

IL-2 역가의 통계적 추정에 관한 연구

박래현** 박석영* 이석훈**

요 약

최근 암의 면역치료 요법에 사용이 활발히 시도되고 있는 IL-2(Interleukin-2)의 역가를 측정하는 문제를 통계적 모형을 통하여 정립하고 그 모형하에서 모수의 함수로 표현되는 역가의 추론과정을 연구하였다. 표준시료와 비교하여, 환자로 부터 얻은 미지의 시료의 역가를 구하기 위하여 선형모형을 제시하고 베이지안 기법을 사용하여 계수들의 함수로 나타내지는 역가의 신뢰구간을 구하였으며 실제 데이터에 적용하여 보았다.

1. 서 론

IL-2(Interleukin-2)는 주로 활성화된 T 임파구로 부터 생성되는 호르몬의 일종으로 T 임파구의 증식촉진, 자연살해세포 및 세포독성 T 임파구의 기능항진, B 세포에 대한 직접 또는 간접적 기능촉진등이 알려져 있다. 이런 다양한 면역학적 기능증강작용과 또한 임포카인 활성화 살해세포(lymphokin-activated-killer cell)의 생성에 관여하는 성질들이 알려짐에 따라 최근 암의 면역치료 요법에 임상적인 이용이 활발히 시도되고 있다. IL-2에 의존하여 증식되는 세포독성 T세포계열(cytotoxic T lymphocyte line, CTLL)을 이용하여 Gillis et al(1978)등이 고안한 생물학적 측정방법이 널리 이용되고 있으며 이 방법의 개요는 다음과 같다.

시료가 가진 미지의 IL-2 역가의 측정은 이 시료가 CTLL의 증식을 촉진시키는 정도를 동위원소가 표시된 핵산 전구물질(예는 H^3 -Thymidine)이 CTLL의 핵산으로 동화되는 정도로 정량적으로 측정하여, 이를 역가가 알려진 표준 IL-2에 세포증식의 정도와 비교하여 시료의 IL-2역가를 추정하는 것이다.

그러나 표준 IL-2역가와 비교하여 시료의 IL-2 역가를 추정하는 방법이 기술통계 수준에서 연구자들 마다 다소 달리 하고있어 합리적인 IL-2 역가추정에 관한 연구가 필요하다고 생각되며, 기술통계 수준이 아닌 통계적 추론에 바탕을 두는 역가 추정이 필요하다고 생각된다. 이에 저자들은 이 연구에서 적절한 통계적 모형을 설정하여 IL-2역가를 통계적으로 추론하는 방법을 제시하였다.

* 카톨릭대학 의학부 내과학 교실

** 충남대학교 통계학과

제 2 절에서는 IL-2 역가추정을 위한 생물학적 실험방법을 자세히 검토하고 이 실험 방법을 설명하는 통계적 모형을 제시하고자 한다. 제 3 절에서는 궁극적으로 추정되어야 하는 모수를 정의하고 이의 추정문제를 다루고자 한다. 제 4 절에서는 우리가 제시한 추정방법을 카톨릭대학 의학부 내과학 교실에서 얻은 실제 자료에 적용하여 보고자 한다.

2. 실험 및 모형

일반적으로 제시되는 IL-2역가추정을 위한 생물학적 실험은 다음과 같이 행하여 진다. 먼저 IL-2 역가가 알려진 표준시료를 단계적으로 희석(serial two-fold dilution)하면서 각각의 희석된 일정량을 정하여진 CTLL(well당 4×10^4 개)에 투입하여 일정시간동안 CTLL을 증식 시킨다. 이룬다음 핵산전구 물질에 동위원소가 표지된 H^3 -thymidine을 작용시켜서 harvest한후 방출되는 β -ray의 갯수를 β -counter를 이용하여 cpm(counts per minute)으로 측정하는데 이 cpm이 CTLL의 증식정도이다. 또한 미지의 시료에 관하여서도 위와 동일한 방법으로 실험하여 CTLL의 증식 정도를 측정한다.

