

INTEPRETANDO LAS TABLAS DE CONTINGENCIA

Carmen Batanero. Carmen Díaz, María del Mar López-Martín y Gustavo R. Cañadas
UNO, 70, 2015

RESUMEN

En este trabajo se describe el sentido estadístico como unión de cultura, razonamiento estadístico y actitudes positivas. Recurriendo a los datos de supervivencia en el hundimiento del Titanic, se presenta un ejemplo que permitirá introducir distribuciones conjuntas, cálculo de probabilidades, representaciones gráficas y procedimientos que ayudan a analizar la asociación entre variables categóricas.

Palabras clave: Sentido estadístico, trabajo con proyectos, tablas de contingencia.

INTRODUCCION

Aunque actualmente encontramos la estadística en todas las etapas educativas, muchas investigaciones nos alertan de que los estudiantes no adquieren suficientes conocimientos estadísticos ni comprenden bien esta materia. Aunque den las definiciones de los conceptos y usen los algoritmos de cálculo con competencia, falta su comprensión profunda. No son capaces de elegir un procedimiento adecuado para analizar o representar los datos y no detectan los errores en los gráficos o información estadística que aparecen en los medios de comunicación.

Como argumentan Batanero, Díaz, Contreras y Roa (2013), es necesario desarrollar en los estudiantes el sentido estadístico, entendido como dominio reflexivo de la estadística. Una forma de lograrlo es utilizar en el aula el trabajo con proyectos o investigaciones, recorriendo todo el ciclo de investigación estadística: planteamiento de un problema, elección de un método de análisis, análisis de los datos e interpretación de los resultados.

Como ejemplo, en este artículo, planteamos una investigación utilizando los datos de supervivencia en el hundimiento del Titanic que nos lleva a trabajar con tablas de contingencia, un tema que aparece en Educación Secundaria y Bachillerato (MECD, 2015) con profundidad creciente:

- En 3º y 4º curso de Educación Secundaria Obligatoria se introducen estas tablas como recurso de cálculo de probabilidades, tanto en las matemáticas orientadas a las enseñanzas académicas, como en las orientadas a las enseñanzas aplicadas.
- En Bachillerato lo encontramos en las Matemáticas aplicadas a las Ciencias Sociales I y II y en las Matemáticas I (Bachillerato de Ciencias). En este caso se sugiere el estudio formal del tema incluyendo las ideas de dependencia e independencia, distribuciones marginales y condicionadas.

Por tanto, presentamos una investigación que puede trabajarse a distintos niveles de profundidad, desde uno elemental, asequible a los chicos y chicas de Educación

Secundaria Obligatoria hasta otro más abstracto y completos propio del Bachillerato. Para comenzar, reflexionamos sobre los componentes del sentido estadístico.

SENTIDO ESTADÍSTICO

Nuestro modelo de sentido estadístico se concibe como la unión de tres componentes relacionados e igualmente importantes: la cultura estadística, el razonamiento estadístico y unas buenas actitudes hacia la estadística.

- La *cultura estadística* implica conocer y comprender las ideas estadísticas fundamentales (Burrill y Biehler, 2011): dato, representación de datos, variabilidad, distribución, asociación y correlación, probabilidad, muestreo e inferencia. Además de su relevancia en la estadística, todas estas ideas son necesarias para interpretar la información estadística en los medios de comunicación, en otras materias que cursa el estudiante y en su futura vida profesional.
- *El razonamiento estadístico* es un tipo específico de razonamiento que parte de los datos para detectar tendencias o fuentes de variabilidad, para extrapolar los datos y realizar inferencias de muestras a poblaciones o para tomar decisiones acertadas en situaciones inciertas. Se basa en la aceptación de la variabilidad aleatoria y su integración en el análisis de la situación; además se esfuerza en detectar las fuentes de variabilidad y explicarlas, medir la variabilidad debida a diversas fuentes y reducirla cuando sea posible. Un instrumento importante del razonamiento estadístico es la *transnumeración*, o capacidad de elegir un cambio de representación de los datos que permita visualizar y descubrir relaciones en los mismos (Wild & Pfannkuch, 1999).
- Unas *actitudes positivas* suponen la valoración del método y estudios estadísticos como instrumento de investigación y descubrimiento, así como el interés y esfuerzo personal por aprender esta materia, la constancia, curiosidad y rigor en el trabajo con la estadística.

