LA CONSTRUCCIÓN DEL SIGNIFICADO EN LA ENSEÑANZA DE LA INFERENCIA ESTADÍSTICA

Juan Antonio García Cruz
Universidad de La Laguna
Sociedad Canaria de Profesores de Matemáticas

"*Los tenues rayos solares de una tarde de abril penetran a través de las persianas levemente cerradas. La tiza y el borrador están preparados y la pizarra ha sido recientemente borrada. Los estudiantes esperan en clase, unos ojean distraídamente el libro de estadística y probabilidad, otros intercambian comentarios. Se abre la puerta y entra el profesor. La clase de hoy trata de intervalos de confianza.*

Si se cumple que la $p(\theta_1 \leq \theta \leq \theta_2) = 1 - \alpha$ para un valor dado de α, diremos que el intervalo $[\theta_1, \theta_2]$ es un intervalo de confianza, $100(1-\alpha)\%$, para el parámetro θ.

Los alumnos están muy ocupados copiando la definición en sus cuadernos. Sin embargo, uno de ellos, aclara la voz mediante un garraspeo de la garganta y pregunta: Profesor,...por favor,¿ podría explicarme que es un intervalo de confianza?

Esta anécdota ilustra un hecho muy común en las clases de matemáticas y de estadística. El profesor expone la definición, abstracta del concepto, como punto de partida. La definición abstracta, formal, tiene una gran ventaja, no es ambigua. Pero al mismo tiempo, tiene un gran inconveniente: oculta el significado de los términos que la componen. Nosotros los profesores, si alguna vez estudiamos estadística inferencial en la universidad, fuimos con casi toda seguridad introducidos a los conceptos correspondientes de esta forma. Ese es el modelo en el que aprendimos. Todavía hoy es el modelo imperante en la universidad española. Los modelos en que nacemos y crecemos son los modelos que transmitimos, por lo general.

La inferencia estadística apareció por primera vez en la enseñanza preuniversitaria allá por la década de los sesenta. El cuestionario para la prueba de madurez de 8 de agosto de 1963 incluía, dentro del temario de clases prácticas, *el muestreo, el texto de una hipótesis estadística y la estimación*. Luego, con la sustitución del *preuniversitario* por el *curso de orientación universitaria*, desapareció hasta la década de los ochenta en que, vuelve a aparecer con el bachillerato experimental y se consolida, a nivel de programa, en el bachillerato LOGSE. Sin embargo, a pesar de los intentos oficiales de que tales contenidos se extiendan en las enseñanzas del bachillerato, y sobre todo en los cursos terminales, no parece que sea un contenido que haya recibido general aceptación por parte del profesorado y de las autoridades educativas *locales*. De hecho en muchos distritos universitarios sólo se contempla el intervalo de confianza en la prueba de acceso a la universidad. En otros, se incluye el test de hipótesis contemplando sólo el error de tipo I. Otra visión se tiene si se observan los libros de texto de matemáticas aplicadas a las ciencias sociales para segundo de bachillerato. Allí se despliega toda la artillería sobre el muestreo, los intervalos de confianza y los test de hipótesis, aunque algunos, pocos, sólo rematan los test de hipótesis con los dos tipos de errores.

Mi objetivo en esta ponencia es reflexionar sobre los significados de los conceptos involucrados en la inferencia estadística, principalmente en el test de hipótesis. Empezaré por una situación que, a mi parecer, es muy adecuada para entender dos cuestiones. Una es la relación íntima entre un modelo teórico y los datos experimentales, la otra es el significado de la variable z en la distribución Normal (0,1).

La colmena (De Bijenkorf)[1]

Por encargo de los grandes almacenes "*De Bijenkorf*" de Amsterdam se llevó a cabo, en 1947, un estudio sobre determinadas medidas del cuerpo femenino. El propósito de tal estudio era mejorar el nivel de confección de la ropa : prendas con medidas más adecuadas y menos desperdicio de dinero en gastos de adaptación. La siguiente tabla recoge la distribución de estatura de las 5001 mujeres que participaron en el estudio.