위 실험에서 관찰되는 양을 변수로 나타내면 아래와 같다.

(X_{1i}, Y_{1i}) ; 표준 시료에서 희석배수의 \log_2 값, X_{1i} 와 X_{1i} 에서 관찰되는 cpm의 \log 값인 $Y_{1i}(i=1, 2, \dots, n_1)$

(X_{2j}, Y_{2j}) ; 미지의 시료에서 희석배수의 \log_2 값, X_{2j} 와 X_{2j} 에서 관찰되는 cpm의 \log 값인 $Y_{2j}(j=1, 2, \dots, n_2)$

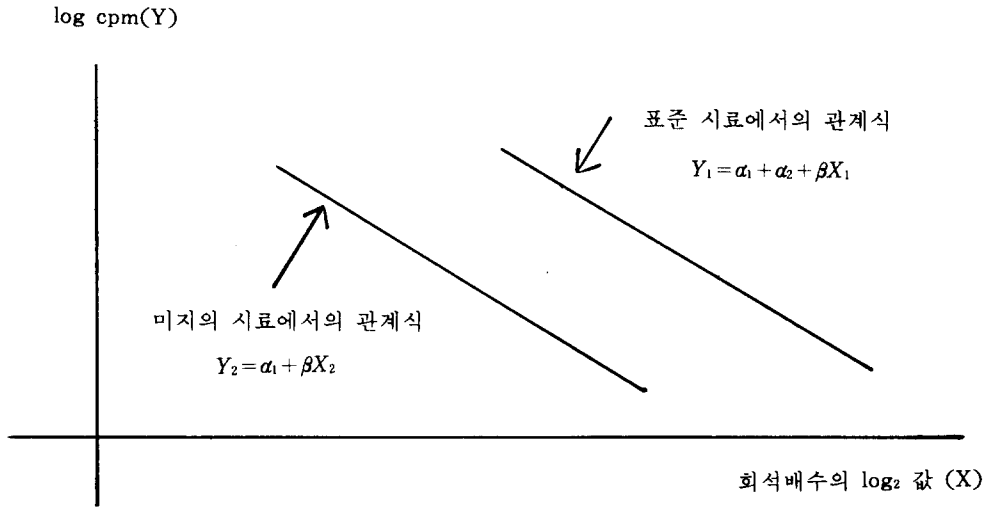
Gillis et al (1989) 및 Stadler 와 Oppenheim(1982) 등의 연구에 의하면 표준시료의 경우 고농도(낮은 희석배수)에서는 \log cpm의 값이 일정(plateau 현상)하나 그 이하의 농도에서는 \log cpm과 희석배수의 \log_2 값 사이에 선형관계가 있으며 대체로 저농도인 미지의 시료에서는 \log cpm과 희석배수의 \log_2 값 사이에 표준 시료의 저농도에서 나타난 직선과 평행인 직선관계가 있다는 것이 알려져 있다. 그러므로 우리는 어느정도 이하의 농도에서는 다음과 같은 통계적 모형을 설정할 수 있다.

$$Y_{ij} = \alpha_1 + \alpha_2 Z_i + \beta X_{ij} + \epsilon_{ij}, \quad i=1, 2 \quad j=1, 2, \dots, n_i \quad (2.1)$$

여기서 X_{ij} 는 우리가 조절가능한 변수, Y_{ij} 는 확률변수로 볼 수 있고 Z_i 는 가변수로 아래와 같다.

$$\begin{aligned} Z_i &= 1, \quad i=1 \text{ 일때} \\ &= 0, \quad i=2 \text{ 일때} \end{aligned}$$

모형(2. 1)을 그림으로 나타내면 아래 그림(2. 1)처럼 된다.



