

## 회귀보완법을 이용한 위암 환자의 수술 후 5년 생존율에 관한 분석

송재기<sup>1)</sup> 이원기<sup>2)</sup> 송명언<sup>3)</sup> 유완식<sup>4)</sup> 정호영<sup>5)</sup>

### 요약

경북대학교병원에서 1985년에서 1994년까지 위암 때문에 위 절제수술을 받은 1,192명의 환자에 대한 자료를 이용하여 5년 생존율에 관해 분석하고자 한다. 일반적으로 위암 진단을 받은 환자가 수술을 받으려고 할 때 또는 수술을 받은 직후에, 환자의 임상적 특성들을 이용하여 수술후 생존시간과 수술후 5년 생존 여부는 큰 의미가 있다. 그러나 많은 경우에 있어서 실제 임상자료는 연구가 진행 중에 있으므로 생존시간이 우측 중도절단된 형태로 관측되어 기존의 판별분석과 로짓분석을 적용할 수 없다. 본 논문에서는 Buckley와 James가 제안한 의사확률변수를 이용하여 수술전과 수술직후, 두 시점에서 중도절단된 자료를 보완하고, 판별분석과 로짓분석을 통하여 수술전과 수술직후에 환자들의 각 특성이 5년 생존 여부에 미치는 영향을 분석을 한다.

### 1. 서론

암환자가 수술을 받았을 때 수술성공 여부는 암의 종류에 따라 그 기준이 다르지만 보통의 경우 수술후 5년 생존여부를 기준으로 두고 있어, 수술을 담당한 의사 및 환자 그리고 환자의 보호자는 수술후 5년 예후에 높은 관심을 보이고 있다. 따라서 환자의 여러 가지 특성과 임상적 소견을 종합하여 암환자가 수술 받기 전과 수술 받은 직후에, 수술후 5년 예후를 알아볼 수 있는 모형이 있다면 수술 전에는 수술을 담당할 의사가 환자에게 수술여부에 대한 판단근거를 제공할 수 있으며, 수술후에는 환자의 생존에 대한 궁금증에 답을 할 수 있는 객관적 자료를 제공할 수 있을 것이다.

본 연구에서는 1985년 1월부터 1994년 12월까지 경북대학교병원 일반외과에서 위절제 수술을 받았던 위암 환자의 자료를 이용하여 수술후 5년 예후를 알아볼 수 있는 모형을 만들고자 한다. 이 기간중 위절제수술을 받은 환자의 총수는 1,282명이었으나 90명의 자료는 여러 요인으로 기록이 불확실하여 이들을 제외한 1,192명의 자료를 연구대상으로 하였다. 자료의 일반적 특성에 따른 분포는 남자 773명(64.8%)명이며 여자는 419명(35.2%)명 이었고, 평균 연령은 54.37세 이었으며, 자료중 713명(59.8%)이 중도절단되어 관측되었다. 그 외 중요한 변수들은 표1에 정리하였다. 원 자료에는 표1의 변수외에도 많은 변수들이 기록되어 있으나 임상적으로나 통계적으로 생존함수에 영향을 주지 않는다고 판단되는 변수들은 제외하였다. 병기 구분은 Intrnational Union Against Cancer(UICC)의 tumor node metastasis(TNM) 분류(1987)를 이용하였다.

- 1) (702-701) 대구광역시 북구 산격동 1370, 경북대학교 통계학과, 교수
- 2) (702-701) 대구광역시 북구 산격동 1370, 경북대학교 통계학과, 박사과정
- 3) (702-701) 대구광역시 북구 산격동 1370, 경북대학교 통계학과, 시간강사
- 4) (700-721) 대구광역시 중구 삼덕동 50, 경북대학교병원 일반외과, 교수
- 5) (700-721) 대구광역시 중구 삼덕동 50, 경북대학교병원 일반외과, 조교수

