

ENCONTRANDO DATOS ANÓMALOS EN LA TRIBUTACIÓN. APLICACIÓN DE LA LEY DE BENFORD EN EL IMPUESTO A LA RENTA EN ECUADOR¹

Ana Lucía Oña Macías*

Sergio Troncoso Igua**

Resumen. La Ley de Benford, conocida también como la ley del primer dígito a lo largo del tiempo, ha sido empleada para detectar anomalías en cifras. De acuerdo a esta ley, la distribución de los números del 1 al 9, como primer dígito es asimétrica. El número 1 tiene mayor probabilidad de aparecer en un conjunto de datos como primer dígito (30,1%) y va decreciendo hasta llegar al 9 con una probabilidad de aparición de apenas 4,6%. En la práctica esta ley permite detectar datos erróneos al no cumplir con la distribución teórica. En este sentido, esta investigación tiene como objetivo aplicar la Ley de Benford a datos fiscales para buscar indicios de actos de evasión y mala declaración. Para esto, se utilizaron las declaraciones de Impuesto a la Renta tanto de personas naturales como sociedades en Ecuador para el año 2014, demostrando que esta ley puede ser empleada de forma confiable para detectar anomalías en las declaraciones de impuestos. La presente investigación se constituye en un indicio para el uso de la ley de manera global, simple y económica previa a controles más exhaustivos y costosos.

Palabras clave: Fraude fiscal; Estadística aplicada; Calidad de datos.

* Universidad de Atacama (UDA), Chile

Contacto: ana.ona@uda.cl

** Universidad Católica de Chile (UC), Chile

Contacto: stroncoso1@mat.uc.cl

1 Versiones anteriores de este artículo se presentaron en el Congreso Interamericano de Estadística en la Universidad Nacional de Rosario, Argentina (2017) y en el Seminario por el quinquenio de la Escuela de Estadística en la Universidad del Valle, Colombia (2018). Esta investigación se llevó a cabo gracias a la información proporcionada por el Centro de Estudios Fiscales del Servicio de Rentas Internas (SRI-Ecuador) y al apoyo de Consultores Estadísticos.

**FINDING ANOMALOUS DATA IN TAXATIO.
APPLICATION OF BENFORD'S LAW IN INCOME TAX IN ECUADOR**

Abstract. Benford's Law, also known as the first digit law over time, has been used to detect anomalies in figures. According to this law, the distribution of numbers from 1 to 9, as the first digit is asymmetric. The number 1 is more likely to appear in a set of data as the first digit (30.1%) and decreases until it reaches 9 with a probability of appearance of just 4.6%. In practice, this law allows the detection of erroneous data by not complying with the theoretical distribution. In this sense, this investigation aims to apply Benford's Law to economic data to search for indications of possible acts of evasion and misrepresentation. For this, the income tax statements of both individuals and companies in Ecuador were used for 2014, demonstrating that this law can be used reliably to detect anomalies in tax returns. The present investigation constitutes an indication for the use of the law in a global, economical and straightforward way before more exhaustive and expensive controls.

Keywords: Tax Fraud; Applied Statistics; Data Quality.

Original recibido el 02/01/2018

Aceptado para su publicación el 24/08/2018

1. Introducción

La evasión es toda reducción de un monto tributario que se ha generado en un país por quienes normativamente están obligados a hacerlo (Villegas, Arguello Velez y Spila García, 1993). Esta es posible por las subdeclaraciones de los contribuyentes que declaran por debajo de sus ingresos reales y las dificultades que presentan las Administraciones Tributarias para observar y controlar las variables que forman parte de la base imponible. De acuerdo al estudio de la CEPAL (Comisión Económica para América Latina y el Caribe) la evasión tributaria en América Latina fluctúa entre el 40% y el 65%, lo cual representa una pérdida de ingresos de 4,6 puntos del Producto Bruto Interno (PBI) en promedio cada año (Gómez Sabaini, Jiménez, y Podestá, 2010). Estas cifras generan preocupación considerando que la presión tributaria interna en promedio en la región es de 13,9% (Centro de Estudios Fiscales, 2017).

De acuerdo a Prats y Macías-Aymar (2008) uno de los grandes causantes de la evasión es que el pacto fiscal que debe existir entre la sociedad y el Estado en los países de América Latina ha sido resquebrajado. En consecuencia, la retribución que perciben los contribuyentes a cambio de sus impuestos por parte del Estado como son bienes y servicios públicos no es clara. La tributación se ve motivada solo por un modelo sancionatorio.

Otros autores como Cullis y Lewis (1997) unen la psicología social y la psicología económica para explicar la evasión tributaria. Ellos determinan que los individuos actúan en base a normas sociales y al comportamiento de los demás individuos. Si en el entorno en el que se desenvuelve el contribuyente la evasión es vista como una práctica común, él tenderá a evadir. Por otro lado, si en su entorno la evasión no es bien vista ni practicada, el contribuyente no optará por esta práctica. Tomando en consideración la opinión de los ciudadanos en Latinobarómetro (2016) se evidencia que la creencia de que evadir impuestos tiene una justificación ha crecido en los últimos años².

Si bien las causas de la evasión pueden ser variadas, desde la estadística se busca construir modelos que puedan evidenciarla. Los pioneros en analizar la evasión tributaria en función al comportamiento de los contribuyentes fueron Allingham y Sandmo (1972). En su modelo se deja de lado la moral y el contribuyente debe tomar la decisión entre evadir o declarar todos sus ingresos. Se concluye que la decisión del contribuyente depende del ahorro fiscal que se genera entre su percepción del riesgo y el costo que tendría ser detectado como evasor.

