

“THE TRIAL OF THE PYX”, UN CONTROL ESTADÍSTICO DE LA CALIDAD CON OCHO SIGLOS DE ANTIGÜEDAD

Autor: Dr. José A. Álvarez Jareño

Universidad: Departamento de Economía Aplicada de la Universitat de València

Dirección: Avda. de los Naranjos, s/n, 46022, Valencia

Teléfono: 96 382 8615

Correo electrónico: Jose.A.Alvarez@uv.es

RESUMEN:

El “*Trial of the Pyx*” es una prueba estadística para el control de calidad de la acuñación de la moneda británica que se lleva realizando desde el siglo XII. En una época en la que la estadística y la probabilidad aún no había establecido sus bases, la monarquía británica estableció un control de la calidad en base a una serie de conceptos de muestreo y estadística que se desarrollarían mucho más tarde. Como precursor de una nueva técnica, también se cometieron algunos errores de importancia. La variación de n unidades se consideraba como n veces la variación de una unidad. Este error perduro durante casi 600 años, hasta que De Moivre en 1730 demuestra que el aumento de variación es proporcional a \sqrt{n} . “La Prueba de la Caja” permitía mucha variabilidad sin incurrir en la ilegalidad, es decir, que la Casa de la Moneda podía fabricar más moneda de la que le exigía el Rey, y por lo tanto, obtener un mayor beneficio.

La persona más ilustre que ha estado al frente de la Casa de la Moneda ha sido Sir Isaac Newton, que fue primero *Warden* desde 1696 a 1699, y posteriormente *Master* hasta su muerte en 1727.

1. Introducción

Aunque el control de la calidad a través de la estadística es una disciplina relativamente moderna¹, se tiene conocimiento de algunos problemas de control de calidad desde la antigüedad. El primero del cual se tiene conocimiento es el que se le encargó a Arquímedes. Hieron, rey de Siracusa, entregó cierta cantidad de oro a un orfebre para que le realizase una corona triunfal digna de su jerarquía. Cuando recibió su nueva y reluciente corona, debió desconfiar del artesano, porque hizo llamar a Arquímedes (287 a. C. – 212 a. C.), y le planteó el problema de si era posible conocer si el orfebre había utilizado todo el oro que se le había entregado, o si bien había agregado otros metales. Arquímedes se marchó con la promesa de pensar en el problema, y un buen día, cuenta la leyenda, que mientras estaba tomando un baño, le llegó la solución al problema, y saliendo completamente desnudo de los baños gritó “¡Eureka, Eureka!”. La solución de este antiguo problema de control de calidad se conoce desde hace muchos años como Principio de Arquímedes: *todo cuerpo sumergido en un fluido experimenta un empuje vertical y hacia arriba igual al peso del volumen del fluido desalojado.*

En la antigüedad no se precisaba de un control de calidad porque los productos eran artesanales y no estaban estandarizados. Todas las vasijas que fabricaba y vendía un artesano era diferentes y no se precisaba que cumplieran con determinadas especificaciones. Sin embargo, el primer objeto que se fabrica en serie son las monedas, que aunque se acuñan por un procedimiento artesanal, deben cumplir con unos determinados estándares como peso o pureza de los metales utilizados.

En este contexto se establecen unas normas para determinar si las monedas acuñadas por la Casa de la Moneda de Londres cumplen con los requisitos establecidos por el Rey en el proceso de producción. Su finalidad² era mantener un estándar y asegurar que se cumple. Lo más sorprendente es que se hace en el año 1279 cuando los conceptos de probabilidad o muestreo aún no habían sido establecidos. Esta prueba se conoce como el “*Trial of the Pyx*” y ha perdurado hasta nuestros días con algunos

!!

¹ Los trabajos de Ronald Fisher en la década de 1920, los de Walter Shewhart, incluyendo los gráficos de control de 1924, o el trabajo conjunto de Jerzy Meyman y Egon Pearson de 1933.

² El ensayo se continua realizando aunque con una finalidad diferente que es inspirar confianza en la moneda y en el sistema.

cambios. Curran-Everett (2009) afirma que el “*Trial of the Pyx*” es el contraste de hipótesis de dos colas más antiguo que se conoce.