표 1.1: 위암수술을 받은 환자 자료의 변수

변수	변수설명
SURVIVAL	Survival duration in years
SEX	성별 ( 1: male, 2: female )
AGE	나이
DEPTH	Depth of invasion ( 1, 2, 3, 4 )
DISTANT	종양의 원격 전이 여부 (0: no, 1: yes)
PS	Pathologic stage (1: a, 2: b, 3: , 4: a, 5: b, 6: )
FS	Final stage (1: a, 2: b, 3: , 4: a, 5: b, 6: )
SIZE	Largest diameter
BORRMANN	Gross type of advanced cancer (0, 1, 2, 3, 4)
NU	UICC 기준에 의한 림프절의 전이 정도 (0, 1, 2)
NN	전이된 림프절의 총수
NF	전이된 림프절의 수 / 제거된 림프절의 수 * 100

일반적으로 암환자의 자료는 우측 중도절단되어 관측되어지므로 외과의들은 수술 받은 환자의 생존함수를 K-M 추정량(Kaplan and Meier, 1958)을 이용하여 추정하고 있으며, 생존에 영향을 미친다고 여겨왔던 임상적 변수들을 평가하거나 새롭게 찾기 위하여 Cox 모형(Cox, 1972)에 적합시켜 왔다. 또한 수술후 5년 예후에 관심이 있는 경우 수술후 5년 이상 생존하였다면 수술에 성공하였다고 간주하고 로짓분석에 적합시켰다. 그러나 Cox 회귀 모형의 경우 비례위험 가정이 깨어지는 경우에 문제가 발생하며 로짓모형의 경우 중도절단된 자료의 처리가 문제이다.

본 연구에서는 이러한 문제점을 해결하기 위하여 암환자가 수술 받기 전과 수술 받은 직후, 두 경우에서 중도절단된 자료를 회귀보완법으로 보완하여 분석하고자 한다. 앞에서 언급하였듯이 자료의 모든 변수들의 값은 수술직후 정확히 기록한 것이나 이들 변수중 SEX, AGE, DEPTH, DISTANT, SIZE, BORRMANN 등은 수술전에도 대략적으로 예측할 수 있는 변수들이므로 이들 변수들을 이용하여, 환자가 수술전에 수술후 생존시간 및 5년 이상 생존여부에 관해 분석하고, 환자가 수술 받은 직후에는 정확히 관측된 모든 유의한 변수들을 이용하여 분석하고자 한다. 2장에서는 수술전과 수술직후, 각 경우에서 위암수술을 받은 환자들의 여러 특성(공변량)을 이용하여 Buckley와 James추정량(이하 B-J추정량)(Buckley & James, 1979; Miller & Halpern, 1982; James & Smith, 1984; Lai & Ying, 1991; Smith & Zhang, 1995; Currie, 1996)을 구하여 중도절단된 생존시간에 대한 자료를 보완하고, 각 공변량의 수술후 생존시간에 대한 영향을 조사하고자 한다. 또한 3장에서는 이렇게 보완된 자료를 이용하여 수술전과 수술직후 시점에서 판별분석과 로짓분석을 통하여 수술후 5년 생존율을 분석하고자 한다.

## 2. 중도절단된 자료의 회귀보완

2장에서는 회귀보완법을 위암수술환자의 자료에 적용하기 위하여 B-J추정량을 고려하였다. 생존분포가 특별한 모수모형이 아닌 경우 주어진 공변량을 이용하여 중도절단된 자료를 보완하기 위해, 선형모형(linear model)가정하에서 Miller(1976)는 정규방정식을 이용

하였으며, Buckley와 James(1979)는 K-M추정량을 이용하여 중도절단된 생존시간의 조건부 기대수명을 추정하는 방법을 제안하였다. Miller와 Halpern(1982)는 이들 두 방법과 비례위험모형(Cox, 1972) 등을 서로 비교하였다. 또한 Lai와 Ying(1991)은 수정된 B-J추정량의 점근적 성질을 조사하였고, 최근에 Smith와 Zhang(1995)은 중도절단된 자료를 B-J추정량을 이용하여 개선된 점도표를 구성하여 분석하였으며, Currie(1996)는 B-J추정량의 회귀계수의 극한값에 대한 식을 얻고 분석하였다. B-J추정량에 대한 개요는 다음과 같다.

