Temas de Bioestadística

por

ANTONINO RULLI

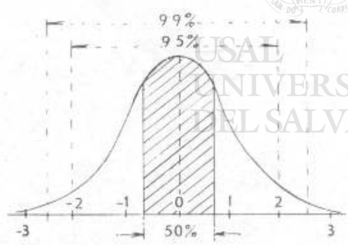
1. Límites probables.

—Profesor: como sabemos que Vd. se ha dedicado últimamente a la Enseñanza de Cálculo Estadístico con orientación biológica, venimos a hacerle, como alumnos que todavía somos, una consulta sobre temas de la materia. Traemos escrito algunos de ellos. Son de ejercicios propuestos por autores muy recomendados (entre ellos HULDAH BANCROFT), que no llevan solución en sus libros y que, creemos, pueden ser muy útiles para nosotros, permitiéndonos un repaso y ampliar y profundizar, si es posible, nuestros —imposible disimularlo— nuestros escasos conocimientos.

—Sean bienvenidos. Estoy a disposición de Vds. en lo que pueda. Veamos si me es posible complacerles.

Se sabe que el 30 % de la población infantil ha tenido ataques de coqueluche. Dada una muestra de 350 escolares, se pide calcular límites probables entre los cuales se encuentra aquella proporción.

—Bosquejando la curva de distribución normal de GAUSS, podemos ver cómo y por dónde quedan esos límites, a cada lado de la media poblacional, debiéndose ellos precisar numéricamente.



Fórmula para el cálculo en unidades:

$$l = np \pm t \sqrt{npq}$$

Es $np = 350 \times 0{,}3$ la media binomial; $\sqrt{npq}$ la desviación standard, o error standard; y t el coeficiente de probabilidad o desvío relativo, que varía con la probabilidad fiducial que se desea y que suele ser 1,96 para P = 95 %. Se obtiene:

$$l = 350 \times 0{,}30 \pm 1{,}96 \sqrt{350 \times 0{,}3 \times 0{,}7}$$

Por comodidad de cálculo y más seguridad en el resultado, sustituiremos a t = 1,96 por 2. Así resulta el intervalo

$$l = 105 \pm 2\ (8{,}6)$$
$$l = 105 \pm 17{,}2$$
$$l = (87,\ 123)$$

redondeando en cifras enteras.

En porcentaje, el cálculo es:

$$l = p \pm t \sqrt{\frac{pq}{n}}$$

$$l = 30 \pm 2 \sqrt{\frac{30.70}{350}}$$

$$l = 30 \pm 2\ (2{,}45)$$
$$l = 30 \pm 4{,}9$$
$$l = (25\ \%,\ 35\ \%)$$

Traducimos las respuestas: entre 87 y 123 niños, o sea entre el 25 y el 35 % de la muestra han tenido ataques de coqueluche con una probabilidad de 95 %.

Si quisiéramos el cálculo, con la probabilidad del 50 %, el intervalo sería menor:

$$l_1 \pm 105 - 0{,}675\ (8{,}6) = 99{,}2 \sim 99$$
$$l_2 = 105 + 0{,}675\ (8{,}6) = 110{,}8 \sim 111$$

es decir:

$$(l_1,\ l_2) = (99;\ 111)$$

Pero, si pretendemos el resultado con probabilidad del 99 %, el intervalo tendría que ser mucho más amplio:

$$l = 105 \pm 2{,}58\ (8{,}6)$$
$$(l_1,\ l_2) = (82;\ 127)$$

—¿Conformes Vds.?

—Sí, profesor, gracias. Tenemos también anotado el siguiente problema:

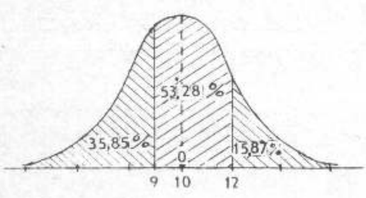
2. Distribución normal.

Los valores X forman una distribución normal gaussiana con un promedio de 10 y desviación standard 2. Se quiere saber cuál es la probabilidad de que sea X mayor que 12, X menor que 9, y X comprendido entre 9 y 12. Es decir:

a) $X > 12$ b) $X < 9$

c) $9 < X < 12$

—Esquemáticamente muestran las soluciones los distintos rayados de la figura presente.



Me permitiré confiar en la adecuada interpretación de Vds. en lo que explique, que será abreviando y evitando la explicación de detalles que crea innecesaria, salvo que ustedes la pidan.

a) Probabilidad de los valores que varían de 12 a $+\infty$, esto es: $P(12, +\infty)$.

