnicas de medición y la teoría de la información influirían sobre la contabilidad en un futuro. Por su parte, en Argentina, hace ya más de medio siglo, la profesora emérita de la FCE-UBA Luisa Fronti de García planteó en su tesis doctoral de 1968 la importancia de la incorporación de contenidos relacionados con las tecnologías de la información por parte del Contador para llevar a cabo sus tareas [Escobar, 2017].

Lineamientos internacionales y locales

Esta tendencia transformadora de la profesión contable no le fue ajena a la Federación Internacional de Contadores, también conocida por sus siglas en inglés como IFAC (International Federation of Accountants), entidad que nuclea a órganos representativos de los contadores de todo el mundo y de la cual la FACPCE (Federación Argentina de Consejos Profesionales de Ciencias Económicas) es miembro. Así, esta organización internacional, en el año 2008, a través de las Normas Internacionales de Formación o IES (International Education Standards) 1-8 emitidas por el Consejo de Normas Internacionales de Formación en Contaduría, declaró, específicamente en la IES 2 (Contenido de los programas profesionales de

formación en contaduría) que la formación en contaduría puede separarse en tres componentes, que son:

- a) contaduría, finanzas y conocimientos relacionados;
- b) conocimiento organizacional y de negocios; y
- c) conocimiento de tecnología de la información y competencias.

Respecto de este tercer aspecto, la IES 2 indica que deben incluirse los siguientes temas:

- (a) conocimiento general de la tecnología de la información;
- (b) conocimiento del control de la tecnología de la información;
- (c) competencias del control de la tecnología de la información;
- (d) competencias del usuario de la tecnología de la información; y
- (e) una o una (sic) mezcla de las competencias correspondientes a las funciones gerenciales, de evaluación y de diseño de los sistemas de información.

Esta misma IES también realiza una declaración sumamente importante: “El contador profesional no sólo utiliza los sistemas de información y sus habilidades en los controles de dicha tecnología, sino también desempeña un papel importante como parte de un equipo en la evaluación, el diseño y la gestión de tales sistemas”. También resulta de interés mencionar el rol axial que la IES 7 (Desarrollo Profesional Continuo: Un Programa de Aprendizaje Permanente y Desarrollo Continuo de la Competencia Profesional) brinda a la capacitación continua.

Como podrá notarse, estos lineamientos están lejos de colocar a las tecnologías de la información como herramientas marginales en la formación de los contadores, sino que las consideran uno de los tres ejes centrales de la currícula, dándoles un rol no sólo como usuarios, sino también protagónico en el diseño de los sistemas de información. Asimismo, resulta evidente de estas prescripciones que el contador debe actuar en forma interdisciplinaria en pos del desarrollo de las metas antedichas.

En línea con las IES mencionadas, el documento base para la acreditación de la carrera de Contador Público de 2013 del Consejo de Decanos de Facultades de Ciencias Económicas de Universidades Nacionales (CODECE) ha resaltado el rol de la información en esta carrera, indicando de forma muy clara en su primer capítulo que “El Contador Público debe tener presente que el objeto de su profesión es la “información” en todas sus formas” y marcando como la primera actividad reservada al título de CP “Diseñar, implementar y dirigir sistemas de registración e información, en todos los segmentos de la contabilidad, que permitan la toma de decisiones para el logro de los objetivos de la organización”. Por su parte, enumera en los contenidos curriculares mínimos 6 áreas temáticas, una de las cuales es “Administración y Tecnología de la Información”.

Existe asimismo la ISACA (siglas originadas en *Information Systems Audit and Control Association*), organización que funciona como bisagra entre la auditoría y las tecnologías de la información. Fundada en 1969 como la EDP¹ Auditors Association [Yang et Guan, 2004], otorga certificaciones (como la COBIT, *Control Objectives for Information and Related Technologies*) en materia de riesgos de tecnologías de información.

Como podrá observarse, tanto los lineamientos internacionales como los sugeridos a nivel local son contundentes a la hora de colocar los conocimientos sobre las tecnologías de la información como fundamentales en la formación de los contadores. Siguiendo estas recomendaciones, Diego Sebastián Escobar, en su tesis de maestría, propuso la ampliación de los contenidos curriculares vigentes en la Facultad de Ciencias Económicas de la UBA en materia de tecnologías de la información en general y en seguridad informática en particular [Escobar, op.cit.]. Resulta, entonces, esencial que los contadores públicos puedan capacitarse en forma permanente en cuestiones relacionadas con las nuevas tecnologías de la información, y la minería de datos no es la excepción.

1 EDP es la sigla de Electronic Data Processing (procesamiento electrónico de datos), en boga unas décadas atrás y hoy en desuso.

Capítulo II: Estadística, informática y minería de datos. Aplicaciones.

Antecedentes

Alrededor de la década de 1980, con el desarrollo de grandes volúmenes de datos, comenzó a tomar forma la minería de datos (*data mining*), llamada también ciencia de datos (*data science*) o descubrimiento del conocimiento a través de los datos (*knowledge discovery in databases*). Una definición muy popular fue formulada por Drew Conway² en 2013, quien considera que la ciencia de datos es una mezcla entre la estadística/matemática, la informática/programación y conocimientos del campo en el que se estén aplicando estas técnicas (*substantive expertise*). Veremos, antes que nada, de manera breve los orígenes de la estadística y la informática.