Más recientemente Bolton y Hand (2002) hacen una revisión de la detección del fraude mediante teorías estadísticas acompañadas de herramientas informáticas. A través de esta combinación se han generado tecnologías eficientes para detectar distintos tipos de fraude como lavado de dinero, uso fraudulento de las tarjetas de crédito, entre otros. Para llegar a identificar estas acciones fraudulentas se pueden utilizar diferentes metodologías estadísticas como por ejemplo el análisis discriminante, modelos logísticos, redes neuronales, métodos clasificadores supervisados o combinación de métodos. Estos métodos estadísticos pueden ser

2 Latinobarómetro es una corporación de derecho privado sin fines de lucro que realizan encuestas de percepción de temas de interés para la región

supervisados o no supervisados. En el caso de los supervisados, se conoce con anticipación los registros fraudulentos y no fraudulentos y se genera un modelo que permita determinar a qué grupo pertenecen las nuevas observaciones. En los modelos no supervisados, se busca comportamientos que denoten anomalías. La Ley de Benford pertenece a este último tipo y sirve para detectar anomalías en las declaraciones.

La aplicación de la Ley de Benford apunta a identificar el fraude fiscal desde la declaración, de manera inmediata, no costosa y de fácil implementación a diferencia de otros métodos estadísticos que, si bien detectan las anomalías, requieren estudios más profundos para ser concluyentes ya que necesitan un análisis específico de la naturaleza de los datos y el número de observaciones (Tsagbey, De Carvalho y Page, 2017). Los antecedentes en esta materia evidencian que la evasión responde a aspectos morales y coercitivos, y su control en algunos casos podría resultar más cara que la misma evasión (Serra, 2000). Por tal motivo, son importantes aquellas herramientas que permitan detectar anomalías antes de iniciar algún control intensivo o auditoría ya que permitiría disminuir costos a la Administración Tributaria, aumentar la percepción de riesgo por parte de los contribuyentes y disminuir la opción de manipular datos.

En este sentido, la Ley de Benford resuelve las dos problemáticas principales que presenta la recolección y generación de información en el presente. Por un lado, minimiza los retos para procesar y analizar los datos; y por otro lado, disminuye el costo en relación a los procesos de control tributario tradicionales, como por ejemplo las auditorías. Estos motivos llevaron a que las Administraciones Tributarias indaguen en estas herramientas que permitan direccionar de mejor forma los controles tributarios para cumplir con su objetivo de recaudación y evitar la evasión tributaria. Bajo esta perspectiva, la Ley de Benford ha sido empleada para detectar anomalías en los datos. El presente documento busca corroborar el uso de la misma. En la siguiente sección se describirá esta ley teóricamente para luego poder abordar su aplicación práctica con un estudio de caso sobre el Impuesto a la Renta en Ecuador.

2. La Ley de Benford (ley del primer dígito)

El desarrollo de esta ley tuvo varios colaboradores a lo largo de la historia que fueron modificándola de acuerdo a su formación y visión para abordarla, los principales son: Simon Newcomb, Frank Benford, Roger Pinkham, Ted Hill y Mark Nigrini. La ley fue enunciada por Newcomb (astrónomo y matemático) mediante la observación de las páginas de las tablas de logaritmos que empleaban para los cálculos en esa época. Las páginas del principio estaban más desgastadas que las del final y ese fue el inicio de esta curiosidad (Newcomb, 1881). En 1938 Benford (físico) estudió diferentes fuentes de información para comprobar la probabilidad de aparición de los dígitos del 1 al 9 en datos reales (Benford, 1938). Posteriormente, el matemático Pinkham sugirió que la distribución de las frecuencias debe ser invariante frente a cambios de escala y demostró que si una ley de frecuencias es invariante en escala, entonces es Ley de Benford (Pinkham, 1961). Fue en 1995 el año en el cual Ted Hill (matemático) demostró la ley empleando el teorema del límite central y le dio el enfoque probabilístico necesario. En 1996, Nigrini (profesor de contabilidad) lo

utilizó para el fraude mediante datos fiscales dando apertura a diversas aplicaciones (Nigrini, 1996).

De acuerdo a la revisión efectuada por Tsagbey *et al.* (2017) esta ley ha sido empleada no sólo para descubrir datos erróneos en el fraude fiscal, sino también en casos de fraude electoral, análisis de sobrevivencia en el comercio de langosta, emisiones tóxicas, calidad de datos en encuesta y su propio estudio en estadísticas oficiales de pesca en Ghana.

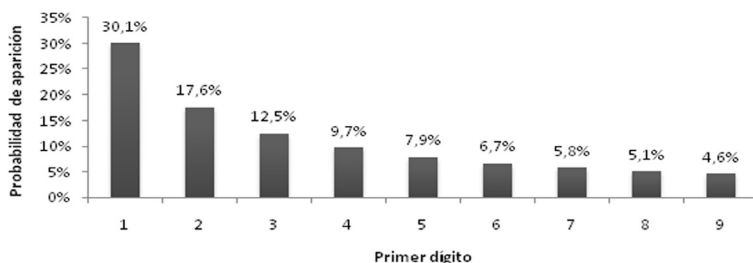
Fewster (2009) asegura que los primeros dígitos generados a partir de cualquier distribución que sea relativamente suave con una muestra mayor a 100 seguirán la Ley de Benford. Sin embargo, la fabricación de datos en contextos reales que cumplan con la Ley de Benford es difícil (Bolton y Hand, 2002).

La frecuencia con la que inician las cifras tiene la siguiente distribución de acuerdo a la ley del primer dígito:

$$P(d)=\log_{10}(1+1/d)$$

donde d simboliza al primer dígito, de 1 al 9. Bajo esta ley de probabilidad, la aparición del número 1 es aproximadamente el 30% mientras que iniciar con el número 2 es de 17% y así va descendiendo como se muestra en la Figura 1.

Figura 1. Distribución del primer dígito de acuerdo a la Ley de Benford



Fuente: Elaboración propia en base a Benford (1938).