$n$ 명의 환자에 대해서 생존시간은  $Y_1, Y_2, \dots, Y_n$ 이고 중도절단된 시간은  $C_1, C_2, \dots, C_n$ 이며 공변량은  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ , ( $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$ ) 일 때 다음과 같은 중도절단된 생존시간  $Z_i$ 를 관측하게 된다.

$$Z_i = Y_i\delta_i + C_i(1 - \delta_i) = \min\{Y_i, C_i\}$$

여기서  $\delta_i$ 는 지시변수로서  $\delta_i = 0$ 인 경우 생존시간  $Y_i$ 가 중도절단 되었음을 가르킨다. 여기서  $p$ 개의 공변량  $x_i$ 와 생존시간  $Y_i$ 에 대해 선형모형을 고려하면 다음과 같다.

$$Y = X\beta + \epsilon, E(\epsilon) = \alpha\mathbf{1}, \text{Var}(\epsilon) = \sigma^2\mathbf{I},$$

여기서 반응변수  $Y$ 는 생존시간으로  $n \times 1$  벡터, 독립변수  $X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^T$ 는 공변량으로  $n \times p$  정칙행렬이며, 회귀계수  $\beta = (\beta_1, \beta_2, \dots, \beta_p)^T$ 는  $p \times 1$  벡터이다. 또한  $\epsilon$ 은  $n \times 1$  오차벡터로  $\epsilon_i$ 의 생존함수는  $S = 1 - F$ 이며, 평균은 절편회귀계수  $\alpha$ 이며 여기서  $\mathbf{1}$ 은 전부 1인  $n \times 1$  벡터이다. 이때  $Y_i^* = Y_i\delta_i + E(Y_i|Y_i > C_i)(1 - \delta_i)$ 라 두면  $E(Y_i^*) = X\beta$ 가 되므로, 회귀계수  $\beta$ 를  $\mathbf{b}$ 로 추정하여, 중도절단된 관측시간  $C_i$ 를  $Y_i^*(\mathbf{b}) = X\mathbf{b}$ 로 보완하고자 한다.

임의의 회귀계수 추정치를  $\mathbf{b} = (b_1, \dots, b_p)^T$ 라 할 때, 관측된 잔차를 순서화하면

$$e_i(\mathbf{b}) = Z_i - \mathbf{x}_i^T \mathbf{b} = e_{(i)}(\mathbf{b})$$

이며, 이 잔차에 대한 K-M추정량을  $\hat{S}$ 이라 하고, 보통의 경우처럼 편의상  $\hat{S}(e_n(\mathbf{b})) = 0$ 이라 하자. 그리고 이 잔차의 순서에 따라  $(Y_i, \delta_i, \mathbf{x}_i)$ 를 재배열하면 새로 보완된 반응변수  $Y_i^*(\mathbf{b})$ 는 다음과 같이 구성되어 진다.

$$Y_i^*(\mathbf{b}) = \mathbf{x}_i^T \mathbf{b} + \left[ e_i(\mathbf{b})\delta_i + \left\{ \sum_{k=1}^n w_{ik}(\mathbf{b})e_k(\mathbf{b}) \right\} (1 - \delta_i) \right],$$

여기서

$$w_{ik}(\mathbf{b}) = \begin{cases} \frac{v_k(\mathbf{b})\delta_k}{\hat{S}\{e_i(\mathbf{b})\}} & \text{if } k > i \\ 0 & \text{otherwise} \end{cases}$$

이며,  $v_k(\mathbf{b})\delta_k$ 는  $e_k(\mathbf{b})$ 에 할당된 K-M추정량  $\hat{S}$ 의 뿔크기(jump size)이다. 또는 행렬형태로 다음과 같이 쓸 수 있다.

$$Y^*(\mathbf{b}) = X\mathbf{b} + W(\mathbf{b})(Z - X\mathbf{b}),$$

여기서  $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_n)^T$ , 그리고

$$W(b) = \text{diag}(\delta) + (w_{ik}(b)) = \begin{pmatrix} \delta_1 & w_{12}(b) & w_{13}(b) & \cdots & w_{1n}(b) \\ 0 & \delta_2 & w_{23}(b) & \cdots & w_{2n}(b) \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & w_{n-1n}(b) \\ 0 & 0 & 0 & \cdots & \delta_n \end{pmatrix}$$

이다.

