

El sentido estadístico y su desarrollo

Carmen Batanero (Universidad de Granada. España)

Carmen Díaz (Universidad de Huelva. España)

José Miguel Contreras (Universidad de Granada. España)

Rafael Roa (Universidad de Granada. España)

Artículo solicitado a los autores por la revista

Resumen

En este trabajo concebimos el sentido estadístico como unión de la cultura y razonamiento estadístico. Se describen las ideas estadísticas fundamentales como componentes de la cultura estadística y los componentes del razonamiento estadístico. También se sugiere que el desarrollo efectivo en los estudiantes se favorece especialmente 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 de razonamiento estadístico, desarrollando su espíritu crítico e iniciativa personal.

Palabras clave

Sentido estadístico, cultura estadística, razonamiento estadístico.

Abstract

In this paper we conceive statistical sense as the union of statistical literacy and statistical reasoning. We describe the fundamental statistical ideas that constitute statistical literacy and the components of statistical reasoning. We also suggest that an effective development of statistical sense is specially favoured with a teaching based on investigations and projects. This method gives sense to the diverse statistical objects and involves students in the research cycle and the statistical reasoning modes, at the times that develop their critical spirit and personal initiative.

Keywords

Statistical sense, statistical literacy, statistical reasoning.

1. Introducción

Aunque la estadística se enseña hoy día en todos los niveles educativos, al ser una herramienta fundamental en la vida personal y profesional, son muchos los estudiantes, que finalizan los cursos de estadística sin comprender correctamente o ser capaces de aplicar los conceptos y procedimientos estadísticos, como se muestra en la amplia investigación sobre el tema (ver, por ejemplo, el resumen de estas investigaciones en Shaughnessy, 2007).

Esta situación paradójica plantea un problema didáctico, pues es claro que la enseñanza actual transmite una estadística sin sentido para los estudiantes. Como solución al mismo, son muchos los investigadores que insisten en la especificidad de la estadística (frente a otras ramas de la matemática). Moore (1991) incluso la considera como una disciplina científica autónoma, con sus métodos específicos de razonamiento. Para ello argumenta que la relación entre estadística y matemáticas no es biunívoca: la estadística toma conceptos matemáticos para el desarrollo de sus métodos, en cambio la matemática no usa conceptos estadísticos. Cabriá (1994) también coincide en que la estadística tiene



un modo propio de razonamiento, el método estadístico, que es necesario enseñar a los estudiantes. La estadística tiene también controversias específicas (por ejemplo, sobre el significado de la probabilidad), y la posición que un estadístico toma sobre ellas tiene un impacto inmediato en su práctica.

La expresión “*sentido numérico*” se encuentra con frecuencia en la investigación en didáctica de la matemática e incluso en los diseños curriculares. Por ejemplo, en el Decreto de Enseñanzas Mínimas de la Educación Primaria (MEC, 2006) encontramos la siguiente descripción del sentido numérico:

“el dominio reflexivo de las relaciones numéricas que se puede expresar en capacidades como: habilidad para descomponer números de forma natural, comprender y utilizar la estructura del sistema de numeración decimal, utilizar las propiedades de las operaciones y las relaciones entre ellas para realizar mentalmente cálculos” (p. 43096).

Observamos en esta descripción y otras semejantes un doble componente de adquisición de conocimiento y razonamiento. Para el caso de la estadística no ha habido una preocupación semejante entre los educadores estadísticos por definir o listar los componentes de lo que sería el “*sentido estadístico*”. Por un paralelismo con la acepción dada a “*sentido numérico*” hemos de aceptar que el sentido estadístico debiera tener un doble componente de conocimiento (o cultura) y razonamiento.

En consecuencia en lo que sigue concebimos el *sentido estadístico* como unión de la *cultura estadística* y el *razonamiento estadístico*. Asimismo, también consideramos que la cultura estadística implica la comprensión adecuada de las ideas estadísticas fundamentales (Burrill y Biehler, 2011), pues estas ideas aparecen en la mayoría de las situaciones en que hay que aplicar la estadística; por tanto son necesarias para enfrentarse con éxito a dichas situaciones. Además, pueden ser enseñadas con diversos niveles de formalización y, por tanto, son asequibles en cualquier nivel educativo, siendo potentes como herramientas de modelización estadística. En segundo lugar, se requiere un razonamiento específico, el razonamiento estadístico que permite tomar decisiones adecuadas o efectuar predicciones a partir de datos y en presencia de incertidumbre. La competencia de análisis de datos se ve hoy en día facilitada por la abundancia de software, tanto para el almacenamiento y transmisión de datos, como para el cálculo y graficación. Sin embargo la competencia en análisis de datos no siempre implica un grado adecuado de razonamiento estadístico, que es más difícil de transmitir.

El objetivo de este trabajo es desarrollar estas ideas y proponer el trabajo con proyectos e investigaciones como estrategia metodológica para desarrollar el sentido estadístico de los estudiantes en los diferentes niveles educativos.