Hill trabajó durante mucho tiempo sobre la teoría de esta ley (Berger y Hill, 2011; Hill, 1995a; Hill, 1995b). Su aporte resulta fundamental ya que en la demostración considera a la distribución del primer dígito como combinaciones de otras distribuciones y además la invarianza³. En conjunto, las frecuencias del primer dígito de las muestras combinadas convergen a una Ley de Benford, aun cuando individualmente no lo hagan en estricto rigor.

Formalmente para corroborar la hipótesis de que el primer dígito de los datos obedece a la distribución de Benford se emplea el estadístico chi cuadrado de Pearson y para saber cuánto se aleja cada dígito de la distribución teórica se emplean los residuos de Pearson. Ambos se encuentran implementados en el software estadístico R (R Development Core Team, 2011) en el paquete Benford Tests (Joenssen, 2015).

³ La ley funciona sin importar la escala en la que se encuentren los datos.

Sea X_i una sucesión de datos y d el primer dígito de esta sucesión; el estadístico de Pearson se define por:

$$\chi^2 = \sum_{(d=1)}^9 (O_d - E_d)^2 / E_d$$

donde O_d y E_d son las frecuencias observadas y esperadas, bajo la Ley de Benford, para el dígito d

$$O_d = \sum_{(i=1)}^n I(D(X_i)=d) \text{ y } E_d = nP_d$$

siendo P_d la proporción teórica de la Ley de Benford.

La hipótesis nula es que las frecuencias observadas del primer dígito son iguales a las frecuencias esperadas (distribución teórica de la Ley de Benford). Para rechazar la hipótesis nula se tiene como referencia el valor del estadístico de una distribución chi cuadrado con 8 grados de libertad. Rechazando esta hipótesis se tiene evidencia de que los primeros dígitos de los datos no cumplen con la Ley de Benford (Tsagbey *et al.*, 2017).

Los residuos de Pearson se definen a partir de la suma cuadrática del estadístico de Pearson:

$$\chi^2 = \sum_{(d=1)}^9 R_d^2 \text{ con } R_d = \{(O_d - E_d)^2 / E_d\}^{(1/2)} \text{ para } d=1, \dots, 9$$

Como muchas teorías y leyes probabilísticas, la ley del primer dígito se puede demostrar de manera formal usando como herramienta principal la teoría de la medida, en particular empleando la teoría de las medidas ergódicas. La razón principal por la cual esta teoría se usa es para entender de manera abstracta la ley del primer dígito debido a la naturaleza iterativa o en términos más formales la naturaleza dinámica de esta ley. En el anexo se deja los indicios de la demostración de esta ley mediante el teorema de convergencia de Birkhoff⁴.

3. Aplicación con datos fiscales de Ecuador

En este tercer apartado se describirá la Ley de Benford, ya no teóricamente, sino mediante una aplicación práctica en las cuentas fiscales de Ecuador. Concretamente, se abordará el Impuesto a la Renta en ese país y sus problemas en torno a la evasión fiscal mediante este análisis estadístico.

El Impuesto a la Renta tiene una marcada importancia en la recaudación tributaria en Ecuador ya que representa aproximadamente el 32% del total de la recaudación (Servicio de Rentas Internas, 2017a). Tiene la característica de ser un impuesto directo, progresivo y que contiene una serie de deducciones y exoneraciones que pueden ser utilizadas por los contribuyentes como mecanismo de evasión del impuesto.

Los controles que lleva adelante la institución son tanto de forma intensiva (auditorías),

⁴ Los detalles de esta demostración desde la teoría ergódica y el teorema de Birkhoff se pueden consultar en Viana y Oliviera (2016).

omisos (contribuyentes que no declaran a tiempo), persuasivos (notificaciones masivas), inconsistencias (valores mal declarados) y por diferencias (contrarrestado con información de terceros) (Servicio de Rentas Internas, 2017b).

A pesar de los esfuerzos de la Administración Tributaria de Ecuador por incrementar la recaudación y hacerla de forma equitativa, existen estrategias utilizadas por los contribuyentes para no pagar impuestos o disminuirlos. Un ejemplo claro de esto es que al 31 diciembre del 2016 del total de la cartera de cobro de la institución, el 50,65% corresponde a obligaciones cuya acción de cobro no se puede realizar, ya sea por impugnaciones, reclamos, excepciones o juicios (Servicio de Rentas Internas, 2017b).

Además, desde el año fiscal 2010 se han detectado transacciones de contribuyentes con empresas "fantasmas". Estas empresas han sido creadas principalmente para simular compras y disminuir la tasa impositiva de otras empresas o personas naturales. Al año 2016, a pesar de los controles, se obtuvo que el 7,95% de todas las compras entre contribuyentes se realizó a este grupo de empresas. A febrero del 2018 el valor recuperado tanto en Impuestos a la Renta como al Impuesto al Valor Agregado asciende a USD 428 millones, pagados por 9,6 mil contribuyentes (Servicio de Rentas Internas, 2018).

3.1 Análisis de la información

Los datos empleados para este estudio corresponden a las declaraciones de Impuesto a la Renta de las sociedades y de las personas naturales que presentaron su declaración a la Administración Tributaria en Ecuador (Servicio de Rentas Internas, SRI) para el año fiscal 2014. Se emplearon 3 formularios de declaración: el formulario 101, 102 y 107. El formulario 101 concentra la declaración de las empresas que se han constituido como sociedades. El formulario 102 incluye los datos de las personas naturales que tienen actividad económica y pueden percibir ingresos por algún trabajo en relación de dependencia. Por último, el formulario 107 abarca solo a los empleados bajo relación de dependencia.

