

(#26) . FOTVOLTAICAS HIPERGEOMÉTRICAS

Hace escasos días, el pasado 22 de enero, [aparecía esta noticia](#) sobre la **decisión del Tribunal Supremo de no indemnizar a los propietarios de las plantas solares** afectados por los recortes del sector en 2010. Al margen de lo justa o no que nos pueda parecer esa decisión (a mí particularmente me irrita), lo que más me llamó la atención de la noticia fuera que el tribunal decidiera **estudiar sólo 5 casos de las más de 150 empresas** que tenían planteados recursos. Así, según reza el artículo, el tribunal escogió 5 casos al azar, como muestra para extrapolar sus conclusiones al resto de la población. **Los informes de esos 5 casos fueron positivos**, es decir, esas plantas solares son rentables, cifrando esa rentabilidad en un 8% anual. Por tanto, en los 5 casos estudiados no se encontró ninguna planta que no fuera rentable, pese a las pérdidas que los propietarios argumentan tras los recortes en las primas a las fotovoltaicas y otros “sablazos” del Gobierno, y donde éstos aseguran que un 66% de las plantas han sido re financiadas por los bancos.

Las preguntas que surgen ante esta forma de proceder del tribunal son muy jugosas: **¿Es correcto escoger una muestra tan pequeña para sacar conclusiones con un mínimo de precisión? ¿Dados esos resultados, se puede decir que realmente la mayor parte de la población de fotovoltaicas es rentable y, de este modo, la sentencia podría considerarse “justa”? ¿Y que hay cierto en la afirmación del alto tribunal sobre que “ninguna de las instalaciones tiene un descenso apreciable de su rentabilidad”?**

En este post vamos a tratar de responder a esas cuestiones de una forma sencilla, empleando la estadística como compañera de viaje. Lo vamos a hacer sin profundizar demasiado, sólo con un esbozo que nos puede dar una visión mucho más adecuada, científicamente hablando, de este delicado tema.

El marco estadístico

Podemos enmarcar este problema a través de la caracterización de un experimento realizado por el tribunal. Supongamos que existen 150 recursos presentados por empresas damnificadas (el artículo no dice el número exacto sino que son “más de 150”), y que, dado que son demasiados casos para analizar uno a uno, escoge una muestra testigo de únicamente 5 casos. Como la muestra es sin reemplazamiento y proveniente de una población finita y pequeña, [Krishnamoorthy \(2006\)](#) nos indica que debemos acercarnos al mundo de **la distribución hipergeométrica** para tratar de calcular probabilidades. Y eso es lo que vamos a hacer.

La distribución hipergeométrica

Tenemos N empresas de las cuales M tienen pérdidas (no son rentables), y el resto $N-M$ son rentables. Tenemos, además, una muestra de n empresas escogida de manera aleatoria y sin reemplazamiento. Finalmente, denotamos con X el número de empresas con pérdidas que se observa en la muestra, siendo k precisamente ese valor observado para cada muestra particular. La variable aleatoria X es considerada una variable aleatoria hipergeométrica con parámetros N y M . De este modo, la probabilidad de observar k empresas con pérdidas en una muestra de tamaño n es la siguiente:



A partir de aquí, podemos empezar ya a realizar cálculos. Para ello vamos a ir planteando preguntas y las vamos a contestar con la [ayuda de StatCalc](#).

1. ¿Cuál es la probabilidad de que no haya ninguna empresa con pérdidas?

Como los 5 informes realizados por los peritos indican que esas 5 plantas son rentables, no hay en la muestra ninguna empresa no rentable. La probabilidad de que no haya ninguna empresa no rentable en la muestra depende del tamaño de la población, el número de empresas con pérdidas en la población, el tamaño de la muestra y el valor observado k en la muestra. Esos es lo que nos dice la fórmula de la función de probabilidad hipergeométrica. Sabemos que $N=150$, es decir, el tamaño de la población. Sabemos que $k=0$, porque no se ha encontrado ninguna empresa con pérdidas, y que la muestra $n=5$, que son los 5 casos testigo. Si suponemos que un 66% de las

empresas tienen muchos problemas económicos (no son rentables porque han sido refinanciadas), es decir, unas 100 empresas, podemos decir que la probabilidad de que $X=0$: $P(X=0)=0.0036$.