El desarrollo efectivo del sentido estadístico en los estudiantes se favorece con una enseñanza basada en investigaciones y proyectos, que permite dotar de sentido a los diversos objetos estadísticos e involucra a los estudiantes en el ciclo de investigación y modos propios de razonamiento estadístico. Además desarrolla su interés por la materia, espíritu crítico, valoración de la evidencia e iniciativa personal.

UNA INVESTIGACIÓN EN EL AULA: EL HUNDIMIENTO DEL TITANIC

Una posible investigación en el aula parte de los datos de supervivencia de los pasajeros del Titanic, disponibles en Internet (por ejemplo, en <http://www.statsci.org/data/general/titanic.txt>). Si trabajamos con alumnos de Educación Secundaria Obligatoria, se le facilitarán las tablas de frecuencia formadas, sin embargo en el caso de Bachillerato, los alumnos podrán trabajar directamente con el fichero y recurriendo a software estadísticos formarán ellos mismos las tablas.

Más concretamente, en este proyecto se persiguen los siguientes objetivos:

- Interpretar una tabla de contingencia. Interpretar las frecuencias dobles, marginales y condicionales.

- Calcular probabilidades sencillas a partir de datos representados en una tabla de contingencia.
- Representar gráficamente datos una tabla de contingencia mediante el diagrama de barras adosadas y gráfico de mosaicos.
- Introducción intuitiva al concepto de asociación con variables cualitativas, diferenciar la asociación de la independencia y comprender las propiedades de la independencia.
- Adquirir algunos procedimientos sencillos para valorar la asociación de los datos en una tabla de contingencia.

Inicio de la actividad

La tragedia del Titanic, se ha hecho famosa por películas y exposiciones, como la existente en Belfast. Se puede comenzar preguntando a los alumnos qué conocen del tema o pidiéndoles que encuentren información en Internet. Una primera actividad sería calcular la proporción global de supervivientes: sólo 451 de 1313 pasajeros (34,3%). No hubo bastantes botes salvavidas y el resto murió ahogado, antes de que el buque más cercano acudiera al rescate.

La segunda tarea sería analizar si la “normal moral” de salvar primero a las mujeres se cumplió en este caso. Se proporcionaría a los alumnos la Tabla 1 para que analicen esta cuestión. Es claro que hubo mujeres que se ahogaron; ¿se puede decir que no hubo preferencia o que la hubo? Se pedirá a los alumnos propuestas para comparar la supervivencia de hombres y mujeres.

Tabla 1. Distribución de género entre supervivientes y ahogados

	Sobrevive	No sobrevive	Total
Mujer	308	154	462
Varón	143	708	851
Total	451	862	1313

Algunos alumnos podrían representar los datos en un diagrama de barras adosado (Figura 1) para visualizarlos con más claridad. El gráfico muestra que hubo más mujeres supervivientes que ahogadas y al contrario entre los varones. Pero al ser el número de hombres y mujeres diferentes, se dificulta la comparación. Se les puede hacer observar que casi 3 de cada 4 mujeres se salvaron y hombres alrededor 1 de cada 8; estas *razones* deberían ser iguales o parecidas, en caso de no preferencia.

