

日 本 計 量 生 物 学 会

1 9 8 7 年 度 年 会

1 9 8 7 年 4 月 2 5 日 (土)

於 統 計 数 理 研 究 所

難病の疫学調査における患者数推定のためのモード予測

高木廣文

聖路加看護大学

I. はじめに

厚生省は、昭和47年以来いくつかの難病を「特定疾患」として指定し、各疾患に対する調査研究班を中心に、各種の研究を実施しているが、全国の患者数を把握するために行う疫学調査もその一つである。調査は、郵送法により全国の特定の診療所や病院に対して、特定年での受診患者数を報告してもらうことにより行われる。調査対象は数千に達し、その回収率は50~70%程度であり、未回収分の患者数を推定する必要が生じる。これまでは、県別の回収率と人口10万人対の受診率（報告受診率）のデータをもとに、単回帰直線による推定を用いることが多かった（柳井他, 1975）。もしくは、回帰直線が原点を通るようにして推定を行う方法などもある（中江他, 1985. 水野他, 1986）。ここでは、調査対象からの報告患者数の分布に着目し、多項分布をもとにして患者数の分布の確率が最大となる場合、すなわちモードを予測する方法を報告する。

II. 方法

調査対象数を M 、回答数を N とし（ $M \geq N$ ）、報告患者数の最大値を k 、患者数が $0, 1, 2, \dots, k$ である回答数を $n_0, n_1, n_2, \dots, n_k$ （ $n_0 + n_1 + n_2 + \dots + n_k = N$ ）とする。（ $M - N$ ）の未回答分について、患者数の分布（ $\delta_0, \delta_1, \delta_2, \dots, \delta_k$ ）を求める必要がある。全対象での患者数の分布を（ $\eta_0, \eta_1, \eta_2, \dots, \eta_k$ ）とすると、 $\eta_i = n_i + \delta_i$ （ $\delta_i \geq 0, i = 0, 1, \dots, k$ ）となるので、 n_i を確定変数とすると、 η_i を直接予測する代わりに δ_i の予測値 $\hat{\delta}_i$ を求めればよいことになる。ただし、

$$\delta_0 + \delta_1 + \delta_2 + \dots + \delta_k = M - N \quad (1)$$

という制約がある。

全対象での患者総数の予測値 $\hat{\eta}$ は、患者数とその分布の予測値との積和により求められる。

各対象からの回答の有無がランダムであると仮定できれば、回答のあった N 対象において、その報告患者数の分布（ $n_0, n_1, n_2, \dots, n_k$ ）は、比率を（ $p_0, p_1, p_2, \dots, p_k$ ）とする多項分布に従うものとみなせる。ここで、 $p_i = n_i / N$ （ $i = 0, 1, \dots, k$ ）は母比率 π_i の最尤推定量である。したがって、未回答の対象（ $M - N$ ）における患者数の分布 δ_i の期待値 $\hat{\delta}_i$ は（ $M - N$ ） p_i となるが、必ずしも期待値が患者数の分布の中で、確率が最大の組合せではない。ここでは、以下のように、患者数の分布の中で確率が最大となるような組合せを求めることにする。

(a) 多項分布におけるモード予測

未回答の対象（ $M - N$ ）について、患者数の分布（ $\delta_0, \delta_1, \delta_2, \dots, \delta_k$ ）が、比率を（ $p_0, p_1, p_2, \dots, p_k$ ）とする多項分布に従うものと仮定する。患者数の分布について、あらゆる場合の組合せの確率を求め、その最大のものを求めればよいが、組合せの数が膨大なものとなるので、ここでは以下のような方法を用いる。

多項分布の確率の対数を取り、制約条件を考慮すると、

$$\ln L = \ln(M - N)! - (M - N) \ln N - \sum_{i=1}^k \ln \Gamma(\delta_i + 1) - \ln \Gamma(M - N - \sum_{i=1}^k \delta_i + 1) + \sum_{i=1}^k \delta_i \ln n_i + (M - N - \sum_{i=1}^k \delta_i) \ln n_0 \quad (2)$$