2. Cultura estadística

La importancia que actualmente recibe la enseñanza de la estadística se debe a la necesidad, reclamada por la UNESCO y otras instituciones de proporcionar una *cultura estadística* que permita al ciudadano participar en la sociedad de la información. Dicha cultura es necesaria en actividades tan habituales como la lectura de la prensa o la interpretación de información en Internet, la participación en encuestas o elecciones o la interpretación de un diagnóstico médico. El término “statistical literacy” ha ido surgiendo de forma espontánea entre los estadísticos y educadores estadísticos, para resaltar el hecho de que la estadística se considera hoy día como parte de la herencia cultural necesaria para el ciudadano educado. Gal (2002) propone que la cultura estadística implica dos competencias relacionadas:

a) Capacidad para interpretar y evaluar críticamente la información estadística, los argumentos apoyados en datos o los fenómenos que las personas pueden encontrar en diversos contextos, incluyendo los medios de comunicación, pero no limitándose a ellos, y b) capacidad para discutir o comunicar sus opiniones respecto a tales informaciones estadísticas cuando sea relevante (Gal, 2002, pp. 2-3).

Watson (2006) propone los siguientes elementos de la cultura estadística: (a) El desarrollo del conocimiento básico de los conceptos estadísticos y probabilísticos; (b) La comprensión de los razonamientos y argumentos estadísticos cuando se presentan dentro de un contexto más amplio de algún informe en los medios de comunicación o en el trabajo; (c) Una actitud crítica que se asume al cuestionar argumentos que estén basados en evidencia estadística. En el proyecto GAISE (Franklin et al., 2007) se concibe la misma como la comprensión del lenguaje básico de la estadística (saber el significado de los términos y símbolos estadísticos y ser capaz de leer gráficos estadísticos) y comprender las ideas fundamentales de estadística.

Vemos que los diferentes autores que la definen incluyen tanto la parte de conocimientos teóricos formales relacionados con la estadística, así como la capacidad de afrontar un problema o situación en el que se pongan en juego estos conocimientos.

3. Ideas estadísticas fundamentales

Los modelos descritos de cultura estadística, además de asumir una actitud crítica frente a la información estadística, coinciden en la necesidad de un conocimiento del lenguaje estadístico y las ideas estadísticas fundamentales. Sería entonces importante identificar cuáles son estas ideas fundamentales. Burrill y Biehler (2011), basándose en un estudio detallado de diversos marcos teóricos educativos y el currículo de estadística en diversos países proponen las siguientes: datos, gráficos, variabilidad aleatoria, distribución, asociación y correlación, probabilidad, muestreo e inferencia.

Datos. Moore (1991) definió la estadística como la ciencia de los datos y señaló que el objeto de la estadística es el razonamiento a partir de datos empíricos, subrayando la importancia del contexto. Mientras que en la enseñanza de otras ramas de matemáticas los datos y contextos suelen ser imaginarios y el interés se centra en los conceptos, el contexto de los datos es esencial en estadística. Además, los estudiantes no están acostumbrados a trabajar con datos de situaciones reales, que frecuentemente requieren de interpretaciones y razonamientos de alto nivel. La aleatoriedad de las situaciones hace que los resultados no sean únicos, presentándose mayor variabilidad en los datos estadísticos que otras áreas de las matemáticas (Sánchez y Batanero, 2011).

Gráficos. Debido a su presencia en los medios de comunicación e Internet, el aprendizaje de los gráficos estadísticos es importante. Por su papel esencial en la organización, descripción y análisis de datos, las tablas y gráficos son un instrumento esencial de transnumeración, uno de los modos esenciales de razonamiento estadístico que consiste en obtener una nueva información de un conjunto de datos al cambiar el sistema de representación (Wild y Pfannkuch, 1999).

Variabilidad aleatoria. Aunque en otras ramas de matemáticas se usan variables, se supone que los datos se ajustan perfectamente a un modelo y no suele haber estudio de la bondad de ajuste o de los residuos del modelo. El estudio de la variabilidad es característico de la Estadística, estudiándose tanto el modelo como los residuos (Engel y Sedlmeier, 2011). La Estadística permite buscar explicaciones y causas de la variabilidad para poder hacer predicciones, por lo que dos fines importantes de la enseñanza de la Estadística son que los estudiantes perciban la variabilidad y manejen modelos que



permitan controlarla y predecirla (Reading y Shaughnessy, 2004).

Distribución. Otra característica esencial del análisis estadístico es que trata de describir y predecir propiedades de los agregados de datos y no de cada dato aislado (Bakker y Gravemeijer (2004). Por ello la enseñanza de la Estadística ha de desarrollar la capacidad de leer, analizar, criticar y hacer inferencias a partir de distribuciones de datos (Shaughnessy, 2007). El razonamiento distribucional implica también conectar los datos (distribución de datos), la población de donde se tomaron (distribución de probabilidad) y las posibles muestras de la misma (distribución muestral).

Asociación y correlación. Mientras que en una dependencia funcional a cada valor de una variable X (independiente) corresponde un solo valor de otra variable Y (dependiente), en el estudio de la asociación a cada valor de X corresponde una distribución de valores de Y , por lo que este concepto amplía el de dependencia funcional. La importancia del concepto de asociación en la toma de decisiones en ambiente de incertidumbre es alta, pero la investigación en psicología muestra que los adultos no suelen emplear las reglas matemáticas, sino estrategias intuitivas, con frecuencia incorrectas en los juicios de asociación. Por ejemplo la creencia infundada en la transitividad del coeficiente de correlación es común en muchos estudiantes (Castro-Sotos, Van Hoof, Van den Noortgate y Onghena, 2009).