2. La “Prueba de la Caja” o el “Ensayo de la Píxide”³

El segundo martes del mes de febrero se realiza desde hace más de 8 siglos el “*Trial of the Pyx*”, y el resultado es leído por el Juez el primer día de mayo. Después de la prueba se lleva a cabo un banquete para celebrar el resultado. La Casa de la Moneda (The Royal Mint) de Londres es la institución encargada de acuñar la moneda en Gran Bretaña, y es independiente de la Corona británica. El Monarca del Reino Unido entrega a la Casa de la Moneda una determinada cantidad de metales preciosos para que esta institución acuñe moneda de curso legal.

El origen del “*Trial of the Pyx*” se puede rastrear hasta el reinado de Henry II (1154 – 1189) tal como expone Craig (1953). En cualquier caso, una prueba pública se llevo a cabo en 1248, y todos los autores están de acuerdo que la prueba se estableció en 1279 por Edward I donde se describe el procedimiento que se debe seguir. Muchas particularidades de ese procedimiento han cambiado a lo largo de los siglos, pero sus premisas básicas se mantienen incólumes.

El procedimiento es más o menos el siguiente, una selección de monedas se guardaban en una caja de madera denominada “*pyx*”⁴ que estaba cerrada. Se depositaba una moneda⁵ por jornada de trabajo de la Casa de la Moneda. Un jurado independiente examina las monedas depositadas en el “*pyx*” para comprobar sus características (diseño, dimensiones, diámetro, composición, peso y pureza) y determinar si la Casa de

+++++

³ El “Ensayo de la Píxide” es la traducción que se le ha dado a esta prueba en la edición en castellano de la novela de Neal Stephenson “El Sistema del Mundo”, perteneciente al Ciclo Barroco, y compuesta por tres volúmenes. En esta obra Sir Isaac Newton se enfrenta a un juicio porque las monedas acuñadas no cumplen con el “ensayo de la píxide”. Este hecho es históricamente cierto aunque el autor lo narra de una forma novelada mezclando la realidad con la fantasía y los personajes históricos con otros completamente ficticios.

“El Sistema del Mundo” también es el título que dio Isaac Newton al tercer libro de su obra “The Principia: Mathematical Principles of Natural Philosophy”.

⁴ Pyx es una palabra de origen romano, procedente del griego “*pyxis*”, caja pequeña. Aunque el uso eclesiástico de la palabra “*pyx*” antes de 1400 hacía referencia al cáliz donde se guardaba el pan sacramentado. Esta podría ser otra acepción de la palabra “*pyx*” porque era en una de las capillas de la Abadía de Westminster donde se guardaban los patrones que servirían de comparación en la prueba.

En castellano, y según el Diccionario de la Real Academia Española de la Lengua, el píxide es el copón o caja pequeña en que se guarda el Santísimo Sacramento o se lleva a los enfermos.

⁵ Una moneda cada 15 libras de oro, y una cada 60 libras de plata.

- b) Si la proporción es inferior al estándar se produce una pérdida de confianza en la moneda, lo que implica inflación en los precios. En este caso, la Casa de la Moneda debía reingresar al monarca una cantidad de metal igual a la diferencia observada.

3. La prueba desde un punto de vista estadístico

Para realizar una comprobación son necesarias dos cosas: un estándar y un método estadístico.

El contraste o comprobación se debe realizar tomando como referencia un estándar, para lo que se guardaba una “placa⁷” de oro en un lugar seguro y que durante los primeros siglos fue una de las capillas de la Abadía de Westminster.

Desde el punto de vista de un estadístico la prueba de la píxide es un ejemplo perfecto de inspección por muestreo para comprobar la calidad de un producto. El problema se puede formular en términos de un contraste de hipótesis. La Corona británica está preocupada por las desviaciones positivas y negativas del estándar establecido para la acuñación de monedas, por lo que precisa un contraste de dos colas. Como no es económico para el jurado pesar todas las monedas acuñadas (e imposible contrastarlas todas, porque el contraste es un proceso destructivo en lo referente a la pureza del metal), es necesario recurrir a una muestra. Como la acuñación es un proceso continuado, se debe implementar un procedimiento de muestreo y contraste.

En primer lugar, se debe establecer un procedimiento de muestreo, a continuación se fijará el intervalo del contraste, y el contraste estadístico, y finalmente se determinará la región crítica de rechazo. Sin embargo, todas estas técnicas estadísticas son muy posteriores a la implantación de la prueba de la píxide, la mayoría se desarrollaron durante el siglo XIX y principios del XX.

