



La variable oculta en la suerte del apostador de loterías The hidden variable in the luck of the lottery gambler

Casanova-Leal Hugo

Escuela Venezolana de Planificación, Caracas

Correo: casanovahu@yahoo.es

Resumen

De todas las suposiciones populares que tienen los apostadores sobre los sorteos de las loterías de cupones, distintas de las metafísicas, como el número de la suerte, las creencias sobre el papel que juega el peso de las bolillas en el sorteo están muy arraigadas. Se cree que salen las bolillas más livianas o pesadas según sea el tipo de aleatorizador; se cree que el peso de las bolillas es la variable que oculta la caja negra del sistema. Este artículo analiza este hecho a través de pruebas estadísticas sobre información real, los sorteos de una lotería de cupones en el año 2000, y descarta tales creencias como atribuciones hechas sobre esperanzas frustradas. Sin embargo, la distribución de los pesos ofrece sesgos que, no logrando incidir en los sorteos, pudiera, a primera vista, hacer verosímil la creencia del apostador. Por lo tanto, se realiza un análisis estadístico en dos niveles, peso estático de las bolillas e influencia del aleatorizador.

Descriptores: Aleatorizador, loterías, azar, análisis exploratorio, análisis descriptivo.

Abstract

Of all the popular beliefs, other than the metaphysical beliefs of bettors on lottery draws, such as lucky numbers, beliefs about the role played by the weight of the balls in the draw are very ingrained. The lighter or heavier balls are believed to be depending on the type of randomizer; It is believed that the weight of the balls is the variable that hides the black box of the system. This article analyzes this fact through statistical tests on real information, lottery draws of coupons in the year 2000 and discards such beliefs as attributions made on frustrated hopes. However, the distribution of weights offers biases that, failing to affect draws, could, at first glance, make credible the belief of the bettor. Therefore, a statistical analysis was performed in two levels, static weight of the balls and influence of the randomizer

Keywords: Randomization, lotteries, chance, exploratory analysis, descriptive analysis.

INTRODUCCIÓN

La Teoría de Juegos expresa una forma corriente de la vida donde maximizar el beneficio usando la astucia, el pensamiento lateral y el conocimiento acerca del otro son herramientas de éxito. En los juegos en donde la cooperación es esencial, llegar a acuerdos o negociar es la solución menos costosa para las partes. Esto justifica la existencia de la Ley y por lo tanto del arbitraje. De allí que la duda y la desconfianza estén siempre presentes en nuestros tratos. El juego es expresión del mundo hobbesiano. Por ello, la trampa, la picardía, el engaño, la astucia, etcétera, son, en muchas ocasiones, socialmente aceptadas o toleradas; la ya famosa *mano de Dios* de Maradona es ejemplo de esto; driblar el balón de básquet con habilidad para confundir al oponente que quiere arrebatarlo, hacer amagues engañosos con el balón de fútbol para superar al contrario, engañar al arquero ante un penalti, robar una base en beisbol, etcétera, son formas sanas de engaño, pues el resultado de la acción es de muy bajo impacto y genera emocionalidad sobre la virtud de la astucia. De la misma manera, dudar sobre la inexistencia de cartas marcadas, dados cargados, dobles del dominó limados, lados duplicados en una moneda, posturas corporales o señas en los juegos de cartas, etcétera, es corriente.

Pero la desconfianza no solo proviene de una actitud vital de protección, sino tiene referentes empíricos que la refuerzan; así, han existido episodios de estafas asociados con ataques a las loterías o aprovechamiento de sus sorteos para cometer delitos, como el fraude ocurrido en el año 1942 con la Lotería Nacional de Argentina (Pigna, 2007); los niños cantores (los que cantaban el guarismo sorteado) sacaron de sus mangas la bolilla “ganadora” n° 31025 en el sorteo del 24 de julio de 1942.

De este modo, en los juegos de loterías, pese a la confianza que se tenga en las casas administradoras del juego y en los controles de los gobiernos, el apostador, centra su atención en el juego y en el aleatorizador sobre los que genera algunas dudas de cómo podría estar trucado. De esta manera esperaríamos que piense que las bolillas de los sorteos de las loterías de cupones tengan pesos y tamaños muy distintos. Tales dudas son legítimas y aun cuando sean secundarias respecto a otras creencias más importantes que pudieran ser verosímiles si los controles de calidad se descuidan. En esta investigación, que es continuación de otra sobre la influencia de la zona de venta de boletos en los sorteos (Fermentum 73 vol. 25 Mayo-Agosto 2015), atacamos esta creencia de influencia del peso de las bolillas en los sorteos sobre información de 35 eventos de una lotería

de cupones en el año 2000. La descripción detallada de tal mecanismo fue detallada en esa ocasión.

METODOLOGÍA Y CONSTRUCCIÓN ESTADÍSTICA DE LA CONJETURA

Cubrir este objetivo implica considerar al mecanismo como un sistema cuyas *entradas* son las bolillas, el *proceso* está constituido por el mecanismo aleatorizador, finalmente, las *salidas* por las bolillas sorteadas (S) y las que se quedan en el bolillero (NS). Este sistema debe tener algunas características básicas; la caja negra o proceso no puede abrirse (conocerse), pues debe construirse para que sea impredecible y no agregue valor a los eventos, esto es, no privilegie algunas características físicas de las bolillas de salida. Lo único que debe hacer es permitir la salida de algunas de ellas (15 de 25) sin que sea posible responder *por qué salen unas bolillas y no otras*. Esta *caja negra* puede tomar múltiples formas y mecanismos internos de funcionamiento. Tal mecanismo debe anular toda racionalidad posible, esto es, debe estar hecho para “embrutecer” al espectador de modo tal que al minimizar la racionalidad o explicación causal se maximice el pensamiento mágico o meta-científico. Diríamos que este mecanismo debe “engañar” a toda variable física concomitante e interviniente para invocar al azar; el espectador solo debe caer en cuenta de la bolilla sorteada, bajo asombro, cuando esta salga del mecanismo. Entonces diremos, por razón insuficiente, que el *azar* ha actuado en el dispositivo. Una segunda característica es que, dado que la caja negra no puede abrirse (hacerse gris o blanca), las bolillas de entrada y las de salida deben ser indistinguibles; no es posible responder la pregunta *¿qué diferencia existe entre las bolillas de entrada y las de salida?* Tal vez la única respuesta sería *ninguna, porque todas tienen la misma probabilidad de salir*; la probabilidad las unificaría frente a la pequeña diferencia de peso, pero esta respuesta induce una cadena de preguntas más difíciles como por ejemplo *¿cómo se garantiza esto más allá del cálculo teórico del marco muestral?*; *construido un aleatorizador cualquiera ¿cómo se sabe que la bolilla que salió sorteada lo hizo con equiprobabilidad?* *¿La dinámica interna de la caja negra unifica los pesos, esto es, las hace invariantes frente a la diferencias de pesos?* Veamos esto con un poco de detalle; es claro que las bolillas van a tener pesos diferentes, pero la dinámica del aleatorizador debe tener el mismo efecto que la ola en la playa sobre los bañistas. El oleaje tiene una fuerza diferencial, si es muy fuerte tumbará a las personas livianas hasta hacerlas caer o golpearlas contra la arena, levantará a las medianas de peso y pasará sobre las pesadas sin mayor efecto, si el oleaje es muy