그러면 새롭게 만들어진 자료  $Y^*(\mathbf{b})$ 를 이용하여 회귀계수  $\beta$ 는 최소제곱법에 의해 다음과 같이 새롭게 추정되어 진다.

$$\mathbf{b} = (X^T X)^{-1} X^T Y^*(\mathbf{b})$$

이러한 절차를 새롭게 추정된 회귀계수가 특정한 값으로 수렴할 때까지 반복하여 B-J 추정량  $\hat{\beta}$ 을 얻을 수 있으며, 그 존재성은 James와 Smith(1984) 그리고 Lai와 Ying(1991)에 의해 밝혀졌으며, 초기추정량으로는 관측된 자료인  $\mathbf{Z}_i$ 를 이용한 최소제곱추정량을 많이 사용한다.

본 연구에 사용한 자료의 공변량은 임상에서 생존시간에 영향을 미친다고 알려졌거나 있을 것으로 짐작되는 변수, 그리고 통계적으로 유의하면서 다중공선성이 심하지 않은 변수들을 선택하였다. 수술전 모형에서는 공변량으로 DEPTH, DISTANT, SIZE, 그리고 BORRMANN을 사용하였으며, 수술직후 모형에서는 NF를 추가하여 사용하였다. 실제로 수술전에 사용한 공변량은 정확도가 수술후에 예측하는 것에 비하여 대략적으로 알 수 있는 정도이나 본 연구에서는 수술후 정밀하게 조사된 자료를 사용하였다. 사용된 공변량 중 DEPTH, DISTANT, BORRMANN은 순위형 자료이지만 대표본이므로 그대로 사용하였고, SIZE와 NF의 분포 역시 비대칭 현상이 있어 Box-Cox 변환을 시도하였으나 변수변환의 필요성이 크게 제기되지는 않아 그대로 사용하였다. B-J추정법으로 추정한 결과는 표 2.1과 같다. 표 2.1에서 보는 것처럼 고려한 공변량에서 공변량들의 값이 증가함에 따라 생존시간이 감소 한다는 것을 알 수 있으며 이것은 여러 임상적 조건에 비추어 볼 때 타당하다고 하겠다. 추정결과 중도절단자료의 생존시간은 평균 6.18 1.98년으로 증가하였으며 공변량에 따른 추정결과는 표 2.2와 같다. 표 2.2의 Size와 NF는 연속자료로 관측하였으나 조건부 기대수명 평균의 개략적인 형태를 알아보하고자 적당히 나누어 나타내었다.

표 2.1: B-J추정법에 의한 추정결과

		상수	DEPTH	DISTANT	SIZE	BORRMANN	NF
수술전 모형	추정치	16.61640	-2.28977	-3.64005	-0.02382	-0.75029	
	표준오차	0.63545	0.20502	0.27794	0.00424	0.15589	
수술직후 모형	추정치	14.69563	-1.78681	-2.06587	-0.01405	-0.65314	-0.06076
	표준오차	0.65931	0.21557	0.30037	0.00442	0.16177	0.00533

표 2.2: 공변량에 따른 보완된 생존시간과 중도절단된 생존시간의 차이

공변량		N	수술전 모형		수술직후 모형	
			평균차이	표준편차	평균차이	표준편차
Depth	1	260	8.85	2.08	7.93	2.04
	2	209	6.49	0.77	5.73	1.09
	3	233	5.81	0.35	4.71	0.26
	4	11	4.80	2.02	4.52	0.53
Distant	0	679	7.18	1.91	6.24	1.99
	1	34	5.48	1.29	4.93	0.91
Size	20미만	140	8.52	2.18	7.51	2.12
	20-30	115	7.58	1.96	6.71	1.97
	30-40	124	7.05	1.89	6.15	2.11
	40-50	96	6.68	1.58	5.80	1.64
	50이상	238	6.24	1.21	5.31	1.35
Borrman	0	260	8.85	2.08	7.93	2.04
	1	50	6.19	0.84	5.26	1.11
	2	247	6.05	0.70	5.09	0.92
	3	148	6.17	0.81	5.33	1.00
	4	8	5.63	1.12	4.41	2.10
NF	0	493			6.58	2.05
	1-10	113			5.69	1.56
	11-30	75			5.08	1.10
	31이상	32			4.27	0.90

