	Aprovações	Reprovações	Sub-totais
Liceus. . . .	88,00	55,00	143
Ens.Téc. . .	32,00	20,00	52
Sub-totais.	120,00	75,00	195

Êstes números foram calculados de acôrdo com a hipótese anterior, isto é, na suposição de que a percentagem de aprovações é a mesma para os dois grupos de candidatos.

A variância de qualquer das entradas na tabela dada é aproximadamente igual ao recíproco da soma dos recíprocos dos valores esperados:

$$\frac{1}{V}=\frac{1}{88}+\frac{1}{55}+\frac{1}{32}+\frac{1}{20}=0,1108$$

$$V=9,003 \quad \text{Desvio padrão}=\sqrt{V}=3,00$$

O desvio observado é $92-88,00=4,00$ mas para julgar da sua significância é conveniente, por uma questão de rigor, aplicar a correcção de Yates, o que dá para o referido desvio o valor final 3,50. A razão entre êste e o seu desvio padrão é, portanto $\frac{3,50}{3,00}=1,17$

Ora uma tabela de áreas da curva normal mostra que êste desvio corresponde a um nível de significância de $P=24\,\%$. Isto é, mesmo que os dois grupos de estudantes fôssem amostras casuais da mesma população seria de esperar, aproximadamente, uma vez em cada quatro, um desvio da proporcionalidade calculada, tão grande ou maior que o revelado pelos dados.

Não há portanto razão para suspeitar que as escolas técnicas sejam menos eficientes do que os liceus no ensino das matemáticas visto os dados não fornecerem evidência que justifique tal conclusão.

ESTATÍSTICA MATEMÁTICA

CONTRIBUCIÓN AL ESTUDIO DE LAS MÉDIAS DE UNA SERIE ESTADISTICA

por *O. Fernández Baños* (professor da Universidade de Madrid)
e *Jimenez Montoya* (licenciado em Ciências)

Es sabido que según la definicion corriente, se llama valor medio de una serie estadistica cuantitativa a todo valor intermedio entre el menor y el mayor de ella, no siendo todos iguales; y, por consiguiente, es obvio que toda serie estadistica cuantitativa de términos no todos iguales admite infinitas medias.

Muchas son las fórmulas ideadas para expresar valores medios de una serie estadistica de números reales; y segun la definicion genérica dicha se presenta naturalmente el problema de hallar una expresion general que represente todas las medias posibles en funcion de un parámetro real, que sea de cálculo sencillo y que, a ser posible, contenga como casos particulares a las usadas habitualmente.

La cuestion queda resuelta mediante la llamada media general:

$$f(m)=\sqrt[m]{\frac{\sum_{i=1}^{n} X_i^m}{n}}=M_m \qquad (1)$$

siendo m un número real, n un número entero y positivo y X_i números positivos reales; condicion esta última que no restringe el valor práctico de la fórmula porque basta un simple cambio de origen en la serie primitiva dada.

La fórmula (1) cumple las condiciones siguientes:

a) el limite de $f(m)$ cuando m decrece indefinidamente es el más pequeño de la serie y cuando crece indefinidamente tiene como limite al mayor de la misma.

En efecto, ordenando los X_i en orden creciente, tendremos:

$$X_i=\alpha_i X_n\,;\ \alpha_i<1 \qquad i=1,2,\cdots n-1$$

$$X_i=\beta_i X_1\,;\ \beta_i>1 \qquad i=2,3,\cdots n$$

$$\log f(m)=\frac{1}{m}\log\left[X_n^m\frac{\alpha_1^m+\alpha_2^m+\cdots+\alpha_{n-1}^m+1}{n}\right]=$$

$$=\log X_n+\frac{1}{m}\log\left[\frac{\alpha_1^m+\alpha_2^m+\cdots+\alpha_{n-1}^m+1}{n}\right]$$

$$\log f(m)=\frac{1}{m}\log\left[X_1^m\frac{1+\beta_2^m+\beta_3^m+\cdots+\beta_n^m}{n}\right]=$$