débil pasará por todos los bañistas sin mayor efecto y si son olas moderadas levantará a todas las personas equitativamente, todas sentirán un efecto similar. En otras palabras la equiprobabilidad, que es un cálculo teórico, ha de materializarse a través del mecanismo de la caja negra, así como se materializa en la simetría y homogeneidad del dado de seis lados. Pero esta dinámica es desconocida por el apostador el que a pesar de la confianza que tenga en las instituciones y empresa de lotería, genera una duda razonable sobre los aleatorizadores aun cuando la caja negra se construya objetivamente y bajo controles de calidad.

De esta manera, el mecanismo de la caja negra debe minimizar sus efectos sobre las características físicas de las bolillas y, además, estas deben maximizar su homogeneidad, ser indistinguibles sobre la actuación de la caja negra. Y confiando en que esto es así, nos permite descartar estos elementos para centrarnos en la duda del apostador. Según su creencia, el peso de las bolillas tendría una influencia en el sorteo, esto implica, en términos más técnicos, que las bolillas de salida constituirían una *partición* de las de entrada (a riesgo del abuso de la palabra, pues toda partición es excluyente, sin embargo, en el imaginario del apostador puede subsistir la creencia según la cual hay un grupo de bolillas que no son sorteadas por ser pesadas) o que el aleatorizador es sesgado. De tal manera que la conjetura del apostador implicaría tres aspectos:

1. Las bolillas de salida son una partición de las de entrada.
2. El mecanismo de la caja negra induciría un sesgo de selección.
3. La partición de las bolillas pudiera beneficiar a la serie ganadora.

El primer aspecto se refiere a la homogeneidad de las bolillas, por lo tanto, es básicamente un estudio sobre la varianza de su peso, que incluye al segundo aspecto,

pues si existiese la partición esta debería explicar el sesgo del mecanismo; finalmente para el tercero se analizaría la pertinencia de una correlación entre los pesos y los guarismos marcados. Igualmente haremos un estudio descriptivo-exploratorio con el objeto de familiarizarnos con las estructuras de los pesos de las bolillas.

CORRELACIÓN PESO-GUARISMO

El peso es una variable métrica (nivel de razón), en cambio, los guarismos (considerados como variable) no son métricos, son nominales, meros signos carentes de contenido factual. La Epistemología de la Ciencia divide los conceptos científicos en dos grandes grupos, los conceptos no métricos y los métricos. Los primeros son conceptos clasificatorios y comparativos que se conocen como *cualitativos*; por otra parte, los conceptos métricos se refieren a magnitudes y se llaman propiamente *cuantitativos*. Esta clasificación procede, según Mosterín (1987), de Hempel (1952), sin embargo, no hay acuerdo unánime sobre esta ni sobre la de las escalas de medición que se desprenden de estos conceptos. Por brevedad y pertinencia usaremos lo más convencional de ellas, procurando no perder generalidad. El problema central es la forma de definir el *isomorfismo*, esto es, la correspondencia entre una estructura numérica y el dominio de objetos que se adecuen a esta estructura. La medición bajo el criterio de verdad por correspondencia es un problema de isomorfismos. Por lo tanto, partiremos de estos posibles "usos" de los números para desarrollar nuestro tema.

Los números pueden tener por lo menos tres usos distintos:

- 1) como rótulos o marcas de identificación;
- 2) como signos que indican la posición de un grado de una cualidad en una serie de grados y
- 3) como signos que indican las relaciones cuantitativas entre cualidades. A veces, desempeñan las tres funciones al mismo tiempo, (Cohen & Nagel, 2000).

Tabla 1. Pesos y números asignados a las bolillas de los dos primeros sorteos de la lotería evaluada en 2000

		Bolillas 1 a 12												
Núm. asignado		1	2	3	4	5	6	7	8	9	10	11	12	
Sorteo1		2.47	2.37	2.44	2.35	2.39	2.43	2.39	2.21	2.4	2.33	2.35	2.39	
Sorteo2		2.45	2.44	2.45	2.42	2.47	2.47	2.48	2.48	2.43	2.43	2.39	2.43	
		Bolillas 13 a 25												
Núm. asignado		13	14	15	16	17	18	19	20	21	22	23	24	25
Sorteo1		2.18	2.42	2.43	2.41	2.33	2.25	2.38	2.47	2.4	2.33	2.41	2.41	2.39
Sorteo2		2.47	2.47	2.47	2.42	2.43	2.5	2.41	2.42	2.41	2.39	2.41	2.41	2.45

Fuente: libro de sorteos de la lotería evaluada. Año 2000

En la aritmética se conocen como números nominales, ordinales y cardinales. Ahora nos preguntamos ¿cómo se asignan los guarismos a las bolillas? Bastaría con observar un par de series para ver qué modalidad se usa. La Tabla 1 muestra la asignación de los números a los dos primeros sorteos del año 2000.

Se puede ver que no hay relación entre los pesos y la enumeración, por lo que la forma de asignar los números es la primera modalidad. Entonces: ¿se podría obtener un índice de correlación guarismo-peso de la bolilla? Claramente es absurdo, el guarismo marcado no es un isomorfismo de alguna característica de la bolilla. Entonces se dice que esta clase de enumeración no cumple con los requisitos de adecuación formal o material (Mosterín, 1987, pág. 14) para variables clasificatorias. Esto es, no refieren a ninguna partición de objetos reales y no subsumen bajo estos guarismos ningún dominio de objetos. Diríamos que son una simple lista de códigos. Estos guarismos (no se les llama *números* para evitar la asociación del signo con algún sistema numérico) son “variables” nominales, pero por debajo del nivel clasificatorio, son meras marcas con mínimo significado solo se refieren al nombre, vaciado de contenido fáctico. Sería como intentar hallar sentido a la relación de nombres propios de personas con su peso. Tampoco podríamos decir que es una *correlación espuria*, pues esta existe en variables métricas solo que no se podría explicar como la correlación entre la postura de huevos de las gallinas y la muerte por accidentes de tránsito, no hay una teoría que sostuviera esto.