— 그림 (2. 1) —

3. 역가의 추정

일반적으로 미지의 시료에서 IL-2 역가추정은 다음과 같이 한다. 표준시료에서 CTLL 세포가 증식된 정도와 같은 만큼 세포가 증식되기 위하여는 미지의 시료를 표준시료의 경우에 비해 어느 정도 희석하여야 하는가가 미지시료의 IL-2역가이다. 즉 다시 말하면 Y의 값이 Y_0 (Y_0 의 값은 보통 plateau 현상을 보이는 곳에서의 cpm의 50%값 혹은 30%값에 log를 취한 값으로 잡아 왔음)로 주어 졌을 때 표준시료에서 회석배수의 \log_2 값을 x_s 라 하고, 미지의 시료에서 회석배수의 \log_2 값을 x_T 라 하면

$$\frac{2^{x_T}}{2^{x_S}} = 2^{x_T - x_S}$$

또는

$$100 \times 2^{x_T - x_S}$$

가 미지 시료의 IL-2 역가이다. 여기서 특기할 만한 사실은 Y_0 의 값을 어떻게 잡아도 $x_T - x_S$ 의 값은 일정하므로 Y_0 의 값을 어떻게 잡느냐는 중요한 문제가 되지 못한다. 결국 우리의 목적은 $2^{x_T - x_S}$ 를 통계적으로 추정하는 것인데 이 문제는 역추정 문제(calibration problem)와 비슷한 문제라 할 수 있다.

(1) 점추정

모형 (2. 1)을 행렬을 이용하여 재표현하면 아래와 같이 된다.

$$y = X\beta + \epsilon \quad \dots\dots\dots(3. 1)$$

여기서

$$y = \begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ \vdots \\ Y_{2n_2} \end{bmatrix}, \quad X = \begin{bmatrix} 1 & 1 & X_{11} \\ 1 & 1 & X_{12} \\ \vdots & \vdots & \vdots \\ 1 & 1 & X_{1n_1} \\ 1 & 0 & X_{21} \\ \vdots & \vdots & \vdots \\ 1 & 0 & X_{2n_2} \end{bmatrix}, \quad \beta = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{1n_1} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{2n_2} \end{bmatrix}$$

이고 우리는 오차항 ϵ 이 $N(0, \sigma^2 I)$ (σ^2 은 미지)를 따른다고 가정하겠다. 그러면 β 의 MLE는 $\hat{\beta} = (\hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta})' = (X'X)^{-1}X'y$ 가 되고 x_T, x_S 의 MLE들은 각각

$$\hat{x}_T = \frac{Y_0 - \hat{\alpha}_1}{\hat{\beta}}, \quad \hat{x}_S = \frac{Y_0 - \hat{\alpha}_1 - \hat{\alpha}_2}{\hat{\beta}}$$

이 되므로 결국 $2^{x_T - x_S}$ 의 MLE는 아래처럼 된다.

$$2^{\hat{x}_T - \hat{x}_S} = 2^{\hat{\alpha}_2 / \hat{\beta}} \Rightarrow 2^{(x_T - x_S)} = 2^{\hat{\alpha}_2 / \hat{\beta}}$$

(2) 구간 추정

x_T, x_S 가 같은 Y_0 를 얻게 하는 값이라는 데 근거하여 $\alpha_1 + \alpha_2 + \beta x_S = \alpha_1 + \beta x_T$ 라는 식을 얻을 수 있고 이 식을 이용하면 $2^{x_T - x_S} = 2^{\alpha_2 / \beta}$ 을 얻게 되며 IL-2역가의 구간 추정은 결국 α_2 / β 의 신뢰구간을 구하면 해결된다. 우리는 α_2 / β 의 구간 추정문제를 Hunter와 Lamboy(1981)가 역 추정 문제를 다룰 때 사용했던 베이지안 방법에 근거하여 해결하였는데 이는 아래와 같다.