Una limitación del estudio para poder detectar a manera de microdato a los evasores fue la cantidad de datos que se tenía por empresa o persona. Por ejemplo, el formulario 101 (de empresas) tiene 580 variables pero aproximadamente el 92% de las variables tienen valores declarados como cero o no declarados, lo que disminuye sustancialmente la información para cada empresa y no permite generar resultados confiables.

Por este motivo, para este estudio se abordó el análisis de las variables relacionadas al giro del negocio de la empresa (activos, pasivos, ingresos, costos, etc.) y las de carácter individual de las personas naturales (gastos e ingresos personales). Además, en Ecuador la normativa permite que los contribuyentes puedan presentar declaraciones sustitutivas en las cuales pueden cambiar los valores declarados. Para evitar problemas de duplicidad de información y contar con los valores más recientes se utilizaron solo las últimas declaraciones de los contribuyentes.

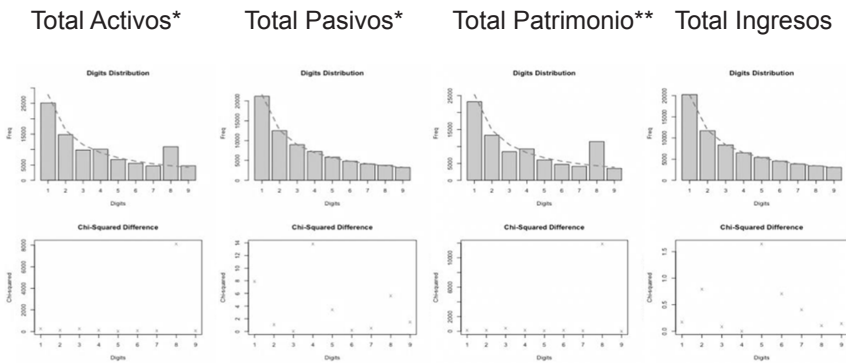
La base de datos de las declaraciones de las empresas (formulario 101) consta de 116.502 observaciones y 580 variables. La de las personas naturales con actividad

económica (formulario 102) tiene 720.387 declarantes (555 variables) y la base de datos de las personas naturales bajo relación de dependencia (formulario 107) alcanza los 3.481.873 registros (55 variables). Los datos faltantes o declaraciones en cero en cada variable no fueron considerados para el análisis. En datos reales el cumplimiento a cabalidad de la Ley de Benford es difícil, se deben analizar varios aspectos para confirmar una anomalía. En el Anexo 1 se detalla información con respecto al número de observaciones, dígitos con mayor diferencia, el valor absoluto de la diferencia, el estadístico y su probabilidad asociada (valor p). La Hipótesis nula es el cumplimiento de la Ley de Benford. Se muestran los niveles de significancia bajo los cuales se rechaza la hipótesis.

3.1.1. Variables formulario 101. Sociedades

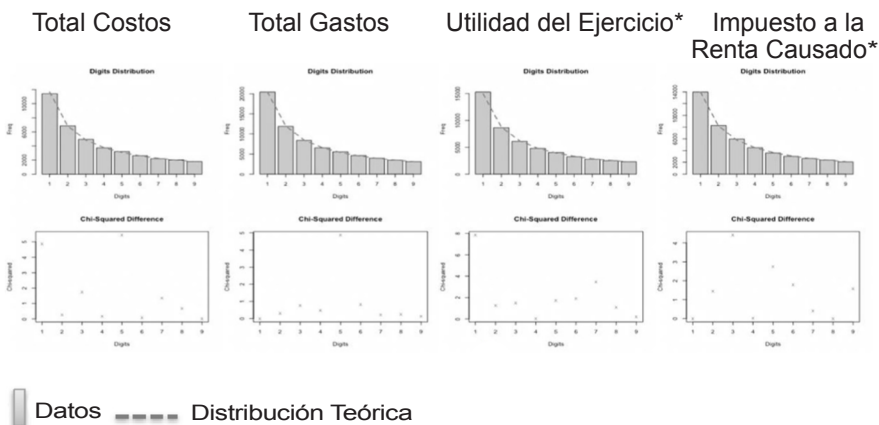
En la declaración del Impuesto a la Renta de sociedades solo se tiene las variables relacionadas con el giro del negocio. En la Figura 2 se puede observar cómo las variables seleccionadas cumplen con la distribución teórica de la Ley de Benford a excepción del Total Activos y el Total Patrimonio. Esta situación se da porque el requisito para constituirse como compañía anónima es poseer un capital mínimo de USD 800 lo cual afecta a las dos variables, dado que el capital se considera en el activo y en la declaración del patrimonio. Otro aspecto importante a considerar es que se requiere de un/a contador/a para reportar esta información. Se puede concluir que al tener una estructura contable por ley y al tener mayor probabilidad de ser auditados, de forma general se evidencia un cumplimiento tributario verificado por esta ley.

Figura 2. Aplicación Ley de Benford formulario 101. Sociedades. Ecuador, año fiscal 2014



Continúa

Figura 2 (continuación).



Nota: * Nivel de significancia 5%.