Estatura(cm)	Frecuencia	Estatura(cm)	Frecuencia
139	1	163	290
140	1	164	294
141	4	165	291
142	3	166	261
143	2	167	222
144	8	168	184
145	4	169	157
146	17	170	167
147	18	171	109
148	32	172	86
149	51	173	65
150	54	174	62
151	71	175	29
152	78	176	49
153	115	177	28
154	149	178	17
155	170	179	5
156	208	180	10
157	208	181	6
158	231	182	3
159	301	183	1
160	302	184	2
161	321	185	0
162	313	186	1

[1] *De Normale Verdeling*. Martin Kindt y Jan de Lang. Educaboek b.v. Culemborg. The Netherlands, 1986. Proyecto Hewet.

Calculando media, moda y mediana, encontramos los valores 162’1, 161, 162’6 respectivamente (valores aproximados). Notamos que los tres valores están muy próximos. Construyamos el intervalo centrado en la media y cuyos extremos son la media más y la media menos una desviación típica (s = 6’5), tenemos (155’5 , 168’5). Un recuento simple nos indica que el número de mujeres cuya estatura cae en dicho intervalo es 3426 que equivale a 68’51% de la población. Importante: En el intervalo (155’6, 168’6) las estaturas **¡distan menos de una desviación típica** de la media o centro del intervalo! Quiero resaltar este aspecto de distancia al centro del intervalo y su relación con la desviación típica.

Si hago los cálculos para los intervalos con centro la media y extresmos más, menos dos desviaciones típicas y más, menos tres desviaciones típicas obtenemos respectivamente: (149’1 , 175’1) y (142’6 , 181’6). Estos intervalos continen 4738 (94’74%) y 4985 (99’68%) de elementos de la población.

Los correspondientes datos para la Distribución Normal son, de forma aproximada, 68’27 %, 95’45% y 99’73% respectivamente. Además la Normal cumple la igualdad entre los tres parámetros centrales, media, moda y mediana. Luego tenemos una distribución, estatura de mujeres, discreta bastante próxima a la Norma. Los datos empíricos se aproximan bastante bien al modelo teórico.

Ahora preguntaremos a los alumnos a qué distancia se encuentra un valor determinado, p.e. 178, de la media. Claramente está a 15’9 cm. Pero le pedimos que nos exprese tal valor en desviaciones típicas, es decir 15’9/6’5, lo que es igual a 2’45. Repitamos la operación ahora con el valor 143, obteniendo 143-162’1=-19’1; y -2’93 desviaciones típicas. Se hacemos lo mismo con más valores obtendremos positivos (por encima) y negativos (por debajo) y en todos realizamos la misma operación: $\frac{x-\overline{x}}{s}$.

Ahora están los alumnos preparados para entender los valores z de la *tabla de la Normal*. Los z no son otra cosa que distancias en desviaciones típicas al centro de la distribución. Positivas a la derecha, negativas a la izquierda.

Áreas bajo la curva normal

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

El Azar y la Divina Providencia

Utilizar la probabilidad para determinar si un efecto pudiera ser debido al azar es tan antiguo como la propia teoría del azar. Uno de los primeros ejemplos es la prueba de Arbuthnot. El primer tratado teórico impreso sobre el azar fue escrito por Christiaan Huygens y publicado por van Schooten en 1657. *De raciociniis in ludo aleæ* utiliza *expectatio*, un antecesor de la esperanza matemática, y no la probabilidad como el concepto central:

> Proposición 3: Sea p el número cualquiera de casos para a, sea q el número cualquiera de casos para b, tomando todos los casos igualmente posibles (*proclivi*), mi *expectatio* es $\dfrac{pa+qb}{p+q}$.

El tratado surge como solución al problema del reparto de la apuesta. La expectatio de Huygens es tanto el valor de la apuesta como el premio a recibir en un juego justo. Este pequeño tratado, pues solo consta de catorce páginas, es traducido al inglés por John Arbuthnot en 1692.

John Arbuthnot nació en 1667 en Inverbervie (Escocia) y murió el 27 de Febrero de 1735 en Londres. En 1696 se graduó como médico por la Universidad de St Andrews. Fue miembro de la Royal Society y del Real Colegio de Médicos. Fue también médico personal de la Reina Ana de Inglaterra. En la actualidad se le considera un escritor satírico, estimado por sus contemporáneos al mismo nivel que Jonathan Swift. Aparte de ser el traductor al inglés del primer tratado teórico sobre el azar, Arbuthnot es famoso por una memoria publicada en *Philosophical Transactions* de la Royal Society, volumen xxvii, correspondiente a los años 1710, 1711 y 1712 titulada An Argument for Divine Providence, taken from the constant Regularity observ'd in the Births of both Sexes[2].