$(\alpha_1, \alpha_2, \beta, \sigma^2)$ 에 가정할 수 있는 정보적 사전분포에 관하여는 물리적으로 근거가 확보되지 못하고 empirical bayesian 기법의 활용도 수리적으로 매우 복잡하여 우리는 사전분포를 비정보적 사전분포(noninformative prior)인 σ^{-2} 에 비례하는 것으로 잡는다. 그러면 (α_2, β) 의 사후분포(posterior)는 다음과 같은 자유도 $\nu = n_1 + n_2 - 3$ 인 다변량 t분포가 된다.

$$P(\alpha_2, \beta \mid \text{데이터}) = \frac{\Gamma\left(\frac{\nu+2}{2}\right)}{\pi \Gamma\left(\frac{\nu}{2}\right) |\beta|^{\frac{1}{2}}} \left\{ 1 + \left(\frac{\alpha_2 - \hat{\alpha}_2}{\beta - \hat{\beta}} \right)' S^{-1} \left(\frac{\alpha_2 - \hat{\alpha}_2}{\beta - \hat{\beta}} \right) \right\}^{-\frac{1}{2}(\nu+2)} \quad \dots\dots(3. 2)$$

여기서 $S = s^2 \begin{pmatrix} c_{22} & c_{23} \\ c_{23} & c_{33} \end{pmatrix}$ 이고 c_{ij} 는 $(X'X)^{-1}$ 의 i 번째 행, j 번째 열의 원소이며 s^2 은 σ^2 의 불편 추정치로 아래와 같다.

$$s^2 = \frac{1}{\nu} (y - X\hat{\beta})' (y - X\hat{\beta})$$

식 (3. 2)에서 변수 변환을 이용하면 $\zeta = \alpha_2/\beta$ 의 사후 분포는 아래와 같이 된다.

① ν 가 짝수 일 때 ;

$$m = (1/2)(\nu + 2) \text{ 라 하면}$$

$$P(\zeta | \text{데이터}) = \frac{k_1}{(m-1)k_2k_4^{m-1}} + \frac{2^m(2m-3)!!}{(m-1)!} \frac{k_1 k_3 k_2^{m-2}}{\Delta^{m-1/2}} \tan^{-1}(k_3/\Delta^{1/2})$$

$$+ \frac{k_1 k_3^2}{(2m-1)k_2k_4^{m-1}} \sum_{k=0}^{m-2} \frac{(2k_2k_4)^k (2m-1)(2m-3)\cdots(2m-2k-1)}{(m-1)(m-2)\cdots(m-k-1)\Delta^{k+1}}$$

② ν 가 홀수 일 때 ;

$$m = (1/2)(\nu + 1) \text{ 이라 하면}$$

$$P(\zeta | \text{데이터}) = \frac{2k_1}{(2m-1)k_2k_4^{(1/2)(2m-1)}} +$$

$$\frac{2k_1k_3^2}{(2m-1)\Delta k_2k_4^{(2m-1)/2}} \left\{ 1 + \sum_{k=1}^{m-1} \frac{(8k_2k_4)^k (m-1)(m-2)\cdots(m-k)}{(2m-3)(2m-5)\cdots(2m-2k-1)\Delta^k} \right\}$$

위에서

$$(2m-3)!! = (2m-3)(2m-5)\cdots\cdots 1$$

$$k_1 = \frac{\Gamma\{(\nu+2)/2\}}{\Gamma(\nu/2)\pi | \nu S |^{1/2}},$$

$$k_2 = \nu^{11} \zeta^2 + 2\nu^{12} \zeta + \nu^{22}, \quad k_3 = -2\nu^{11} \zeta \hat{\alpha}_2 - 2\nu^{12} \zeta \hat{\beta} - 2\nu^{12} \hat{\alpha}_2 - 2\nu^{22} \hat{\beta}$$

$$k_4 = 1 + \nu^{11} \hat{\alpha}_2^2 + 2\nu^{12} \hat{\alpha}_2 \hat{\beta} + \nu^{22} \hat{\beta}^2, \quad \Delta = 4k_2k_4 - k_3^2$$

이고 ν^j 는 $(\nu S)^{-1}$ 의 i 번째 행, j 번째 열의 값이다.

α_2/β 의 사후 분포에서 α_2/β 의 $100(1-\alpha)\%$ HPD(Highest Posterior Density) 신뢰구간만 얻으면 $2^{\alpha_2/\beta}$ 의 $100(1-\alpha)\%$ HPD신뢰구간은 쉽게 얻을 수 있다.