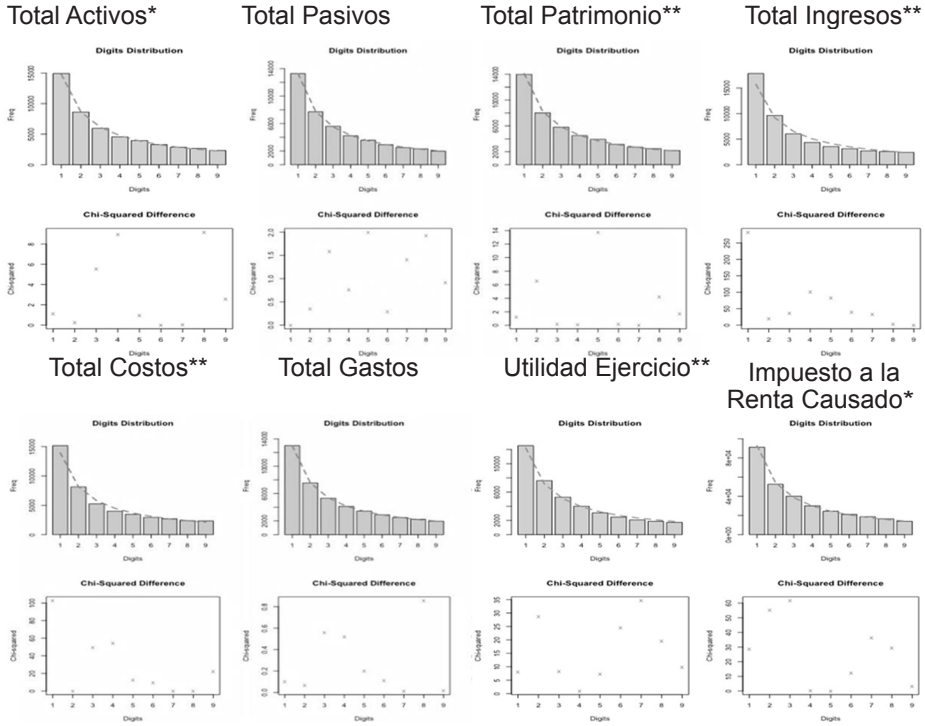
** Nivel de significancia 1%.

Fuente: Elaboración propia mediante Servicio de Rentas Internas de Ecuador.

3.1.2. Variables formulario 102. Personas naturales.

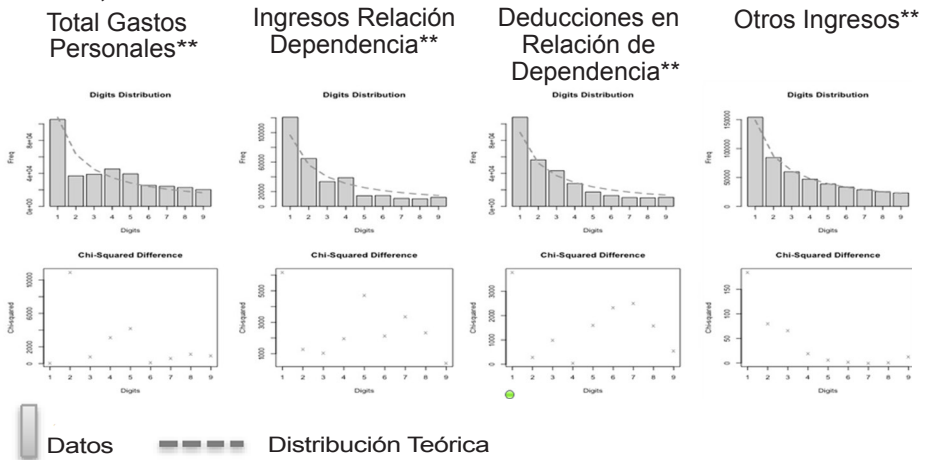
En las declaraciones del Formulario 102 del Impuesto a la Renta de personas naturales se presentan los dos grupos de variables consideradas: la primera que agrupa a las variables relacionadas con la actividad económica (Figura 3) y la segunda la que agrupa a las variables de ingresos y egresos por otras fuentes (Figura 4). En la Figura 3 se puede observar que el primer dígito de las variables Total de Pasivos y Total de Gastos siguen la Ley de Benford. Las demás variables no cumplen esta ley; sin embargo sus diferencias no son grandes como se visualiza en los residuos de Pearson. Si se contempla el segundo grupo de variables (Total Gastos Personales, Ingresos y Deducciones en Relación de Dependencia (RD) y Otros Ingresos), todas ellas presentan diferencias significativas con la distribución teórica. Se evidencian anomalías en las declaraciones especialmente en el Total de Gastos Personales. Para analizar con mayor detalle esta situación se consideran las declaraciones de las personas naturales en relación de dependencia, las cuales no presentan otros ingresos de actividad económica sino solo de empleo.

Figura 3. Aplicación Ley de Benford formulario 102. Personas Naturales. Ecuador, año fiscal 2014



Fuente: Elaboración propia en base a Servicio de Rentas Internas de Ecuador.

Figura 4. Aplicación Ley de Benford formulario 102. Personas Naturales. Ecuador, año fiscal 2014



Nota: * Nivel de significancia 5%.
 ** Nivel de significancia 1%.

Fuente: Elaboración propia en base a Servicio de Rentas Internas de Ecuador.

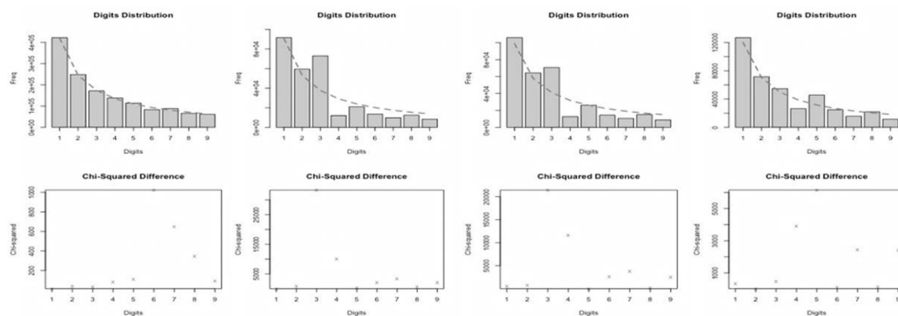
3.1.3. Variables formulario 107. Personas naturales en relación de dependencia

Cuando se analiza las variables de la declaración del Impuesto a la Renta de las personas en relación de dependencia, exceptuando la variable Impuesto a la Renta empleador, el primer dígito de las demás variables no siguen la distribución de la Ley de Benford⁵. Las mayores diferencias se dan en Total Ingresos, aporte al Instituto Ecuatoriano de Seguridad Social (IESS) y los valores de las deducciones de gastos personales como sucede en las declaraciones del formulario 102. Considerando solo las variables relacionadas con los gastos personales, la variable Deducción por Salud es la que más se aproxima visualmente a la distribución. De las cinco variables permitidas para la deducción de gastos personales, es la única que no presenta un tope. Esto evidencia que la existencia de un tope máximo de declaración como existen en las otras deducciones (alimentación, vivienda, vestimenta y educación) motiva a que las declaraciones de las personas se agrupen en montos muy cercanos al tope.