La memoria comienza así:

> "Entre las innumerables huellas de la Divina Providencia que se encuentran en las obras de la Naturaleza, hay una sobresaliente entre todas que muestra el equilibrio exacto que se mantiene entre el número de hombres y mujeres; con ello se garantiza que las especies ni fallen ni perezcan, ya que cada macho tendrá su hembra, y de una edad proporcionada. Esta igualdad entre machos y hembras no es el efecto del azar sino de la Divina Providencia, que trabaja para un buen fin, lo cual paso a demostrar…"

Arbuthnot argumenta que la mano conductora del ser divino puede observarse en la razón casi constante de varones y hembras cristianizados en la ciudad de Londres, en el período de 82 años comprendido entre 1629 y 1710. Los datos presentados por Arbuthnot muestran que, durante tal período, el número anual de varones cristianizados fue consistentemente bastante más alto que el número de hembras cristianizadas, aunque nunca mucho más alto. La razón más alta de varones respecto de las hembras (4748:4107=1'156) ocurrió el año 1661, y equivale a una proporción de varones nacidos igual a 0'536; la proporción más baja (1'011) es la correspondiente al año 1703, que equivale a una proporción de varones nacidos igual a 0'503.

Tales observaciones llevaron a Arbuthnot a plantearse la siguiente cuestión:

¿Está el sexo gobernado por el Azar?

Veamos sus argumentos.

Arbuthnot imagina un dado con dos caras M y F (masculino y femenino respectivamente). Para determinar todas las *chances* de cualquier número determinado

[2] Un argumento en favor de la Divina Providencia, tomado de la constante regularidad observada en los nacimientos de ambos sexos.

de tal dado, eleva el binomio M+F a la potencia igual al número dado de dados. Los coeficientes de los términos resultantes serán las *chances* buscadas. Por ejemplo, si disponemos de cuatro dados, tenemos: $\mathbf{M^4+4M^3F+6M^2F^2+4MF^3+F^4}$. Luego las *chances* son 1:4:6:4:1.

Para n dados tenemos:

$$\mathbf{M^n+\frac{n}{1}M^{n-1}F+\frac{n}{1}\times\frac{n-1}{2}M^{n-2}F^2+\frac{n}{1}\times\frac{n-1}{2}\times\frac{n-2}{3}M^{n-3}F^3+\ldots}$$

Ahora señala que, si el número de dados es par, entonces el término medio del desarrollo anterior contempla las *chances* para igual número de M que de F. Todos los demás términos tiene o más M o más F.

Si ahora alguien se compromete a obtener, con cuatro dados, tantas M como F, entonces tiene todos los términos salvo el central en su contra; y su *suerte* (*lot*) es la suma de todas las *chances*, como el coeficiente en el término central es a la potencia de 2 elevada al exponente igual a número de dados utilizados. Así, en dos dados tenemos $\mathbf{\frac{2}{4}=\frac{1}{2}}$, en cuatro dados $\mathbf{\frac{6}{16}=\frac{3}{8}}$, etc…

Para hallar este término central cuando utilizamos cualquier número de dados, bastará con calcular el término $\mathbf{\frac{n}{1}\times\frac{n-1}{2}\times\frac{n-2}{3}\times\ldots}$ hasta que el número de términos sea igual a $\mathbf{\frac{1}{2}n}$. Por ejemplo, el coeficiente del término central para la potencia décima es 252, la potencia décima de 2 es 1024, luego si A se compromete a, en una sola tirada con diez dados, obtener igual número de M que de F, tiene 252 chances en 1024, es decir su suerte es $\mathbf{\frac{252}{1024}=\frac{63}{256}}$, que es menor que $\mathbf{\frac{1}{2}}$.