ANÁLISIS DESCRIPTIVO/EXPLORATORIO DE LOS PESOS DE LAS BOLILLAS

La estadística descriptiva clásica consiste en la realización de pruebas *analíticas* de los estimadores o estadísticos muestrales con el objeto de hipotetizar la estructura de la población. Recientemente con el desarrollo de la informática el análisis descriptivo evoluciona a análisis exploratorio que pretende profundizar en la información, mostrarla a plenitud y para ello recurre a la *visualización de datos*, nueva corriente que está en pleno desarrollo y que rebasa los linderos de la estadística. Palmer (1999) adelanta un lema para el análisis exploratorio: “El análisis descriptivo debería empezar con los datos, no con los resúmenes de datos”. Lo compartimos, pero debería corregirse a: el análisis descriptivo halla un orden en el caos de datos para luego explorar la estructura hallada. Esta tendencia nueva la inicia Tukey, Tufte y otros, en los años sesentas con los diagramas de caja y tallos y hojas, entre los más usados. La visualización de datos, de preferencia gráfica, se ve

apoyada por la informática, sin embargo, ambos tipos de análisis son complementarios o secuenciales, pues la escuela exploratoria, al ser más *analógica*, es más globalizante haciendo perder detalles que la clásica escuela descriptiva, al ser más *analítica*, podría descubrir; pero lo analítico al ser reduccionista podría hacer perder la visión total que elementos intuitivos del investigador podrían observar. Es aquella vieja relación entre el bosque y el árbol.

El análisis descriptivo tiene como objetivo construir analíticamente la estructura de la población a través de tres dimensiones estadísticas; el *centro* de la distribución, las *partes* y la *variación* de los datos respecto del centro, ello a través de diversos estimadores de localización, forma, posición y variabilidad. No bastaría con ver uno o dos estadísticos de localización o centrales, pues estos nos conducen a la estructura de las partes y estas a la variabilidad de los datos. Así, los estadísticos centrales nos ofrecen distintas perspectivas del significado de los datos; la media nos habla de la carga ideal de estos, del valor físico donde la mayoría se ajusta o se acerca; por su parte, la mediana nos habla del valor que divide la distribución en dos partes iguales y entre ambas obtenemos la relación peso-posición indicándonos la direccionalidad de la cola de la distribución; finalmente, la moda nos habla de la repetición de los valores que para efectos de las distribuciones de probabilidad sería buena su coincidencia con la media y única. De esta manera se tiene el estudio de las partes en forma de cuantiles, percentiles o medidas de posición, que hacen el mismo trabajo que la mediana, pero dividiendo en dos o más segmentos iguales la distribución. Su objetivo, como se verá en el estudio de cajas y bigotes, es evaluar la densidad de los datos en cada segmento (especie de escaneo de datos), esto es, responde por ejemplo, preguntas del tipo: ¿hay datos muy agrupados en los segmentos o se hallan más alongados? Determinar la existencia de grupos intermedios y su forma de agrupamiento es cualitativamente importante. Finalmente, la variabilidad es el tema central de los estudios más desarrollados como el ANOVA, etcétera. Por su parte, el análisis exploratorio de datos, EDA, por sus siglas en inglés, explora gráficamente la estructura construida. Aquí se usarán dos de los clásicos, el diagrama de tallos y hojas (DTH) y el diagrama de cajas y bigotes (DCB) desarrollados por Tukey.

EVALUACIÓN DEL CENTRO DE LOS DATOS

La Tabla 2 muestra los estadísticos descriptivos. Un total de 575 pesos, donde el menor es 2 y el máximo 2.62 subsisten en un espacio de cobertura 0.62. Se evaluarán los

Tabla 2. Estadísticos descriptivos de los pesos de 575 bolillas usadas en 35 sorteos de la lotería evaluada usando SPSS 19

Estadísticos descriptivos	gramos		
Media	2.4295	Mínimo	2.00
IC (Media,95%) Lim Inf	2.4230	Máximo	2.62
IC (Media, 95%) Lim Sup	2.4360	Rango	0,62
Media recortada al 5%	2.4354	Amplitud intercuartil	0,0600
Mediana	2.4500	Asimetría	-1.539
Moda	2.4700	Curtosis	3.744
Varianza	0.006	Total	575
Desv. Típ.	0.07882		

Fuente: Libro de juegos de la lotería evaluada en el año 2000. Procesados mediante SPSS 19

valores centrales (media, mediana y moda), que dan estabilidad a la estructura bajo la premisa según la cual en la Media Aritmética al ser, topológicamente un punto de concentración o de densidad la variabilidad sería menor que en los extremos. Un indicador de esto es que los valores centrales se muestran iguales. Esta es la primera referencia, buscar simetría estadística. Así, al observar la media de los pesos como punto de equilibrio se produce ahí, necesariamente, un efecto balanza. Sin embargo, esto no garantiza la simetría, pues al ver a la mediana notamos que es más alta. En la misma Tabla 2 vemos la media recortada a 5%; esta, eliminando la dispersión extrema de 5% de valores en cada lado debería acercarse más a la media a la mediana, pero se nota que este ajuste es muy precario. Veamos qué efecto tiene esto.

La mediana reparte los valores equitativamente en su entorno. Si la media es mayor que la mediana (media > mediana) nos indicaría que unos valores altos atraerían a la media hacia la derecha de la distribución. Esto se debería a que la media, al depender de una suma es sensible a los valores altos. Pero esto no es lo que sucede, la media está por debajo de la mediana y la única explicación es que una cierta cantidad de valores pequeños acumulan suficiente carga para empujar la media hacia la izquierda, indicando la existencia de una cola en esa dirección.

Por otra parte, la moda se halla hacia la derecha, en la relación media-mediana-moda indicando que la mayor elevación de la curva al estar por encima de la mediana y la media como valor más bajo de los tres delata un sesgo negativo. Asimismo, los estrechos límites de confianza para la media y el reducido rango de la distribución dan muestra de la gran densidad de datos alrededor de este valor central. Esto es, en un Rango general de 0.62 se concentra 100% de valores y en un reducido Rango Intercuartil (0.06) se halla 50% de los valores centrales. Esto indica la existencia de muchos valores similares. Sin embargo, las diferencias entre los extremos (Max

y Mín) respecto de la mediana (med-min = 0.45; Max-med = 0.17; ó 2.65 veces más largo el lado izquierdo) nos muestra el largo de la cola a la izquierda. De esta manera, se tiene una primera imagen intuitiva de la estructura de la distribución y una primera explicación, la cobertura de la distribución es muy reducida, el centro se caracteriza por una alta densidad de valores alrededor de la media y existe una cola importante a la izquierda. A primera vista este hecho daría soporte empírico a las creencias de los apostadores, pero ¿es justificada?

Ahora exploraremos las partes de la distribución a través de un diagrama de Cajas y Bigotes (Box Plot) en adelante DCB y uno de Tallos y Hojas (Stem and Leaf plot), en adelante DTH. Ambos análisis exploratorios, que al ser gráficos, enriquece la intuición de lograr visualizar las diferencias en el análisis descriptivo. Esto persigue Tukey con el análisis exploratorio. A continuación se explican brevemente estos métodos en el siguiente punto.