Se incluyeron en la Figura 5 las variables de los décimoterce y décimocuarto sueldos. El décimocuarto sueldo es un salario mínimo que reciben todas las personas en base a sus meses trabajados (USD 340 en el año 2014 para Ecuador)⁶. Por su naturaleza es evidente que la acumulación se dé en el dígito 3. De esta forma se deja asentada la importancia de la selección de las variables, ya que la herramienta es muy oportuna pero debe ir de la mano del conocimiento del impuesto y variables que se quieran analizar.

Figura 5. Aplicación Ley de Benford formulario 107. Personas Naturales en relación de dependencia. Ecuador, año fiscal 2014

Total Ingresos** Aporte IESS** Base Imponible** Valor retenido**

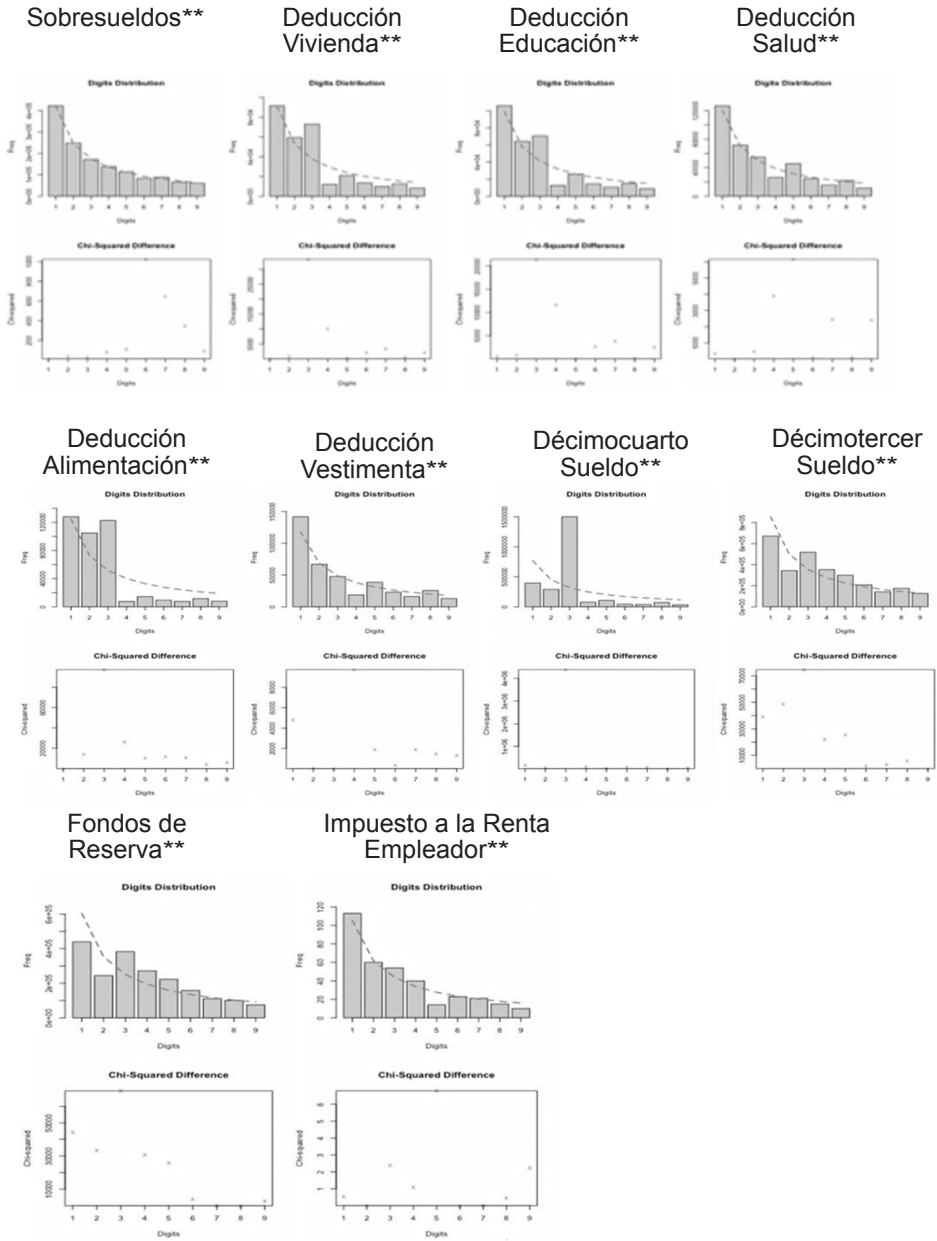


Continúa

5 A un nivel de significancia del 1%.

6 En Ecuador, el décimoterce sueldo se paga en el mes de diciembre y el décimocuarto sueldo en el mes de agosto con el objetivo de suplir los gastos por inicio de las actividades escolares.

Figura 5 (continuación).



 Datos  Distribución Teórica

Nota: * Nivel de significancia 5%.

** Nivel de significancia 1%.

Fuente: Elaboración propia en base a Servicio de Rentas Internas de Ecuador.

