DINÁMICA DE POBLACIONES: EL ANÁLISIS DE SU ESTIMACIÓN Y LA DIFICULTAD DE SU ENSEÑANZA

F. J. Zamudio S.[1]

[1] **Profesor-Investigador de tiempo completo de la Universidad Autónoma Chapingo, División de Ciencias Forestales, Departamento de Estadística, Matemática y Cómputo.**

RESUMEN

El trabajo presenta con detalle el desarrollo de la estimación de una población que cambia a través del tiempo (dinámica de poblaciones), con la adicional dificultad de la estimación de la característica de la población que se desea estudiar. Las ideas involucradas son más complejas de lo que aparentan, haciéndolas difíciles en su uso para quienes las necesitan y estudian en sus cursos de licenciatura y maestría, así como la dificultad del educador para trasmitir los conocimientos esenciales que le permitan al estudiante adecuarlas a situaciones análogas que pueden analizarse con esta técnica.

PALABRAS CLAVES: Dinámica de poblaciones, enseñanza, estimación.

POPULATION DYNAMICS: ANALYZING ESTIMATION AND TEACHING DIFFICULTIES

SUMMARY

This work presents with detail the development to estimate a population that change with time (dynamics of populations), complicated with the estimation of the characteristic under study. The involved ideas are more complex than they appear, making them difficult in their use to whom need and study this topic in undergraduate and graduate programs, besides the problems of teaching to explain the essential knowledge to the student which allow him to figure it out how to use it in similar situations whose analysis could be with this technique.

KEY WORDS: Dynamics of populations, teaching, estimation.

INTRODUCCIÓN

En el presente artículo se analiza un ejemplo de estimación para determinar distribuciones de edades en poblaciones salvajes. La intención original donde se encuentra este ejemplo (Searle, 1982: pp. 133-134) es muy simple, exhibir el uso de la matriz inversa en la solución de un sistema de ecuaciones simultáneas. Ahora, la intención es mostrar las ideas subyacentes al ejercicio que, sintetizado en una ecuación, involucra conocimientos de la matemática, la probabilidad y la estadística, al menos entre otras disciplinas. Con lo anterior se busca que el lector observe la distancia que dista de una ecuación, de apariencia sencilla, a la comprensión de la misma. Hay para quienes esta distancia es tan larga que sucumben en las primeras ideas, abandonando el propósito de comprender cabalmente su significado y conformándose con usar dicha ecuación en situaciones parecidas (cuando les es posible descubrir el parecido), mecánicamente y, suprimiéndose la posibilidad de trasladar dicho conocimiento a situaciones distintas donde el manejo teórico adecuado de las ideas que lo rodean permitiría también su uso. Otro objetivo es mostrar los problemas inherentes de la enseñanza, cuando los maestros sólo repiten a sus alumnos lo que los libros muestran, por carecer ellos mismos de las teorías implícitas de lo que enseñan, omitiendo ideas sustantivas de lo que se quiere trasmitir. Por último y como complemento de lo anterior, ver las vicisitudes del maestro para compartir con sus estudiantes el conocimiento científico que semeja altísimos edificios, en los cuales, con frecuencia, no se sabe el nivel en el que se está.

Recibido: 18 de octubre, 2004
Aceptado: 19 de noviembre, 2004

En seguida aparece el ejemplo, motivo del análisis, traducido del inglés.

Distribuciones de edad en poblaciones salvajes. Un problema que ocurre al estudiar la dinámica de poblaciones salvajes es el de determinar las distribuciones de edades. Suponer para alguna población animal siendo estudiada, digamos venados, que hay *n* grupos de edad y la proporción de la población en el grupo de edad *i* es r_i para $i=1,2,\ldots,n$. Para una población de venados salvajes, estos valores nunca pueden ser conocidos exactamente porque no todos los venados pueden ser localizados, ni pueden sus edades ser conocidas con exactitud. Pero las edades de los animales que pueden ser localizados pueden ser estimadas, de las cuales estimaciones $\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_n$ de las proporciones poblacionales pueden ser calculadas. Venados cuyas edades exactas son conocidas también pueden estar disponibles, en la forma de una colonia nacida y criada en cautiverio, o de venados nacidos en cautiverio, marcados con su fecha de nacimiento, liberados y posteriormente recapturados. De tales animales uno puede determinar la eficacia del personal cuya tarea normal es estimar la edad de venados capturados, de la población salvaje, alcanzando esto al hacer que tal personal estime las edades de un número de venados cuyas edades verdaderas estén disponibles. De esta manera es posible poner un valor sobre p_{ij}, la probabilidad que un animal estimado como de edad *i* es realmente de edad *j*. Entonces, por ejemplo, la fracción estimada en la edad 1 en la población, $\hat{r}_1$, es la suma de las fracciones verdaderas, $r_1, r_2, \ldots, r_n$ cada una multiplicada por la probabilidad de que tal animal es estimado como de edad 1. Así,

$$\hat{r}_1 = p_{11}r_1 + p_{12}r_2 + \ldots, p_{1n}r_n \,, \qquad (*)$$

y para todos los grupos de edad juntos

$$\begin{bmatrix} \hat{r}_1 \\ \hat{r}_2 \\ \vdots \\ \hat{r}_n \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix}$$