Una vez presentados los anteriores cálculos, Arbuthnot formula el siguiente problema: *A apuesta contra B a que cada año nacerán más varones (M) que hembras (F). Encontrar la suerte de A o el valor de su expectativa.*

Por todo lo dicho queda claro que para cada año su suerte es menor que $\mathbf{\frac{1}{2}}$. Pero otorguemos el valor igual a $\mathbf{\frac{1}{2}}$. Si se compromete con lo mismo para un período de 82 años consecutivos, entonces su suerte será $\mathbf{\left(\frac{1}{2}\right)^{82}}$ cuyo cálculo mediante la tabla de logaritmos es $\mathbf{\frac{1}{4836000000000000000000000}}$. Luego sigue que es la divina providencia y no el azar la que regula el sexo.

El artículo finaliza con el siguiente escolio*: De aquí sigue que la poligamia es contraria a la Ley Natural y a la Justicia y a la Propagación de la Raza Humana; pues allí donde los varones igualan en número a las féminas, si un hombre toma veinte esposas, diecinueve hombres deben vivir en celibato, lo que es repugnante al Designio de la Naturaleza; tampoco es probable que veinte mujeres sean preñadas tan bien por un solo hombre como por veinte.*

Podemos sintetizar el argumento utilizado por Arbuthnot de la forma siguiente:
Sea H la hipótesis el Azar gobierna el sexo. Sea E el suceso observar mayor proporción de varones que de hembras en los nacidos a lo largo de 82 años consecutivos. Ahora se evalúa p(E/H), y se observa que es prácticamente cero. Luego se rechaza la hipótesis H.
El suceso E es un suceso *extraño*, pues tiene poca probabilidad, bajo la hipótesis H. Este es un concepto clave para entender el test de hipótesis. Veamos una situación ideada para desarrollar en el aula en la que se pone en juego el significado anterior.

Un jugador de baloncesto tiene un promedio de acierto en tiros libres del 60%. Asegura haber mejorado su tasa y como prueba de ello lanza diez tiros libres y acierta en nueve ¿Podemos concluir que efectivamente ha mejorado?

Primera solución mediante simulación. Utilizando la tabla de números aleatorios obtenemos 20 resultados equivalentes a diez tiros libres. Previamente hemos realizado la siguiente asignación: números 0-1-2-3-4-5 equivalen a enceste, lo contrario a fallo. De esta forma simulamos un lanzador con promedio de 60% de acierto en tiros libres.
La situiente tabla muestra los resultados obtenidos:

3	8	8	6	9	7	7	0	6	4	**30%**
4	0	4	7	4	0	4	3	3	2	**90%**
1	0	8	1	9	6	2	1	8	5	**60%**
1	1	6	7	9	3	0	1	5	9	**60%**
6	9	7	2	3	4	8	4	9	2	**50%**
2	9	7	8	5	0	7	3	4	6	**50%**
2	2	8	7	9	8	4	8	2	0	**50%**
6	5	6	7	2	3	7	1	1	2	**60%**
3	7	3	4	0	4	3	9	3	2	**80%**
4	7	3	8	3	6	4	7	1	6	**50%**
7	8	7	1	8	7	4	7	1	5	**40%**
8	8	7	1	5	1	0	9	8	4	**50%**
7	4	2	2	9	6	5	6	7	9	**40%**
2	2	0	3	9	3	1	9	4	3	**80%**
5	8	3	4	3	9	8	9	7	8	**40%**
7	3	3	6	4	0	5	0	9	6	**60%**
8	6	9	3	5	6	9	8	3	2	**40%**
5	5	0	5	8	9	5	7	9	0	**60%**
1	3	7	2	2	8	9	6	4	1	**60%**
6	3	5	1	8	9	7	1	9	7	**40%**

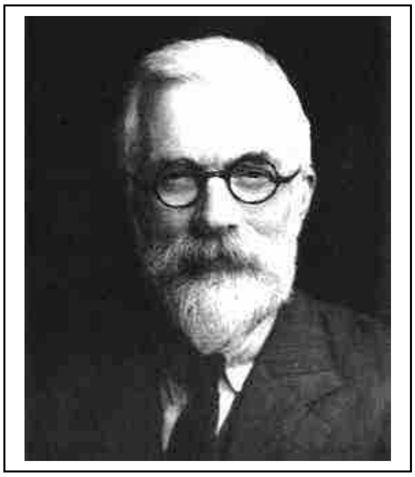
De los 20 resultados sólo en uno hemos obtenido un 90% de acierto, y dos veces un 80%. Así que podemos concluir que es *extraño*, por poco frecuente, obtener tales aciertos con tal promedio. Luego no parece que los nueve aciertos de diez lanzamientos sean debidos al azar y podríamos concluir que efectivamente ha mejorado su promedio de tiros libres.