4. Conclusiones

El presente estudio brinda una evidencia clara que la Ley de Benford puede ser empleada de manera confiable para un primer análisis de evasión. Al utilizar las declaraciones de Impuesto a la Renta tanto de personas naturales como de sociedades en Ecuador, se comprueba que la distribución teórica del primer dígito de los valores declarados es una herramienta que representa la realidad de las declaraciones.

La aplicación de la Ley de Benford da indicios que en las declaraciones de personas naturales existen datos anómalos. En la información que reporta el contribuyente de forma personal se perciben más anomalías frente a las variables que están sujetas a una estructura contable. Hay que considerar que para aplicar la ley es importante tener un vector de información con datos completos y abundantes. Adicionalmente, la Ley de Benford es muy útil para detectar anomalías en la distribución de los datos, pero se debe analizar más a profundidad para asegurar que esas anomalías se traducen en fraude.

Para poder desarrollar una herramienta de uso común es importante entender el giro del negocio y las necesidades específicas en cada caso. Existen temas de depuración de la base de datos que debe ser considerados antes de realizar un análisis y la presencia abundante de valores nulos da luces de que existen casilleros que no son muy utilizados. A eso hay que sumarle aspectos normativos como el valor inicial para constituirse como sociedad o los topes máximos de variables como los décimos sueldos o las deducciones, las cuales generan de por sí un incumplimiento de la distribución teórica que no representa directamente anomalías de los datos.

Este análisis se puede expandir a bases de datos más voluminosas u otras variables relevantes de acuerdo al criterio del experto en el impuesto analizado. Se pueden considerar casos más puntuales como: declaraciones sustitutivas, con un mismo contador, consistencia entre años, entre otras. La aplicación de esta ley permite analizar de forma rápida, oportuna, confiable y económica grandes cantidades de datos, lo cual es muy útil en temas que no son sólo tributarios.

Referencias bibliográficas

- Allingham, M. G. y Sandmo, A. (1972). Income tax evasion: A theoretical analysis. *Journal of public economics*, 1 (3), 323–338.
- Benford, F. (1938). The Law of Anomalous Numbers. *American Philosophical Society*, 78(4), 551–572.
- Berger, A. y Hill, T. P. (2011). A basic theory of Benford's law. *Probability Surveys*, 8, 1–126.
- Bolton, R. J. y Hand, D. J. (2002). Statistical Fraud Detection: A Review. *Statistical Science*, 17(3), 235–255.
- Centro de Estudios Fiscales (2017). Presión tributaria del gobierno central. Recuperado de <https://cef.sri.gov.ec/mod/page/view.php?id=11057>

- Cullis, J. G. y Lewis, A. (1997). Why people pay taxes: From a conventional economic model to a model of social convention. *Journal of economic psychology*, 18 (2), 305–321.
- Fewster, R. M. (2009). A simple explanation of Benford's law. *The American Statistician*, 63(1), 26–32.
- Gómez Sabaini, J. C., Jiménez, J. P. y Podestá, A. (2010). Tributación, evasión y equidad en América Latina y el Caribe. En J.P. Jiménez, J. C. Gómez Sabaini y A. Podestá (Comps.), *Evasión y equidad en América Latina* (11-67). Santiago de Chile, Chile: CEPAL.
- Hill, T. P. (1995a). Base-invariance implies Benford's law. *American Mathematical Society*, 123(3), 887–895.
- Hill, T. P. (1995b). A Statistical Derivation of the Significant-Digit Law. *Statistical Science*, 10(4), 354–363.
- Joenssen, D. W. (2015). Benford Test: Statistical Test for Evaluating Conformity to Benford's Law, R package version 1.2.0. Recuperado de <https://cran.r-project.org/web/packages/benford.analysis/index.html>
- Latinobarómetro (2016). *Informe Latinobarómetro 2016*. Recuperado de <http://www.latinobarometro.org/latContents.jsp>
- Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers. *American Journal of Mathematics*, 4(1), 39–40.
- Nigrini, M. J. (1996). A taxpayer compliance application of Benford's Law. *The Journal of the American Taxation Association*, 18(1), 72-91.
- Pinkham, R. S. (1961). On the distribution of first significant digits. *The Annals of Mathematical Statistics*, 32(4), 1223–1230.
- Prats, J. y Macías-Aymar, I. (2008). El pacto fiscal en los Andes. Desafección, deslegitimidad e ineficiencia tributaria en Bolivia, Ecuador y Perú. *Revista Fiscalidad*, 2, 11–58.
- R Development Core Team. (2011). *R: A language and environment for statistical computing*. Recuperado de <http://www.R-project.org>.
- Serra, P. (2000). Evasión tributaria: ¿cómo abordarla? *Revista de Estudios Públicos*, (80), 193-228.
- Servicio de Rentas Internas (2017a). *Estadísticas de recaudación*. Recuperado de <http://www.sri.gob.ec/web/guest/estadisticas-generales-de-recaudacion>
- Servicio de Rentas Internas (2017b). *Informe de gestión enero-diciembre 2016*. Recuperado de <http://www.sri.gob.ec>
- Servicio de Rentas Internas (2018). *Empresas fantasmas*. Recuperado de <http://www.sri.gob.ec/web/guest/empresas-inexistentes>

Tsagbey, S., De Carvalho, M. y Page, G. L. (2017). All data are wrong, but some are useful? advocating the need for data auditing. *The American Statistician*, 71(3), 231-235.

Viana, M. y Oliviera, K. (2016). *Foundations of ergodic theory*. Cambridge, EEUU: Cambridge University Press.

Villegas, H., Arguello Velez, G. y Spila García, R. (1993). La evasión fiscal en la argentina. *Revista Derecho Fiscal*.

