한국통계학회논문집 제2권 2호
1995년 12월 pp. 434-444

전문가용 한국형 통계패키지 개발연구 I 1)

- 생존분석, 베이지안분석, 보험통계를 중심으로 -

이정진, 강근석, 이윤오, 김지현, 이창수, 김성철2)

요 약

외국산 통계패키지를 대체할만한 국산 통계패키지를 개발하려는 움직임이 현재 국내에서 활발히 전개되고 있다. 본 논문은 일반인을 대상으로하는 통계패키지 개발연구에 대한 경험을 토대로 통계전문가들이 많이 사용하는 모듈인 생존분석, 베이지안 분석, 보험통계 분야 등에 관한 한국형 통계패키지 모듈에 대한 설계를 소개하고자 한다.

1. 서론

다량의 정보를 신속 정확하게 처리하여 분석하는 통계패키지의 사용은 최근 급성장하여 요즈음은 국가 공공기관이나 대학, 연구소, 대기업, 그리고 웬만한 중소기업에서도 사용하고 있다. 하지만 현재 국내에서 많이 사용되고 있는 통계패키지들은 SAS, SPSS 등 모두 외국산 제품들이어서 그 구입과 임대에 많은 외화가 낭비되고 있다. 그래서 이들 외국산 통계패키지를 대체할만한 국산 통계패키지를 개발하려는 움직임이 현재 국내에서 활발히 전개되고 있다. 본 논문에서는 '통계학 교육용 한글 소프트웨어 개발연구'(이정진 외 [2])와 '일반인용 한글 통계패키지 개발연구'(이정진 외 [3])에 대한 경험을 토대로 통계전문가들이 많이 사용하는 모듈인 생존분석, 베이지안 분석, 보험통계 분야 등에 관한 한국형 통계패키지의 설계 개발을 소개하고자 한다.

통계패키지의 개발은 통계계산(statistical computing) 분야에 관심있는 사람에게는 통계이론적(이론이나 알고리즘등)으로는 이미 거의 공개되어 있어 큰 문제점이 없다 (안현순 [1], Cooke et al [5], Kennedy and Gentle [10], Maindonald [12] 참조). 다만 이러한 이론을 '사용자들에게 쉽고 유용하게끔' 시스템이나 결과화면을 설계하여 구현하는데 있어서 여러 가지 방법으로 시도해서 수정하는 'trial-and-error'에 많은 시간과 인력이 필요하다. 통계패키지 개발과정중에 제일 어려움을 겪는 부분이 이와 같은 설계에 관한 문제인데 최근 전산통계 분야의 중요한 이슈가 되고 있다

2절에서는 기존에 연구되었던 일반인을 대상으로한 통계패키지의 설계를 소개하면서 새로운 모듈과의 연결을 설명하였다. 3절에서는 생존분석 모듈에 대한 설계, 4절은 베이지안 분석 모듈의 설계, 5절은 보험통계에 관한 모듈 설계를 설명하고, 6절에서는 결론및 향후과제에 관한 제안을 하였다.

---

1) 이 논문은 1993년도 한국학술진흥재단의 대학부설연구소 연구과제 연구비에 의해 연구되었음.
2) (156-743) 서울특별시 동작구 상도동 1-1 숭실대학교 통계학과

2. 통계패키지의 개발방향

통계패키지 개발의 장기적인 목적은 세계시장에서 SAS, SPSS 등과 경쟁할만한 패키지를 만드는 것이다. 하지만 이 제품들은 수 십년간에 걸친 연구개발과 운용경험이 축적되어 나온 우수한 제품들이기 때문에, 당장에 모든 분야에서 이들보다 우수한 제품을 만드는 것은 거의 불가능한 일이다. 따라서 한국형 통계패키지의 기본적인 개발방향은 외국산 패키지보다 우수한 기능을 순차적으로 만들어 나가는 것이다. 이와같은 개발방향으로 일반인을 대상으로 하는 국산 통계패키지 개발에 사용한 전략은 다음과 같다.

-- 한글의 입출력이 자유로운 패키지가 되어야 한다. 외국산 패키지가 우리나라 사용자들에게 가장 불편한 점은 한글로서 입출력이 불가능하다는 점일 것이다.
-- 국산화 기술축적을 위해 통계분석에서 자주 이용되는 모듈의 개발을 우선적으로 한다.
-- 패키지의 운용시스템은 풀다운메뉴(pull-down menu) 형식이, 자료의 입출력방법은 스프레드쉬트(spreadsheet) 방식을 이용하는 소프트웨어가 사용자에게 편리하다.
-- 대형컴퓨터용보다 PC용, PC중에서도 우선은 매킨토시용보다는 우리나라에서 제일 많이 사용되는 IBM/PC 호환기종 소프트웨어를 우선 개발한다.
-- IBM/PC의 운영시스템중 우선은 DOS시스템에서 그래픽환경을 구사하는 패키지를 개발한 다음, 한글 WINDOWS를 이용한 패키지를 개발한다.
-- 그래픽 카드는 IBM/PC의 호환기종에 가장 많이 보급되어 있는 Hercules와 VGA카드를 지원하면 PC 사용자의 90%이상이 이 시스템을 사용할 수 있다.
-- 결과출력은 그래프를 많이 이용하여 사용자가 쉽게 분석을 할 수 있도록 설계한다.

이러한 개발전략은 전문가용 통계패키지 개발에도 유용할 것으로 판단되지만 전문가들이 사용하는 패키지는 일반인들이 이용하는 것과 달리 그 통계적 분석기법이 매우 다양하다. 따라서 이번 개발에는 외국산 패키지가 갖고 있는 통계적 분석기법중 전문가들의 이용빈도가 많은 모듈을 우선적으로 개발하거나, SAS SPSS 등에 없는 베이지안분석, 보험통계 기법을 개발하였다.