EVALUACIÓN DE LAS PARTES

Estas las podemos evaluar con los centiles (o percentiles), en donde se agrupan los datos en clúster porcentualmente iguales (cuartiles, quintiles, deciles, etcétera) para ver el comportamiento de estos; el DCB usa cuatro clúster o segmentos iguales en cantidad, pero seguramente difieren en su modo de organización, esta información es vital. El DTH muestra los valores extremos (Li y Ls) según el criterio mostrado en las ecuaciones siguientes donde se utiliza también el DCB y se exponen los valores a detalle con mínima reducción. Este diagrama es de suma utilidad, pues muestra gráficamente la huella empírica de la distribución.

Por su parte, el DCB es un análisis centrado en los cuartiles de los datos; este puede verse como un escaneo de datos que evalúa y describe sección por sección la masa de datos; divide la masa de valores en cuatro secciones, desde el valor mínimo hasta el cuartil 25 o C_1 ,

desde el cuartil 25 hasta el 50 o C_2 , desde este hasta 75% o C_3 y desde esta hasta el máximo. Las siguientes relaciones establecen los límites superior (L_s) e inferior (L_i) para los valores extremos.

$$L_i = C_1 - a(C_3 - C_1) \quad (1)$$

$$L_s = C_3 + a(C_3 - C_1) \quad \text{con } a = \{1,5 \text{ o } 3\}$$

Donde L_i y L_s son, respectivamente límites inferior y superior de los extremos; $C_3 - C_1 = RI$, rango intercuartil. Con $a = 1,5RI$ extremos o $a = 3RI$ muy extremos

Nótese que esta división en clúster de igual tamaño (25%) permite observar la densidad de los valores en la caja y en los extremos, pues los límites de los intervalos los determina el porcentaje de valores que es fijo; en el histograma los límites son preestablecidos. Evidentemente la decisión de cuáles son valores extremos es arbitraria. El análisis corriente indica que se toman aquellos valores que, desde RI disten 1.5 y 3 veces su tamaño, pero pueden ser otros. La Tabla 3 muestra los estadísticos empleados por DCB y los límites ya calculados con las ecuaciones 1.

Así, el análisis comienza con la Tabla 3 que muestra los estadísticos del DCB. Este análisis toma como referencia la mediana, ya que es más robusta que la media, es decir, resiste más el incremento de valores extremos sin mucho desplazamiento, cosa que no pasa con la media. De la Tabla 3 se deben observar las diferencias entre los intervalos, L_i -Min; C_1 - L_i ; C_2 - C_1 ; C_3 - C_2 ; L_s - C_3 ; Max - L_s .

Tabla 3. Estadísticos del diagrama de Cajas y Bigotes

Estadístico	Signo	Peso
Mínimo	Min	2.00
Límite inferior	L_i	2.32
Primer cuartil	C_1	2.41
Segundo cuartil	C_2	2.45
Tercer cuartil	C_3	2.47
Límite superior	L_s	2.56
Máximo	Max	2.62

Fuente: Libro de juegos de la lotería evaluada en el año 2000. Procesados mediante SPSS 19

Los intervalos C_1 -Min; C_2 - C_1 ; C_3 - C_2 y Max - C_3 , son respectivamente los cuatro cuartiles y contienen 25% de datos cada uno. Obviamente mientras más pequeños los intervalos, mayor será la densidad de valores. Esta precisión analítica puede verse visualizada en el DCB de la Figura 1. Se puede observar la larga cola izquierda (de valores livianos) formada por valores extremos de 1.5RI (círculos) y muy extremos 3RI (asteriscos) justificando, a primera vista, la conjetura del apostador de existencia de un grupo de pesos livianos, cola algo ex-

tendida o una cola de bajo grosor hacia la derecha. Igualmente puede verse la elevada concentración de valores en RI, particularmente C_2 - C_1 menos denso que C_3 - C_2 más denso, en esa región se hallan 50% de los pesos. Esto corrobora lo estudiado analíticamente en el análisis descriptivo. El DTH (Figura 2) muestra los detalles de esta distribución.

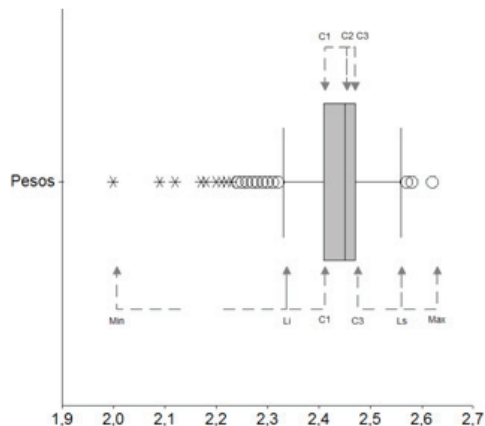


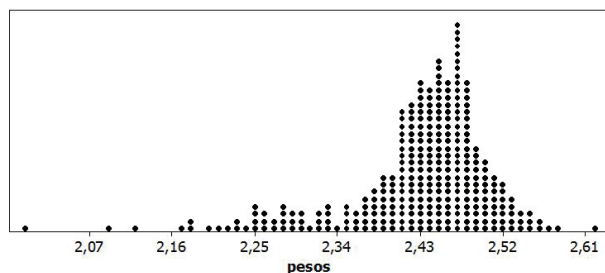
Figura 1. Diagrama de Cajas y Bigotes de 575 bolillas de los sorteos de la lotería evaluada en el año 2000

Fuente: Libro de juegos de la lotería evaluada en el año 2000. Procesados mediante SPSS 19

Asterisco: valores ± 3 RI. Círculo: Valores $\pm 1,5$ RI

La Figura 2 muestra el DTH. Este muestra los tallos 23, 24 y 25; sobre todo que el tallo modal es 24 o clúster de datos de 2.40 a 2.49 gr (con 380 valores). Igualmente los 60 valores son extremos inferiores, obtenidos según las ecuaciones 1. Igualmente vemos que el RI analizado antes, muestra un espacio muy reducido donde se halla 50% de los valores; esto puede verse en el DTH, desde el C_1 (2,41;

Diagrama de puntos



Each symbol represents up to 2 observations.

Figura 2. Diagramas de Puntos y Tallos y Hojas de 575 bolillas de los sorteos de la lotería evaluada en el año 2000

Fuente: Libro de juegos de la lotería evaluada en el año 2000

5^o tallo) hasta el C₃ (2.47; 8^o tallo) se hallan valores muy parecidos. El análisis exploratorio sigue al descriptivo en una nueva etapa de profundización de la información, pero se necesita la estructura que determina el análisis descriptivo. Con esta combinación de lo analítico (descriptivo) con lo analógico (Exploratorio) la estadística se convierte en una poderosa herramienta de interpretación; como decía Tukey; de trabajo detectivesco.