Por tanto, si $\hat{\mathbf{r}}$ denota el vector de proporciones estimadas, **r** el vector de proporciones verdaderas y **P** la matriz de probabilidades, $\hat{\mathbf{r}} = \mathbf{P} \cdot \mathbf{r}$. Estimaciones mejoradas de las proporciones pueden ser obtenidas resolviendo esta ecuación para **r** usando la inversa de P, es decir, $\mathbf{r} = \mathbf{P}^{-1}\hat{\mathbf{r}}$.

Hasta aquí la traducción.

La Ecuación (*), cuando se explica en clase, a unos les parece plausible, otros hacen como que les parece, pero veamos que hay detrás de tal ecuación para determinar qué tan plausible puede ser si profundizamos en ella y encontramos todo lo que se debió conocer y recordar para entenderla.

Desarrollo teórico

Se tiene una población $P = \{x_1, \cdots, x_n\}$, por ejemplo, venados en un bosque. Los elementos de esta población poseen características, por ejemplo su edad, cuya distribución se desea conocer. Digamos que $h(\cdot)$ es la característica bajo estudio, en el ejemplo $h(x_i)$ significa la edad del venado x_i. En general, las características como la edad, son expresadas como números reales, *i.e.*,

$$h : P \to \Re$$

Para conocer la distribución de la característica $h(\cdot)$ en *P*, es suficiente conocer la proporción de valores $h(\cdot)$ que pertenecen a cada una de las partes de una partición adecuada en $\Re$. Digamos, si $\{C_i\}_{i=1,\ldots,p}$ es una partición sobre $\Re$, sólo es necesario conocer la proporción de $h(\cdot)$ que pertenecen a cada una de las C_i, $i = 1,\ldots,p$. Si hacemos,

$$P_i = \{x : h(x) \in C_i\}, \quad i = 1,\ldots,p \qquad (1)$$

y denotamos por $n_i = \#(P_i)$, el número de elementos *x* en P_i, entonces la proporción r_i de elementos *x* en *P* cuya característica $h(x)$ está en C_i es

$$r_i = \frac{n_i}{n} = \frac{\#(P_i)}{\#(P)} = \Pr(P_i), \quad i = 1,\ldots,p \qquad (2)$$

que puede verse como la probabilidad de ocurrencia de P_i, i.e., $Pr(P_i)$.

En general, como en el ejemplo de la distribución de la edad de los venados en un bosque, no es posible acceder a cada elemento y, a los que se accede, no es posible con precisión conocer el valor de la característica bajo estudio $h(\cdot)$, como puede verse cuando hablamos de la edad de un venado.

Entonces, para estimar esta distribución nos tenemos que conformar con:

- El valor que le dé un experto al verdadero valor de la característica $h(\cdot)$. Es decir, los valores de la característica de los elementos en esta muestra, $h(y_1),\ldots,h(y_m)$, son desconocidos con precisión, pero el experto les dará un valor (los estimará) a cada uno de ellos; los valores estimados se pueden denotar como: $\hat{h}(y_1),\ldots,\hat{h}(y_m)$.
- Una muestra aleatoria de *P*, digamos $M=\{y_1,\ldots,y_m\}$

Para proceder tenemos que evaluar la experiencia o precisión del experto.

Suponer que se tiene un conjunto de elementos $L=\{z_1,\ldots, z_r\}$ de los cuales se conoce con precisión $h(z_1),\ldots,h(z_r)$. Es decir, en el ejemplo, un grupo de *r* venados de los cuales se conoce su edad.

Sin proporcionarle al experto información sobre las $h(z_i)$, se le pide que determine el valor $h(z_i)$, i=1, …,r, o que estime $h(z_i)$ y obtenga los valores $\hat{h}(z_i)$, i=1, …,r.

Obtenidos los $\hat{h}(z_i)$ y conocidos los $h(z_i)$ se puede construir la siguiente matriz:

$$\mathbf{N}=\{n_{ij}\}_{\substack{i=1,\ldots,p\\ j=1,\ldots,p}}\text{, con } \sum_{i,j} n_{ij}=r \tag{3}$$

n_{ij}: el número de veces que el experto estimó $h(\cdot)$ como perteneciente a la categoría C_i cuando en realidad pertenece a la C_j