### 3. 5년 생존율에 관한 분석

이 장에서는 위암환자의 수술시 보통 수술후 5년 예후에 관심을 두고 있으므로, 5년 생존율에 관한 자료분석을 하였다. 이러한 분석을 위한 통계적 방법으로는 판별분석과 로짓 분석을 이용할 수 있으며 이와 같은 분석을 적용하기 위해서는 자료가 완전하게 관측되어야 한다. 앞서 보았듯이 다수의 자료가 중도절단되어 관측되므로 이러한 분석을 적용할 수 없어 B-J추정량을 이용하여 완전한 자료로 전환하여 사용하고자 한다. 그러나 이 경우에도 B-J추정량은 주어진 공변량의 함수로 나타나 서로 종속적인 상관성이 있을 수 밖에 없으므로 이론적으로 문제가 있으나, 중도절단 자료를 버리는 것보다는 보다 많은 정보를 사용할 수 있다는 이점을 살려 적용해보고자 한다.

먼저 5년후 생존여부를 예측할 수 있는 판별분석을 하였다. 판별분석의 목적은 수술후 5년 생존여부를 결정해 줄 수 있는 독립변수를 찾아 기준을 세우는데 있다. 판별분석을 하기 위해서는 사용될 독립변수들이 정량적으로 측정되어야 하고 다중정규분포가정을 만족하여야 하지만 여기서 사용된 독립변수는 최소한 순위형 자료이며 대표본이므로 이러한 가정은 무시하였다. 다양한 독립변수를 선택하여 분석을 하였으나 민감도와 특이도에 크게 차이가 나지 않아 본 분석에서는 B-J추정량에 사용한 공변량 변수를 독립변수로 사용하였다(B 방법). 또한 만약 B-J추정량을 적용하지 않고 중도절단된 자료를 버릴 경우(A 방법)와 비교해 보았다. 중도절단된 자료를 버릴 경우 5년 이상의 시간에서 중도절단된 경우에는 5년이상 생존한 것이 확실하므로 생존한 것으로 간주하였다.

술전과 수술직후에 각 시점에서 수술후 5년 생존여부를 판별해 줄 판별함수의 추정계

수는 표 3.1과 같다. 여기서 A 방법은 5년 미만 중도절단자료를 제외한 841명의 자료로 분석하였으며, B 방법은 B-J추정법으로 1,192명의 자료를 모두 사용한 것이다.

표 3.1: 판별함수의 추정계수

공변량	수술전 모형		수술직후 모형	
	A 방법	B 방법	A 방법	B 방법
DEPTH	1.16350	1.17983	1.04000	1.03877
DISTANT	1.19518	0.04749	1.41754	1.21063
SIZE	0.01529	0.01309	0.01220	0.00912
BORRMANN	0.33830	0.07185	0.32053	0.06130
NF			0.03257	0.04691

표 3.1에 의하면 모든 경우에 있어 각 독립변수 값이 증가할수록 사망 그룹으로 분류된다는 것을 알 수 있으며, B-J추정량을 사용한 경우와 그렇지 않은 경우 수술 전에는 DISTANT와 BORRMANN에서 크게 차이가 남을 알 수 있고, 수술 후에는 BORRMANN에서만 다소 차이를 보이고 있다. 이들 판별함수를 이용한 분류결과는 표 3.2와 같다.