ANÁLISIS DE LA CONJETURA DEL APOSTADOR

La información necesaria para comenzar este análisis aparece en la Tabla 4. Esta muestra los valores descriptivos de las bolillas que resultaron sorteadas (S) y las que no fueron sorteadas (NS). Un total de 345 bolillas sorteadas de peso medio 2.423 se hallan en un rango de 0.5. Obsérvese que la media es menor a la mediana (Media < Mediana) por lo tanto, esperamos una cola hacia la izquierda de la distribución. La media recortada apenas logra incrementar el valor de la media por lo que esta cola de pesos livianos es alargada. Igual suerte tie-

nen los pesos no sorteados. Un total de 230 se hallan en un espacio de 0.62, donde la media al ser menor que la mediana, revela la existencia de una cola hacia los pesos bajos. De tal manera que fenomenológicamente, pero no analíticamente, las distribuciones son similares entre ellas respecto al total de pesos ya analizados. En este nivel del análisis es perentorio recordar que los pesos analizados son “estáticos”, uno a uno, sin considerar la riqueza de la dinámica del caos en el bolillero por el chorro de aire del mecanismo de la caja negra. Pero tiene validez en tanto que evalúa dos estados, el inicial de todos los pesos y el final de bolillas sorteadas y no sorteadas, la influencia del mecanismo en este último estado queda anulada o se diría que los efectos del mecanismo están en la información de las bolillas sorteadas y las no sorteadas. La diferencia que pueda existir entre ambos grupos, que haga dudar del mecanismo, es lo que proponemos.

Profundizando en nuestra visualización de los pesos, agregamos los DTH para cada grupo de pesos S y NS. La Figura 3 es elocuente al ver en detalle que salvo

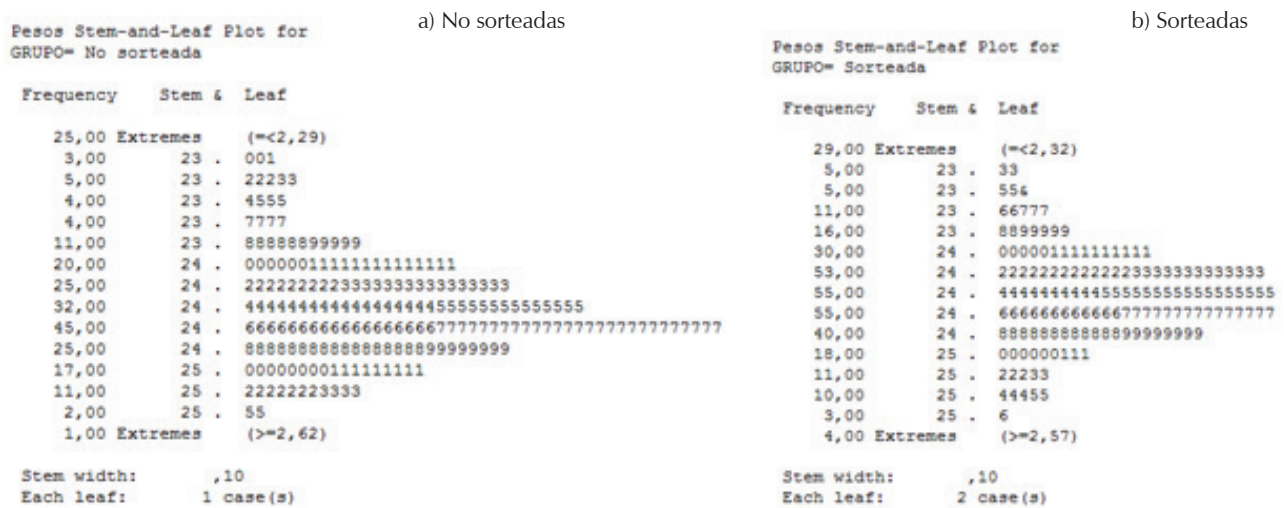


Figura 3. Diagramas de tallos y hojas de las bolillas no sorteadas, sorteadas y diagramas de caja y bigotes de ambos grupos. Lotería evaluada en el año 2000

Fuente: Libro de juegos de la lotería evaluada en el año 2000. Procesados mediante SPSS 19

Tabla 4. Estadística descriptiva de las bolillas sorteadas y no sorteadas Lotería evaluada en el año 2000

Grupo	Sorteadas		No Sorteadas	
	Estadístico	Error típ.	Estadístico	Error típ.
Media	2.4231	0,00582	2.4338	0.00386
Media recortada al 5%	2.4308		2.4382	
Mediana	2.4450		2.4500	
Mínimo	2.00		2.12	
Máximo	2.62		2.62	
Rango	0.62		0.50	
Total No sorteadas	230		345	

Fuente: Libro de juegos de la lotería evaluada en el año 2000. Procesados mediante SPSS 19

el incremento de algunos pesos en los tallos modales (sobre todo el tallo 24) no habría variaciones estructurales internas. Notemos la presencia de mayor cantidad de hojas en algunos tallos, sobre todo sobresalen las hojas 6 y 7 del tallo 24 en ambos.

Ahora se verán ambas distribuciones en una perspectiva más globalizante, la Figura 4 (A y B), muestra el diagrama de puntos de ambos grupos de pesos superpuestos (A) y las curvas normales igualmente superpuestas de ambos grupos de pesos (B) modeladas con Minitab 14. En "A" podemos ver la totalidad de los puntos de pesos divididos según los grupos S y NS, por lo que se observan diferencias internas, podemos observar que no existen agrupamientos de S y NS separados sino que cubren todo el rango. La distribución de los NS (puntos negros) se ve más alta que la de los S (cuadrados rojos) porque el gráfico los apila; notamos la existencia de más cuadrados rojos, pues representan a los "S" que son superiores, recordemos que se seleccionan 15 bolillas de 25. De tal modo que salen de la caja negra valores similares a los que se han quedado. Según la conjetura del apostador se esperan agrupamientos de puntos rojos hacia la izquierda y muy pocos o ningún punto negro.

De manera que el análisis exploratorio sirvió para hacer el trabajo de detective que propone Tukey, hallar evidencia refutable contra la conjetura del apostador; evidencia que se pone a prueba mediante análisis más completos como pruebas de rachas, T para muestras relacionadas o independientes, ANOVA.

En este sentido, el gráfico de curvas normales "B" ya es un gráfico analítico, una modelación normal en el que se nota que las curvas no se desplazan en sus centros por lo similar de las medias, lo que da como resultado que

ambos grupos proceden de una misma población. El apuntalamiento de la curva S (rayas rojas cortadas) es mayor que el de las NS (línea negra continua) por las frecuencias, las S son más que las NS, pero no parece haber variaciones importantes en sus parámetros, las desviaciones típicas son cercanas. Evidentemente esta conjetura debe probarse, pues, la subjetividad es engañosa; lo que parece poco a escala humana puede ser mucho en las escalas de las variables y solamente pruebas de bondad de ajuste puede permitirnos decidir sobre la igualdad o semejanza estadística entre ellas. Esta hipótesis es la que usaremos en el ANOVA.