Probabilidad. La característica principal de la Estadística es hacer uso de modelos aleatorios, a diferencia de otras ramas de la matemática donde se usan modelos deterministas. Pero, al contrario que para otros conceptos matemáticos, no hay una única acepción de la probabilidad. Las tres aproximaciones principales para la escuela son:

- En la *concepción clásica*, se define la probabilidad de un suceso como el cociente entre el número de casos favorable al suceso y el número de todos los casos posibles, siempre que todos sean equiprobables.
- En el *enfoque frecuencial* se obtiene una estimación experimental de la probabilidad. Su valor teórico sería el límite de la frecuencia relativa de aparición del suceso al realizar la experiencia un número infinito de veces en las mismas condiciones. Un aspecto importante en este enfoque es comprender la diferencia entre probabilidad (valor teórico constante que nunca alcanzamos) y frecuencia relativa (estimación experimental de la probabilidad, que puede cambiar de una estimación a otra). También hay que entender que los resultados de una experiencia son impredecibles, pero se puede predecir el comportamiento general de un gran número de resultados (Batanero, Henry y Parzysz, 2005).
- En el *enfoque subjetivo*, la probabilidad no es una propiedad objetiva de los sucesos, sino una percepción o grado de creencia en la verosimilitud de la persona que asigna la probabilidad sobre la plausibilidad de ocurrencia del suceso. Muchos problemas de toma de decisión o elaboración de un juicio son abiertos o tienen más de una posible decisión y en su solución intervienen tanto factores matemáticos como extra matemáticos. Esta concepción de la probabilidad sería adecuada para modelizar este tipo de situaciones.

Muestreo e inferencia. Relacionar las características de las muestras con las de la población que representan es el principal fin de la estadística y nos sirve para decidir qué datos recoger y para obtener conclusiones con algún grado de probabilidad. Varios autores sugieren que es posible una comprensión informal de la inferencia, desde la secundaria. Para ello se comenzaría por la discriminación entre la posición central y variabilidad en las distribuciones de datos y el uso de estas dos características para decidir cuándo dos distribuciones son iguales o diferentes (Rubin, Hammerman y Konold, 2006). Más adelante se puede hacer una aproximación al contraste de hipótesis en la forma siguiente (Rossman, 2008): (a) Comenzar con una hipótesis sobre los datos; (b) usar la simulación para concluir que los datos observados son poco plausibles si la hipótesis es cierta;

y (c) rechazar la hipótesis inicial basándose en los resultados. Con apoyo de la simulación también puede llevarse al estudiante a comprender el concepto de estimación por intervalo.

4. Pensamiento y razonamiento estadístico

Además de la comprensión de las anteriores ideas, su uso adecuado en la resolución de problemas estadísticos requiere del desarrollo del razonamiento estadístico¹. Uno de los modelos para describir este razonamiento es debido a Wild y Pfannkuch (1999), quienes lo conciben como la suma de cuatro dimensiones: (a) El ciclo de investigación, que consiste en la serie cíclica de pasos a seguir desde que se plantea un problema estadístico hasta que se resuelve o bien se modifica y que es bastante similar al proceso general de resolución de problemas; (b) los modos fundamentales de razonamiento estadístico; (c) el ciclo de interrogación, que se aplica constantemente en la solución de problemas estadísticos, tanto a nivel global como en cada posible paso y consiste en la búsqueda y comprobación sucesivas de explicaciones, hipótesis o preguntas, desde los datos, los análisis realizados o los resultados; y (d) una serie de actitudes, como el escepticismo, la mentalidad abierta, la perseverancia, el espíritu crítico o la curiosidad. Los modos fundamentales de razonamiento estadístico, según estos autores son los siguientes:

- *Reconocer la necesidad de los datos:* Mientras que en otras ramas de las matemáticas los datos son anecdóticos, la base de la investigación estadística es el hecho constatado de que muchas situaciones de la vida real sólo pueden ser comprendidas a partir del análisis de datos que han sido recogidos en forma adecuada. La experiencia personal o la evidencia de tipo anecdótico no es fiable y puede llevar a confusión en los juicios o toma de decisiones. Se trata de basarse en la evidencia proporcionada por los datos empíricos.
- *Transnumeración:* Este término indica la comprensión que surge al cambiar la representación de los datos; por ejemplo, al pasar de una lista desordenada de datos al histograma se visualiza el rango y la moda. En general puede hablarse de tres tipos de transnumeración: (1) la que produce al definir una medida que “captura” las cualidades o características de un cierto fenómeno; por ejemplo, cuando utilizamos una escala de actitudes para “medir” la actitud hacia la matemática de un conjunto de estudiantes; (2) al pasar de los datos brutos a una representación tabular o gráfica que permita extraer sentido de los mismos, como el ejemplo citado del histograma; (3) al “traducir” el significado que el estadístico ha captado y que surge de los datos, en forma que sea comprensible a otras personas; por ejemplo, cuando explicamos en forma simplificada las conclusiones que se deducen de un gráfico de caja o de los resultados de un contraste estadístico.
- *Percepción de la variación.* Como hemos indicado, la variabilidad aleatoria es una idea fundamental en estadística. Un componente del razonamiento estadístico es la identificación de las fuentes que producen dicha variación, que pueden ser la propia medida, los datos, el muestreo, el análisis, ser producida por factores específicos (como la diferencia de altura en chicos y chicas). El trabajo estadístico también asume como variación aleatoria aquella cuyas fuentes no quedan explicadas. El razonamiento estadístico permite buscar explicaciones y causas para la variación y realizar inferencias y predicciones, con un cierto margen de error, teniendo en cuenta la variación no explicada o aleatoria.
- *Razonamiento con modelos estadísticos.* Al igual que en otras ramas de las matemáticas, el trabajo estadístico es esencialmente un proceso de modelización. La principal diferencia es la presencia de aleatoriedad, así como la relevancia que adquieren los modelos probabilísticos como por ejemplo, la curva normal. También se utilizan modelos matemáticos como