Segunda solución utilizando la distribución binomial. La cuestión a plantear es, tenemos una B(10, 0’6), ¿Cúal es la probabilidad de obtener p(x=9)? Observamos que p(x=9)=0’04. Luego como es muy pequeña, rechazamos la hipótesis de que p(acierto)=0’6 y concluimos que sí ha mejorado su tasa de aciertos en tiros libres.

La simulación nos familiariza con el azar, mediante el proceso de muestreo y la observación y comparación de los porcentajes particulares de cada muestra de diez lanzamientos. La teoría nos refuerza la intuición generada por la simulación. Muchos ejemplos como este son necesarios hasta clarificar el significado de hipótesis, suceso y resultado *extraño*. Pero esto todo es necesario para comprender la esencia del test de hipótesis.

Las matemáticas de una catadora de té

En el desarrollo moderno de la estadística jugó un papel esencial Sir Ronald Aylmer Fisher (1890-1962), al punto de ser considerado uno de sus fundadores. Entre las importantes contribuciones de Fisher cabe destacar el diseño de experimentos y el test de hipótesis. Según Fisher, el razonamiento inductivo nunca puede confirmar una hipótesis. En todo caso, a lo más que se puede llegar es a rechazar la hipótesis por insostenible a la luz de los datos aportados por los experimentos.



La obra *The World of Mathematics* (Newman, 1956) contiene un artículo de R. A. Fisher paradigmático sobre la prueba de significación de una hipótesis estadística titulado *Las matemáticas de una catadora de té*. El artículo comienza con la siguiente afirmación, *Una dama afirma que al probar el té con leche puede distinguir qué fue lo primero que se echó en la taza, el té o la leche*.

Fisher considera el problema de plantear un experimento mediante el cuál se llegue a demostrar la afirmación de la catadora de té, y que constituye toda una metodología sobre la prueba de significación de una hipótesis estadística.

La cuestión de partida es ¿Cómo realizar el experimento y por qué?

Experimento

Se preparan 8 tés con leche, cuatro de una forma y cuatro de la otra. Luego se llevan las tazas, escogidas al azar, a la catadora para que indique su opinión. Deberá separar las tazas en dos grupos según su modo de formación.

Interpretación de los resultados posibles

Es necesario que consideremos todos los resultados posibles y qué interpretación corresponde a cada uno de ellos. Al separar las tazas en dos grupos, caben los siguientes resultados y sus probabilidades:

Resultado	**Probabilidad**
Cuatro aciertos	1/70
Tres aciertos y un fallo	16/70
Dos aciertos y dos fallos	36/70
Un acierto y tres fallos	16/70
Ningún acierto	1/70

Una persona que no tuviese la habilidad de la catadora de té conseguiría acertar las cuatro tazas con una probabilidad de 1/70, es decir, un poco más de un 1%. La posibilidad de éxito es tan pequeña que habría que atribuir tal resultado al azar.

La tabla anterior proporciona los resultados del experimento bajo la hipótesis de que un sujeto no distingue entre dos objetos (Hipótesis nula). Bajo esa hipótesis se ha construido la tabla de probabilidades de los resultados posibles del experimento, es decir, la hipótesis nula proporciona la distribución de probabilidad para el contraste de la hipótesis.

Una objeción: ¿Por qué ocho tazas y no, por ejemplo, seis tazas con tres de cada tipo? Aclarar esta objeción no lleva al clarificar la prueba de significación.

Prueba de significación

¿Cuándo consideraremos que el resultado del experimento no es debido al azar y que, por lo tanto, es significativo estadísticamente? Esta cuestión nos lleva a fijar el nivel de significación de la prueba. Fisher fija el nivel base o tipo de significación en 1 sobre 20 (*5%, por corriente y conveniente*).

Si hubiéramos utilizado seis tazas, tres de cada tipo, entonces la probabilidad de dividir correctamente el conjunto de seis tazas es 1/20, es decir el sujeto obtendría un éxito en la prueba, por puro azar, en el 5% de las veces que la realizara. Esta es la razón por la que se descarta el experimento compuesto por seis y se opta por el de ocho tazas.