Esquema de Muestreo. Desde el punto de vista estadístico el muestreo debería ser aleatorio, pero el concepto de muestra aleatoria es un concepto moderno. Los primeros documentos no indican cómo se hacía el muestreo, pero es imposible de creer que las diferentes (y muy desconfiadas) partes quedarán satisfechas con una selección que ellos no vieran como esencialmente aleatoria según expone Stigler (1999). Aunque

⁷ Los “plates” era una fina placa de oro o de plata de una pureza determinada que se utilizaba como estándar para la comparación de la moneda acuñada. Se podría identificar como la hipótesis nula de un contraste.

matemáticamente, la muestra no era aleatoria en un sentido moderno, por lo menos, la selección debía ser imparcial a los ojos de todas las partes.

Se pesaban conjuntos de monedas para evitar variaciones en los pesos individuales. Las monedas eran pesadas de forma agregada, 100 monedas de una vez. El peso medio de las monedas comprobadas era comparado con el estándar. Desde el punto de vista del diseño experimental, esto era admirable, ya que el peso agregado sólo está sujeto a un único error de medida frente a los 100 de las monedas pesadas individualmente. Sería una aplicación del Teorema Central del Límite antes de su exposición y demostración.

Tolerancia de la prueba o variaciones permitidas. No se puede asumir una excesiva variabilidad en la cantidad de metal utilizado en la acuñación. Como la tecnología para acuñar moneda no era perfecta, se daba por sentado por todos que algún tipo de ajuste se debía hacer con la variabilidad. La finalidad de la prueba es que la variabilidad de la moneda sea lo más baja posible, es decir, que el error del intervalo en torno al estándar prefijado sea lo más pequeño posible. Como este depende del nivel de significación y del número de observaciones, ambos parámetros deberán ser elevados para reducir el error del intervalo.

Se precisan dos estándares, el del peso y la tolerancia. Los estándares no tienen utilidad sin los métodos estadísticos que fueron inventados con ese propósito. Es cierto que los barones británicos tenían nociones de muestreo, de agregación y de ajustes de la variabilidad. Pero no sabían como ponerlo todo en conjunto. Esencialmente, ellos especificaban el ajuste para una moneda individual, luego lo extrapolaban multiplicando para obtener el ajuste para la cantidad agregada. Este es el error más importante en todo el proceso. La variación de n unidades se consideraba como n veces la variación de una unidad.

De Moivre demostró en 1730 que el incremento de la variación es proporcional a \sqrt{n} y no a n como se estaba realizando en el “ensayo del píxide”. Permitía mucha variabilidad sin incurrir en la ilegalidad.

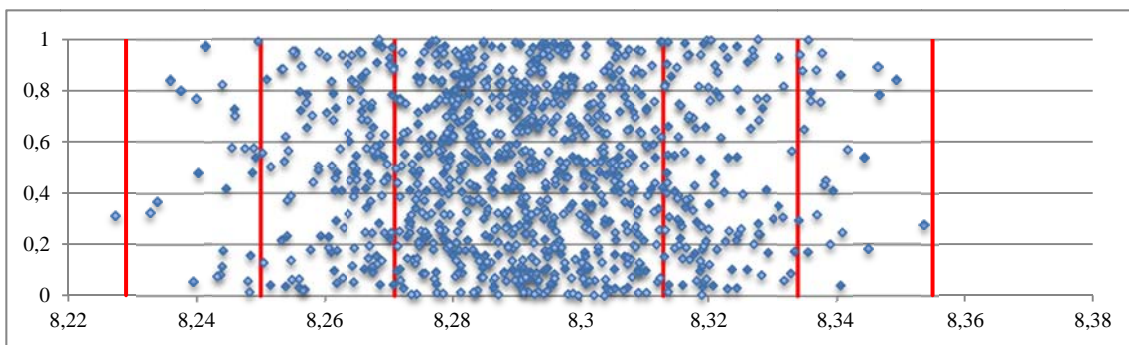
Existen evidencias de la gradual explotación de estos márgenes en momentos económicamente difíciles para la Corona británica. Incluso no han sido raras las acusaciones entre ingleses y franceses de haber explotado fraudulentamente esta

diferencia, tal como afirma Stigler (1999). Aunque es difícil de demostrar que fuera por este error en la variabilidad, como a continuación se indica.