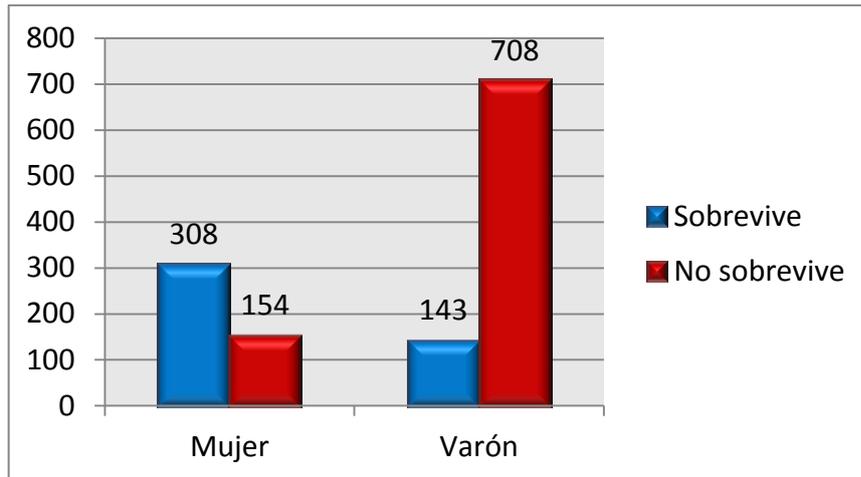


Figura 1. Supervivencia de mujeres y hombres en el Titanic

En términos de probabilidades, si llamamos M y H a los sucesos “ser hombre”, “ser mujer” y S al suceso “supervivencia”, a partir de la Tabla 1, podemos calcular tres tipos diferentes de probabilidades:

- Probabilidad simple; por ejemplo, de supervivencia $P(S) = \frac{462}{1313} = 0,352$
- Probabilidad condicional; por ejemplo supervivencia en las mujeres. $P(S|M) = \frac{308}{462} = 0,667$ y en los hombres $P(S|H) = \frac{143}{851} = 0,168$.
- Probabilidad conjunta: de ser mujer y salvarse $P(S \cap M) = \frac{308}{1313} = 0,235$.

El profesor puede hacer notar a los chicos la diferencia entre estas tres probabilidades y pedirles que calculen otras; también puede resaltar que un error común es confundir $P(S|M)$ y $P(M|S)$, que son distintas, pues $P(M|S) = \frac{308}{451} = 0,683$. El diagrama en árbol (Figura 2) ayuda a visualizar el hecho de que de todos los que sobreviven la mayoría son mujeres. Como vemos mediante este diagrama, o incluso con la misma tabla de contingencia, el cálculo de la probabilidad inversa $P(M|S)$ se reduce a aplicar la regla de Laplace, en vez del teorema de Bayes.



Figura 2. Diagrama en árbol representando los datos

Es claro también la mayor probabilidad de salvarse siendo mujer que siendo hombre; de hecho la *razón de probabilidades*:

$$\frac{P(S|M)}{P(S|H)} = \frac{0,352}{0,168} = 2,09$$

Quiere decirse que la probabilidad de salvarse fue doble para las mujeres que para los hombres o que de cada tres personas salvadas dos fueron mujeres.

Para trabajar con porcentajes, el profesor puede pedir a los chicos calcular el porcentaje de supervivencia de hombres (16,8%) y mujeres (66,7%) y comparar con el porcentaje global (35,2%) calculado anteriormente. Si hubiese *independencia* entre el género y la supervivencia, estos tres valores deberían ser iguales. Es claro de la comparación que se favoreció a las mujeres frente a los hombres. Un diagrama de barras apiladas (Figura 3) ayuda a visualizar las dos distribuciones de datos. En los chicos mayores, se puede aprovechar este momento para introducir la idea de distribución condicional (supervivientes y ahogados en hombres o en mujeres) y distribución marginal (supervivientes y ahogados globalmente). Si hubiese independencia las dos distribuciones de frecuencias relativas (o porcentajes) condicionales deben coincidir entre sí y con la marginal.