여기서 한가지 언급할 것은 Buonaccorsi 와 Gatsonis(1988)가 선형모형에서 두 회귀 계수의 비(ζ)에 대한 정보가 있을 때 ζ 에 대해 직접 사전분포를 주고 ζ 의 구간 추정을 한 방법을 변용할 수 있겠으나 문제점은 ζ 의 사전 분포가 정보적이어야 한다는 점과 정보적 사전 분포하에서도 ζ 의 사후분포를 구체적으로 구할 수 없다는 점이다. 이와 같은 이유 때문에 본 연구에서는 Hunter와 Lamboy(1981)의 접근 방법을 채택했다.

4. 예제 및 토의

제 3 절에서 논의된 추정방법을 이용하여 실제 실험 데이터로부터 IL-2역가를 추정하여 보고자 한다. 본 절에서 사용하고자 하는 데이터는 카톨릭대학 내과학 교실에서 실험한 데이터로서 3개의 동일한 표준시료, IL-2(Amersham Corperation, Illinois, U.S.A. 1,200,000 unit/mg)를 2¹배 부터 2¹⁸배까지 18회 희석시켜가면서 각 희석배수에서 cpm을 관찰하였다. 또한 3개의 동일한 미지의 시료를 2¹배부터 2⁴배까지 4회 희석시키면서 각 희석배수에서 cpm을 관찰하였다. 이와 같이 얻은 관찰치는 총 66개로 그림(4. 1)에서 log cpm과 희석배수의 log₂값으로 도표화하여 보았다. 이 도표에서 우리는 표준시료의 경우 ('1'로 표시됨), Gillis et al(1978)가 지적한 바와 같이 고농도 즉 2¹배 부터 2⁵~2⁶배까지 희석된 시료에서의 log cpm은 plateau 현상이 일어난 것을 볼 수 있다. 따라서 우리는 그들의 의견을 바탕으로, 2⁶배 이상 희석된 시료로 부터 관찰된 log cpm값을 반응 변수 Y로 하고 희석배수의 log₂ 값을 X로 가정하였다. 즉 X_{1j}(j=1, ..., 39)는 6에서 18까지, X_ℓ(ℓ=1, ..., 12)는 1에서 4까지의 값을 취하게 된다. 이들 자료와 우리가 가정한 모형 (2. 1)을 통해서 우리는 아래의 값들을 구할 수 있다.

$$(X'X)^{-1} = \begin{bmatrix} 0.0945, & -0.0410, & -0.0045 \\ -0.0410, & 0.2698, & -0.0169 \\ -0.0045, & -0.0169, & 0.0018 \end{bmatrix}$$

$$x^2 = 0.0394, \quad \nu = 48$$

$$V = \begin{bmatrix} 1.2698, & -0.0169 \\ -0.0169, & 0.0018 \end{bmatrix} 1.8912$$

$$V^{-1} = \begin{bmatrix} 4.7727, & 44.9198 \\ 44.9198, & 716.8725 \end{bmatrix}$$

$$\hat{\beta} = \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} 5.3582 \\ 1.7765 \\ -0.2416 \end{bmatrix}$$

신뢰수준	$x_T - x_S$ 의 신뢰구간	$2^{x_T - x_S} \times 100$ 의 신뢰구간
95%	(-7.9241, -6.7736)	(.4117, .9140)
90%	(-7.8345, -6.8697)	(.4381, .8551)
80%	(-7.7261, -6.9816)	(.4723, .7913)

한편 이 결과를 우리가 제시한 점추정치와 비교하여 보면 점추정치 $x_T \hat{x}_S$ 는 -7.3531, 따라서 $2\sqrt{\hat{x}_S} \times 100 = .6116$ 이 신뢰구간들의 거의 중앙에 약간 하한쪽에 있음을 알 수 있다. 이로써 우리는 그간 기술통계적으로 추정되어 오던 방식을, 통계적 모형을 사용하여 신뢰구간을 구할 수 있는 방법으로 발전시켜 제시하였다.