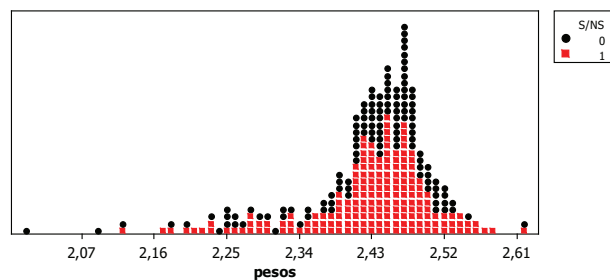
La Tabla 5 muestra, en sus columnas 3 y 4, las bolillas que fueron sorteadas y el total de ellas según los valores extremos de 60 bolillas livianas (columna 4) solo salieron sorteadas 29 (columna 3); esto contrasta con los valores centrales del RI que de 288 bolillas fueron sorteadas 167 en los 23 sorteos. Igualmente bolillas de peso más alto, como las 135 contenidas en el bigote superior, resultaron sorteadas 80. Finalmente de las 8 más pesadas ubicadas como extremo superior resultaron sorteadas 7. Esto elimina la conjetura del apostador.

ANÁLISIS CONFIRMATORIO DE LOS PESOS

El estudio final se confirma para el mismo tipo de aleatorizador y sorteos que mantengan las mismas condiciones experimentales. Ya se discutió que cada caja negra puede resultar distinta y que estos estudios pueden generalizarse hasta que otra prueba experimental diga otra cosa. El esquema de inferencia con el que trabajamos es de refutación (en sentido popperiano de *modus tollens* o negación del consecuente); esto es, sostenemos una hipótesis, que queremos anular ($\rightarrow 0$),

Fuente: Libro de juegos de la lotería evaluada en el año 2000. Procesados mediante Minitab 14
(■ 1): Sorteada (● 0): No Sorteada

Diagrama de puntos pesos de las bolillas: 1 S, 0 NS



Each symbol represents up to 2 observations.

Figura 4. a). Diagramas de puntos y b) curvas normales de las bolillas no sorteadas y sorteadas. Lotería evaluada en el año 2000

Fuente: Libro de juegos de la lotería evaluada en el año 2000. Procesados mediante Minitab 14
(— 0): No Sorteada. (- - -) 1: Sorteada

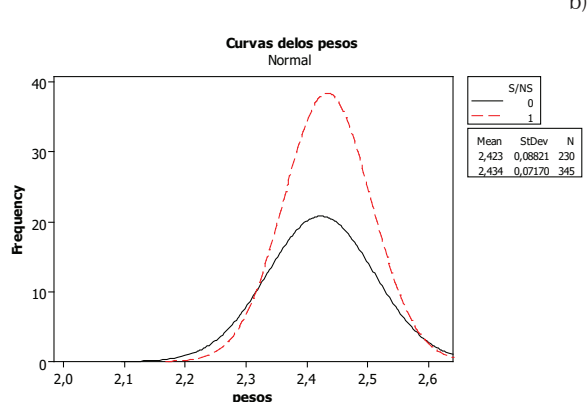


Tabla 5. Rangos del diagrama de cajas y bigotes, tamaño de los intervalos y número de bolillas sorteadas. Lotería evaluada en el año 2000

Rango de la caja	Límites de los Intervalos y amplitud	Número de bolillas sorteadas y porcentaje	Número total de bolillas y porcentaje	Total de bolillas por cuartil
Del Mínimo a Li	2.00-2.32 = 0.32	29 (8.41%)	60 (10.43%)	25%
Del Li al C1	2.32-2.41 = 0.09	53 (15.36%)	84 (14.61%)	
Del C1 al C2	2.41-2.45 = 0.04	89 (25.80%)	144 (25.04)	
Del C2 al C3	2.45-2.47 = 0.02	87 (25.22%)	144 (25.04)	
Del C3 al Ls	2.47-2.56 = 0.09	80 (23.19%)	135 (23.48)	
De Ls al Máximo	2.62-2.56 = 0.06	7 (2.03%)	8 (1.40)	25%
Total	0.62	345	575	100

Fuente: Libro de juegos de la lotería evaluada en el año 2000.

hasta que sea efectivamente refutada; este es el esquema que sigue la inferencia inductiva.

La prueba que necesitamos para confirmar lo que los análisis previos nos han mostrado, es la de igualdad de medias pues esta nos dirá si las muestras de S y NS proceden de la misma población. La propiedad física más importante que usa la caja negra de las bolillas es su peso *ceteris paribus* o considerando similares sus tamaños, elasticidad etcétera. Por lo tanto, tenemos una prueba de un factor y dos subpoblaciones, los pesos sorteados y los no sorteados. De esta manera, se puede aplicar un ANOVA de un factor o una prueba T de igualdad de medias para muestras independientes. Como se sabe, ambas pruebas arrojan los mismos resultados. Pero previamente debemos evaluar dos elementos, la homogeneidad de las poblaciones y su normalidad. Para lo primero, se revisan los diagramas de cuantiles Q-Q para luego aplicar una prueba KS (Kolmogorov-Smirnov) bajo la hipótesis nula de normalidad.

PRUEBAS DE NORMALIDAD Y HOMOGENEIDAD

Los gráficos Q-Q se pueden considerar exploratorios, pues, aunque su construcción requiere analiticidad sobre la muestra, la evaluación del resultado es visual, de tal forma que "La construcción del gráfico de probabilidad normal se realizará a través de los cuantiles de la normal estándar, de forma que aceptaremos la hipótesis de normalidad de los datos, siempre que los puntos en el gráfico tengan un comportamiento suficientemente rectilíneo" (Castillo y Lozano, 2007). Los valores de la muestra una vez normalizados se aplican a una función de distribución $F(x)$ y graficando los "puntos de posición gráfica" o *Plotting Positions*, que en el caso del gráfico normal Q-Q son las coordenadas, $(x_i, y_i) = (p_i, F(z_i))$ y en el gráfico Q-Q sin tendencia son las coordenadas $(x_i, y_i) = (p_i, F^{-1}(z_i) - z_i)$. En ambos casos, los puntos deben estar muy cerca (o sobre) las líneas rectas. La forma de calcular

p_i que representa la posición del cuantil ha variado en el tiempo; originalmente se presentaba como i/n , pero debido a problemas de graficación (Castillo y Lozano, 2007) se modificó por Hanzen en 1930 por $p_i = (i - 0.5)/n$, donde 0.5 es una corrección de continuidad. Esta propuesta aparece en SPSS como Rankit.