$$=\log X_1+\frac{1}{m}\log\left[\frac{1+\beta_2^m+\beta_3^m+\cdots+\beta_n^m}{n}\right]$$

$$\lim_{m\to\infty}\log f(m)=\log X_n\,;\quad \lim_{m\to\infty} f(m)=X_n$$

$$\lim_{m\to-\infty}\log f(m)=\log X_1\,;\quad \lim_{m\to-\infty} f(m)=X_1$$

b) Dando valores enteros a m se obtienen las medias mas usuales en Estadistica. En efecto:

$$
\begin{aligned}
f(-1) &= \text{la media armónica} = M_{-1} \\
f(1) &= \text{» » aritmética} = M_1 \\
f(2) &= \text{» » cuadrática} = M_2 \\
f(3) &= \text{» » cúbica} = M_3 \\
&\dots\dots\dots
\end{aligned}
$$

y el limite de $f(m)$ cuando $m \to 0$ es la media geométrica.

Como para $m=0$ la funcion toma la forma no definida $\sqrt[0]{1}$ aplicaremos el procedimento ordinario de hallar su limite y tendremos segun la regla de l'Hopital

$$\lim_{m\to 0} \log f(m) = \lim_{m\to 0} \frac{1}{m} \log\left[\frac{X_1^m + \cdots + X_n^m}{n}\right] = \frac{0}{0}$$

$$\lim_{m\to 0} \log f(m) = \lim_{m\to 0} \frac{X_1^m \log X_1 + \cdots + X_n^m \log X_n}{X_1^m + \cdots + X_n^m} = $$

$$= \frac{\sum_1^n \log X_i}{n}, \tag{2}$$

de donde:

$$\lim_{m\to 0} f(m) = \sqrt[n]{X_1^m X_2^m \cdots X_n^m} = M_0 = \text{media geométrica.}$$

c) La $f(m)$ además de ser continua en todo el campo real sin otra singularidad que la dicha del punto cero — lo cual es elemental en Matemáticas — es funcion creciente; y por consiguiente esta propriedad en unión de las dos anteriores permite afirmar que a cada valor de m finito corresponde un valor medio de la serie estadistica dicha y reciprocamente a cada valor medio de esta un valor de m.

Para esto basta demostrar que la derivada de $f(m)$ es positiva. En efecto:

$$y = \log f(m) = \frac{1}{m}\left[\log\left(X_1^m + X_2^m + \cdots + X_n^m\right) - \log n\right]$$

$$\frac{dy}{dm} = \frac{1}{m^2}\left[m \frac{X_1^m \log X_1 + \cdots + X_n^m \log X_n}{X_1^m + X_2^m + \cdots + X_n^m} - \right.$$

$$\left. - \log \frac{X_1^m + \cdots + X_n^m}{n}\right] \tag{4}$$

que es positiva, porque:

$$m\frac{X_1^m \log X_1 + \cdots + X_n^m \log X_n}{X_1^m + \cdots + X_n^m} =$$

$$= \frac{\log\left[(X_1^m)^{X_1^m} (X_2^m)^{X_2^m} \dots (X_n^m)^{X_n^m}\right]}{X_1^m + X_2^m + \cdots + X_n^m} > \log \frac{1}{n} \sum_{i=1}^n X_i^m$$

o sea:

$$\frac{\log\left[(X_1^m)^{X_1^m} (X_2^m)^{X_2^m} \dots (X_n^m)^{X_n^m}\right]}{\log\left[\frac{X_1^m + X_2^m + \cdots + X_n^m}{n}\right]^{X_1^m + X_2^m + \cdots + X_n^m}} > 1$$

o sea:

$$\frac{(X_1^m)^{X_1^m} (X_2^m)^{X_2^m} \dots (X_n^m)^{X_n^m}}{\left[\sum_{i=1}^n \frac{X_i^m}{n}\right]^{X_1^m} \left[\sum_{i=1}^n \frac{X_i^m}{n}\right]^{X_2^m} \cdots \left[\sum_{i=1}^n \frac{X^m}{n}\right]^{X_m^m}} > 1$$