Con estas frecuencias se puede construir la matriz

$$\mathbf{Q}=\{q_{ij}\}_{\substack{i=1,\ldots,p\\ j=1,\ldots,p}}=\begin{bmatrix}\mathbf{q}_1'\\ \vdots\\ \mathbf{q}_p'\end{bmatrix}\text{, donde } \mathbf{q}_i'=\lfloor q_{i1}\, q_{i2}\ldots q_{ip}\rfloor \tag{4}$$

q_{ij}: Probabilidad de que el experto clasifique una $h(\cdot)$ en C_i cuando en realidad pertenece a C_j y están dadas por

$$q_{ij}=\frac{n_{ij}}{n_{\cdot j}}$$

$$q_{ij}=\Pr\big[\overset{\text{experto}}{h(\cdot)\in C_i}\mid \overset{\text{realidad}}{h(\cdot)\in C_j}\big] \tag{5}$$

Por supuesto que, a medida que *r* sea grande, los valores q_{ij} estarán más cercanos a los verdaderos valores asociados al experto.

Así **Q** nos da las probabilidades condicionales asociadas al experto, de clasificar una característica como perteneciente a la categoría C_i dado que en realidad pertenece a la categoría C_j. En otras palabras, **Q** representa la habilidad del experto. Note que si es diagonal el experto es perfecto, de otro modo, los términos fuera de la diagonal nos hablan de sus imprecisiones.

De esta evaluación de la destreza del experto respecto a la realidad se generan dos particiones de L,

1. Una partición que habla del lugar real a la que pertenece cada valor $h(z_i)$, respecto a los C_i, *i.e.*,

$$L_i=\{z\in L: h(z)\in C_i\}\quad i=1,\ldots,p \tag{6}$$

$$\begin{aligned}&L_i\subset L\\ &L_i\cap L_j=\varnothing\quad \forall\, i\neq j\\ &\bigcup_{i=1}^{p} L_i=L\end{aligned} \tag{6a}$$

2. Una partición que habla del lugar estimado a la que el experto asignó a cada valor de la característica $h(z_i)$, que denotamos por $\hat{h}(z_i)$, respecto a los C_i, *i.e.*,

$$\hat{L}_i=\{z\in L: \hat{h}(z)\in C_i\}\quad i=1,\ldots,p \tag{7}$$

$$\begin{aligned}&\hat{L}_i\subset L\\ &\hat{L}_i\cap\hat{L}_j=\varnothing\quad \forall\, i\neq j\\ &\bigcup \hat{L}_i=L\end{aligned} \tag{7a}$$

Así, $\{L_i\}_{i=1,\ldots,p}$ y $\{\hat{L}_i\}_{i=1,\ldots,p}$ son dos particiones de L.

Si al azar se selecciona un elemento de L, digamos z, la probabilidad de que z caiga en la parte L_j, sería la probabilidad de ocurrencia de L_j y estaría dada por

$$\Pr(L_j)=\frac{\#(L_j)}{r} \tag{8}$$

Para determinar la probabilidad de ocurrencia de $\hat{L}_i$ para alguna i=1, …, p, se puede describir a $\hat{L}_i$ en términos de la partición $\{L_j\}_{j=1,\ldots,p}$, es decir,

$$\hat{L}_i \cap L = \hat{L}_i \cap \left(\bigcup_{j=1}^{p} L_j \right) = \bigcup_{j=1}^{p} \left(\hat{L}_i \cap L_j \right)$$

$$\Pr\left(\hat{L}_i\right) = \Pr\left(\bigcup_{j=1}^{p} \left(\hat{L}_i \cap L_j \right) \right) = \sum_{j=1}^{p} \Pr\left(\hat{L}_i \cap L_j \right) = \sum_{j=1}^{p} \Pr\left(\hat{L}_i \mid L_j \right) \Pr\left(L_j \right)$$

Observe cuidadosamente, respecto de los n_{ij} , lo siguiente:

$$\Pr\left(\hat{L}_i \mid L_j\right) = \frac{\Pr\left(\hat{L}_i \cap L_j\right)}{\Pr\left(L_j\right)} = \frac{n_{ij}/r}{n_{\cdot j}/r} = \frac{n_{ij}}{n_{\cdot j}} = q_{ij}$$

del denominador, $\Pr\left(L_j\right) = \frac{\#\left(L_j\right)}{r} = \frac{n_{\cdot j}}{r}$

De este modo se ha establecido una ecuación que relaciona la probabilidad de ocurrencia de una parte estimada $\hat{L}_i$ con las probabilidades de ocurrencia de las partes reales $L_j, \quad j = 1, \ldots, p$, haciendo uso de la habilidad del experto expresada en $\mathbf{Q}$, *i.e.*,

$$\Pr\left(\hat{L}_i\right) = \sum_{j=1}^{p} q_{ij} \Pr\left(L_j\right) \tag{9}$$

Note que

$$P\left(\hat{L}_i\right) = \sum_{j=1}^{p} \frac{n_{ij}}{n_{\cdot j}} \frac{n_{\cdot j}}{r} = \frac{n_{i\cdot}}{r} = \frac{\#\left(\hat{L}_i\right)}{r} \tag{9a}$$