Al fijar tal nivel, también se descarta el resultado tres aciertos y un fallo, pues su probabilidad es algo superior al 20%.

Luego los posibles resultados del experimento se dividen en dos grupos cuyas interpretaciones son opuestas: los que no muestran una desviación significativa respecto de la hipótesis y aquellos que sí muestran diferencias significativas, regiones de aceptación y rechazo de la hipótesis, respectivamente.

El examen de los resultados posibles del experimento nos proporciona una prueba estadística significativa. Sin embargo, Fisher hace una observación importante: *El hecho de que 1 acontecimiento sobre 70 (cuatro aciertos) sea significativo estadísticamente no basta para demostrar experimentalmente cualquier fenómeno natural. Para afirmar que un fenómeno natural es experimentalmente demostrable se necesita un verdadero método de realización y no sólo una referencia aislada*.

Cuando sometemos una hipótesis a un test, nuestra intención es concluir que la misma es verdadera o falsa. Sin embargo, la lógica nos impide llegar a conclusiones tan categóricas.

Veamos desde el punto de vista lógico el significado de la observación anterior de Fisher.

Sea H la hipótesis “la catadora posee la habilidad que afirma”. Sea S el suceso “dadas 8 tazas, la catadora divide correctamente el conjunto en dos mitades”. Tenemos pues que si H es verdadera también lo es S. Ahora bien, S es verdadera (*ocurre* es el significado de *verdad* para nuestro caso), ¿qué podemos concluir respecto de H

$$\begin{array}{ll} \mathbf{H \to S} & \\ \hline & \mathbf{S} \\ H & \end{array}$$

Si concluimos que H es verdadera hemos incurrido en lo que se denomina la falacia del consecuente. Sabemos que tal conclusión no es lógicamente válida. El hecho de que ocurra S hace más creíble H, pero no se deduce de ello. Sin embargo, si disponemos del siguiente argumento lógico (*modus tollens*):

$$\begin{array}{ll} \mathbf{H \to S} & \\ \hline & \neg\mathbf{S} \\ \neg\mathbf{H} & \end{array}$$

Es decir, si S es *falsa* (*no ocurre*), entonces también es falsa H. Esta es la base lógica del test de Fisher y su principal argumento contra el razonamiento inductivo. A lo sumo podemos rechazar una hipótesis, nunca confirmarla mediante tal razonamiento.

Lo visto suministra la base argumental del siguiente apartado.

La hipótesis nula

Para Fisher la hipótesis nula nunca puede demostrarse y sí refutarse. En este sentido cada experimento existe para darnos la oportunidad de refutar la hipótesis nula.

¿Qué ocurre entonces con la hipótesis contraria? (hipótesis alternativa).

En su argumentación Fisher explícita por primera vez cuál es la hipótesis nula: *El sujeto es incapaz de distinguir dos tipos de objetos*. Obsérvese que es lo contrario de lo que afirma la catadora de té. Entonces, se podría objetar que si un experimento excluye la hipótesis nula debe probar la hipótesis contraria. Contra esta objeción Fisher argumenta que aunque la hipótesis contraria sea razonable no interesa como hipótesis nula por inexacta. Pues si fuera posible asegurar que un individuo nunca se equivoca en sus juicios, deberíamos tener de nuevo una hipótesis exacta y esta se podría refutar por un simple fallo y en cambio nunca podría ser demostrable por una cantidad finita de experiencias. Tal hipótesis nula debe ser exacta porque suministra el estadístico de prueba y su distribución, a partir de la cuál se construye la prueba de significación.

Según esta argumentación, el test de significación de Fisher no es más que una forma de inferencia incierta. Disponemos de los datos suministrados por una muestra y de una hipótesis. La decisión se toma entre dos alternativas: o bien ha ocurrido un resultado extraño y sorprendente (la catadora acierta) o bien la hipótesis no es cierta. Lo que aquí cabe es cuantificar, probabilísticamente hablando, lo extraño y sorprendente del hecho de que la catadora presente una división exacta de las ocho tazas. Esta cuantificación se obtiene al fijar, como hemos visto, el nivel de significación.