Estadística

La estadística es una ciencia consolidada hacia el siglo XIX, en base a los aportes de distintos campos de estudio. Fausto Toranzos, un destacado erudito argentino, distinguió cuatro líneas diferentes en su consolidación, que serán vistas sucintamente.

Una importante corriente que ayudó a consolidar la estadística fue el desarrollo de la técnica censal, con antecedentes en el Antiguo Egipto, el pueblo hebreo, Grecia y Roma y que gana un fuerte impulso académico en Alemania hacia el siglo XVIII. Cabe mencionar el rol de uno de los profesores de la universidad de Goetingen, Gottfried Achenwall (1719-1772), miembro de la escuela conocida como “Estadística Universitaria”, de fuerte cariz descriptivo. Achenwall habría acuñado la palabra “estadística”, definiéndola como “Ciencia de las cosas que pertenecen al Estado, llamando Estado a todo lo que es una sociedad civil y al país en que ella habita, con todo lo que se encuentra de activo y de efectivo; la Estadística se ocupa de los fenómenos que pueden favorecer o defender la prosperidad del Estado”, agregando asimismo que “la política enseña cómo deben ser los Estados, la Estadística explica cómo son realmente” [Toranzos, 1997]. Aquí se encuentran las raíces de la llamada

² <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>

estadística descriptiva que, como su nombre lo indica, intenta describir o recontar datos.

Una segunda línea de gran gravitación fue la iniciada por John Graunt (1620-1674) y continuada por autores como William Petty (1627-1678) y Johann Peter Süssmilch (1707-1767). Trabajando con material proporcionado por los registros parroquiales dieron un gran impulso a la demografía, descubriendo leyes y realizando incluso estimaciones del tamaño de las poblaciones.

Por otro lado cabe resaltar el papel del análisis del azar, llamado luego probabilidad, que puede remontarse al siglo XVI, con los aportes de autores italianos como Niccolò Fontana, más conocido como Tartaglia y Girolamo Cardano, y que se consolida hacia el siglo XVII con estudiosos como Fermat y Pascal. Estos matemáticos analizaban exhaustivamente los juegos de dados, aunque también es dable mencionar que algunos autores, como Alberto Landro (2010) enfatizan en las tempranas aplicaciones que poseían para la resolución de problemas económico-actuariales; siendo, según esta visión, que los estudiosos asimilaban los juegos de dados a los contratos aleatorios, en particular a los contratos de seguros sobre embarcaciones, muy desarrollados en el Mar Mediterráneo en aquella época. Otro autor destacado en esta evolución es nada menos que Luca Pacioli, quien en su *Summa de Arithmetica, geometria, proportioni e proportionalità* de 1494 no sólo mencionó el método de los comerciantes venecianos para registrar sus operaciones (la llamada Partida doble, conocida también como Principio de Pacioli en su honor), sentando así las bases de la contabilidad tal como la conocemos hoy en día, sino que también incursionó en la teoría del azar [Sriraman et Lee, 2014].

Posteriormente, con los aportes de Gauss, comenzó a desarrollarse la teoría de los errores o del ajuste, y ya hacia fines del siglo XIX y del siglo XX, con autores de la talla de Pearson, se consolidó la llamada estadística inferencial, que a diferencia de la estadística descriptiva intenta no ya analizar datos poblacionales, sino que en base a información incompleta se propone extraer conclusiones sobre datos desconocidos. Esta rama otorga, sin dudas, uno de los sustentos conceptuales de lo que hoy se conoce como minería de datos.

Argentina ha producido una gran cantidad de especialistas en estadística. Cabe mencionar el rol de Fausto Toranzos o de José González Galé; quienes fueron

asimismo docentes de esta Casa de Altos Estudios. Asimismo vale la pena mencionar que existen universidades públicas donde puede cursarse la carrera de grado de estadística, entre ellas la Universidad Nacional de Rosario (UNR) y la Universidad Nacional de Tres de Febrero (UNTREF), y diversos postgrados como el de la UNTREF y la UBA en su Facultad de Ciencias Exactas y Naturales.

El desarrollo de la informática

Ya hacia el siglo XIX existían los llamados computadores, que eran personas organizadas de forma de realizar cálculos sobre importantes volúmenes de datos. Uno de los mayores equipos de computadores del siglo XVIII fue producto de una reforma tributaria. Al barón Gaspard de Prony, junto a grandes matemáticos como Legendre o Cournot, en tiempos de la Revolución Francesa se le encargó una reforma de la imposición a la propiedad inmueble, que requería la elaboración de tablas logarítmicas para rearmar el catastro.

Así surgió el proyecto de las *tables du cadastre*, cuya organización estaba inspirada en las ideas sobre la separación del trabajo de Adam Smith, precursor del pensamiento de los clásicos de la administración como Taylor o Fayol. En la cúpula se encontraban matemáticos que diseñaban los algoritmos, había un mando medio que distribuía las tareas y en la base operativa realizaban los cálculos personas que no eran expertas en matemática, en algunos casos peluqueros que habían quedado desempleados ante los cambios de las modas tras el fin del *Ancien Régime* [Campbell-Kelly, 2018]. El programa terminó interrumpido por los avatares políticos, pero inspiró a Charles Babbage en su infructuosa idea de crear computadoras mecánicas.