¹ La adquisición de una competencia mínima de análisis de datos, no es hoy día un problema, pues se ve favorecida por la abundancia de software “amistoso”.



gráficos, o funciones (por ejemplo, en regresión); todos ellos han de contemplarse como representaciones de la realidad e instrumentos para comprenderla, diferenciando el modelo de la realidad.

- *Integración de la estadística y el contexto:* Debido a la importancia que adquiere el contexto, la capacidad de integrarlo es también un componente esencial del razonamiento estadístico. Este tipo de razonamiento aparece especialmente en las fases iniciales (planteamiento del modelo) y finales (interpretación del modelo en la realidad) del ciclo de modelización.

El desarrollo del razonamiento y la producción de sentido son resaltados para el caso de la estadística por Shaughnessy, Chance y Kranendonk (2009), sugiriendo las siguientes competencias de razonamiento para reforzar la comprensión de las ideas fundamentales: (a) análisis de problemas, búsqueda de patrones y relaciones en los datos, percepción de la estructura y planteamiento de conjeturas; (b) elegir y evaluar estrategias, aplicando el ciclo interactivo de investigación estadística; (c) buscar y utilizar conexiones entre las conclusiones y el contexto; (d) reflexionar sobre si la solución es razonable y suficiente.

5. Desarrollo del sentido estadístico a través del trabajo con proyectos

El sentido estadístico, como unión de la cultura y razonamiento estadísticos debe construirse en forma progresiva desde la educación primaria, secundaria, bachillerato y hasta la universidad. En este sentido, las nuevas propuestas curriculares proporcionan una oportunidad de introducir gradualmente ideas estadísticas desde la educación primaria, aumentando el nivel de formalización progresivamente. Pensamos que la mejor forma de ayudar al estudiante a desarrollar su sentido estadístico es basar las clases de estadística en el trabajo con proyectos, bien planteados por el profesor o escogidos libremente por los alumnos. En lugar de introducir los conceptos y técnicas descontextualizadas, o aplicadas únicamente a problemas tipo, difíciles de encontrar en la vida real, se trataría de presentar las diferentes fases de una investigación estadística: planteamiento de un problema, decisión sobre los datos a recoger, recogida y análisis de datos y obtención de conclusiones sobre el problema planteado.

A continuación analizamos un ejemplo de proyecto adaptado de Shaughnessy, Chance y Kranendonk (2009), quienes a la vez se inspiraron en un debate surgido a partir de una noticia en el Daily News el 12 de Marzo de 2012, con el título “Will women some day run faster than men in a marathon? The answer might surprise you”². El proyecto comenzaría presentando a los alumnos la citada noticia y el debate surgido y planteando la siguiente pregunta: ¿Llegarán a superar las mujeres a los hombres en las pruebas de 200 metros lisos en las olimpiadas? ¿En qué año?

Les proporcionaríamos los datos de los tiempos alcanzados por los ganadores de la medalla de oro en las olimpiadas, disponibles en la Tabla 1. Un primer análisis de estos datos puede servir para trabajar las ideas de distribuciones bivariantes y el uso de estadísticos descriptivos para la descripción de cada uno de los grupos. Si se observa, la mejora de las marcas en los hombres se ha pasado de 22,2 segundos a 19,32 en 112 años; para las mujeres el cambio ha sido de 24,4 a 21,8 en 64 años; en mucho menos tiempo que los hombres han ganado aproximadamente 3 segundos.

Sin embargo, la disminución de los tiempos no es sistemática; se observan fluctuaciones y cuando hay decrecimiento de una olimpiada a otra, no es constante. Incluso el mismo atleta (ver el ejemplo de Usain Bolt) puede empeorar su marca de una olimpiada en la siguiente. Nos encontramos

² Ver noticia en <http://www.nydailynews.com/sports/more-sports/running-doc-women-run-faster-men-marathon-answer-surprise-article-1.1037443>

delante de un problema en que *necesitamos datos* y donde *percibimos una variación aleatoria* que tratamos de explicar; pero el listado de datos en sí mismo es poco explicativo. Necesitamos representar los datos de un modo adecuado para observar mejor la tendencia; es decir, hemos de realizar un proceso de transnumeración.