주어진 데이터로부터 가정된 모형의 적합성을 살펴보기 위하여 Plateau 현상 까지를 설명하는 모형으로

$$E(Y) = r_1 + r_2Z + r_3XV + r_4V \quad \dots\dots\dots(4. 1)$$

을 가정하여 보았다. 여기에서 V 는 회석배수가 6회 이상 인지를 보이는 의사변수이다.

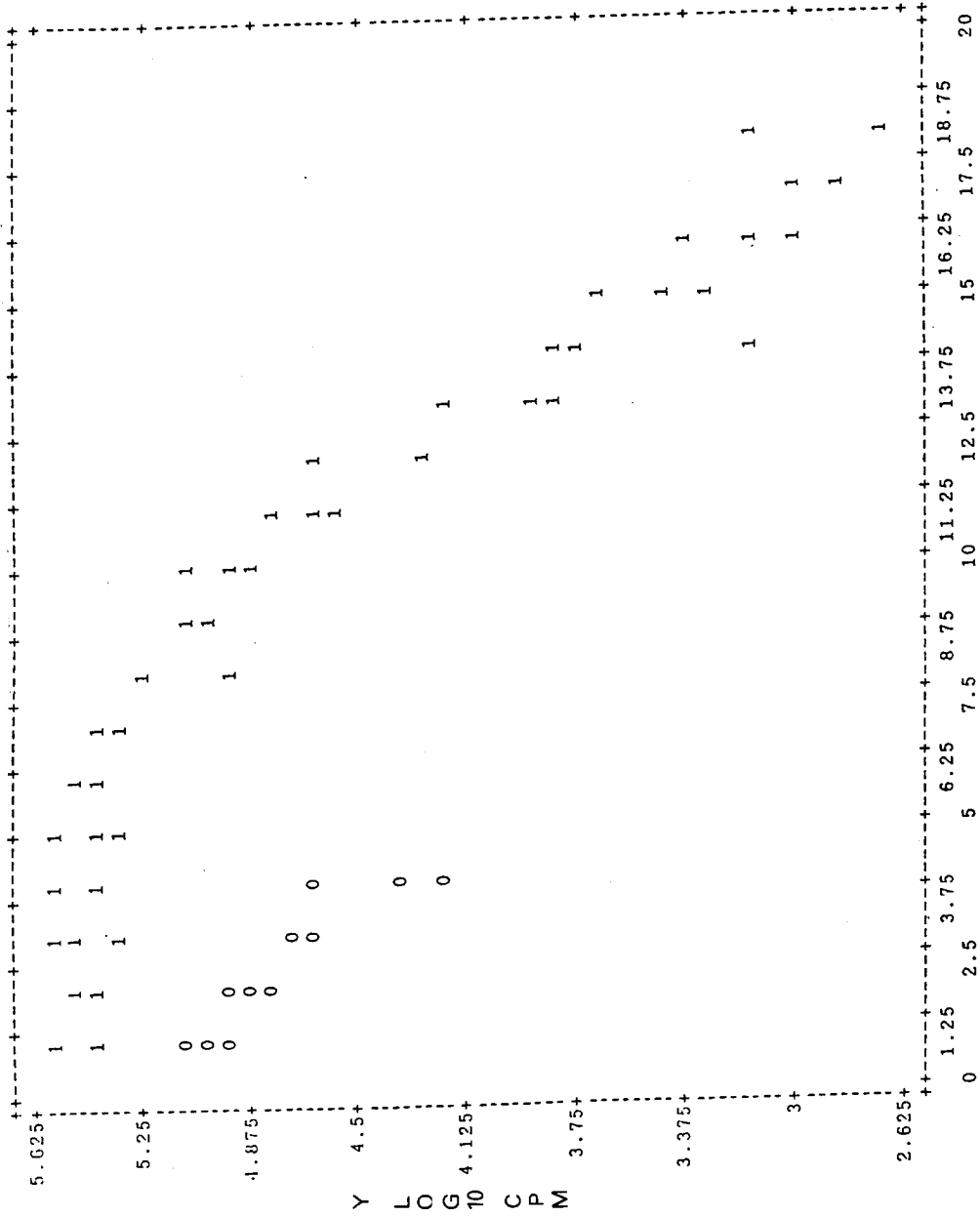
전체 66개의 관찰치로부터 얻은 추정치들과 관련된 검정결과는 F 값 605.1814(p -value $< .0001$)를 포함하여 모두 대단히 유의하다는 사실을 보여주었다. 또한 표준시료에서 1회 부터 5회 회석된 시료에서 얻은 관찰치를 제외하고 모형 (2. 1)에 적용시킨 결과 F 값이 446.4866(p -value $< .0001$)로 높은 값을 얻었다.