Haciendo $\mathbf{v}' = \left[\Pr\left(L_1\right)\Pr\left(L_2\right)\ldots\Pr\left(L_p\right)\right]$ se tiene $\Pr\left(\hat{L}_i\right) = \mathbf{q}'_i\, \mathbf{v}$ y $\hat{\mathbf{v}}' = \left[\Pr\left(\hat{L}_1\right)\Pr\left(\hat{L}_2\right)\ldots\Pr\left(\hat{L}_p\right)\right]$, finalmente las relaciones mencionadas quedan en la ecuación

$$\hat{\mathbf{v}} = \mathbf{Q}\mathbf{v} \tag{10}$$

Una vez obtenida **Q**, ésta se puede utilizar para estimar la distribución de la característica $h(\cdot)$ en *P*, de la siguiente manera:

Se obtiene una muestra de *P*, la que anteriormente denotamos como $M = \{y_1, \ldots, y_m\}$. Se presume que es una muestra tomada de modo que todos los individuos en *P* tuvieron la misma posibilidad de ser seleccionados, en otras palabras, la probabilidad de seleccionar cualquiera de las *x* en *P* fue $1/n$. Lo anterior pretende lograr que la distribución de la característica $h(\cdot)$ en $M = \{y_1, \ldots, y_m\}$ sea la misma que la distribución de la característica $h(\cdot)$ en *P*.

Lo anterior significa lo siguiente:

Sea (vea la Ecuación (6)),

$$M_j = \{y \in M \mid h(y) \in C_j\}, \quad j=1,\ldots,p \tag{11}$$

y $\#(M_j)=m_j$, entonces la proporción de elementos en *M* cuya característica está en C_j, la que denotaremos por s_j, es (vea la Ecuación (8))

$$s_j = \frac{m_j}{m} = \frac{\#\left(M_j\right)}{\#\left(M\right)} = \Pr\left(M_j\right) \quad j = 1, \ldots, p \tag{12}$$

El muestreo aleatorio pretende lograr que s_j $=Pr(M_j)= \hat{\Pr}(P_j) = \hat{r}_j$, $j=1,\ldots,p$, donde $s_j = \hat{r}_j$, $j=1,\ldots,p$, significa que la proporción verdadera de sujetos en la muestra pertenecientes a la clase de edad C_j, *i.e.*, s_j , está muy cerca (es un buen estimador de r_j, lo que se denota por $\hat{r}_j$) de la proporción verdadera de sujetos en la población pertenecientes a la clase de edad C_j. Lo anterior se abrevia diciendo que la muestra es adecuada.

Ahora bien, las *h(y)* en *M* no son conocidas, como en general no lo son en *P*. Con la muestra *M* el experto estima $\hat{h}(y_1), \ldots, \hat{h}(y_m)$ y estas estimaciones, junto con la partición $\{C_j\}_{j=1,\ldots,p}$, inducen una partición en el conjunto $\hat{M} = \{\hat{h}(y_1), \ldots, \hat{h}(y_m)\}$ del siguiente modo (vea la Ecuación (7)):

$$\hat{M}_j = \{y \in M : \hat{h}(y) \in C_j\}, \quad j = 1, \ldots, p \tag{13}$$

Si el experto actúa sobre *M* del mismo modo que lo hizo sobre *L* (**hipótesis: Q** no cambia a **Q***), entonces la relación establecida entre $\Pr\left(\hat{L}_j\right)$ y las probabilidades $\Pr\left(L_j\right)$, *j=1,…,p*, se mantiene entre la $\Pr\left(\hat{M}_j\right)$ y las probabilidades $\Pr(M_j)$, *j=1,…,p*, es decir (ver la Ecuación (9) haciendo **Q**=**Q***, en la muestra *M*),

$$\Pr(\hat{M}_i)=\sum_{j=1}^{p} q_{ij}\Pr(M_j),\quad i=1,\dots,p \qquad (14)$$

Así, si la **hipótesis** del muestreo se satisface se tiene (ver la Ecuación (9) haciendo **Q**=**Q*** en la muestra *M*),

$$P(\hat{M}_i)=\sum_{j=1}^{p} q_{ij}\Pr(M_j)=\sum_{j=1}^{p} q_{ij}s_j=\sum_{j=1}^{p} q_{ij}\hat{\Pr}(P_j)=\sum_{j=1}^{p} q_{ij}\hat{r}_j \qquad (15)$$