Una posible representación de los datos sería un diagrama de dispersión (Figura 1), en el cuál se coloree o se represente con un carácter diferenciado las marcas de las mujeres y de los hombres. Observamos claramente que las mujeres tardan más que los hombres en todas las olimpiadas y que las dos series de puntos son decrecientes; se trataría de una relación inversa entre el tiempo en segundos tardado en la prueba de 100 metros y el tiempo de calendario (año de celebración de la olimpiada); conforme pasa el tiempo mejoran las marcas. Asimismo se observa una pendiente más pronunciada en la serie de puntos correspondiente a las mujeres.

Año	Atleta	País	Tiempo	Atleta	País	Tiempo
1900	Walter Tewksbury	USA	22,2			
1904	Archie Hahn	USA	21,6			
1908	Robert Kerr	Canadá	22,6			
1912	Ralph Craig	USA	21,7			
1920	Allan Woodring	USA	22			
1924	Jackson Scholz	USA	21,6			
1928	Percy Williams	Canadá	21,8			
1932	Eddie Tolan	USA	21,12			
1936	Jesse Owens	USA	20,7			
1948	Mel Patton	USA	21,1	Fanny Blankers	NED	24,4
1952	Andy Stanfield	USA	20,81	Marjorie Jackson	AUS	23,89
1956	Bobby Morrow	USA	20,75	Betty Cuthbert	AUS	23,55
1960	Livio Berruti	ITA	20,62	Wilma Rudolph	USA	24,13
1964	Henry Carr	USA	20,36	Edith McGuire	USA	23,05
1968	Tommie Smith	USA	19,83	Irena Szewinska	Poland	22,58
1972	Valeriy Borzov	USSR	20	Renate Stecher	GDR	22,4
1976	Don Quarrie	JAM	20,23	Barbel Eckert	GDR	22,37
1980	Pietro Mennea	ITA	20,19	Barbel Wockel	GDR	22,03
1984	Carl Lewis	USA	19,8	Valerie Brisco	USA	21,81
1988	Joe DeLoach	USA	19,75	Florence Griffith	USA	21,34
1992	Mike Marsh	USA	20,01	Gwen Torrence	USA	21,81
1996	Michael Johnson	USA	19,32	Marie-Jose Perec	FRA	22,12
2000	Konstantinos Kenteris	GRE	20,09	Marion Jones	USA	21,84
2004	Shawn Crawford	USA	19,79	Veronica Campbell	JAM	22,05
2008	Usain Bolt	JAM	19,3	Veronica Campbell	JAM	21,74
2012	Usain Bolt	JAM	19,32	Allyson Felix	USA	21,88

Tabla 1. Datos sobre tiempos de los ganadores de las pruebas olímpicas de 200 metros



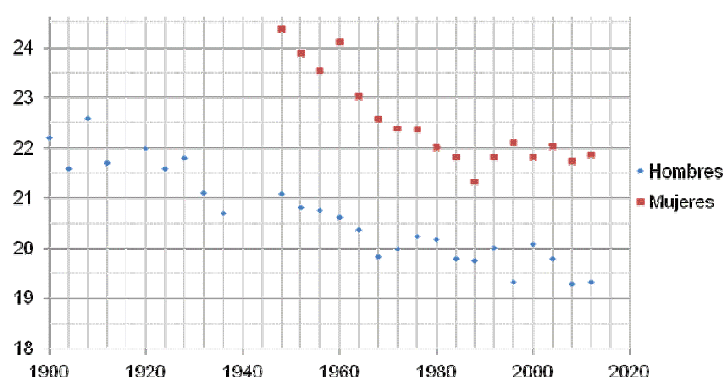


Figura 1. Marcas olímpicas en las pruebas de 200 metros

Aunque el proyecto se puede resolver con calculadora y papel, otra posibilidad de trabajo es con Excel (u otro software estadístico como SPSS o R), para que los alumnos aprendan a introducir los datos en la hoja de cálculo y representarlos. La Figura 1 se ha obtenido con Excel, con la cual se puede representar automáticamente un diagrama de dispersión o superponer (como hemos hecho en este caso) dos diagramas correspondientes a dos series de datos. Entre las muchas opciones de Excel nos encontramos con una serie de facilidades para realizar un estudio de regresión. Utilizando la opción “añadir línea de tendencia”, se puede elegir entre varias posibles familias de funciones para ajustar al diagrama de dispersión (siguiendo el criterio de mínimos cuadrados). Nosotros hemos comenzado probando el ajuste lineal; Excel también proporciona la opción de añadir al gráfico la ecuación (en este caso, la ecuación de la recta de regresión).

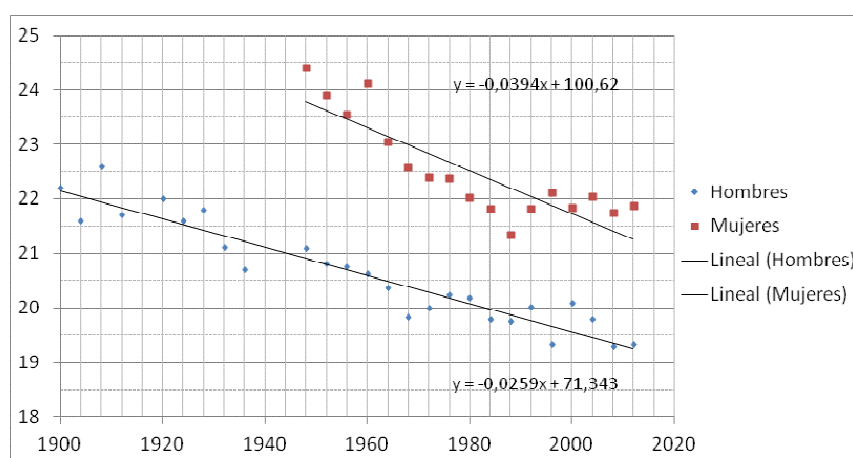


Figura 2. Rectas de regresión ajustadas a los datos de mujeres y hombres

Nosotros hemos utilizado dicha opción, obteniendo las dos ecuaciones siguientes (Figura 2):