となる。ここで、 $\ln \Gamma(z)$ は対数ガンマ関数である。

(2)式による $\ln L$ は、分布のパラメータ π_i ではなく、変数 δ_i の関数となっているので、ここでは

「対数確率」と呼ぶことにする。

患者数の分布のモードを求めるには、(1)式の制約のもとで $\ln$ を最大にする δ_i ($i=0,1,\dots,k$)の予測値 $\hat{\delta}_i$ を求めればよい。なお、 δ_i は実際には整数値であるが、予測値 $\hat{\delta}_i$ は連続量として扱うことにする。

(b) 対数確率の偏微分

対数確率 $\ln$ を δ_x ($x=1,2,\dots,k$)で偏微分すると、

$$\frac{\partial \ln}{\partial \delta_x} = \ln(n_x/n_0) - \psi(\delta_x+1) + \psi(M-N - \sum_{i=1}^k \delta_i + 1), \quad \psi(z) = \frac{d \ln \Gamma(z)}{dz} = \int_0^\infty \left(\frac{e^{-t}}{t} - \frac{e^{-zt}}{1-e^{-t}} \right) dt \quad (3)$$

となる。上式の右辺が0となるような δ_x ($x=1,2,\dots,k$)を、数値解析により求めることにする。

2次の導関数は、

$$\frac{\partial^2 \ln}{\partial \delta_x^2} = -\psi'(\delta_x+1) - \psi'(M-N - \sum_{i=1}^k \delta_i + 1), \quad \frac{\partial^2 \ln}{\partial \delta_x \partial \delta_y} = -\psi'(M-N - \sum_{i=1}^k \delta_i + 1) \quad (4)$$

となる。ここで、 $x, y=1,2,\dots,k$; $x \neq y$ であり、

$$\psi'(z) = \frac{d\psi(z)}{dz} = \int_0^\infty \left(\frac{te^{-zt}}{1-e^{-t}} \right) dt \quad (5)$$

である。なお、上記の導関数を用いてニュートン・ラフソン法などの数値解析を行うと、途中の計算で予測値が負となることがある。これを避けるには、 $z_x = \ln(\delta_x)$ ($x=1,2,\dots,k$)のように変換し、 z_x について解けばよい。

(c) 平均2乗誤差

M対象中、患者数が i ($i=1,2,\dots,k$)である対象数 η_i の予測値 $\hat{\eta}_i$ は $(n_i + \hat{\delta}_i)$ により求められる。ただし、 $\hat{\delta}_i$ は δ_i の予測値である。したがって、 $\hat{\eta}_i$ の分散は δ_i の分散 $V(\delta_i)$ と等しい。

平均2乗誤差を考えると、

$$\begin{aligned} V(\hat{\eta}_i) &= E\{(\hat{\eta}_i - \eta_i)^2\} = E\{(\hat{\delta}_i - \delta_i) - (\delta_i - \delta_i)\}^2 = (\hat{\delta}_i - \delta_i)^2 + E\{(\delta_i - \delta_i)^2\} \\ &= \{(M-N)P_i - \hat{\delta}_i\}^2 + (M-N)P_i(1-P_i) \end{aligned} \quad (6)$$

となる。

患者数を i とする対象数の予測値が $\hat{\eta}_i$ のとき、 $\hat{\eta}_i$ 施設での患者総数 $\hat{e}_i$ は $i \cdot \hat{\eta}_i$ となるので、その分散は、

$$V(\hat{e}_i) = i^2 V(\hat{\eta}_i) = i^2 V(\delta_i) \quad (7)$$

により求められる。全国患者数の予測値 $\hat{e}$ の分散は、

$$V(\hat{e}) = \sum_{i=1}^k i^2 V(\hat{\eta}_i) - (M-N) \sum_{i \neq j} i j P_i P_j + \sum_{i \neq j} i j (\hat{\delta}_i - \delta_i)(\hat{\delta}_j - \delta_j) \quad (8)$$

となる。

III. 応用例

厚生省特定疾患神経皮膚症候群調査研究班疫学調査分科会ならびに難病の疫学調査研究班事務局が、昭和61年に実施した全国調査（一次調査）のデータに適用した結果を例示する。

IV. 参考文献

柳井晴夫，他：疫学調査における有病率補正の方式，日本公衛誌，22(2)，71-81，1975

中江公裕，他：膠原病四疾患第一次全国調査成績；全国患者数の推定，厚生省特定疾患難病の疫学調査研究班，昭和59年度研究業績報告書，135-150，1985

水野正一，他：全国調査成績を用いての有病数の一推計方法，日本公衛誌，33(10)，255，1986