Denotando $\Pr(\hat{M}_i)=\hat{s}_i$, la anterior ecuación es:

$$\hat{s}_i=\sum_{j=1}^{p} q_{ij}\hat{r}_j,\quad i=1,\dots,p \qquad (16)$$

Haciendo (ver la Ecuación (12) y (16))

$$\hat{\mathbf{s}}=\begin{bmatrix}\hat{s}_1\\ \hat{s}_2\\ \vdots\\ \hat{s}_p\end{bmatrix},\quad \hat{\mathbf{r}}=\begin{bmatrix}\hat{r}_1\\ \hat{r}_2\\ \vdots\\ \hat{r}_p\end{bmatrix}=\begin{bmatrix}s_1\\ s_2\\ \vdots\\ s_p\end{bmatrix}=\mathbf{s} \qquad (17)$$

finalmente se encuentra la ecuación (ver la Ecuación (10)),

$$\hat{\mathbf{s}}=\mathbf{Q}\mathbf{s}=\mathbf{Q}\hat{\mathbf{r}} \qquad (18)$$

siempre que la muestra sea adecuada, i.e., $\mathbf{s}=\hat{\mathbf{r}}$.

Observe que $\Pr(\hat{M}_i)$ no es más que la probabilidad de que ocurra $\hat{M}_i$, o la proporción de los *m* elementos en la muestra *M* que el experto clasifica en la clase C_i.

Como $\hat{\mathbf{s}}$ es conocido, y también **Q**, entonces se puede obtener de la ecuación anterior un valor para $\hat{\mathbf{r}}$, siempre que **Q** no sea singular, por

$$\hat{\mathbf{r}}=\mathbf{Q}^{-1}\hat{\mathbf{s}} \qquad (19)$$

lo que estima la proporción de elementos de la población *P* en cada parte C_j, *j*=1,…,*p*. Esto es lo que se deseaba.

Los errores que pueden ocurrir son, en las dos hipótesis consideradas, a saber:

a) El experto actúa sobre *M* como lo hizo sobre *L*, es decir, ya que se tiene

$$\hat{\mathbf{v}}=\mathbf{Q}\mathbf{v}$$

se debe tener $\hat{\mathbf{s}}=\mathbf{Q}\mathbf{s}$.

b) El muestreo aleatorio debe lograr $s_i=\hat{r}_i$, lo que significa que $\hat{r}_i$ sea muy cercano a, si no igual, r_i, *i*=1,…,*p*, de modo que $\mathbf{s}=\hat{\mathbf{r}}$ y se debe tener $\hat{\mathbf{s}}=\mathbf{Q}\hat{\mathbf{r}}$.

En general, las hipótesis anteriores no se cumplen cabalmente y se tienen errores.

Para estimar a **r** se usará la ecuación $\hat{\mathbf{s}}=\mathbf{Q}\hat{\mathbf{r}}$, de donde se obtiene la Ecuación (19)

$$\hat{\mathbf{r}}=\mathbf{Q}^{-1}\hat{\mathbf{s}}$$

En general, el experto no actuará igual sobre *L* que sobre M, de modo que las q_{ij} obtenidas de *L* no serán las correctas en la Ecuación (18)

$$\hat{\mathbf{s}}=\mathbf{Q}\mathbf{s}$$

sino se generarán otras probabilidades q_{ij}^{*}, *i*=1,…,*p*, *j*=1,…,*p*, que conformarán una nueva matriz **Q*** la cual hará de la ecuación anterior una identidad, i.e.,

$$\hat{\mathbf{s}}=\mathbf{Q}^{*}\mathbf{s} \qquad (20)$$

Sustituyendo este valor en la Ecuación (19) se tiene

$$\hat{\mathbf{r}}=\mathbf{Q}^{-1}\hat{\mathbf{s}}=\mathbf{Q}^{-1}\mathbf{Q}^{*}\mathbf{s} \qquad (21)$$

Por último, el vector **s** que contiene las proporciones de sujetos en la muestra que caen en cada una de las categorías de edad C_i, *i*=1,…,*p*, será diferente al vector **r** que contiene las proporciones de sujetos en la población que caen en cada una de las categorías de edad correspondientes. Note que **r** es un vector de valores fijos, mientras que, **s** es un vector aleatorio. Así, **r** puede escribirse como:

$$\mathbf{r}=\mathbf{s}+\mathbf{d} \qquad (22)$$

o de modo equivalente **s**=**r**-**d**, dónde **d** es un vector aleatorio de diferencias entre **r** y **s.**

Esto substituido en (21) nos proporciona:

$$\hat{\mathbf{r}} = \mathbf{Q}^{-1}\hat{\mathbf{s}} = \mathbf{Q}^{-1}\mathbf{Q}^{*}\mathbf{s} = \mathbf{Q}^{-1}\mathbf{Q}^{*}(\mathbf{r}-\mathbf{d})$$

Al usar la estimación $\hat{\mathbf{r}}$ para **r** producirá el vector de errores

$$\mathbf{e} = \hat{\mathbf{r}} - \mathbf{r} = \mathbf{Q}^{-1}\mathbf{Q}^{*}(\mathbf{r}-\mathbf{d}) - \mathbf{r} \tag{23}$$

Note cuidadosamente que si $\mathbf{Q}^{*}=\mathbf{Q}$, es decir, el experto actúa sobre *L* del mismo modo que sobre *M*, y **d**=**0**, es decir, **s**=**r**, el muestreo no sólo es adecuado sino perfecto, entonces **e**=**0**. De la habilidad del experto depende que $\mathbf{Q}^{*}$ sea muy cercana a **Q** y del tamaño de la muestra depende que **d** sea muy cercano a **0**.