lo cual reduce la cuestion a demostrar que la expresion $a_1^{\alpha_1} a_2^{\alpha_2} \dots a_n^{\alpha_n} = Z$ sometida a la condicion de que: $a_1 + a_2 + \cdots + a_n = \alpha_1 + \alpha_2 + \cdots + \alpha_n = \varphi$ toma su valor máximo cuando: $a_i = \alpha_i$ $i = 1, 2, \cdots n$. Para demostrarlo aplicamos el método de Lagrange a las primeras derivadas respecto a las a_i y tendremos:

$$\frac{d}{da_i} \log Z - \lambda \frac{d\varphi}{da_i} = \frac{\alpha_i}{a_i} - \lambda = 0 \text{ para } i = 1, 2, \cdots n; \text{ o sea:}$$

$$\frac{\alpha_1}{a_1} = \frac{\alpha_2}{a_2} = \cdots = \frac{\alpha_n}{a_n} = \lambda = \frac{\alpha_1 + \alpha_2 + \cdots + \alpha_n}{a_1 + a_2 + \cdots + a_n} = 1.$$

Finalmente, para comprobar que a tales valores corresponde un máximo y no un minimo de la funcion Z tomemos la forma cuadrática formada por las segundas derivadas y veremos que es negativa porque:

$$\frac{d^2}{da_i^2} \log Z = -\frac{\alpha_i}{a_i^2} \text{ para } i = 1, 2, \cdots n$$

$$\frac{\partial^2}{\partial a_r \partial a_s} \log Z = 0 \text{ para } r \neq s$$

recordando que las α_i y sus iguales las a_i son, por hipótesis, números mayores que cero en nuestro caso.

La curva representativa de la funcion $y = f(m)$ tiene como asintotas las rectas paralelas al eje de de abscisas m a las distancias X_1 y X_n entre las cuales se conserva y crece con un punto de inflexion correspondiente al valor de la media geométrica.

Es obvio que no dependiendo de m el valor de la mediana de la serie — la qual es convencional cuando el número de términos es par — no existe un valor fijo de m para el qual siempre corresponda en $f(m)$ la mediana de la serie; aunque siempre se verifica que, dada la serie estadistica de termos positivos no todos iguales hay un número real m para el cual $f(m)$ toma el valor de la mediana por la forma dicha de la curva.

A las formas particulares de la funcion $f(m)$ cuando m toma valores enteros corresponden propiedades caracteristicas de cada una que las hacen útiles o perjudiciales para ciertos fines, tales como son, verbi gratia, la construccion de números indices.

La propiedad práctica mas importante de la $f(m)$ — aparte la de tomar todos los valores entre el menor y el mayor de la serie — es que estabelece una ordenacion sencilla de magnitud entre todas las medias de la cual resulta que siempre es menor la media armónica que la geométrica, ésta menor que la aritmética, ésta menor que la cuadrática, etc.

Es obvio que, conteniendo $f(m)$ todos los valores medios de la serie si se hace alguna clasificacion entre todas las medias utilizando alguna propriedad caracteristica de la clasificacion no es posible que tal propiedad pertenezca a $f(m)$ en general. Así sucede, por ej. a la clasificacion en erráticas y fijas (Furlan V. «Sur une formule generale de la moyenne», «Metron» marzo 1928) siendo estas las que cumplen la propiedad de que la media de la serie total no altere al substituir un número de terminos de la serie por su media parcial respectiva. Esta propiedad es evidentemente cumplida por la media aritmética y no por la mediana y al analizar la $f(m)$ tendremos que si ha de cumplirse tal propiedad, se verificará: $\left[\dfrac{X_1^m+\cdots+X_n^m}{n}\right]^{\frac{1}{m}}=$ $=\left[k\left(\dfrac{X_1^m+\cdots+X_k^m}{k}\right)^{\frac{1}{m}}+X_{k+1}^m+\cdots+X_n^m\right]^{\frac{1}{m}}$ siendo $k<n$ lo cual solo se verifica cuando m toma los valares 0 y 1 lo cual valora las medias geométrica e aritmética en orden al estudio de la variabilidad de las series estadisticas.