Para Fisher la alternativa de rechazar una hipótesis nula no es aceptar la hipótesis alternativa, sino medir el riesgo de equivocarnos. Es decir, controlar la probabilidad de que rechacemos una hipótesis nula cierta, valor que es proporcionado por el nivel de significación o error de tipo I: $\alpha = p(\text{rechazar } H_0 / H_0 \text{ cierta})$.

El artículo de Fisher muestra las debilidades del razonamiento inductivo y, al mismo tiempo, clarifica el papel que en un test de hipótesis juega la hipótesis nula al suministrar el modelo de azar para el contraste y valorar el riesgo de rechazar la hipótesis nula mediante el nivel de significación.

Como conclusión me haré la siguiente pregunta:

¿En qué consiste y qué prueba un experimento estadístico?

El test de Arbuthnot muestra como históricamente se confunde azar con equiprobabilidad y como se utiliza una inferencia válida (no equiprobabilidad) para aceptar una hipótesis mediatizada por las creencias religiosas y sobre la cuál no es posible disponer de datos empíricos. Además, la historia muestra un ejemplo válido de razonamiento sobre la ley de Bernoulli que permite superar el obstáculo de equiprobabilidad-azar.

El Test de Fisher muestra algunas ideas básicas que conviene aclarar en el estudio y aplicación de la estadística a situaciones reales. En primer lugar, el papel de la hipótesis nula y la distribución de probabilidad que genera, utilizada como elemento de contraste para los datos empíricos proporcionados por el experimento. En segundo lugar, la alternativa que consiste en establecer una medida apropiada al riesgo que correríamos al rechazar la hipótesis nula en el supuesto de que sea cierta. Esto nos lleva al concepto de nivel de significación, cuyo profundo significado está en la respuesta a la cuestión tantas veces planteada en situaciones de incertidumbre: ¿Cuán raro es este hecho? El nivel de significación permite una cuantificación del mismo y es, no lo olvidemos, una decisión que debe tomar el diseñador del experimento.

En resumen, con la exposición de estos dos episodios he querido, por un lado mostrar cómo las ideas emergentes del cálculo de probabilidades sirvieron históricamente para organizar el fenómeno del azar. Por otro, observar cómo suministran un ejemplo metodológico para presentar el tema en el aula de secundaria (o en la universidad).

Una forma vívida de iniciar el estudio de la inferencia estadística es partir de una situación realista en la que sea fácil involucrar a los alumnos en una controversia. En otro lugar (García Cruz, 2000) he presentado, en forma de diálogo entre varios personajes, un problema cuya discusión permite aflorar concepciones y creencias sobre la probabilidad y, mediante la guía del profesor, desembocar en la construcción de una distribución de probabilidad *ad hoc* que permite tomar una decisión sobre lo extraño de un resultado. El diálogo y la dramatización se ha sugerido como recurso útil para la presentación del nacimiento y desarrollo de los conceptos matemáticos (Hirchcock, 1996).

La propuesta metodológica

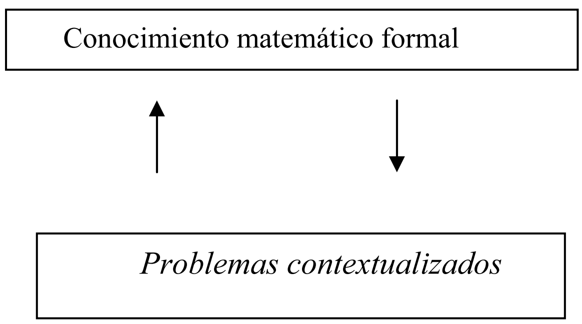
Simplificando, existen dos formas de mirar al conocimiento matemático, cada una corresponde a un estatus epistemológico diferente y con una concepción diferente también de la enseñanza.

Si consideramos que la matemática es primordialmente un sistema formal acabado con aplicación general, entonces la enseñanza consiste en descomponer el conocimiento matemático formal en procedimientos de aprendizaje que, el estudiante aprende a aplicar más tarde.

Si, por el contrario, concebimos la matemática como un proceso en el que prima la actividad de resolución de problemas, la enseñanza se concibe, en correspondencia, como la actividad de hacer matemáticas, donde resolver problemas cotidianos es parte del trabajo y esencia en la construcción del conocimiento.

Ambas concepciones se diferencian fundamentalmente en cómo se estructura el proceso de enseñanza.