Aplicación

La siguiente aplicación es “ficticia” en el sentido que seleccionaremos una población donde la variable por estudiar tiene una distribución totalmente conocida para quien conduce el estudio. También, es “ficticia” en que, la variable que hay que estudiar es la edad de un grupo de personas, la cual se puede conocer con sólo preguntarle a cada sujeto el año de su nacimiento, pero no se hará así, sino un grupo de “expertos” estimará esas edades, obviamente sin conocerlas y usando sólo su experiencia visual.

La población es pequeña, 31 personas, y los expertos no conocen la distribución de edades en esta población.

Así, $P = \{x_1, x_2, \ldots, x_{31}\}$, y $h(x_i)$ es la edad del sujeto *i*=1,2,…,31. Se desea categorizar las edades de esta población en tres grupos:

$$C_1 = (-\infty, 18)$$
$$C_2 = [18, 24]$$
$$C_3 = (24, \infty)$$

Observe que *p*=3, en este ejemplo.

Donde, $\{C_i\}_{i=1,2,3}$ es una partición de $\Re$, la cual induce naturalmente una partición sobre *P*, *i.e.*, (ver la Ecuación (1)),

$$P_i = \{x \in P : h(x) \in C_i\}, \qquad i = 1,2,3$$

Siendo una población finita la variable edad es discreta y la probabilidad de que ocurra P_i, *i*=1,2,3, es la frecuencia relativa dada por el número de sujetos en P_i dividido entre el número total de sujetos.

El que conduce el experimento sí conoce la distribución de edades y de ahí calcula las probabilidades antes descritas, encontrando que (ver la Ecuación (2))

$$\Pr(P_1) = \frac{\#(P_1)}{\#(P)} = \frac{n_1}{n} = \frac{1}{31} = r_1$$
$$\Pr(P_2) = \frac{\#(P_2)}{\#(P)} = \frac{n_2}{n} = \frac{23}{31} = r_2$$
$$\Pr(P_3) = \frac{\#(P_3)}{\#(P)} = \frac{n_3}{n} = \frac{7}{31} = r_3$$

Esta es la distribución de probabilidades que los expertos tratarán de estimar a través de una muestra. Como el conductor del experimento conoce las edades (en el caso de venados esto se lograría conociendo la edad de animales en cautiverio) toma a ocho sujetos y los presenta ante quince expertos para que determinen su edad y así construir la matriz $\mathbf{Q} = \{q_{ij}\}_{\substack{i=1,2,3\\ j=1,2,3}}$, donde:

q_{ij}: Probabilidad de que los expertos clasifiquen una $h(\cdot)$ en C_i cuando en realidad pertenece a C_j, las que están dadas por $q_{ij} = \frac{n_{ij}}{n_{\cdot j}}$, donde n_{ij} es el número de veces que los expertos estiman $h(\cdot)$ como perteneciente a la categoría C_i cuando en realidad pertenecen a C_j, *i,j*=1,2,3.

Los resultados se muestran en seguida:

Sujeto	Categoría Estimada			Categoría Real
	C_1	C_2	C_3	
M05	0	14	1	C_2
L05	1	14	0	C_2
L02	0	15	0	C_2
L03	10	5	0	C_1
L09	0	15	0	C_2
L16	1	14	0	C_2
M06	0	7	8	C_3
M14	0	10	5	C_2
Total	12	94	14	

La matriz $\mathbf{N} = \{n_{ij}\}_{\substack{i=1,2,3\\ j=1,2,3}}$ es (ver la Ecuación (3))

$$\begin{array}{cccc} & \multicolumn{3}{c}{\text{Realidad}} & \\ & C_1 & C_2 & C_3 & \\ \mathbf{N} = & \left[\begin{array}{ccc} 10 & 2 & 0 \\ 5 & 82 & 7 \\ 0 & 6 & 8 \end{array}\right] & \begin{array}{c} C_1 \\ C_2 \\ C_3 \end{array} & \text{Estimación} \end{array}$$

Note que al usar quince expertos, cada sujeto sobre el que se estima su edad aparece quince veces. En otras palabras, se usaron sólo ocho sujetos, pero al usar quince expertos, cada sujeto aparece quince veces y es “equivalente” a considerar una muestra de 120 (15 x 8) sujetos.