Otro esquema similar de procesamiento de la información fue implementado por la Bankers' Clearing House de Londres, donde se procesaban los cheques y se calculaban los saldos resultantes de los bancos. También existían en este tiempo las computadoras analógicas u *orreries*³, que mediante engranajes mecánicos permitían predecir la posición de los astros, que fueron de gran aplicación en navegación.

3 Son llamados así en honor a Charles Boyle, cuarto conde de Orrery, mecenas de la construcción de este tipo de aparatos a principios del siglo XVIII. Lord Kelvin inventó en 1876 uno que permitía predecir las mareas, de gran utilidad para evitar que los barcos encallaran en los puertos.

No puede dejar de mencionarse el armado de tablas para el uso de diferentes disciplinas (astronómicas, financieras, de seguros de vida, trigonométricas), entre las que desempeñaron un papel destacado durante varios siglos las ya mencionadas tablas logarítmicas, que a través de las propiedades de la exponenciación permitían convertir en sumas y restas las multiplicaciones y divisiones y en productos y cocientes las potencias y raíces y, por lo tanto, acelerar en gran medida los cálculos. También fue célebre el Nautical Almanac, que gracias al aporte del erario público británico producía (e incluso hoy produce) tablas con información sobre las mareas para evitar naufragios y otros accidentes marítimos. Como podrá notarse, la recaudación impositiva, el sistema bancario, el mercado asegurador y el comercio internacional, todos campos de profunda injerencia del Contador Público y los profesionales y estudiantes de ciencias económicas en general, fueron motores indispensables a la hora de impulsar la informática en sus comienzos.

El siglo XIX vio la aparición de calculadoras, orientadas a la creciente demanda producto de la sofisticación de las oficinas (bancos, aseguradoras, ferrocarriles, entes públicos, etc.). Hacia la década de 1930 y durante la Segunda Guerra Mundial se multiplicaron las computadoras humanas (muchas de ellas mujeres) y las incipientes computadoras digitales, realizadas por gobiernos y empresas como AT&T, RCA, Remington Tand, NCR e IBM, no sólo para tablas sino también para cálculos balísticos y tareas de inteligencia como ataques criptoanalíticos. La más conocida fue la IBM Automatic Sequence Controlled Calculator, construida por IBM para la Universidad de Harvard entre 1937 y 1943.

Estas computadoras digitales suplantaban a la hora de realizar los cálculos a las personas, quienes en un principio fueron administradoras de estas máquinas. Fue hacia el año 1946 que nació la que es considerada la primera computadora. Valga la pena mencionar que este desarrollo en los dispositivos físicos fue acompañado por un importante aporte de teóricos como Turing o Von Neumann [Copeland et Proudfoot, 1996].

Fue hacia las décadas de los años 1950, 1960 y 1970 que la informática comenzó a vincularse cada vez más con el mundo de los negocios. Es justamente en la década de 1950 que surgió el lenguaje COBOL, *Common Business Oriented Language* o

“Lenguaje común orientado a negocios” [Arranga, 1996], que como su nombre lo indica fue concebido para su uso por parte del ámbito contable y administrativo.

En Argentina, por su parte, es dable mencionar el fuerte impulso que recibió este campo gracias a la Universidad de Buenos Aires. Así, en 1960 arriba a la Facultad de Ciencias Exactas y Naturales la computadora “Clementina”, proyecto dirigido por el Dr. Manuel Sadosky en el que se destacó una graduada de la FCE, la contadora pública Cecilia Berdichevsky y estuvieron involucrados quienes serían luego docentes de la Facultad de Ciencias Económicas, como el economista Roberto Frenkel. Otros proyectos pioneros de la informática fueron el de la CEFIBA (Computadora Electrónica de la Facultad de Ingeniería de Buenos Aires), de 1962, que logró fabricar una computadora en el país, y el CEUNS (Computadora Electrónica de la Universidad Nacional del Sur), que no tuvo el mismo éxito [Jacovkis, 2013].

Relacionado con la profesión contable, y más específicamente con la arista fiscal, vale la pena destacar que en el año 1992, a través de un crédito del BIRF (el hoy Banco Mundial, en ese entonces Banco Internacional de Reconstrucción y Fomento) se inició un proceso de modernización de la administración tributaria argentina, en una década en la que también en otros países latinoamericanos se evidenció el mismo proceso [Murga, 2005]. Pronto se contempló el surgimiento de aplicativos como el SICOSS (Sistema de Cálculo de Obligaciones de la Seguridad Social), la creación de la CUIT (Clave Única de Identificación Tributaria, cuyo uso se extiende a multitud de profesiones y excede por lejos el ámbito impositivo) y otras herramientas que hoy son parte cotidiana de la labor de los contadores públicos.

Bases de datos

Las bases de datos, que acumulan grandes magnitudes de datos, son hoy en día almacenadas utilizando distintos soportes, que pueden ser propios o ajenos, lo cual se ha convertido en una práctica habitual en el mundo de los negocios. Esto se debe a que, en los últimos años, han proliferado los servicios en la nube (*cloud*), a través de los cuales los datos son enviados a un tercero que se encarga de almacenarlos en sus centros de datos (*data centers*), en los que cuenta con servidores y personal capacitado para la tarea. De más está decir que esta operatoria requiere de un notable esfuerzo en materia de seguridad informática y que esto se ha generalizado

gracias a la velocidad que poseen las conexiones en la actualidad y a economías de escala que les permiten minimizar costos.