- $y = -0,0394x + 100,62$; para la recta de mejor ajuste a los datos de los tiempos de los hombres;
- $y = -0,0259x + 71,343$ para la recta de mejor ajuste a los datos de los tiempos de las mujeres;

Es visible en la gráfica que las rectas no son paralelas; por lo tanto llegarán en algún momento a encontrarse. El valor x en que las dos rectas se cortarían corresponderá a un año; este sería el año en que, si los ajustes que hemos realizado son correctos y las tendencias continúan, ¡hombres y mujeres

llegarían a correr a la misma velocidad en la prueba de los 200 m. lisos! ¿Cuál sería este año? Lo podemos determinar algebraica o gráficamente. Para determinarlo algebraicamente, bastaría resolver el sistema de ecuaciones formado por las expresiones algebraicas de las dos rectas, esto es;

$$\begin{cases} y + 0,0394x = 100,62 \\ y + 0,0259x = 71,343 \end{cases}; \text{ restando estas ecuaciones, obtenemos: } 0,0135x = 29,276, \text{ de donde}$$

$$x = \frac{29,277}{0,0135} = 2168,66$$

Es decir, a partir del año 2168, se produciría el empate; podemos también observar la misma solución (Figura 3) utilizando la opción de Excel de extrapolación de las ecuaciones de regresión. En dicha gráfica, aparentemente a partir de esta fecha, las mujeres correrían más que los hombres, dando la razón al periodista. Observamos que, hasta este momento, hemos utilizado dos componentes nuevos del razonamiento estadístico: el razonamiento con modelos (en este caso, el modelo de recta de regresión) y la integración de la estadística y el contexto, interpretando lo que los gráficos y ecuaciones nos quieren indicar sobre la realidad (marcas olímpicas en las pruebas de 200 metros).

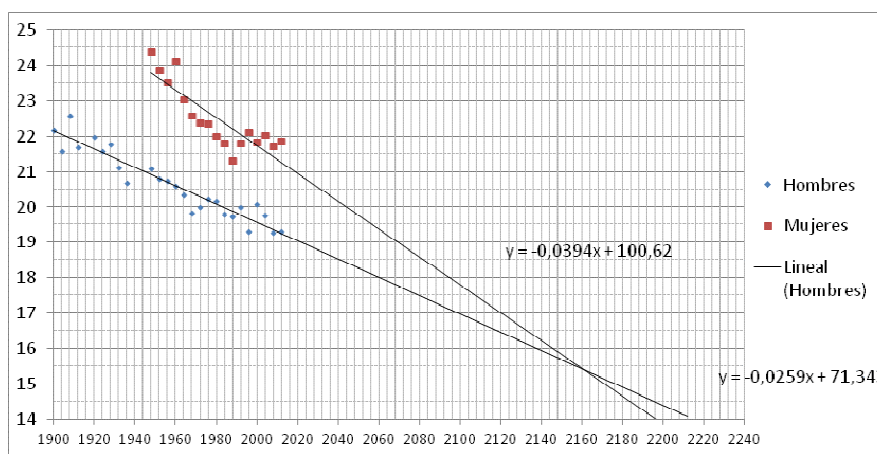


Figura 3. Extrapolación de las rectas de regresión

Hasta acá hemos utilizado los cinco tipos de razonamiento estadístico fundamental; también hemos realizado un ciclo completo de investigación estadística (problema, datos, análisis, conclusión). Hemos trabajado muchas de las ideas estadísticas fundamentales: datos, gráficos, variabilidad, asociación, distribución, inferencia. Nos faltaría ejercitar nuestro espíritu crítico y escepticismo; es decir, el componente actitudinal de la cultura y razonamiento estadístico.

Hemos de recordar en este punto la diferencia entre datos (en este caso las marcas olímpicas) y modelo (las rectas de regresión) y hacer observar que el modelo no es exactamente igual que los datos y que podrían existir otros modelos que explicasen los datos. Será importante, entonces complementar el análisis realizado hasta el momento, con alguna medida de la bondad del ajuste y, en caso necesario, probar otros modelos que se ajusten mejor a los datos.

Otra información importante proporcionada por Excel es el coeficiente de determinación R^2 , siendo R el coeficiente de correlación (en este caso, de correlación lineal). El coeficiente de determinación, R^2 nos indica la proporción de varianza de la variable dependiente en la recta de regresión Y (en este caso esperanza de vida) explicada por el modelo lineal utilizado (por la ecuación de la recta de regresión). En el proyecto que estamos analizando, obtenemos $R^2=0,8971$ ($R=0,947$) para el caso de los hombres, que es muy alto, pues como sabemos, el máximo valor absoluto del



coeficiente de correlación es 1 (correspondiente a $R=1$ o $R=-1$), que son los casos en que los puntos del diagrama de dispersión se alinearían perfectamente sobre una línea recta. En la Figura 1 podemos ver que, en efecto los tiempos de los hombres se sitúan casi sobre la recta, que al ser decreciente, correspondería a un coeficiente de correlación $R=-1$. Para los datos sobre las mujeres obtenemos $R^2=0,7313$ ($R=0,855$) algo menor que el anterior, pero también alto. También podemos observar la mayor distancia de estos datos a la recta de regresión.