La matriz **Q** es (ver la Ecuación (4))

$$\mathbf{Q}=\begin{bmatrix}10/15 & 2/90 & 0\\ 5/15 & 82/90 & 7/15\\ 0 & 6/90 & 8/15\end{bmatrix}=\begin{bmatrix}60/90 & 2/90 & 0\\ 30/90 & 82/90 & 42/90\\ 0 & 6/90 & 48/90\end{bmatrix}$$

Si denotamos con *L* al conjunto de sujetos clasificados por los expertos, entonces, la partición de *L* que habla del lugar real a la que pertenece cada valor $h(x)$ es (ver la Ecuación (6)):

$$L_i=\{x\in L: h(x)\in C_i\},\quad i=1,2,3$$

y se tiene (ver la Ecuación (8))

$$\Pr(L_1)=\frac{\#(L_1)}{8}=\frac{1}{8}=\frac{15}{120}$$
$$\Pr(L_2)=\frac{\#(L_2)}{8}=\frac{6}{8}=\frac{90}{120}$$
$$\Pr(L_3)=\frac{\#(L_3)}{8}=\frac{1}{8}=\frac{15}{120}$$

La partición que habla del lugar estimado a la que los expertos asignaron a cada edad, denotada por $\hat{h}(x)$, respecto a los C_i es (ver la Ecuación (7)):

$$\hat{L}_i=\{x\in L: \hat{h}(x)\in C_i\}\quad i=1,2,3$$

Obviamente, de la tabla anterior se tiene (ver la Ecuación (9a)):

$$\Pr(\hat{L}_1)=\frac{12}{120}$$
$$\Pr(\hat{L}_2)=\frac{94}{120}$$
$$\Pr(\hat{L}_3)=\frac{14}{120}$$

Recuerde que se denotó por $\hat{\mathbf{v}}'=[\Pr(\hat{L}_1)\Pr(\hat{L}_2),\Pr(\hat{L}_3)]$ y $\mathbf{v}'=[\Pr(L_1)\Pr(L_2),\Pr(L_3)]$, y debe cumplirse, como se cumple, que (ver la Ecuación (10)):

$$\hat{\mathbf{v}}=\mathbf{Q}\mathbf{v}$$

$$\begin{bmatrix}12/120\\94/120\\14/120\end{bmatrix}=\begin{bmatrix}10/15 & 2/90 & 0\\ 5/15 & 82/90 & 7/15\\ 0 & 6/90 & 8/15\end{bmatrix}\begin{bmatrix}1/8\\6/8\\1/8\end{bmatrix}$$

Obtenida Q, se puede utilizar ahora con una muestra.

Se tomó una muestra de *P* de tamaño ocho a la que denotamos por *M*.

Los resultados de la muestra se muestran en seguida:

Sujeto	Categoría Estimada			Categoría Real
	C_1	C_2	C_3	
M11	0	7	8	C_3
L08	1	14	0	C_2
L04	0	15	0	C_2
L15	4	11	0	C_2
M10	0	0	15	C_3
M04	0	5	10	C_3
M07	0	9	6	C_2
M15	0	15	0	C_2
Total	5	76	39	

Aquí las particiones de *M* son: $M_i=\{x\in M: h(x)=C_i\}$ y $\hat{M}_i=\{x\in M: \hat{h}(x)\in C_i\}$, *i*=1,2,3.

Así, el vector $\hat{\mathbf{s}}$ está dado por (ver la Ecuación (17)):

$$\hat{\mathbf{s}}=\begin{bmatrix}5/120\\76/120\\39/120\end{bmatrix}=\begin{bmatrix}\Pr(\hat{M}_1)\\ \Pr(\hat{M}_2)\\ \Pr(\hat{M}_3)\end{bmatrix}$$

y el vector $\mathbf{s}'=[\Pr(M_1),\Pr(M_2),\Pr(M_3)]$ está dado por (ver las Ecuaciones (12) y (17)):

$$\mathbf{s}=\begin{bmatrix}0\\5/8\\3/8\end{bmatrix}=\begin{bmatrix}0\\75/120\\45/120\end{bmatrix}$$

si **s** fuera muy cercano a **r**, una estimación de **r**, a la que se ha denotado como $\hat{\mathbf{r}}$ sería:

$$\hat{\mathbf{r}} = \mathbf{Q}^{-1}\hat{\mathbf{s}} = \begin{bmatrix} 307/202 & -4/101 & 7/202 \\ -60/101 & 120/101 & -105/101 \\ 15/202 & -15/101 & 405/202 \end{bmatrix} \begin{bmatrix} 5/120 \\ 76/120 \\ 39/120 \end{bmatrix} = \begin{bmatrix} 5/101 \\ 315/808 \\ 453/808 \end{bmatrix} = \begin{bmatrix} 40/808 \\ 315/808 \\ 453/808 \end{bmatrix}$$

De la tabla anterior obtenida con la muestra, se puede obtener $N^* = \{n_{ij}^*\}_{\substack{i=1,2,3 \\ j=1,2,3}}$ donde n_{ij}^*: número de veces que los expertos clasifican un sujeto en la muestra como perteneciente a C_i cuando en realidad pertenece a C_j.