Resulta de interés notar que se puede hablar de una clasificación dicotómica de las bases de datos. Por un lado, están las bases de datos OLTP (Online Transaction Processing) y, por el otro, las OLAP (Online Analytical Processing). Las OLTP son aquellas bases de datos de carácter dinámico que utilizan información en tiempo real, siendo en todo momento objeto de consultas, eliminaciones y modificaciones. Resulta por cuestiones técnicas muy difícil utilizarla con fines de análisis. Por el contrario, las OLAP son bases de datos estáticas, que trabajan con información histórica de las OLTP y suelen actualizarse de manera periódica. Dado que resultan mucho más manejables, tienden a ser utilizadas para el armado de estadísticas y, como puede uno imaginarse, para los procesos de toma de decisiones por parte de gerentes, funcionarios y otros responsables de la conducción y el control de las organizaciones [Anguiano Morales, 2014].

Piénsese, por ejemplo, que un banco podría contar al mismo tiempo con una base de datos de tipo OLTP y una de tipo OLAP. La de tipo OLTP sería la destinada a consultar y modificar minuto a minuto el saldo de sus clientes, teniendo carácter operativo. Sin embargo, nada obsta que, a fin de que los equipos técnicos de la entidad financiera puedan realizar análisis de solvencia o liquidez con el objeto de decidir a qué clientes conviene brindar créditos y a qué tasa, al mismo tiempo desarrollen una base de datos de tipo OLAP para realizar estos análisis estadísticos.

Entonces, ¿Qué es la minería de datos?

Es difícil responder a esta pregunta, ya que se trata de un campo del conocimiento que aún no ha sido institucionalizado. Prueba de ello es que, en nuestro país, de hace unas décadas se trabaja en estos temas en cursos de posgrado pero recién en los últimos años aparecieron cursos de grado vinculados específicamente a la temática.

Más claro todavía es el hecho de que en cualquier medio gráfico se pueden encontrar numerosos términos como análisis de datos (*data analysis*), ciencia de datos (*data science*), analítica de negocios (*business analytics*) o datos masivos (*big*

data) sin que existan límites claros, sintiéndose muchas veces un aroma más comercial que científico. Prueba de ello es que el sitio de internet Proofreader.com ha reseñado 14 diagramas de Venn diferentes que intentan abarcar la definición de ciencia de datos.

Sí se puede mencionar la profunda comunión que existe entre la minería de datos y las ciencias económicas. La labor por antonomasia del auditor, la detección de fraudes (*fraud analysis*), fue lo que empujó hace unos 30 años el desarrollo de técnicas informáticas a medida que un número creciente de transacciones contó con un respaldo en soporte digital. Por citar un ejemplo, un opúsculo de Zhu Tianqing presentado ante el prestigioso IEEE (Institute of Electrical and Electronic Engineers) utiliza técnicas provenientes de la oceanografía para aplicarlas en el campo de los fraudes [Zhu, 2006]. Hoy en día se encuentran muy extendidas, además, las herramientas que permiten predecir la tasa de abandono o *churn rate* (proporción de clientes perdidos por la compañía) y analizan patrones de conducta de los consumidores.

Lo que también resulta prístino es que se trata de un campo en formación que requiere un profundo trabajo interdisciplinario. Así, será de gran utilidad para un contador que quiera abocarse a estos temas relacionarse con matemáticos, estadísticos, actuarios, economistas u otros graduados o estudiantes vinculados con el mundo de la estadística y las probabilidades y, al mismo tiempo, tender puentes con ingenieros, científicos y técnicos versados en el mundo de la informática, tanto en el mantenimiento y la seguridad de bases de datos como en el diseño de algoritmos que puedan resultar relevantes en su campo de acción.

Capítulo III. Auditoría, estadística e informática. La ley de Benford

Generalidades

Un caso paradigmático del uso del instrumental estadístico en el campo de la contabilidad es la aplicación de la ley de Benford para detectar comportamientos sospechosos en el marco de un programa de auditoría. En base a esta ley, que ha dado un salto cualitativo con el avance de la potencia de cálculo de los ordenadores, se pueden analizar los primeros dígitos de las operaciones de cualquier ente o

persona física y llegar a determinar, con un alto grado de confiabilidad, la presencia de movimientos irregulares en las operaciones de una organización.

Simon Newcomb (1835-1909), astrónomo canadiense-estadounidense, notó que las primeras hojas de los libros de logaritmos estaban más gastadas que las últimas. Esto significaba que los números de mantisa baja (es decir, aquellos que comenzaban con dígitos menores) eran más frecuentes que aquellos de mantisa alta. El ingeniero y físico estadounidense Frank Albert Benford Jr. (1883-1948) realizó, varias décadas después, la misma observación. Por lo tanto se puso a relevar los primeros dígitos de todo tipo de fenómenos naturales y sociales (medidas de lagos, domicilios, constantes científicas, etc.), y llegó a la conclusión de que, cuando los números se distribuyen de forma exponencial, la probabilidad de que un número comience con 1 es mucho más alta de que empiece con 2, y de que comience con 2 es más alta de que empiece con 3, y así sucesivamente siendo 9 el dígito menos frecuente. Se trata de una propiedad que algunos eruditos, como Samuel Goudsmit, vieron pocas décadas atrás con profunda desconfianza, sin creer que pudiera tener utilidad práctica [Nigrini, 2012].