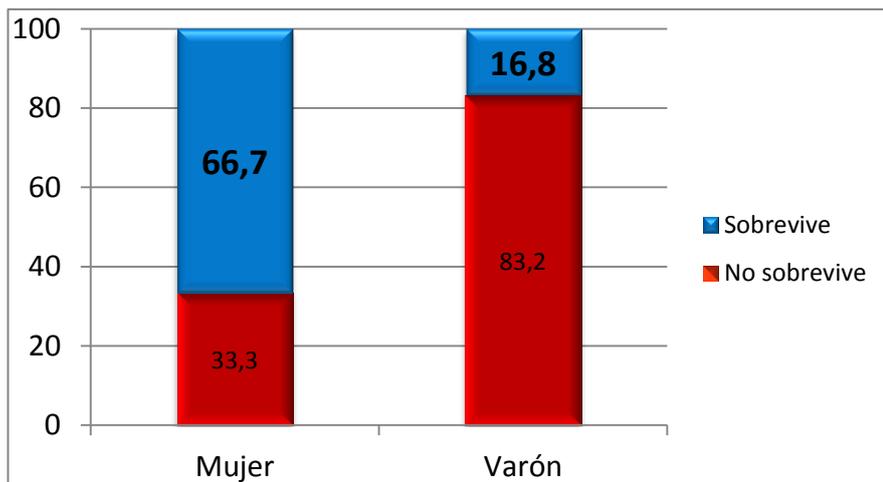


Figura 3. Distribuciones condicionales de supervivencia o no en mujeres y hombres

Los alumnos pueden realizar también un gráfico de mosaicos a mano con ayuda de una regla graduada (Figura 4). Este gráfico es una representación mediante áreas de las proporciones: la longitud del eje vertical se divide proporcionalmente al número de mujeres y hombre, formando dos rectángulos de la misma base. Cada uno se divide en proporción al número de supervivientes y ahogados, usando dos colores para diferenciarlos. En este gráfico representamos la distribución marginal de hombre y mujeres (eje vertical), las distribuciones condicionales de mujeres y hombres supervivientes o no (dos rectángulos horizontales, divididos a su vez) y la distribución conjunta (dada por los rectángulos pequeños).

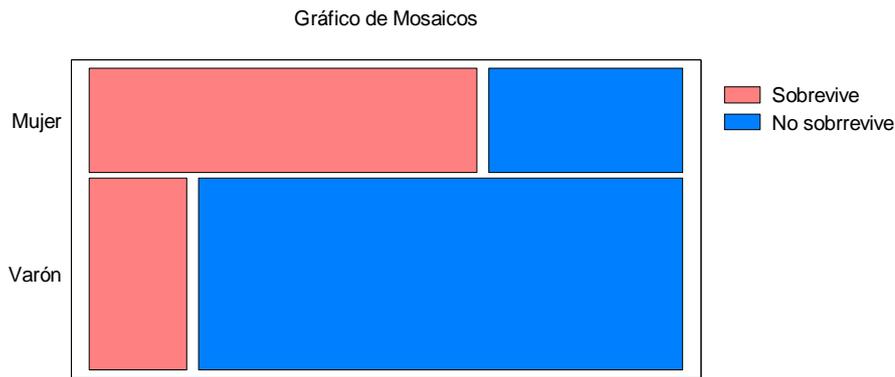


Figura 4 Gráfico de mosaicos de la supervivencia de hombres y mujeres

Podríamos trabajar a continuación con los datos de supervivencia según la clase social. El profesor podría facilitar a los estudiantes la Tabla 2 y solicitar las *frecuencias esperadas* en las celdas si la supervivencia *fuese independiente* de la clase social, obteniendo como resultado la Tabla 3.

Tabla 2. Distribución de pasajeros según supervivencia y clase social

	Sobrevive	No sobrevive	Total
Primera clase			322
Segunda clase			280
Tercera clase			711
Total	451	862	1313

Puesto que global de supervivencia fue $\frac{451}{1313} = 0,352$, debiera repetirse en cada clase social. Así, para la primera clase el número esperado de pasajeros sería: $322 \times 0,352 = \frac{451 \times 322}{1313} = 110,6$, que redondeado serían 111 pasajeros. De igual modo se calcula el número esperado en el resto de las celdas, observando la siguiente relación:

$$\text{Frecuencia esperada} = \frac{\text{Total fila} \times \text{Total columna}}{\text{Total datos}}$$

Se puede hacer notar que en cada fila sólo hay que calcular una celda, pues el total de las filas no varía. Además, la última fila se puede deducir cuando se conocen las dos primeras.

Tabla 3. Distribución esperada de supervivencia por clase social en caso de independencia

	Sobrevive	No sobrevive	Total
Primera clase	111	211	322
Segunda clase	96	184	280
Tercera clase	244	467	711
Total	451	862	1313

Al comparar con los *datos observados* de supervivencia (Tabla 4) vemos una gran diferencia con los esperados (Tabla 3), que se hace más visible en el gráfico de mosaicos (Figura 5). Sobreviven más de primera clase y muchos menos de tercera respecto a lo esperado. Por tanto hubo discriminación y la supervivencia *estuvo asociada* a la clase social.