En este programa surgen otras propuestas, particularmente para la distribución normal aparece, por defecto, la de Blom G. hecha en 1958 y que tiene la forma

$$p_i = \frac{i - 1/3}{n + 1/4}, \text{ con } i = 1, 2, \dots, n$$

Esta breve discusión se realiza, pues los valores de p_i no van a coincidir en los distintos programas estadísticos y cada uno tiene varios. Sin embargo, estas modificaciones que se hacen para mejorar la graficación o para corregir la continuidad en el hallazgo de valores pequeños, no van a cambiar la realidad de la masa de datos; si estos carecen de normalidad estas propuestas no se la van a dar y dado el caso de contradicción entre ellas cabría la duda sobre el rechazo de la hipótesis nula. En este trabajo se utilizó SPSS 19 y el valor por defecto de p_i , el uso de otras propuestas no cambia la decisión.

La Figura 5a de cuantiles muestra que los valores aun cuando concentrados al centro de las rectas se alejan de ellas en el borde inferior, haciendo muy difícil sostener una hipótesis de normalidad; igualmente sucede con la Figura 5b que muestra las diferencias entre la distribución normal y los valores tipificados; en esta gráfica Q-Q sin tendencia, los valores deben distribuirse sin mostrar o sugerir alguna forma geométrica, debiendo aparecer alrededor de la línea como un enjambre distribuido sobre la recta, por lo que se espera que la prueba KS confirme esta conjetura.

La Tabla 6 muestra los resultados de una prueba KS que testa la hipótesis nula de normalidad de la distribución (H_0) frente a la alterna de su rechazo (H_1) con el estadístico de prueba de máxima diferencia $D = \max |F_n(x) - F_0(x)|$. Este procedimiento contempla los mismos

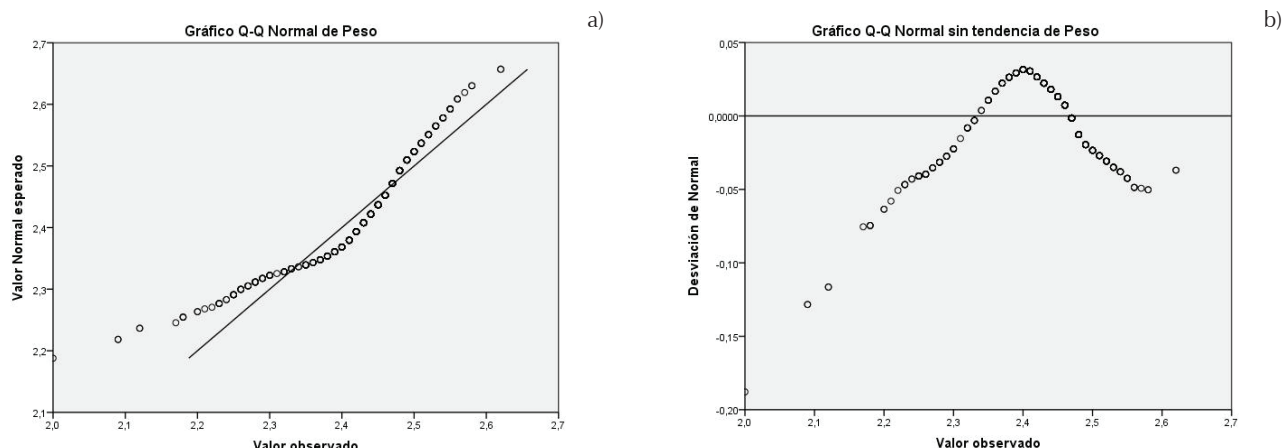


Figura 5. a) Q-Q Normal y b) sin tendencia para los pesos de 575 bolillas. Lotería evaluada en 2000

Fuente: Libro de juego año 2000. Lotería Evaluada. Procesados con SPSS 19

elementos que los gráficos Q-Q, es decir, la normalización de los valores y su comparación contra la función de distribución bajo hipótesis nula $F(x)$. Como puede verse en la Tabla 6, con una Z de KS de 0.169 y un p-value nulo se debe rechazar, sin lugar a dudas, la normalidad de la distribución de los pesos de las bolillas; igualmente la prueba de Shapiro-Wilk, que rechaza la normalidad de los pesos de las bolillas.

A continuación se verá la prueba de Levine para igualdad de varianzas obtenidas mediante SPSS 19 y MINITAB 14. Esta prueba tiene diversas expresiones en los programas estadísticos. Consiste en aplicar un ANOVA a la desviación media, caso de SPSS, $D = |x_i - \bar{x}|$ o la desviación mediana en el caso de Minitab 14 $D = |x_i - Me$

diana!. La Tabla 7 muestra que el estadístico de Levene (una F en ambos casos) para SPSS es mucho mayor que el de MINITAB; el estadístico tomado como desvíos respecto de la media, rechaza la hipótesis de nulidad ($0.015 < 0.05$; 95%), mientras que el estadístico tomado como desvíos respecto de la mediana no la rechaza ($0.088 > 0.05$; 95%). Por lo anterior, se realizó una prueba t de igualdad de la media (2.4295) y la mediana (2.45) siendo rechazada absolutamente ($t = -6.238$; p-value= 0.000). De este modo los datos son sensibles a la metódica estadística.

Si bien se prefiere el análisis de MINITAB, ya que se basa en la mediana, que es un estadístico más robusto, no haremos este análisis en este momento, pues rebasa el

Tabla 6. Prueba de Kolmogorov-Smirnov para los pesos de las bolillas usadas en los sorteos del año 2000 de la lotería evaluada

	Pruebas de normalidad					
	Kolmogorov-Smirnov			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
Peso	0.169	575	0.000	0.882	575	0.000

Fuente: Libro de juegos de la lotería evaluada. Procesados con SPSS 19. Año 2000
a. Corrección de la significación de Lilliefors

Tabla 7. Prueba de homogeneidad de varianzas del SPSS y Minitab para los pesos de las 575 bolillas de la lotería evaluada

	Estadístico de Levene	gl1	gl2	Sig.
SPSS	5.919	1	573	0.015
MINITAB	2.96			0.086

Fuente: Libro de juegos de la lotería evaluada. Procesados con SPSS 19 y Minitab 14. Año 2000

Tabla 8. Anova de un Factor. Pesos de las bolillas y grupo (S, NS) para las bolillas de la lotería evaluada

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	0.016	1	0.016	2.541	0.111
Intra-grupos	3.550	573	0.006		
Total	3.566	574			

Fuente: Libro de juegos de la lotería evaluada. Procesados con SPSS 19. Año 2000

objetivo del trabajo. Por el momento, se tiene que los datos no son normales y sensibles a los métodos de homogeneidad por lo que cabría duda a este respecto; pero en descargo de esta situación, la duda se inclina a favor de la homogeneidad pues esta debe rechazarse sin duda alguna. La Tabla 8 muestra el análisis de igualdad de medias hecho a través de un ANOVA de un solo factor. El estadístico de contraste no rechaza la igualdad de medias ($p\text{-value} = 0.111 > 0.05$; 95%), de modo que las muestras S y NS procederían de la misma población, esto es, que tienen las mismas características estadísticas, con lo cual podemos hipotetizar que el aleatorizador no interviene en la selección de las bolillas.