따라서 우리는 Gillis, Stadler 등의 연구결과와 주어진 자료가 일치된 결과를 보여준다고 받아들일 수 있었다.

우리가 문제를 베이지안 기법으로 택한것을 Fieller의 정리가 자료에 의존하여 때로 매우 넓은 신뢰구간을 제공하며, 또한 특정한 조건을 만족하여야 하는 반면 우리의 베이지안 접근은 임의의 자료에 대하여 정해진 신뢰도를 갖는 합리적인 신뢰구간을 제공하는 잇점이 있다고 보여진다.

본 연구에 있어서 극복되어야 할 몇가지 문제점을 지적한다면 첫째는 표준시료의 고농도에서 나타나는 Plateau 현상인데 이를 포함한 모형으로는 역가를 추정할 수 없었다. 이점을 해결하기 위한 연구가 기대된다. 둘째는 앞에서 언급한 사실의 연장인데 결국 Plateau 현상을 어디까지로 보는가 하는 문제이다. 본 연구에서는 6회 회석할 때 까지로 받아들였으나 그 객관적 기준을 제시하지는 못하였다. 그러나 이 시점을 미지로 하여 모형을 세울경우는 비선형 모형이 되어서 대단히 복잡한 계산상의 문제가 제기된다. 이에 대하여는 새로운 방법의 개발을 모색하여야 하지 않은가 생각되어진다. 마지막으로 사후 분포를 얻기 위하여 사용되는 비정보적 사전분포에서 기울기 β 가 음수를 갖는다는 학문적 이론을 수용하지 못하였다. 이점 역시 수리적으로 계산가능한 사후분포를 얻기위한 희생이었는데 이에 대한 연구가 필요로 된다.

PLOT OF Y WITH X BY Z



X : Log₂ Dilution of Sample

그림 (4. 1)

참 고 문 헌

- (1) Buonaccorsi, J.P. and Gatsonis, C.A. (1988). "Bayesian inference for ratios of coefficients in a linear model." *Biometrics*. 44, 87-101.
- (2) Gillis, S., Ferm, M.M., Ou, W. and Smith, K.A. (1978). "T cell growth factor : Parameters of production and a quantitative microassay for activity." *J. Immunol.* 120 : 20-27.
- (3) Hunter, W.G. and Lamboy, W. F. (1981). "A Bayesian analysis of the linear calibration problem (with Discussion)" *Technometrics* 23, 323-350.
- (4) Stadler, B.M. and Oppenheim. J.J. (1982). "Human interleukin-2 biological studies using purified IL-2 and monoclonal anti-IL-2 antibodies in Lymphokines vol. 6, Mizel, S.B." p. 117-135, *Academic press, New York*.

A Study on Statistical Inference on IL-2 Titer

Park Lai Hyun, Park Suk Young, Lee Suk Hoon

Abstract

This article deals with statistical inference on Interleukin-2 titer of which the clinical applications to cancer immunotherapy and some immunodeficiency diseases have been widely tried. A Linear model and the Bayesian approach are used to explain the bioassay which performs the measurements of IL-2 activity from an patient and an inference procedure including confidence intervals for the IL-2 titer of the patient through comparison with the Standard IL-2 is suggested and a real case of example is illustrated.