Operamos con desvíos relativos:

$$P\left(\frac{12\text{-}10}{2}, +\infty\right) = P(1, +\infty)$$

Usando tablas de probabilidad de $-\infty$ a $+\infty$, hallamos:

$$P(1, +\infty) = 1 - P(-\infty, 1) = 1 - 0{,}8413 = 0{,}1587$$

También:

$$P(1, +\infty) = P(-\infty; -1) = 0{,}1587 \text{ (Resp. a)}$$

b) Probabilidad de los valores X que varían de $-\infty$ a 9:

$$P\left(-\infty, \frac{9\text{-}10}{2}\right) = P(-\infty, -0{,}5) = 0{,}3085 \text{ (Resp. b)}$$

c) La probabilidad de X entre 9 y 12, es:

$$P\left(\frac{9\text{-}10}{2}, \frac{12\text{-}10}{2}\right) = P\left(-\frac{1}{2}\ 1\right) =$$

$$= P(-\infty, 1) - P(-\infty, -0{,}5)$$

$$= 0{,}8413 - 0{,}3085 = 0{,}528 \quad \text{(Resp. c)}.$$

—En realidad, profesor, el problema que nos ha preocupado más y que nos parece muy especialmente importante, es el siguiente:

3. Distribución binomial de probabilidad.

Un cirujano desarrolla una técnica quirúrgica nueva para una enfemedad en la cual la mortalidad postoperatoria usual es de 20 %. Aplica la técnica en 10 pacientes y no tiene casos fatales.

a) ¿Cuál es la probabilidad de que tratando 10 pacientes, no haya muertes respecto de la enfermedad de la que debe esperarse una mortalidad del 20 %?

b) ¿Cree Vd. que el cirujano está autorizado para creer que su técnica constituye un progreso con respecto al método del tratamiento anterior?

c) ¿Cuál es el menor número de pacientes que tendría que tratar, sin ningún caso fatal antes de poder afirmar que su técnica representa un adelanto significativo respecto de la técnica antigua?

—Verdaderamente el problema toca fundamentos. El desarrollo binomial podrá darnos la solución a las tres partes del problema.

El desarrollo es:

$$(q+p)^n = q^n + C_n^1 q^{n-1} p + \dots + C_n^k q^{n-k} p^k + \dots + C_n^n q^0 p^n$$

es decir:

$$(0{,}2 + 0{,}8)^{10} = (0{,}2)^{10} + 10\,(0{,}2)^9\,(0{,}8) + \dots + 10\,(0{,}2)\,(0{,}8)^9 + (0{,}8)^{10}$$

a) El último término da la probabilidad de la variante $k = n = 10$. Es decir:

$$P = (0,8)^{10}$$

$$\log P = 10 \log 0,8 = 10\ (\bar{1},9031) =$$

$$= 10\ (-\ 0,0969) = -\ 0,969 = \bar{1},031$$

de donde

$$P = 0,1074 = 10,74\ \% \qquad \text{(Resp.)}$$

Esta es la probabilidad de que no haya muertes en 10 intervenciones.

b) Hay acuerdo muy generalizado entre los estadísticos en considerar, en toda distribución normal, como significativos los desvíos unilaterales superados por la probabilidad de 2,5 %, aceptando el intervalo fiducial de 95 %. Sentada esta condición, corresponde rechazar el resultado del problema como significativo, vale decir como superior al usual.

c) Para hallar el número n de pacientes que llegan al umbral de la R.C. (región crítica) de probabilidad 2,5 % bastará resolver la ecuación exponencial

$$(0,8)^{n} = 0,025$$

Se obtiene:

$$n \log 0,8 = \log 0,025$$

$$n = \frac{\log 0,025}{\log 0.8} = \frac{\bar{2},3979}{\bar{1},9031}$$

$$n = \frac{1,6021}{0,0969} = 16,53$$

Aceptamos como respuesta el número $n = 17$ de pacientes intervenidos sin casos fatales, como número mínimo para que la nueva técnica esté en condiciones de competir con la usual.

—¿Por qué no es de 5 % la R.C. como la hemos visto aplicar otras veces?

—Porque esta vez no consideramos los dos extremos del desarrollo, sino solamente el último de sobrevivientes, que es el único que interesa.

—Nuestra última pregunta, profesor, si nos la permite. Y, solamente para verificar o corregir nuestro dudoso tratamiento numérico. Se trata del siguiente problema:

4. Grado de asociación de atributos. Método de Ji cuadrado.

Fueron inyectados con un germen infeccioso, 50 ratones de laboratorio para experimentar una vacuna preventiva. Con ésta se vacunaron 40, se observaron reacciones positivas en 28, y de los no vacunados sobrevivieron 2. Se pide:

a) El grado de asociación entre los atributos componentes, vacunación y contaminación mediante el cuadro tabular 2×2 de G. U. Yule.

b) La ampliación del cuadro real anterior con las clases binarias teóricas de componentes independientes, verificando el necesario grado de asociación que los comprometen. Y, finalmente:

c) Por el método de Ji cuadrado, determinar si hay asociación significativa entre la vacunación y la enfermedad.