En este segundo caso, queda más de un 25% de la varianza de los datos no explicada por el modelo y podríamos tratar de mejorar la predicción, eligiendo para ello una función de regresión diferente, siempre que el valor del coeficiente de determinación aumente (que otro modelo explique mayor proporción de varianza). Por suerte Excel proporciona diferentes opciones para la línea de tendencia: lineal, polinómica, potencial, logarítmica, etc. (que se pueden obtener pulsando sobre dicha línea). El alumno puede ensayar con diferentes opciones, obtener para cada una la gráfica, ecuación y coeficiente de determinación y decidir, entre todas ellas, cuál proporciona un mejor ajuste. Como ejemplo, si usamos una función polinómica de grado cuatro, conseguimos subir el coeficiente de determinación a $R^2=0,93$, suficientemente elevado, pues se explica el 93% de la variación de los datos (Figura 4)

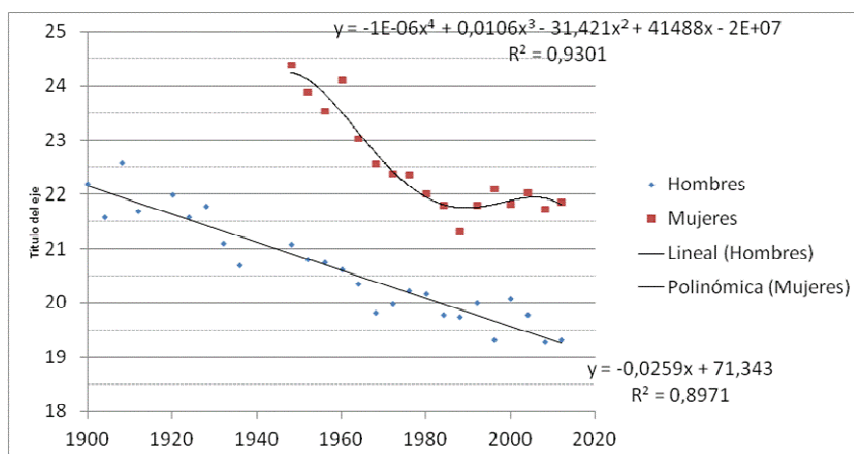


Figura 4. Ajuste de un modelo polinómico a los datos de mujeres

Concluimos que en este segundo modelo, no es tan clara la conclusión de que las mujeres lleguen a superar a los hombres en las olimpiadas. Es verdad que conseguiríamos encontrar el punto de corte entre las dos funciones, pero la tendencia ya no se conserva decreciente en las mujeres, sino que aparecen oscilaciones que podrían o no continuar en el futuro. Será entonces necesario esperar a tener nuevos datos para ajustar otra vez el modelo de regresión y obtener una conclusión definitiva. Por otro lado, recordamos que modelo y realidad no siempre coinciden y que las tendencias observadas podrían variar en el futuro.

6. Reflexiones finales

Como señalaron Anderson y Loynes (1987), la estadística es inseparable de sus aplicaciones, y su justificación final es su utilidad en la resolución de problemas externos a la propia estadística. En el ejemplo de proyecto analizado hemos trabajado los dos componentes del sentido estadístico: cultura y razonamiento.

Respecto a la cultura, se han trabajado multitud de ideas fundamentales: comenzando con la

recopilación de unos datos reales, hemos analizado su distribución bivalente, comparando dos distribuciones diferentes, analizando en cada una su tendencia y variabilidad a través de estadísticos y gráficos. Diferentes modelos de regresión (lineal, polinómico) nos han permitido describir los datos y hacer predicciones; complementadas mediante el coeficiente de determinación para analizar la bondad del ajuste. Otras propiedades de la correlación y regresión (como son por ejemplo el signo y el valor del coeficiente, el signo de la pendiente según sea la relación directa o inversa, la relación entre pendiente y el coeficiente de correlación), aunque han quedado implícitas podrían analizarse con más detalle con los estudiantes según los objetivos del profesor.

En el proyecto también se han ejercitado actitudes propias de la cultura y el razonamiento estadístico como la creatividad, espíritu crítico e imaginación. Se utilizaron también todos los tipos fundamentales de razonamiento estadístico y se involucró a los estudiantes en un ciclo completo de investigación y modelización estadística. Con todo ello los estudiantes adquieren competencia en conocimiento estratégico, es decir ser capaz de aplicar un conocimiento. La habilidad para aplicar los conocimientos matemáticos es frecuentemente mucho más difícil de lo que se supone, porque requiere no sólo conocimientos técnicos (tales como preparar un gráfico o calcular un promedio), sino también conocimientos estratégicos (saber cuándo hay que usar un concepto o gráfico dado). Los problemas y ejercicios de los libros de texto sólo suelen concentrarse en los conocimientos técnicos, mientras que los proyectos incluyen también conocimientos estratégicos, a la vez que aumentan la motivación del estudiante.