Para encontrar multitud de relaciones entre las diversas medias consideremos la funcion:

$$Q(m)=\frac{\sum_1^n X_i^m}{\sum_1^n X_i^{m-1}}=\frac{M_m^m}{M_{m-1}^{m-1}} \tag{6}$$

para $m>0$, y

$$Q(-m)=\frac{\sum_1^n X_i^{-m}}{\sum_1^n X_i^{-m-1}}=\frac{M_{-m}^{-m}}{M_{-m-1}^{-m-1}}=\frac{M_{-m}^{-m}}{M_{-(m+1)}^{-(m+1)}} \tag{7}$$

para $m<0$

$$Q(1)=f(1)=M_1,\ Q(0)=f(-1)=M_{-1},\ f_{(m)}^m=M_m^m \tag{8}$$

tenemos que la funcion M_m^m es continua en todo el campo real con limites cero e infinito al variar m de menos a mas infinito cuando las X_i son mayores que la unidad o con los mismos limites invertidos cuando las X_i son menores que la unidad en el supuesto de ser todas positivas por lo cual no cumple la condicion básica de valor medio de una serie estadistica excepto cuando m se conserva dentro del rango en que tal funcion toma valores entre el menor y el mayor de la serie.

De las anteriores relaciones resulta inmediatament

$$M_m^m=Q(m)\,M_{m-1}^{m-1}=Q(m)\,Q(m-1)\,M_{m-2}^{m-2}=$$
$$=Q(m)\,Q(m-1)\cdots Q(2)\,M \quad \text{para } m \text{ positivo} \quad (\!$$

$$M_{-(m+1)}^{-(m+1)}=\frac{1}{Q(-m)}M_{-m}^{-m}=\frac{1}{Q(-m)\,Q(-m+1)}M_{-m+1}^{-m+1}=$$
$$=\frac{1}{Q(-m)\,Q(-m+1)\cdots Q(-1)\,M_{-1}}$$

o sea:

$$M_{-m}^{-m}=\frac{1}{Q(-m+1)\,Q(-m+2)\cdots Q(-1)\,M_{-1}} \quad (10$$

para m negativo.

Estas fórmulas permiten escribir un número indefinido de relaciones entre las diversas medias tales como:

$$M_2^2=M_1\,Q(2) \tag{11}$$
$$M_3^3=M_1\,Q(2)\,Q(3) \tag{12}$$
$$M_{-2}^2=Q(-1)\,M_{-1} \tag{13}$$

que nos dicen, respectivamente, que la media cuadrática es media geométrica entre la aritmética y la antiarmónica; la media cúbica es media geométrica entre la aritmética, la antiarmónica y la $Q(3)$ que no recibe nombre especial; la media $f(-2)$ es media geométrica entre la armónica y la $Q(-1)$, etc., etc.

La forma sencilla de la funcion $Q(m)$ permite compararlas fácilmente entre si, cuando m toma valores enteros consecutivos, para ver cuanto es una mayor que su consecutiva. En efecto:

$$\frac{Q(m)}{Q(m-1)}=\frac{\sum_1^n X_i^m \sum_1^n X_i^{m-2}}{\left|\sum_1^n X_i^{m-1}\right|^2}=$$
$$=\frac{\sum_1^n X_i^{2(m-1)}+\underset{h\neq k}{S}X_h^m X_k^{m-2}}{n^2\left|\frac{1}{n}\sum_1^n X_i^{m-1}\right|^2}=$$
$$=\frac{\left[\sum_1^n X_i^{m-1}\right]^2+\underset{h\neq k}{S}(X_h-X_k)^2 X_h^{m-2}X_k^{m-2}}{n^2\left[\frac{1}{n}\sum_1^n X_i^{m-1}\right]^2}=$$
$$=1+\frac{\underset{h\neq k}{S}(X_h-X_k)^2 X_h^{m-2}X_k^{m-2}}{n^2 M_{m-1}^{2(m-1)}}$$

que puede verse en Mortara, «Lezioni di Statistica Metodologica» pg 108.