—Muy bien. El problema pertenece al tipo del que pensaron resolver los estadísticos norteamericanos cuando los astronautas del Apolo XI volvieron a la Tierra después del maravilloso viaje a la Luna. Con el material traído del suelo lunar, supuestamente con microbios, gérmenes o virus patógenos, los científicos prepararon una solución que inyectaron a cierto número de roedores. Pero éstos no mostraron signo alguno de contaminación, para suerte de los habitantes de la Tierra que, desde entonces, con otras pruebas valiosas, pudieron fundadamente considerarse fuera de todo peigro de contagio y, por lo tanto, de una temida propagación universal.

Vamos al problema que Vds. han traído. Comenzaremos por hacer la clasificación dicotómica que ofrece a simple vista una distribución lógica, clara y total de los datos, con notación conocida:

$$
N = 50 \left\{ \begin{array}{l} (A) = 40 \left\{ \begin{array}{l} (AB) = 28 \\ (A\beta) = 12 \end{array} \right. \\ \\ (\alpha) = 10 \left\{ \begin{array}{l} (\alpha B) = 2 \\ (\alpha\beta) = 8 \end{array} \right. \end{array} \right.
$$

a) Tabla de contingencia 2 × 2, en parte con frecuencias reales conocidas:

	B	β	
A	28 (24)	12 (16)	40
α	2 (6)	8 (4)	10
	30	20	50

La fórmula de Yule:

$$
Q = \frac{(AB) \cdot (\alpha\beta) - (A\beta) \cdot (\alpha B)}{(AB) \cdot (\alpha\beta) + (A\beta) \cdot (\alpha B)}
$$

da el grado de asociación pedido:

$$
Q = \frac{28 \cdot 8 - 12.2}{28 \cdot 8 + 12.12} = + 0,806
$$

altamente favorable teniendo en cuenta que la asociación perfecta es de grado $Q = +1$.

b) Ahora, calculemos las frecuencias de clase con la hipótesis de atributos componentes independientes. La expresión que acusa la frecuencia de vacunados sobrevivientes, es:

$$(AB)_0 = \frac{(A)\ (B)}{N}$$

Su cálculo da

$$(AB)_0 = \frac{40.30}{50} = 24$$

Las otras pueden calcularse de modo semejante, pero también por substracción como sigue:

$$\begin{aligned} (A\beta)_0 &= 40 - 24 = 16 \\ (\alpha B) &= 30 - 24 = 6 \\ (\alpha\beta) &= 20 - 16 = 4 \end{aligned}$$

Estas frecuencias teóricas, figuran como se ha pedido, ampliando el cuadro anterior. El grado de asociación que corresponde a los atributos componentes, en este caso, también respondiendo a pregunta del enunciado, es como debe esperarse:

$$Q = \frac{24 \cdot 4 - 16 \cdot 6}{24 \cdot 4 + 16 \cdot 6} = \frac{96 - 96}{96 + 96} = 0$$

c) Sólo queda, aplicar el método de Ji cuadrado, de K. PEARSON, para obtener un resultado más preciso, y llevando consigo, además, la probabilidad del suceso.

Los valores observados (O), los teóricos ya calculados (C), las diferencias (O-C), $(O\text{-}C)^2$, y los cocientes $(O\text{-}C)^2/C$ forman el siguiente cuadro, necesario para efectuar este último cálculo:

O	C	O-C	$(O\text{-}C)^2$	$(O\text{-}C)^2/C$
28	24	4	16	0,67
12	16	–4	16	1,00
2	6	–4	16	2,67
8	4	4	16	4,00
50	50			8,34 **

En la tabla de R. A. FISHER de Ji cuadrado, con un grado de libertad y umbrales de significación 5 y 1 % de probabilidad, se encuentra:

$$\chi^2 \begin{cases} n' = 1;\ 5\,\% \ \ldots\ldots\ 3{,}841 \\ n' = 1;\ 1\,\% \ \ldots\ldots\ 6{,}635 \end{cases}$$

El resultado del cálculo: $\chi^2 = 8{,}34$ se ubica en la R.C. de la distribución tabular, muy inferior al nivel de probabilidad 1 %. Es, pues, altamente significativo. Vale decir que, las diferencias entre los valores observados por efecto de la vacuna experimentada y los teóricos calculados con la hipótesis de independencia de los atributos componentes, son óptimamente satisfactorios.

—Profesor: esto es todo y más de lo que era nuestra intención de consultarle hoy. Si nos autoriza volveremos otro día para aclarar otros puntos obscuros para nosotros. Por ejemplo, de Correlación y Regresión, de Comparación de pequeñas muestras por el método de "STUDENT", y, algún otro que, como los anteriores, traeremos escrito para ser precisos.

—Desde luego, están autorizados. He tenido y tendré gran placer en atenderlos cuando gusten. Los espero pues.

—Muchas gracias, profesor.