Todo ello se pone de manifiesto en el ejemplo presentado. El lector interesado puede encontrar otros ejemplos de cómo es posible desarrollar un currículo de estadística para la educación secundaria o universitaria mediante proyectos debidamente secuenciados (ver Batanero y Díaz, 2011).

Agradecimientos: Proyecto EDU2010-14947 (MCINN-FEDER) y grupo FQM126 (Junta de Andalucía).

Bibliografía

- Anderson, C. W. y Loynes, R. M. (1987). *The teaching of practical statistics*. New York: Wiley.
- Bakker, A. y Gravemeijer, K. P. E. (2004). Learning to reason about distribution. En J. Garfield y D. Ben Zvi (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp 147-168). Dordrecht, The Netherlands: Kluwer.
- Batanero, C. y Díaz, C. (Eds.) (2011). *Estadística con proyectos*. Granada: Departamento de Didáctica de la Matemática
- Batanero, C., Henry, M., Parzysz, B., (2005). The nature of chance and probability. En G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 15-37). New York: Springer.
- Burrill, G., & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. En C. Batanero, G. Burrill y C. Reading (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education - A joint ICMI/IASE study* (pp. 57-69). Dordrecht: Springer.
- Cabriá, S. (1994). *Filosofía de la estadística*. Servicio de Publicaciones de la Universidad de Valencia.
- Castro-Sotos, A. E., Vanhoof, S., Noortgate, W. y Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review*, 2 98-113
- Engel, J. y Sedlmeier (2011). Correlation and regression in the training of teachers. En C. Batanero, G. Burrill, C. Reading y A. Rossman (Eds.), *Teaching statistics in school mathematics- challenges for teaching and teacher education. A joint ICMI/IASE study* (pp. 247-258). New York: Springer.
- Franklin, C., Kader, G., Mewborn, D. S., Moreno, J., Peck, R., Perry, M. y Scheaffer, R. (2007). *A curriculum framework for K-12 statistics education. GAISE report*. Online:



- <http://www.amstat.org/education/gaise/>.
- Gal, I. (2002). Adult's statistical literacy: Meaning, components, responsibilities. *International Statistical Review* 70(1), 1-25.
- Moore, D. S. (1991). Teaching statistics as a respectable subject. En F. Gordon y S. Gordon (Eds.), *Statistics for the twenty-first century* (pp. 14-25). Mathematical Association of America.
- Reading, C. y Shaughnessy, J. M. (2004). Reasoning about variation. En J. Garfield y D. Ben-Zvi (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 201-226). Dordrecht, The Netherlands: Kluwer.
- Rossman, A. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal*, 7(2), 5-19. Online: <http://www.stat.auckland.ac.nz/serj>.
- Rubin, A., Hammerman, J. K. L & Konold, C. (2006). Exploring informal inference with interactive visualization software. En B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics*. Cape Town, South Africa: International Association for Statistics Education. Online: www.stat.auckland.ac.nz/~iase/publications.
- Sánchez, E. y Batanero, C. (2011). Manejo de la información. En E. Sánchez (Coord.), *Enseñanza y aprendizaje de las matemáticas, Casos y perspectivas* (pp. 64-92). México, D. F.: Secretaría de Educación Pública.
- Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. En F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957-1010). Greenwich, CT: Information Age y NCTM.
- Shaughnessy, J. M., Chance, B., & Kranendonk, H. (2009). *Focus in high school mathematics: Reasoning and sense making in statistics and probability*. Reston, VA: NCTM.
- Watson, J. M. (2006). *Statistical literacy at school: growth and goals*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Wild, C. y Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67 (3), 221-248.

Carmen Batanero Bernabeu, Facultad de Ciencias de la Educación, Universidad de Granada. Fue miembro del Comité Ejecutivo de ICMI (International Comisión on Mathematical Instruction y Presidenta de IASE (International Association for Statistical Education). Ha coordinado varios congresos y proyectos de educación estadística. batanero@ugr.es

Carmen Díaz Batanero, Facultad de Ciencias de la Educación, Universidad de Huelva, es doctora en Psicología. Fue becaria del Plan de Formación del Profesorado Universitario y es actualmente Profesora Contratado Doctor en área de Metodología de las Ciencias del Comportamiento. Su línea de investigación es la enseñanza y aplicaciones de la estadística, tema en el que ha codirigido dos tesis doctorales. Ha publicado trabajos sobre inferencia estadística y didáctica de la estadística. carmen.diaz@dpsi.uhu.es

José M. Contreras García, nacido en Granada, fue becario de Plan de Formación del Personal Investigador y es actualmente profesor ayudante doctor de la Universidad de Granada. Licenciado en Ciencias Matemáticas y en C.C. y T.T. Estadísticas, Máster en Estadística Aplicada y doctor en Didáctica de la Matemática. Ha realizado publicaciones en didáctica de la estadística. jmcontreras@ugr.es

Rafael Roa Guzmán, es doctor en Didáctica de la matemática y profesor Titular de Universidad de la Universidad de Granada. Ha participado en varios proyectos de investigación en educación estadística y ha publicado artículos y comunicaciones en congresos sobre esta temática.