Siendo más específicos, la probabilidad está dada por la siguiente expresión:

$$Prob(D_1 = d_1) = \log \left(\frac{1}{1 + d_1} \right)$$

Donde los cálculos se realizan en base al logaritmo decimal o *briggsiano*; es decir, en base 10. Esta misma propiedad puede ser llevada a las primeras dos, tres o cuatro cifras. Así, siguiendo con el razonamiento ya explicado, en una población tenderán a aparecer más números que comiencen con 11 que aquellos que empiecen con 19. El uso de más de un dígito resulta de notoria importancia cuando quieren realizarse exámenes exhaustivos.

Existen varias propiedades de relevancia. Una de ellas es el teorema de Pinkham, conocido también como el teorema de la invariancia de escala, que indica que si se multiplica un conjunto de datos que se distribuyen respetando la ley de Benford entonces el conjunto de datos resultante va a seguir ajustándose a la ley antedicha. Esto resulta de interés para la profesión, ya que si se desea expresar ciertas operaciones en otras divisas o practicar algún tipo de indexación el resultado sería, a estos efectos, indistinto.

Otras propiedades similares, algunas de ellas atribuidas a Adhikari y Sarkar, surgen si se multiplican los números por $1/x$, o cuando números que responden a una distribución cualquiera comienzan a multiplicarse entre sí (salvo excepciones), dándose un fenómeno análogo al del teorema central del límite, por el que todas las distribuciones tienden a distribuirse según la ley de Benford [Nigrini, op. cit.]. Por otro lado, a fin de establecer el nivel de ajuste a la distribución antedicha, Nigrini sugiere utilizar el desvío medio absoluto, conocido por sus siglas en inglés como MAD (Mean Absolute Deviation), un algoritmo similar en su cálculo a la varianza pero que utiliza el módulo (es decir, la diferencia absoluta) en vez del cuadrado de la diferencia de los distintos valores respecto de la media.

El uso de la ley de Benford en contabilidad y economía

Dentro de las ciencias económicas, uno de los pioneros en descubrir la utilidad de la Ley de Benford fue George Joseph Stigler (1911-1991), ganador del Premio Nobel de Economía de 1982 y uno de los principales referentes de la conocida como Escuela de Chicago. Dentro del campo de la contabilidad, se debieron esperar algunos años más para que esta ley de distribución se popularizara. Así, el matemático y contador sudafricano Mark Nigrini, a fines de la década de 1980 y principios de la de 1990, llegó a la conclusión de que resultaría de gran utilidad en la detección de fraudes [Nigrini, op. cit.].

Así, si por ejemplo, se repite de forma desproporcionada el primer dígito 9, esto puede ser un indicio de que existe un umbral que está siendo deliberadamente eludido. Es decir, si las operaciones a partir de \$1000 requieren de la autorización de un superior, es muy probable que, si se encuentra un número excesivo de operaciones de \$900, se trate en realidad de que se están generando sobrepagos teniendo como criterio no exceder el antedicho límite, o que un solo contrato mayor a \$1000 está siendo fraccionado con tal de esquivar el sistema de control y distribución de responsabilidades de la organización. Como podrá imaginarse, si se utiliza esta misma prueba pero utilizando dos dígitos los resultados terminan siendo todavía más ricos a la hora del análisis.

Hay diversos ejemplos de su utilización en materia de auditoría; por ejemplo, en el sector público. Su utilización en grandes volúmenes de datos de dos estados brasileños demostró que existe una sobrerrepresentación de los primeros dígitos 6 y 7, lo cual según Costa es un indicio de que se dividen compras para evitar pasar el límite de 8000 reales y beneficiar a un proveedor elegido discrecionalmente, ya que el antedicho monto es el punto a partir del cual se vuelve necesario realizar una licitación [Costa et al., 2012].

Aplicación simple con datos de Argentina

A fin de realizar una pequeña demostración del uso de la ley de Newcomb-Benford con datos de la economía local se descargaron del portal de datos abiertos del Estado Nacional⁴ sendos archivos correspondientes a las operaciones con *warrants* de soja y arroz del año 2017. Se trata de archivos en formato .csv (*Comma-Separated Values*) que representan lo que en la jerga se denomina un conjunto de datos (*dataset*), materia prima *par excellence* de un trabajo de minería de datos.

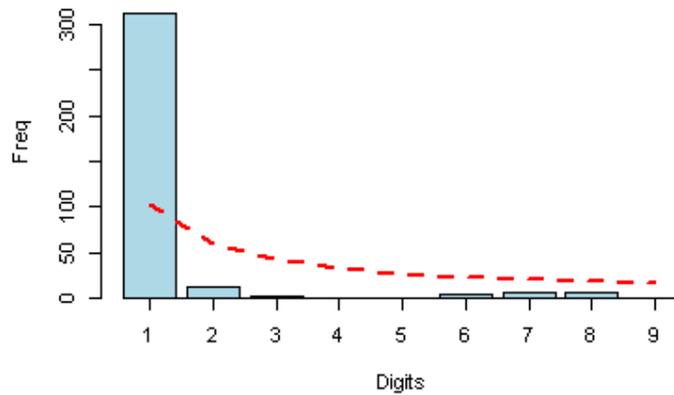
Luego de un pequeño procesamiento (había signos “\$” en el archivo de datos de soja que entorpecían la operatoria) quedaron listos para su uso por parte de un aplicativo. El lenguaje que se decidió utilizar fue R, uno de los más consolidados en el terreno estadístico (concebido específicamente para este propósito) y se decidió trabajar con el entorno RStudio, que posee herramientas que facilitan la rápida visualización de la información.