Esto es muy interesante, porque indica que si lo que dicen los demandantes es verdad, la probabilidad de que se observe ese resultado en la muestra es extremadamente pequeña. Es decir, probablemente o mienten los propietarios o los informes periciales. Fijaos que si, en lugar de haber un 66% de empresas no rentables, hubiera sólo, por ejemplo, un 10%, es decir, 15 de las 150, entonces: $P(X=0)=0.59$. Por tanto, **los resultados de los informes periciales (de no ser erróneos) son mucho más consistentes con la situación de que hubiera muy pocas empresas con pérdidas**, es decir, que la mayoría fueran rentables. Evidentemente, a medida que bajamos el valor de M , es decir, el de número de empresas no rentables en la población, el valor de $P(X=0)$ aumenta.

2. ¿Cuál es la precisión de la estimación?

Aquí ya se empieza a complicar el asunto. El valor de probabilidad es un valor de estimación puntual. Pero esas estimaciones están sujetas a error, porque estamos empleando muestras en lugar de toda la población. Por tanto, para hacernos una mejor idea de lo "buenos" que son nuestros cálculos debemos dar alguna medida de precisión. La precisión es un concepto fácil de entender si pensamos en que el tribunal supremo ha escogido 5 casos al azar, pero si hubiera realizado ese proceso muchas veces, esos 5 casos serían diferentes unos de otros, por lo que habría una muestra diferente para cada experimento. Esa variabilidad muestral significa que tenemos una distribución de estimaciones puntuales, por lo que habrá una media y una varianza. Es por eso que X es una variable aleatoria cuyos momentos se pueden calcular. Existen varios métodos para computar los intervalos de confianza, algo en lo que no vamos a entrar aquí. Aplicando el procedimiento indicado en [Krishnamoorthy \(2006\)](#), el valor del **intervalo de confianza de dos colas al 95%** es de $(0 ; 0.43)$.

La interpretación de este cálculo es muy interesante, porque nos está diciendo que **la precisión de la estimación es muy**

baja, ya que cubre desde 0 a 0.43, es decir, casi la mitad del rango posible de valores de probabilidad. Si multiplicamos los valores del intervalo por el valor de la población N, nos da un intervalo de confianza para M, es decir, para **el número de empresas no rentables, que sería: (0 ; 64.5)**. Dicho de otro modo, con los resultados que el Tribunal Supremo obtiene del peritaje la precisión está en un rango entre 0 y aproximadamente 65 empresas con pérdidas. Esto no quiere decir que ese intervalo contenga siempre el parámetro a estimar, eso es una falacia en la interpretación de los intervalos de confianza, sino que simplemente, 95 de cada 100 veces que repitamos el experimento el valor poblacional estará contenido en el intervalo que salga (que será diferente cada vez). Como esto resultará quizá un poco engorroso de entender para algunos lectores, simplemente la conclusión que sacamos es que **con una muestra de 5 casos la amplitud del intervalo de confianza es tan grande que la estimación se vuelve muy imprecisa.**

3. ¿Ninguna de las fotovoltaicas es no rentable?

Esa afirmación que el artículo atribuye al Tribunal Supremo no tiene ningún tipo de base estadística. Con los datos de la estimación no se puede decir que en la población no haya ninguna empresa no rentable. En la muestra sí, claro, pero no en la población. Podemos computar el caso hipotético en el que el Tribunal podría haber llegado a esa conclusión. Para ello, y admitiendo un nivel de confianza del 95%, y con un error máximo admisible de 0.05 unidades, lo que sería un 5% en la escala de probabilidad $[0,1]$, **habría que haber muestreado aproximadamente 38 casos, y que de esos 38 casos todos hubieran sido rentables, es decir, que de 38 casos testigo, no hubiera ningún informe pericial indicando problemas de rentabilidad.** Es entonces en estas circunstancias cuando se podría afirmar que existen evidencias estadísticas de que ninguna de las 150 empresas afectadas tiene problemas de rentabilidad.