표 3.2: 판별분석에 의한 판별결과

		판별분석에 의한 결과											
		수술전 모형						수술직후 모형					
		A 방법			B 방법			A 방법			B 방법		
		생존	사망	합계	생존	사망	합계	생존	사망	합계	생존	사망	합계
관측결과	생존	268 (69.61)	117 (30.39)	385	524 (71.20)	212 (28.80)	736	312 (81.04)	73 (18.96)	385	627 (85.54)	106 (14.46)	733
	사망	65 (14.25)	391 (85.75)	456	80 (17.54)	376 (82.46)	456	90 (19.74)	366 (80.26)	456	127 (27.67)	332 (72.33)	459
	합계	333	508	841	604	588	1192	402	439	841	754	438	1192

표 3.2에 의하면 수술전 모형에서 A, B 방법으로 판별한 결과 옳게 판별한 비율은 각각 78.4%, 75.5%였으며, 수술직후 모형에서는 각각 80.6%, 80.5%였다.

먼저 수술전 모형에 비해 수술직후 모형의 옳게 판별한 비율이 크게 증가하지 않았는데, 이는 수술전 모형에서는 원래 대략 예측된 공변량 값을 사용하여야 하나 본 분석에서는 수술후 정확히 관측된 공변량 값을 사용하였기 때문인 것 같다.

그리고 두 경우 모두에서 B 방법(B-J추정량을 사용한 경우)의 판별력이 다소 떨어지는 것 같다. 그러나 실제 수술을 받은 암환자의 사망은 1-2년 내에 대부분 발생하며, 그 이후의 사망률은 급격히 감소한다고 알려져 있다. 그러므로 수술후 생존시간이 5년 미만에서 중도절단된 자료들 중 많은 자료들이 2년 이후에서 중도절단되어지며 이들 중 상당수가 5년 이상 생존할 가능성이 훨씬 높을 것이라고 예상되어 진다.

표 3.2의 관측결과에서도 나타났듯이 중도절단된 자료의 대부분은 5년이상 생존할 것으로 예측되었다. 본 연구에 사용된 자료의 경우 5년 미만에서 중도절단된 자료의 공변량이 대개 좋았으며, 중도절단된 자료의 관측 수명이 긴 관계로 B-J추정량을 이용한 조건부 기대수명이 크게 나타났다. 이는 중도절단자료의 경우 상당수는 생존할 가능성이 많다는

것을 단면적으로 보여준다. 이러한에도 불구하고 이들 5년 미만에서 중도절단된 자료들을 분석에서 제외한 A 방법은 표본이 모집단을 대표할 수 없게 되어 편의현상을 불러올 수 있으므로, 판별함수에 사용된 독립변수의 계수추정에 문제가 있어 추후 자료에 대한 예측능력을 저하시킬 것으로 생각된다. 따라서 A 방법의 판별력이 B 방법보다 다소 좋다고 해도 그대로 사용하기에는 문제가 있다.

또한 지금의 자료는 연구기간이 상당히 길어서 수술후 5년 이상 생존한 중도절단 자료가 상대적으로 많이 나타나지만 보통의 이와 유사한 연구에서는 연구기간이 길지 않기 때문에 버려지는 5년 이하 중도절단된 자료가 상대적으로 많아져서 더 많은 문제가 있으리라 생각되어 진다.

다음으로는 B-J추정량을 이용한 보완자료를 이용하여 로짓분석을 하였으며 분석 결과는 표 3.3에 정리하였다.

표 3.3: 로짓분석 결과

			상수	DEPTH	DISTANT	SIZE	BORRMANN	NF
수술 전 모형	A방법	추정치	-4.1952	1.0089	2.5330	0.0162	0.3562	
		표준오차	0.3738	0.1473	0.4103	0.0037	0.1083	
		승산비		2.742	12.591	1.016	1.428	
	B방법	추정치	-4.8440	1.1793	1.5083	0.0114	0.2605	
		표준오차	0.3439	0.1212	0.2143	0.0028	0.0903	
		승산비		3.252	4.519	1.012	1.298	
수술 직후 모형	A방법	추정치	-3.8820	0.7786	1.9589	0.0134	0.3207	0.0423
		표준오차	0.3750	0.1555	0.4256	0.0039	0.1133	0.0064
		승산비		2.178	7.092	1.013	1.378	1.043
	B방법	추정치	-4.5599	0.9704	0.9363	0.0085	0.2180	0.0433
		표준오차	0.3489	0.1283	0.2370	0.0030	0.0956	0.0049
		승산비		2.639	2.551	1.009	1.244	1.044