Las guineas de 1710 debían tener un peso de 128 granos (aproximadamente 8,3 gramos), un diámetro de 25 mm., y un contenido en oro de 0,9134. La desviación permitida era de 1/400, es decir, 0,32 granos (0,02 gramos) por cada moneda. El error que cometían era que aplicaban la desviación de forma proporcional al número de monedas, siendo la desviación permitida para 100 monedas de 32 granos (2,075 gramos), cuando en realidad la desviación permitida debía ser de 3,2 granos (0,2075 gramos). De lo anteriormente expuesto se deduce que de cada 100 monedas se podría ahorrar un 1 gramo de metal, sin temor a no superar el “*Trial of the Pyx*”, por lo que de cada 830 monedas se podría fabricar una adicional (1,2 por mil).

Si se realiza una simulación de Monte Carlo sobre la base que la acuñación de las guineas sigue una distribución Normal, y no se quiere incurrir en un rechazo por parte del jurado, no se podrá fijar un peso inferior a 8,292 gramos por moneda, lo que sería lo mismo que ahorrar 0,008 gramos por moneda. Es decir, de cada tonelada de oro que el Rey entregaba a la Casa de la Moneda, esperaba que se pusieran en circulación 120.482 guineas, sin embargo, alguien con un precisión desconocida para la época podría fabricar hasta 120.598 sin incurrir en problemas. La diferencia son 116 guineas, exactamente 0,962 por mil, sin embargo, el jurado era capaz de detectar un error en la pureza de 1 una parte entre 1.000, como se verá en el siguiente apartado. Y todo esto, sin que se desperdicie ni un solo miligramo de oro en el proceso de producción que estaba más cerca de la alquimia en los siglos XV y XVI que de un proceso industrial.

Gráfico 1. Generación de una muestra de 1.000 monedas con distribución de probabilidad $N(8,292; 0,021)$



En el Gráfico 1, se puede observar una muestra generada aleatoriamente mediante una distribución normal con media 8,292 gr. y desviación 0,021 gr. Las líneas rojas representan 2-sigma, 4-sigma y 6-sigma (siguiendo la terminología de control de calidad). Con esta distribución, y tomando muestras de 100 monedas, se superaría el “*Trial of the Pyx*”. No obstante, y como se ha indicado, la Casa de la Moneda realiza una inspección previa por lo que se descartaran todas aquellas monedas que presenten un peso inferior a la tolerancia, teniendo en cuenta la precisión de los elementos de medida. Si la Casa de la Moneda realiza un pesaje individual de las monedas en su inspección para dar el visto bueno a la circulación de una moneda, la situación hubiera sido beneficiosa para ella, debido al error que se cometía en la tolerancia por el desconocimiento de la proporción con respecto al número de monedas.

En condiciones normales, 2,7 de cada 1.000 monedas estarían fuera de las líneas que delimitan 6-sigma, un error superior a la ventaja que se puede obtener con un correcto conocimiento de la tolerancia de la distribución de probabilidad de una población Normal.

4. Sir Isaac Newton, the Master of the Royal Mint, y Abraham De Moivre

La persona más ilustre que ha estado al frente de la Casa de la Moneda ha sido Sir Isaac Newton, que fue primero *Warden* desde 1696 a 1699, y posteriormente *Master* hasta su muerte en 1727. Abraham De Moivre se exilió en Londres por motivos religiosos y fue elegido miembro de la Royal Society en 1697. Mantuvo cierta relación con Newton, e incluso se dice que lo recomendaba para la resolución de problemas matemáticos por considerarlo más competente que él mismo en ese campo. También fue designado en 1710 para la comisión de la Royal Society que debía estudiar la disputa entre Leibniz y Newton sobre quién era el descubridor del cálculo diferencial. Newton había fallecido cuando se hizo pública la demostración de De Moivre en 1730.

Newton pasó su primer “*Trial of the Pyx*” en 1701. Sin embargo la prueba más importante fue la de 1710, el jurado de orfebres rechazó las monedas por ser inferiores al estándar en 1 parte por 1.000, pero Newton demostró satisfactoriamente que los estándares de la prueba usados desde 1707 eran demasiado puros (24 quilates) para una comparación justa. El problema surgió de la calidad del oro utilizado en las “placas de la prueba”, no por encontrar diferencias en el peso de las monedas. Como describe

En Estados Unidos la Comisión Anual del Ensayo fue suspendida por el Presidente Carter en 1977 al considerarla una ceremonia innecesaria.