Antes de comenzar se descargó una biblioteca (*library*) denominada “benford.analysis”, que posee herramientas específicas para trabajar con esta distribución. Se diseñó un código (*script*), por cierto muy sencillo, y se obtuvieron, por un lado, gráficos y, por el otro, salidas con información sobre distintos valores estadísticos al estilo de los paquetes econométricos.

Se presenta el gráfico que muestra la frecuencia de los dígitos en el conjunto de datos de *warrants* de soja.

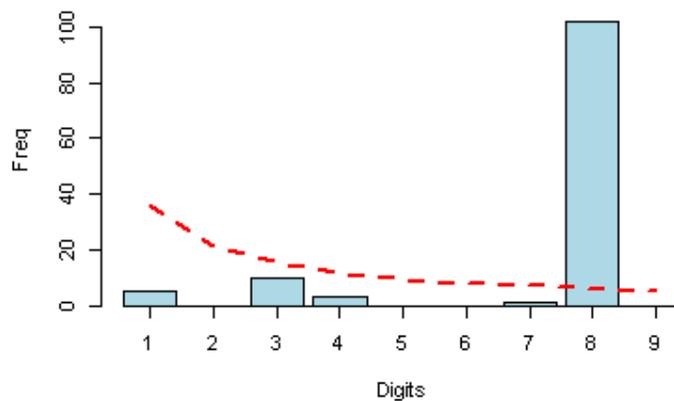
⁴ <https://datos.gob.ar/>

Digits Distribution



Se muestran los datos correspondientes a los movimientos en el mercado de arroz:

Digits Distribution



Las conclusiones son las que uno podría esperar. Tanto los gráficos, de manera intuitiva, como el estadístico del desvío medio absoluto mostraron un mejor ajuste de los *warrants* de soja a la ley de Benford que los de arroz. Por un lado, se puede pensar en que se debe a que el universo de contratos de soja es más grande que el de contratos de arroz, razón por la que la probabilidad de tender a la distribución de Benford es más alta. También podría pensarse en que, al ser un mercado más grande, existen menos posibilidades de que un agente o un grupo de ellos pueda influir en el proceso de formación de precios. De todos modos, se trata de un mero ejemplo con datos de la realidad argentina a título ilustrativo. Desde ya que, para poder extraerse conclusiones serias, se requiere un estudio hondo de las características de la cadena agroalimentaria que excede el propósito de esta tesina.

Conclusiones

A lo largo del presente trabajo resultó claro que la disciplina contable ha vivido una fuerte transformación a partir de las últimas décadas con la irrupción de la informática. Esto representa, por un lado, una amenaza y, al mismo tiempo, una oportunidad.

Una amenaza porque aquellos profesionales que no se actualicen o se concentren en tareas repetitivas sin cultivar un espíritu crítico corren el riesgo de ser desplazados en el mercado laboral por otros graduados más dinámicos y por algoritmos que avanzan a pasos agigantados. Y, sin embargo, al mismo tiempo se puede hablar de una oportunidad porque, al capacitarse en materia tecnológica e integrarse a equipos de trabajo junto a expertos de otras especialidades, al contador se le abre un ilimitado y colorido abanico de nuevos campos, y no sólo en los terrenos en los que tradicionalmente desempeña sus tareas.

Lejos de los estereotipos del contador que, vestido de saco y corbata, va a todas partes llevando con una mano una calculadora y con la otra una maleta de cuero llena de papeles, los lineamientos locales definen al contador como un profesional de la información “en todas sus formas” (sic). Así, en tiempos en que la misma desempeña un papel central en la toma de decisiones, resulta de enorme importancia que el contador público sume sus conocimientos a la hora de generar y controlar información para la alta gerencia tanto en el ámbito público como en el privado como en el tercer sector.

Dentro de esta revolución tecnológica, se mencionó la importancia que posee la minería de datos, que implica la explotación de grandes volúmenes de información combinando conocimientos matemático-estadísticos, de computación y del campo de aplicación (economía, contabilidad, sociología, biología, etc.). Las posibilidades que se abren a través de la combinación intensiva de estas áreas son infinitas.

Se enfatizó en la estrecha relación entre las ciencias económicas y el desarrollo de la informática. Así, se empezó relatando la “prehistoria” de la computación, en épocas en las que se desarrollaban tablas de logaritmos, financieras o de seguros de vida para facilitar los cálculos de contadores, actuarios, economistas, administradores o recaudadores de impuestos hasta la actualidad, en la que los

ordenadores se han vuelto omnipresentes. Hoy resulta casi imposible desarrollar una carrera profesional, ya sea como auditor independiente, administrativo en relación de dependencia, docente o investigador sin sentarse delante de un monitor (o, en tiempos más recientes, de la pantalla de un dispositivo móvil).

Asimismo, se remarcó que fueron las necesidades propias de las ciencias económicas las que trajeron desafíos y problemas a resolver que impulsaron y condicionaron el crecimiento de la computación. No por nada existen desde hace 60 años lenguajes pensados para su aplicación en el mundo de los negocios y, respecto de la minería de datos en particular, fue su uso en la detección de fraudes uno de los primeros objetos a los que se abocaron los especialistas.