5. Anexos:

Anexo 1. Test chi cuadrado de las variables analizadas

Formulario	Variable	Numero observaciones	3 primeros dígitos con mayores diferencias				Test Chi-cuadrado			
			Dígito	Diferencia absoluta	Dígito	Diferencia absoluta	Estadístico	Valor P		
101-Sociedades	Total Activos**	92.517	8	6.203,5	1	2.750,4	3	1.720,0	9.201,2	0,00
	Total Pasivos*	71.692	1	413,4	4	310,3	8	143,8	34.292,0	0,04
	Total Patrimonio**	84.062	8	7.156,0	1	2.111,2	3	2.055,6	13.151,0	0,00
	Total Ingresos	67.078	2	97,0	5	93,5	1	60,2	40.955,0	0,85
	Total Costos	38.755	1	238,4	5	129,3	3	92,0	14.667,0	0,07
	Total Gastos	67.932	5	162,1	3	90,3	6	61,2	78.767,0	0,45
	Utilidad del Ejercicio*	49.682	1	342,9	2	105,7	7	100,2	19.102,0	0,01
	Impuesto a la Renta Causado*	46.421	3	160,2	2	108,7	5	100,7	12.448,0	0,03
	Total Activos**	49.171	4	206,2	3	184,4	8	151,8	28.594,0	0,00
Total Pasivos	44.144	3	99,7	5	83,6	8	65,9	92.381,0	0,32	
Total Patrimonio**	46.854	2	233,6	5	226,0	1	132,5	28.049,0	0,00	
Total Ingresos**	52.236	1	2.108,4	4	716,2	5	588,1	600,3	0,00	
Total Costos**	46.382	1	1.198,6	3	596,9	4	489,9	252,3	0,00	
Total Gastos	43.077	3	55,0	4	46,6	8	43,5	24.588,0	0,96	
Utilidad del Ejercicio**	40.589	2	452,6	1	315,5	7	286,8	142,1	0,00	
Impuesto a la Renta Causado**	308.425	2	1.734,0	1	1.695,2	3	1.543,8	227,3	0,00	
Total Gastos Personales**	359.176	2	26.285,8	5	10.936,0	4	10.428,3	21.899,0	0,00	
Ingresos en Relación de dependencia**	320.243	1	24.409,3	5	10.947,2	2	8.514,0	23.444,0	0,00	
Deducciones en Relación de dependencia**	297.901	1	18.416,9	6	6.812,5	7	6.583,9	13.709,0	0,00	
Otros Ingresos**	494.517	1	5.242,6	2	2.644,1	3	2.018,3	372,6	0,00	
Total Ingresos**	3.340.691	4	418.743,4	1	198.204,1	2	164.909,9	648.000,0	0,00	
Aporte IESS**	3.262.174	1	217.779,2	2	155.068,3	3	139.549,1	213.050,0	0,00	
Base Imponible**	3.458.185	1	240.182,4	2	200.133,9	4	94.289,3	234.980,0	0,00	
Valor retenido**	459.009	5	3.849,1	7	3.560,8	3	2.502,0	1.297,3	0,00	
Sobresueldos**	1.392.503	6	9.764,6	7	7.258,0	8	4.996,0	2.406,1	0,00	
Deducción Vivienda**	301.463	3	35.434,6	4	17.181,8	7	7.653,4	52.876,0	0,00	
Deducción Educación**	328.342	3	29.682,4	4	19.240,6	7	8.478,2	43.493,0	0,00	
Deducción Salud**	399.421	5	13.945,4	4	12.336,9	7	7.515,2	16.057,0	0,00	
Deducción Alimentación**	414.000	3	70.919,4	4	32.544,8	2	31.985,2	181.960,0	0,00	
Deducción Vestimenta**	391.336	1	23.833,1	4	19.220,4	5	7.704,5	21.684,0	0,00	
1to sueldo**	2.549.967	3	1.181.912,3	1	371.883,5	4	170.568,3	5.053.500,0	0,00	
13er sueldo**	2.835.557	1	182.790,7	3	162.546,1	2	155.990,8	221.530,0	0,00	
Fondos de reserva**	2.005.906	1	164.041,9	3	131.844,6	2	108.604,5	211.780,0	0,00	
Impuesto a la renta empleador	350	5	13,7	3	10,3	1	7,6	13.648,0	0,09	

Notas: ** Rechazo H0 al nivel de significancia 1%

*Rechazo de H0 al nivel de significancia 5%.

Fuente: Elaboración propia en base a Servicio de Rentas Internas de Ecuador.

Anexo 2. Descripción de la demostración empleando teoría ergódica

Puntualmente el teorema de convergencia de Birkhoff es el que permite esta demostración.

Intuitivamente si se tiene una serie de números del 1 al 9, parecería más fácil decir que el 1 tiene mayor probabilidad de aparecer que el 9. Esto no es tan evidente si se tiene otro tipo de series, como por ejemplo una sucesión de números generada al iterar potencias de 2, es decir: $\{1, 2, 4, 8, 16, \dots, 2^n, \dots\}$. Sin embargo, el comportamiento de los primeros dígitos es el esperado por la Ley de Benford como se muestra a continuación. La frecuencia del primer dígito (d) de esta sucesión está expresada por:

$$\lim_{(n \rightarrow \infty)} \frac{1}{n} |\{j=0, 1, \dots, n-1 \mid \exists r \in \mathbb{N} \text{ tal que } d10^r \leq 2^j < d10^{r+1}\}|$$

la desigualdad $d10^r \leq 2^j < d10^{r+1}$ indica que el primer dígito de la iteración j de $\{1, 2, 4, 8, 16, \dots, 2^n, \dots\}$ es exactamente d . El valor de este límite converge a:

$$\int_{\log_{10} d}^{\log_{10}(d+1)} 1 dx = \log_{10}(d+1) - \log_{10}(d) = \log_{10} \left(1 + \frac{1}{d}\right)$$

De esta forma se evidencia el cumplimiento de la Ley de Benford en otro tipo de sucesiones y como una aplicación del teorema de Birkhoff.