4. ¿Qué se puede decir con los datos que tenemos?

Llegados a este punto hay que ser prudentes con las conclusiones estadísticas. Con sólo una muestra de 5 casos, sobre una población de 150, y no habiendo ningún informe que

indique que las empresas no son rentables, lo que quizá podríamos decir es esto:

1. **El tamaño de la muestra es muy pequeño para obtener una estimación precisa, lo que indica que hay una evidencia similar de que no existe ninguna empresa con pérdidas y de que existen varias decenas de ellas que sí que las tienen.** Eso es decir muy poco, la verdad.

2. Sí que podemos concluir que menos de un 66% de empresas no tienen problemas de rentabilidad, es decir, hay evidencias suficientes para sostener que no hay 100 empresas (de las 150) con problemas de rentabilidad ya que el intervalo de confianza de $M=(0 ; 65)$, es decir, no incluye el 100, por lo que **con los datos en la mano no se confirma la hipótesis de que hay al menos un 66% de empresas con problemas.** Son menos empresas, según los datos.

3. **No se puede decir en ningún caso que no existan empresas con problemas de rentabilidad en la población.** Es más, con la precisión tan baja de la estimación lo más probable es que haya varias decenas de ellas. Para decir eso, habrían sido necesarias 38 empresas estudiadas con 38 informes positivos, es decir, todas rentables...pero en $n=38$, no en $n=5$.

4. Se podría calcular el tamaño de muestra necesario para haber obtenido estimaciones mucho más precisas. Con un margen de error de 0.05 unidades y al 95% de confianza, y si el Tribunal Supremo hubiera partido de la hipótesis de que un 66% son no rentables, entonces **lo adecuado hubiera sido escoger una muestra de ¡104 empresas!** Como eso sería prácticamente imposible, quizá el Tribunal debería de haber partido de una hipótesis más conservadora, por ejemplo que sólo el 10% no eran rentables. Aún así, el tamaño mínimo de muestra para una estimación precisa **tendría que haber sido de ¡72 casos!**

Conclusión

¿Se puede tomar una decisión judicial de ese calibre con datos tan poco precisos? Pues que cada uno valore la respuesta en base a lo que acabo de escribir en el post. Desde luego que **si alguno de nosotros (investigadores universitarios) enviamos**

un estudio a una revista científica con este tipo de estimaciones imprecisas, el editor nos rechazaría el artículo al instante, es más, quizá se sintiera insultado ante tal “despropósito”.

Desconozco si judicialmente este tipo de evidencias estadísticas se toma de otra manera. Al fin y al cabo las conclusiones derivadas de esa muestra tan pequeña no son del todo inútiles. Al menos se puede decir que el número de empresas no rentables es menor que las 100 que dicen que han sido refinanciadas. Pero poco más. En todos nuestros cálculos hemos supuesto que los informes periciales son correctos, algo de lo que desconfían los afectados. Y es lógico que lo hagan, porque si realmente hay tantas empresas con problemas de rentabilidad como dicen, el que en 5 casos hayan salido todos rentables, la probabilidad de que eso ocurra sería extremadamente baja, del orden de 0.0036; es decir, de cada 1000 veces que se repitiera el juicio, saldría menos de 4 veces ese resultado.

Sería posible, pero muy poco probable. Este problema se podría analizar con otro enfoque, a través de estadística bayesiana. Pero con la estadística frecuentista parece evidente pensar que hay demasiada imprecisión para tomar una decisión tan importante solamente por criterios estadísticos. Quizá no haya sido así, y el Tribunal Supremo haya tenido esta evidencia estadística como un factor más para la toma de decisiones. En cualquier caso, es comprensible que los afectados estén muy enfadados por el tratamiento “estadístico” realizado, dicho todo esto con el máximo de los respetos a la Justicia, claro está. □

Nota final: Todo este post se ha basado en el artículo referenciado al comienzo. Si esa información de partida es inexacta los resultados de mis cálculos obviamente variarán.