로짓분석의 결과를 보면 5년 미만 중도절단 자료를 분석에서 제외할 경우 DISTANT의 승산비가 매우 높게 나타남을 알 수 있다. 이에 비해 B-J추정량을 사용한 경우 각 변수의 승산비는 거의 일정하게 추정됨을 알 수 있으며, 또한 DISTANT의 승산비가 중도절단자료를 제외하였을 때보다 상당히 떨어짐을 알 수 있다. 이는 중도절단 자료를 제외하고 분석함으로써 종양의 원격 전이가 있을 경우 DISTANT의 계수를 과추정하여 사망확률을 지나치게 크게 한다는 사실을 알 수 있다.

#### 4. 결론

통계적 자료분석을 위한 이론은 대부분의 경우 완전한 자료의 형태를 요구한다. 그러나 생존자료의 경우 중도절단되어 관측되는 자료가 다수 있어 통계적 분석에 한계를 낳는다. 비록 생존분석법에 의한 자료분석이 많이 개발되어 있지만 아직도 임상 의들이 알고 싶은 많은 문제점들을 해결해 주지는 못한다. 이러한 경우 중도절단된 자료를 회귀보완법에 의해 보완하여 완전한 자료화가 가능하다면 기존의 이론들을 바로 활용할 수 있는 이점이 있다. 그러나 중도절단된 자료를 회귀보완법에 의해 보완하면 이들 자료는 실제 관측된 생존

시간과 공변량에 대해 종속되어지므로 이론적으로는 아직 문제점이 존재한다고 하겠다.

### 참고문헌

- [1] Buckley, J. J. and James, I. R. (1979). Linear regression with censored data, *Biometrika*, **66**, 429-436.
- [2] Cox, D. R. (1972). Regression models and life-tables (with discussion), *Journal of the Royal Statistical Society, Ser B*, **34**, 187-202.
- [3] Currie, I. D. (1996). A note on Buckley-James estimators for censored data, *Biometrika*, **83**, 912-915.
- [4] James, I. R. and Smith, P. J. (1984). Consistency results for linear regression with censored data, *Annals of Statistics*, **12**, 590-600.
- [5] Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations, *Journal of American Statistical Association*, **53**, 457-481.
- [6] Lai, T. L. and Ying, Z. (1991). Large sample theory of a modified Buckley-James estimator for regression analysis with censored data, *Annals of Statistics*, **19**, 1370-1402.
- [7] Miller, R. G. and Halpern, J. (1982). Regression with censored data, *Biometrika*, **69**, 521-531.
- [8] Smith, P. J. and Zhang, J. (1995). Renovated scatter plots for censored data, *Biometrika*, **82**, 447-452.

[ 1998년 11월 접수, 1999년 5월 최종수정 ]



## Analysis of 5-year Survival Rate of Gastric Cancer Patients Using Pseudo Random Variable

Jae-Kee Song<sup>1)</sup> Won Kee Lee<sup>2)</sup> Myung Unn Song<sup>3)</sup>  
Wansik Yu<sup>4)</sup> Ho Young Chung<sup>5)</sup>

### ABSTRACT

In the period 1985-1994, 1,192 patients with gastric cancer underwent surgery at the Department of Surgery, Kyungpook National University Hospital. In this paper, we analyze the 5-year survival rate with the stomach data via the discriminant analysis and logistic regression analysis. To remedy heavily-censored data the Buckley and James's pseudo random variable(1979) is used and the incomplete survival times are adjusted in applying regression model.

---

1) Professor, Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea.

2) Ph. D. candidate, Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea.

3) Part-time Lecturer, Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea.

4) Professor, Department of Surgery, Kyungpook National University Hospital, Taegu, 700-721, Korea.

5) Assistant Professor, Department of Surgery, Kyungpook National University Hospital, Taegu, 700-721, Korea.