6. Conclusiones

La Corona británica precisaba de un método estadístico para el control de la calidad de la moneda por dos motivos: la variabilidad del proceso y la destrucción de la muestra para comprobar la pureza. No se pueden comprobar todas las monedas, y se hace necesario, elegir una muestra. La estadística no juega ningún papel en el desarrollo del “*Trial of the Pyx*”, pero claramente adelanta conceptos que permitirán el desarrollo de la estadística. Disponer de un estándar, las “placas” para el ensayo, es similar a establecer una hipótesis nula para contrastarla posteriormente. El establecimiento de una región aceptación y otra de rechazo (dos colas) en función de la dispersión aceptada de la variable, que se fijaba a través de la tolerancia.

De todas formas, el Ensayo del Píxide tenía algunas deficiencias que desde el punto de vista actual serían criticables. Al no existir el concepto de aleatoriedad, no se aplicaba este principio a la selección de la muestra, y era sustituido por una selección que a la vista de las diferentes partes fuera imparcial, es decir, no beneficiase a nadie. La muestra que se utilizaba no estaba lo suficientemente definida y implicada diferentes fuentes de sesgo. Los cambios que se produjeron en la tolerancia se llevaron a cabo primero en América en 1830, un siglo después de la demostración de De Moivre, lo que implica que el ensayo siguió jugando su papel político y no fue necesaria una revisión metodológica de su aplicación a corto plazo.

Aunque Sir Isaac Newton disponía de los conocimientos y mantenida cierta relación con De Moivre, que fue el matemático que demostró el error que se estaba produciendo en la tolerancia, no se tienen evidencias que se beneficiase de esta situación. Todos los esfuerzos de Newton estaban encaminados a reducir la variabilidad de las monedas acuñadas, aportando mejoras en el proceso de acuñación debido a que la metalurgia fue, junto con la alquimia, de gran interés para él.

7. Bibliografía

Alfaro, Esteban; José Luís Alfaro, José Mondejar y Manuel Vargas (2004). Control Estadístico de la Calidad: Una reseña histórica. Documentos de Trabajo de la Facultad de Ciencias Económicas y Empresariales de la Universidad de Castilla-La Mancha. http://www.uclm.es/AB/fcee/D_trabajos/2-2004-1.pdf

Ballarín, Ángel (2008). “Una aproximación a la historia de la cuantificación del riesgo: de la antigüedad al siglo XVIII”. VI Encuentro de trabajo sobre Historia de la Contabilidad. Ponencias y Comunicaciones.

http://www.aeca.es/vi_encuentro_trabajo_historia_contabilidad/pdf/01_ballarin.pdf

Craig, John (1953). The Mint. Cambridge: Cambridge University Press.

Curran-Everett, Douglas (2009). “Explorations in statistics: hypothesis tests and P values”. Advances in Physiology Education, Vol. 33, pp. 81-86.

Hilgard, Julius E. (1872). “An application of an exponential function”. Proceedings of the Social Sciences, editado por David L. Sills. New York: Macmillan and Free Press.

Newman, E. G. V. (1975). “The Gold Metallurgy of Isaac Newton”. Gold Bulletin, Vol. 8, Núm. 3, págs. 90-95.

Murray, Athol L. (1997). “Sir Isaac Newton and the Scottish recoinage, 1707-10”. Proceedings of the Society of Antiquaries of Scotland, Vol. 127, págs. 921-944.

Stephenson, Neal (2004). The System of the World. Arrow Books, Berkshire.

Stigler, Stephen M. (1996). “Statistics and the Question of the Standards”. Journal of Research of the National Institute of Standards and Technology. Vol. 101, Núm. 6, pp. 779-789. <http://nvl.nist.gov/pub/nistpubs/jres/101/6/j6stig.pdf>

Stigler, Stephen M. (1997). “Eight Centuries of Sampling Inspection: The Trial of the Pyx”. Journal of the American Statistical Association. Vol. 72, Núm. 359, pp. 493-500. <http://www.jstor.org/pss/2286206>

Stigler, Stephen M. (1999). Statistics on the Table: The History of Statistical Concepts and Methods. Harvard University Press, Cambridge, Massachusetts.

Watts, M. (1992). “The Trial of the Pyx”. Reserve Bank of New Zealand Bulletin, vol. 55, N° 3, pp. 221-223.