$$\begin{array}{c} \text{Realidad} \\ \begin{array}{ccc} C_1 & C_2 & C_3 \end{array} \end{array}$$

$$\mathbf{N}^* = \begin{bmatrix} 0 & 5 & 0 \\ 0 & 64 & 12 \\ 0 & 6 & 33 \end{bmatrix} \begin{array}{l} C_1 \\ C_2 \\ C_3 \end{array} \ \text{Estimación}$$

La matriz $\mathbf{Q}^*$ estaría dada por

$$\mathbf{Q}^* = \begin{bmatrix} 0 & 5/75 & 0 \\ 0 & 64/75 & 12/45 \\ 0 & 6/75 & 33/45 \end{bmatrix} = \begin{bmatrix} 0 & 6/90 & 0 \\ 0 & 76.8/90 & 24/90 \\ 0 & 7.2/90 & 66/90 \end{bmatrix}$$

Por supuesto que se debe de cumplir:

$$\hat{\mathbf{s}} = \mathbf{Q}^*\mathbf{s}$$

$$\begin{bmatrix} 5/120 \\ 76/120 \\ 39/120 \end{bmatrix} = \begin{bmatrix} 0 & 5/75 & 0 \\ 0 & 64/75 & 12/45 \\ 0 & 6/75 & 33/45 \end{bmatrix} \begin{bmatrix} 0 \\ 5/8 \\ 3/8 \end{bmatrix}$$

Esta es la misma ecuación que la obtenida para calibrar la habilidad de los expertos, sólo que allá se usó el conjunto *L* de sujetos y aquí el conjunto *M*. Ahora *M* no es para este propósito sino para estimar **r**, el vector que contiene la proporción poblacional en cada una de las clases. Si aquí se calcula es para exhibir los efectos cuando hay problemas con los expertos y la muestra no es adecuada. Los errores ocasionados por ambas hipótesis, están dadas por el vector de errores (ver la Ecuación (23))

$$\mathbf{e} = \hat{\mathbf{r}} - \mathbf{r} = \begin{bmatrix} 40/808 \\ 315/808 \\ 453/808 \end{bmatrix} - \begin{bmatrix} 1/31 \\ 23/31 \\ 7/31 \end{bmatrix} = \begin{bmatrix} 432/25048 \\ -8819/25048 \\ 8387/25048 \end{bmatrix}$$

Este vector como señala la Ecuación (23) está dado por

$$\begin{aligned} \mathbf{Q}^{-1}\mathbf{Q}^*\mathbf{s} - \mathbf{r} &= \begin{bmatrix} 307/202 & -4/101 & 7/202 \\ -60/101 & 120/101 & -105/101 \\ 15/202 & -15/101 & 405/202 \end{bmatrix} \begin{bmatrix} 0 & 5/75 & 0 \\ 0 & 64/75 & 12/45 \\ 0 & 6/75 & 33/45 \end{bmatrix} \begin{bmatrix} 0 \\ 5/8 \\ 3/8 \end{bmatrix} - \begin{bmatrix} 1/31 \\ 23/31 \\ 7/31 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 71/1010 & 3/202 \\ 0 & 90/101 & -45/101 \\ 0 & 39/1010 & 89/202 \end{bmatrix} \begin{bmatrix} 0 \\ 5/8 \\ 3/8 \end{bmatrix} - \begin{bmatrix} 1/31 \\ 23/31 \\ 7/31 \end{bmatrix} \\ &= \begin{bmatrix} 40/808 \\ 315/808 \\ 453/808 \end{bmatrix} - \begin{bmatrix} 1/31 \\ 23/31 \\ 7/31 \end{bmatrix} \\ &= \begin{bmatrix} 432/25048 \\ -8819/25048 \\ 8387/25048 \end{bmatrix} \end{aligned}$$

el cual, si $\mathbf{Q} = \mathbf{Q}^*$ y $\mathbf{s} = \mathbf{r}$, sería un vector nulo.

Para terminar este análisis, y siguiendo con el desarrollo dejado en la Ecuación (23), si se desean hacer inferencias sobre **r**, el vector de interés, sólo resta suponer, por ejemplo, una distribución sobre el vector de errores **e**, la cual comúnmente es una distribución multinormal con vector de medias igual al vector **0** y matriz de varianzas y covarianzas igual a $\sigma^2\mathbf{I}$, donde la dimensión del vector y la matriz es la de la matriz **Q**. A partir de ese momento se está en el ámbito de la estadística.

Dejamos hasta aquí el análisis, esperando se logre alguno de los propósitos mencionados y sirva a quienes tienen interés especial en saber cómo es que se sabe.

LITERATURA CITADA

SEARLE, S. R., 1982, Matrix algebra useful for statistics, John Wiley and Sons, New York, 438 p.