Así, dentro del uso de estas herramientas informáticas y estadísticas, se reseñó la ley de Newcomb-Benford, de enorme importancia para la labor del contador en general y del auditor en particular. Esta ley de distribución de los primeros dígitos que implica que las cifras más bajas tiendan a aparecer más habitualmente que las más altas se convierte, así, en un elemento adminicular de notable importancia a la hora de realizar un primer paneo rápido que permita detectar posibles irregularidades en grandes bases de datos en cuestión de segundos. En su momento estuvo quien creyó que se trataba de una mera curiosidad, pero con el paso de los años quedó demostrado que, gracias a la potencia de cálculo de los ordenadores contemporáneos, se puede convertir en una verdadera piedra angular en un informe de auditoría y en otros segmentos de las ciencias económicas.

Se puede concluir que, sin lugar a dudas, los contadores públicos, si quieren seguir realizando aportes a la sociedad, deben realizar un enorme esfuerzo. El esfuerzo de lanzarse al camino sin retorno de la capacitación continua. El esfuerzo de familiarizarse con los avances en la informática, la estadística y otras ramas del saber. Un esfuerzo que, sin pasión, es muy difícil de sostener a lo largo de los años.

Pero, sobre todo, el esfuerzo de tomar el primer tomo de un diccionario (o conectarse con uno en línea) y buscar la etimología del término “auditor”, cuya raíz latina significa “el que oye”. Hoy más que nunca, en esta era de renovación permanente, en la que se necesita amalgamar el entusiasmo y el manejo tecnológico de los jóvenes y la trayectoria de los mayores, el contador debe saber oír y jamás despreciar el aporte de los estudiantes y jóvenes graduados, que tienen

mucho para sumar al campo académico y profesional. También debe oír con humildad y buena predisposición, dejando de lado prejuicios y viejas rivalidades inconducentes, a todo aquel especialista de las más diversas disciplinas que pueda traer ideas útiles para la labor cotidiana. El contador que oye o, mejor todavía, que escucha. El único contador posible en el tercer milenio.

Bibliografía

Anguiano Morales, J. D. (2014). *Características y tipos de bases de datos*. IBM DeveloperWorks.
https://www.ibm.com/developerworks/ssa/data/library/tipos_bases_de_datos/index.html

Arranga, E. C., & Coyle, F. P. (1996). *Object-oriented COBOL*. Cambridge University Press.

Campbell-Kelly, M. (2018). *Computer: A History of the Information Machine*. Routledge.

Copeland, B. J., & Proudfoot, D. (1996). On Alan Turing's anticipation of connectionism. *Synthese*, 108(3), 361-377.

Costa, José Isidio Freitas, Josenildo Dos Santos, and Silvana Melo Travassos. "Análise de conformidade nos gastos públicos dos entes federativos: estudo de caso de uma aplicação da Lei de Newcomb-Benford para o primeiro e segundo dígito em dois estados brasileiros." *Revista Contabilidade & Finanças-USP* 23.60 (2012).

Escobar, D. S. (2017). Formación del Contador Público en Tecnología y Seguridad de la Información. Propuesta de reforma curricular. Tesis de Maestría. *Maestría en Seguridad Informática. Facultad de Ciencias Económicas, Ciencias Exactas y Naturales e Ingeniería*.

Federación Internacional de Contadores (IFAC), "Formas Internacionales de Formación"; 2008, [consultada el 23 de abril de 2018]. Disponible en: "http://www.ifac.org/system/files/downloads/Spanish_Translation_Normas_Internacionales_de_Formacion_2008.pdf"

Jacovkis, P. M. (2013). *De Clementina al siglo XXI: Breve historia de la computación en la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires*. Eudeba.

Landro, A. H. (2010). *Acerca de la probabilidad. Ediciones Cooperativas.*

Murga, Y. (2005). Modernización de los sistemas tributarios de América del Sur durante los años noventa. *Ponencia presentada en octubre de.*

Nigrini, M. (2012). *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection. Wiley Corporate F&A.*

Sriraman, B., & Lee, K. (2014). Commentary on Perspective I: The Humanistic Dimensions of Probability. In *Probabilistic Thinking* (pp. 117-119). Springer, Dordrecht.

Toranzos, F. I. (1997). *Teoría Estadística y Aplicaciones. Ediciones Macchi.*

Yang, D. C., & Guan, L. (2004). The evolution of IT auditing and internal control standards in financial statement audits: The case of the United States. *Managerial Auditing Journal*, 19(4), 544-555.

Zhu, T. (2006, December). Suspicious financial transaction detection based on empirical mode decomposition method. In *2006 IEEE Asia-Pacific Conference on Services Computing (APSCC'06)* (pp. 300-304). IEEE.