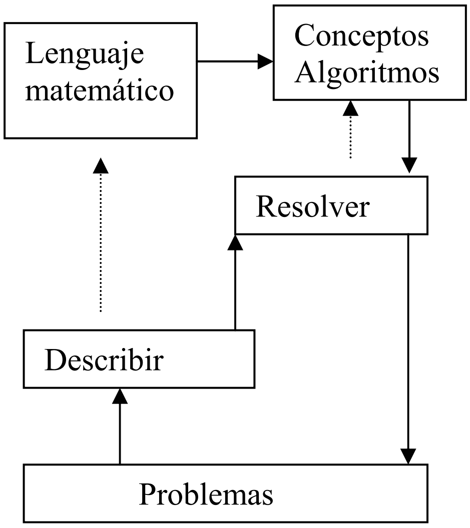
Si contemplamos la matemática como un sistema formal entonces su aplicabilidad está garantizada por el carácter general de sus conceptos y procedimientos. Luego lo primero que hay que hacer es adaptar este conocimiento abstracto para resolver problemas planteados en la realidad.



Primero se traduce el problema a términos matemáticos. Luego se resuelve el problema con ayuda de los medios matemáticos disponibles. Finalmente, la solución matemática se traduce al contexto original. En primer lugar se imparte el conocimiento abstracto que luego, si cabe, será utilizado para resolver ciertas situaciones de aplicación.

Por el contrario, si elegimos enseñar matemáticas como una actividad, la resolución de problemas cobra un significado diferente. La enseñanza se centra en los problemas, lo que significa que la actividad desarrollada para resolver el problema se convierte en el objetivo central del aprendizaje, en vez de ser simplemente el lugar donde utilizar las herramientas matemáticas.

Resolver un problema en este nivel menos formal difiere grandemente de la aplicación de un procedimiento formal.



Se parte de un problema y se comienza con una descripción del mismo que incluye tanto la forma de entender el problema en sí, como el enfoque que se le dará a su solución. La descripción de los problemas desarrolla un lenguaje informal y pone en juego las concepciones y formas de entender la situación por los alumnos. Al mismo tiempo, el lenguaje informal evoluciona en un lenguaje cada vez más formal debido al proceso de simplificación y formalización guiado por el profesor. A largo plazo, la resolución de un tipo particular de problemas se convierte en una rutina, al condensarse y formalizarse el procedimiento en el curso del tiempo, dando lugar a la constitución de los objetos mentales involucrados que precederían, en el sentido señalado por Freudenthal en su fenomenología didáctica, a los conceptos matemáticos formales.

El tiempo que disponemos para trabajar con los alumnos en clase es muy limitado. No lo desperdiciemos dedicándolo solo a la enseñanza de rutinas. El alumnado aprenderá muy poco y nosotros seguiremos anclados en un proceso sin sentido.

Referencias

Fisher, R.A. (1976). 'Las matemáticas de una catadora de té' en James R. Newman (editor) *Sigma. El mundo de las matemáticas*. Tomo 3, pp 194-201. Ediciones Grijalbo. Barcelona.

Freudenthal, H. (1983). *Didactical Phenomenology of Mathematical Structures*. Mathematics Education Library. D. Reidel Publishing Company.

García Cruz, J.A. (2000). El caso de los despedidos de la empresa Westvaco. *Uno*. *Revista de Didáctica de las Matemáticas*,23, 121-128.

Hitchcock, G. (1996). 'Dramatizing the Birth and Adventures of Mathematical Concepts: Two Dialogues', en R. Calinger (editor), *Vita Matemática. Historical Research and Integration with Teaching.* MAA Notes. No. 40, pp 27-41.

Kindt, M. y Jan de Lang (1986). *De Normale Verdeling*. Educaboek b.v. Culemborg. The Netherlands, 1986. Proyecto Hewet.

Newman, J. R. (1976). *Sigma. El Mundo de las Matemáticas*. 5 volúmenes. Ediciones Grijalbo. Barcelona. (Traducción de la obra: *The World of Mathematics*. Simon and Schuster. New York. 1956).

Shoesmith, E. (1987). The continental controversy over Arbuthnot's argument for divine providence. *Historia Mathematica.* **14** (2), pp 133-146.