Tabla 4. Distribución observada de pasajeros según supervivencia y clase social

	Sobrevive	No sobrevive	Total
Primera clase	194	128	322
Segunda clase	119	161	280
Tercera clase	138	573	711
Total	451	862	1313

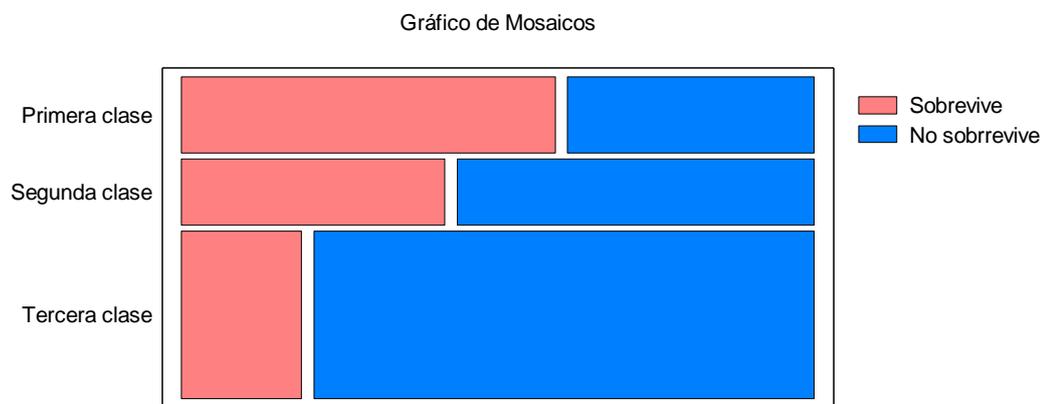


Figura 5. Gráfico de mosaicos de supervivencia según clase social

PARA FINALIZAR

En la investigación propuesta se ha trabajado con muchas de las ideas fundamentales en estadística: partimos de *datos* bivariantes, consideramos varios tipos de distribución (conjunta, condicional, marginal) y sus relaciones. Nos introducimos en las ideas de asociación e independencia, utilizando varios procedimientos para analizarla (comparación de probabilidades y proporciones, razón de probabilidades, comparación de frecuencias observadas y esperadas en caso de independencia).

El razonamiento estadístico se fomenta con la transnumeración, al pasar de la tabla de datos a diferentes tipos de gráfico que nos permite visualizar las relaciones en los datos y las fuentes de variabilidad. Además se ha recorrido todo el ciclo de un estudio estadístico, comenzando con un problema, eligiendo y utilizando modelos matemáticos y estadísticos para resolverlo y enfatizando la interpretación de los resultados. Finalmente, el tema es motivador para los alumnos ya que les lleva a observar la utilidad de la estadística para realizar investigaciones y responder preguntas y fomenta su actitud crítica frente a la información. El profesor interesado puede ampliar lo expuesto sobre tablas de contingencia en Batanero, Cañadas, Contreras & Arteaga (2012).

REFERENCIAS

BATANERO, C.; CAÑADAS, C.; CONTRERAS, J. M.; ARTEAGA, P. (2012): «¿Es sencilla la interpretación de las tablas de contingencia?». *Gamma*, 12, pp. 27-34.

BATANERO, C.; DÍAZ, C.; CONTRERAS, J. M.; ROA, R. (2013): «El sentido estadístico y su desarrollo». *Números* 83, pp. 7-18.

BURRILL, G.; BIEHLER, R. (2011): «Fundamental statistical ideas in the school curriculum and in training teachers. En C. BATANERO; G. BURRILL; C. READING (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education - A joint ICMI/IASE study* (pp. 57-69). Dordrecht: Springer.

MECD (2015): *Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato*. Madrid: Autor.

WILD, C.; PFANNKUCH, M. (1999): «Statistical thinking in empirical enquiry». *International Statistical Review*, 67(3), pp. 221-248.