Anexos

Anexo I. Código analizando el ajuste a la ley de Benford de las operaciones con warrants de soja (Argentina, 2017)

El texto siguiente a un numeral (“#”) es un comentario utilizado para aclarar los pasos seguidos en las instrucciones.

```
#Cálculo de los primeros dígitos con warrants de soja
library(benford.analysis)
#Cargamos la biblioteca benford.analysis. Debe ser instalada previamente con el
comando install.packages("benford.analysis").
warrants.vigentes.soja.2017.sin..signo.peso <- read.csv("warrants-vigentes-soja-
2017-sin signo peso.csv")
#Cargamos el archivo descargado del portal de datos abiertos del Estado nacional.
Se debe completar con la ruta en la cual se haya guardado el archivo.
sojabenford<-benford(warrants.vigentes.soja.2017.sin..signo.peso$cant_mon_pesos)
#Creamos un vector en base a los montantes de todas las operaciones en pesos de
warrants de soja y estudiamos su ajuste a la ley de Benford.
plot(sojabenford)
#Generamos un gráfico con el antedicho vector (plot).
sojabenford
#Solicitamos a R que genere una salida con los datos, al estilo de las de los distintos
paquetes estadísticos y econométricos.
```

Anexo II. Código Analizando el ajuste a la ley de Benford de las operaciones con warrants de arroz (Argentina, 2017)

```
#Análisis de los primeros dígitos con warrants de arroz. Los pasos son análogos a
los de los warrants de soja.
library(benford.analysis)
warrants.vigentes.arroz.2017 <- read.csv("C:/Users/Lidia/Dropbox/Descargas Google
Chrome/Tesis CPN/warrants-vigentes-arroz-2017.csv")
arrozbenford<-benford(warrants.vigentes.arroz.2017$cant_mon_pesos,1)
plot(arrozbenford)
arrozbenford
```

Anexo III. Salida de los datos de las operaciones con warrants de soja (Argentina, 2017)

Benford object:

```
Data: warrants.vigentes.soja.2017.sin..signo.peso$cant_mon_pesos
Number of observations used = 340
```

Number of obs. for second order = 29
First digits analysed = 1

Mantissa:

Statistic Value
Mean 0.205
Var 0.025
Ex.Kurtosis 12.390
Skewness 3.352

The 5 largest deviations:

	digits	absolute.diff
1	1	210.65
2	2	47.87
3	3	41.48
4	4	32.95
5	5	26.92

Stats:

Pearson's Chi-squared test

data: warrants.vigentes.soja.2017.sin..signo.peso\$cant_mon_pesos
X-squared = 623.03, df = 8, p-value < 2.2e-16

Mantissa Arc Test

data: warrants.vigentes.soja.2017.sin..signo.peso\$cant_mon_pesos
L2 = 0.7527, df = 2, p-value < 2.2e-16

Mean Absolute Deviation (MAD): 0.1376796
MAD Conformity - Nigrini (2012): Nonconformity
Distortion Factor: -54.62739

Remember: Real data will never conform perfectly to Benford's Law. You should not focus on p-values!

Anexo IV. Salida de los datos de las operaciones con warrants de arroz (Argentina, 2017)

Benford object:

Data: warrants.vigentes.arroz.2017\$cant_mon_pesos
Number of observations used = 121

Number of obs. for second order = 4
First digits analysed = 1

Mantissa:

Statistic Value
Mean 0.849
Var 0.029
Ex.Kurtosis 5.307
Skewness -2.515

The 5 largest deviations:

	digits	absolute.diff
1	8	95.81
2	1	31.42
3	2	21.31
4	5	9.58
5	4	8.73

Stats:

Pearson's Chi-squared test

data: warrants.vigentes.arroz.2017\$scant_mon_pesos
X-squared = 1568.1, df = 8, p-value < 2.2e-16

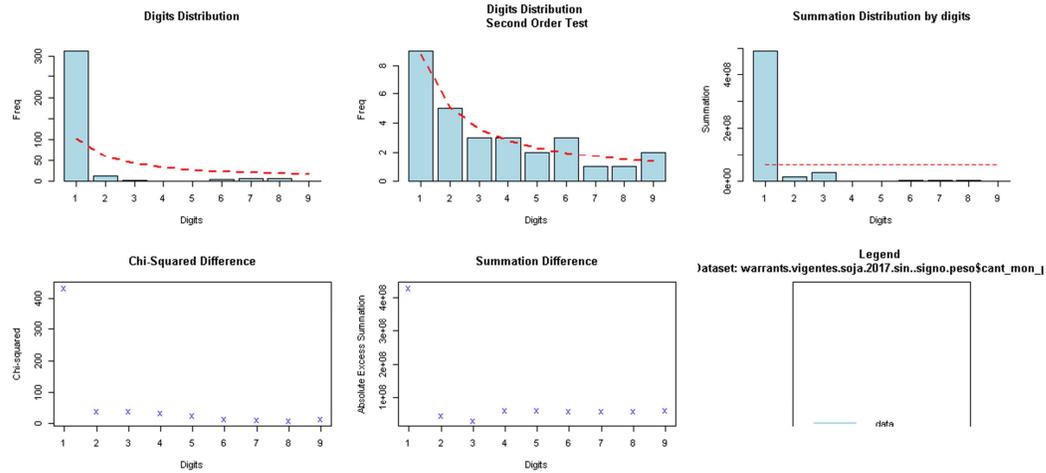
Mantissa Arc Test

data: warrants.vigentes.arroz.2017\$scant_mon_pesos
L2 = 0.59551, df = 2, p-value < 2.2e-16

Mean Absolute Deviation (MAD): 0.1759606
MAD Conformity - Nigrini (2012): Nonconformity
Distortion Factor: 91.39137

Remember: Real data will never conform perfectly to Benford's Law. You should not focus on p-values!

Anexo V. Gráfico correspondiente a los datos de la operatoria con warrants de soja



Anexo VI. Gráfico correspondiente a los datos de la operatoria con warrants de arroz