DISCUSIÓN Y CONCLUSIONES

La conjetura del apostador según la cual el aleatorizador tendría un sesgo a favor de las bolillas de peso ligero tiene justificación en tanto pertenece al mundo del sentido común, dudar como práctica de protección, más aún si se trata de un juego que solo satisface a unos pocos durante el año. Como se sabe, el mercado de las loterías de cupones ofrece más combinaciones de números por sorteo (${}_{25}C_{15} = 3.268.760$ posibles combinaciones de números en este caso) que apostadores (alrededor de un millón en este caso), por lo tanto, la mayoría de las veces, no habría ganador, acumulándose el premio, pero sí elevando las expectativas. Sobre este evento se generaría una duda racional, pues lo que se cree es que siempre debería salir un ganador. Por lo tanto, el juego puede terminar en frustración definitiva respecto del juego (¡hay truco!) o una frustración pasajera (¡No salió mi número de la suerte!). Esta frustración tendría salida en la atribución de sesgo sobre la caja negra. El apostador no renunciaría a darse una explicación racional, pues es difícil creer que siendo juego de azar no beneficie a algún jugador; esto es, no se juega tantas veces y hay pocos ganadores, algo debe pasar, por lo que se atribuye esto al truco. Sin embargo, teniendo una justificación de sentido común no la tiene científicamente. El análisis descriptivo expresa que así como resultan sorteadas bolillas livianas, también lo hacen las más pesadas, esto es, que tienen estructuras similares contra la conjetura del apostador que saldrían más las livianas.

También expresa que esta estructura ofrece una cola fina hacia la izquierda y la de la derecha algo gruesa. La izquierda por fina, es abundante en valores más dispersos obligando a la media a acercarse a ese extremo, por debajo de la mediana. Al haber más dispersión a la izquierda que a la derecha de la distribución, las cargas son mayores de ese lado obligando a la media a estar por debajo de la mediana. Esto podría redundar en favor de la duda del apostador; pero el sesgo por sí solo no argumenta a favor de alguna intencionalidad humana. Habría que recordar que la medida de estos pesos es estática, no sabemos qué ocurre en la caja negra, lo que más se podría hacer es ejercer mayor control sobre los pesos de las bolillas que participan, seleccionándolas en intervalos más reducidos y en hacer más pruebas sobre el proceso del bolillero, para garantizar que las distribuciones de los pesos de entrada y los de salida sean estadísticamente similares, cosa que, como vimos, no se pudo demostrar mediante las pruebas de Levine. El análisis exploratorio, al profundizar en la visualización de las estructuras halladas, muestra una densidad importante de pesos alrededor de la mediana y al superponer ambas estructuras (S y NS) se nota una similitud importante; esto es, no es posible observar grupos de pesos altos o bajos formando un clúster tan grande que nos haga dudar del mecanismo; los pesos en ambas distribuciones lucen similarmente repartidos, pero el sesgo no sería suficiente para dudar de la aleatoriedad del juego. De esta manera, la conjetura del apostador se rechaza empíricamente mediante estos análisis descriptivo/exploratorio y el ANOVA. Por su parte, el análisis de Levine hecho con los desvíos medios y con los medianos ofrece resultados contradictorios sobre la homogeneidad de las subpoblaciones N y NS; la masa de datos no procede de una distribución Normal, según KS y los diagramas Q-Q, pero el ANOVA no puede rechazar la hipótesis de igualdad de medias y aunque pierda cierta validez sin los supuestos, la F es robusta ante el incumplimiento de estos. Sin embargo, esto no tiene mucha importancia, pues los pesos se pueden seleccionar en el proceso de control operativo, por lo que la pregunta es más por un mejor control que justifique con más propiedad el rechazo de la duda del apostador.

REFERENCIAS

- Bar A.R. Un aporte a la discusión del status metodológico de las variables y escalas de medición. (U.d. Chile, Ed.), *Cinta de Moebio* (número 7), 2000.
- Binmore K. *Teoría de Juegos*, Madrid, McGraw-Hill, 1994.
- Castillo-Gutierrez S. y Lozano-Aguilera E.D. Q-Q Plot Normal. Los puntos de posición gráfica. (U. d. Jaen, Ed.), *Iniciación a la Investigación*, volumen 2 (número 9), 2007.
- Cohen M.R y Nagel E. *Introducción a la lógica y al método científico*, Vol. II, (N. Míguez, Trad.), Buenos Aires, Argentina, Amorrortu, 2000.
- Hempel C. *Fundamentals of concept formation in empirical science*, Chicago, The University Chiago Press, 1952.
- Hildebrand D.K y Ott R.L. *Estadística aplicada a la administración y a la economía*, México, Addison Wesley Iberioamericana, S.A., 1998.
- Mosterin J. *Conceptos y teorías en la ciencia*, Madrid, Alianza Universidad, 1987.
- Palmer-Pol A.L. *Análisis de Datos*, Madrid, España, Pirámide, 1999.
- Pigna F. *Clarín.com*. [en línea] [Fecha de consulta: 30 de diciembre de 2007] Recuperado el 5 de enero de 2017. Disponible en: <http://edant.clarin.com/suplementos/zona/2007/12/30/-04001.htm>.
- Prensa N. *La Verdad.com*. [en línea] [Fecha de consulta: 22 de octubre de 2012] Recuperado el 5 de enero de 2017.

Citación sugerida:**Citación estilo Chicago**

Casanova-Leal, Hugo. La variable oculta en la suerte del apostador de loterías. *Ingeniería Investigación y Tecnología*, XIX, 03 (2018): 343-354.

Citación estilo ISO 690

Casanova-Leal H. La variable oculta en la suerte del apostador de loterías. *Ingeniería Investigación y Tecnología*, volumen XIX (número 3), julio-septiembre 2018: 343-354.

SEMBLANZA DEL AUTOR

Casanova-Leal Hugo Delfin. Licenciado Cs. estadísticas UCV en 1995, Doctorado en Estadística. CEAP-UCV. Cuenta con un Diplomado en Investigación por UPEL, 2013. Fue Auditor de Lotería del Táchira año 2000, Jurado Premio Nacional de Ciencia 2015 Ciencias Sociales, Libro: Estadística Tabular. ISBN: 978-980-7440-71-4. Publicación: Falacia de división en el comportamiento social del apostador en las loterías de cupones. *Fermentum* 73. Vol 25 Mayo-Agosto 2015. Su línea de investigación es: Estadística y Estado, así como consultor institucional.