3. 생존분석

생존분석은 수명 자료의 분석을 목적으로 한다. 수명 자료 중에서 확률적 우측 중도절단 자료(randomly right censored data)에 대한 대표적 비모수통계 분석 방법으로, 분포함수(또는 생존함수)의 추정을 위한 Kaplan-Meier 추정량과 회귀분석 모형인 Cox의 비례위험모형(Cox's proportional hazards model)을 들 수 있다. 이 두 방법에 대한 모듈을 참고문헌 Cox [6], Hall and Wellner [8], Kim [11], Wang and Hettmansperger [16] 등의 이론을 근거로 하여 우선적으로 개발하였는데 그 주요 기능은 다음과 같다.

1) 생존함수의 Kaplan-Meier 추정량
   - 생존함수의 추정과 각 시점에서의 신뢰구간
   - 추정된 생존함수와 신뢰구간을 그래프로 표시

- 둘 이상의 층에 대한 생존함수와 신뢰구간을 한 그래프에 그리는 기능

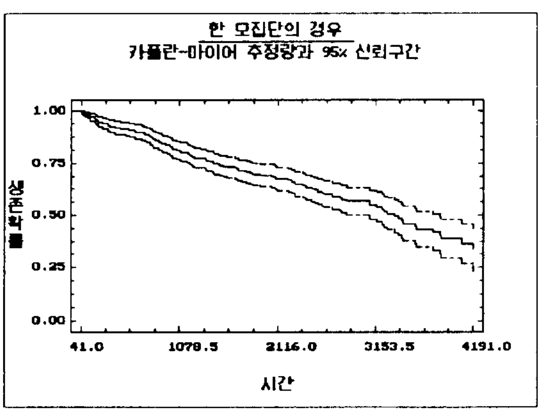
2) Cox의 비례위험모형

- 모수 $\underline{\beta}$의 최우추정치(maximum partial likelihood estimate)
- $\underline{\beta}$의 추정치의 표준오차, 각 추정치의 유의성 검정
- $\underline{\beta}$의 추정치에 대한 공분산 행렬
- 귀무가설 $H_0:\underline{\beta}=\underline{0}$ 에 대한 검정통계량의 값과 p값
- 선택사항으로 최우추정치를 얻기까지의 수렴과정을 출력할 수 있도록 하였다.

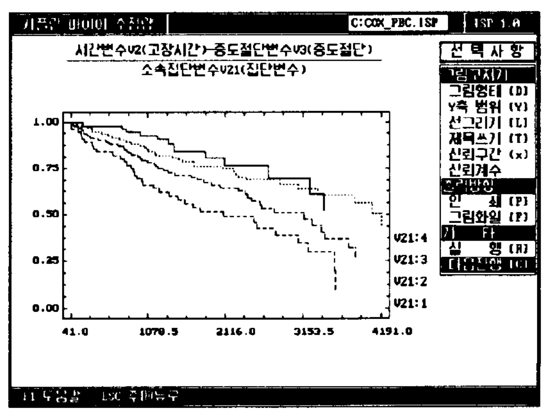
이러한 기능은 SAS 등에도 있지만 결과를 분석하기 위하여 그래픽화면을 이용한 그림을 효과적으로 설계한 것이 장점이라고 볼 수 있다.

3.1 Kaplan-Meier 추정량

확률적 우측 중도절단 자료가 주어졌을 때, 분포함수의 일반화 최우 추정량 (generalized maximum likelihood)인 Kaplan-Meier 추정량의 정의와 성질에 대해서는 잘 알려져 있다. 먼저, 한 모집단에서 추출한 표본일 때, Kaplan-Meier 추정량을 구하고 Greenwood 식 (Greenwood's formula) 에 의한 분산을 이용하여 신뢰구간을 구한다. <그림 3.1>은 그 결과를 그래프로 나타낸 것이다. 신뢰구간의 표시 여부는 선택사항으로 두었다. 둘 이상의 층 또는 모집단의 경우, 각 층에 대한 분포함수의 추정량을 <그림 3.2>와 같이 한 화면에 함께 나타낼 수 있도록 하였다. 이 때 신뢰구간도 선택적으로 표시할 수 있도록 하였다.



<그림 3.1>



<그림 3.2>

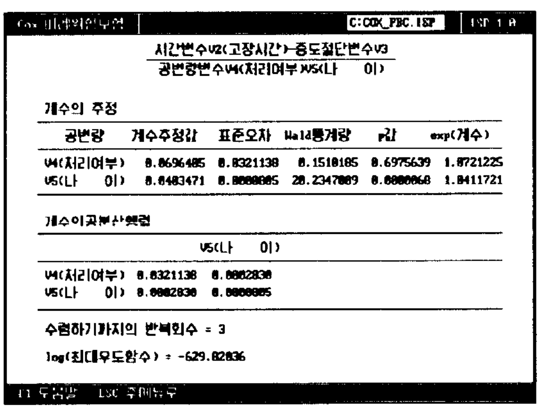
Kaplan-Meier 추정량과 관련하여 앞으로 개발되어야 할 부분은, 생존함수의 신뢰대 (Hall and Wellner (1980) 참조) 및 생존함수의 분위수에 대한 신뢰구간 (Wang and Hettmansperger (1990), Kim (1993) 참조) 등이 있다.

3.2 Cox 비례위험모형

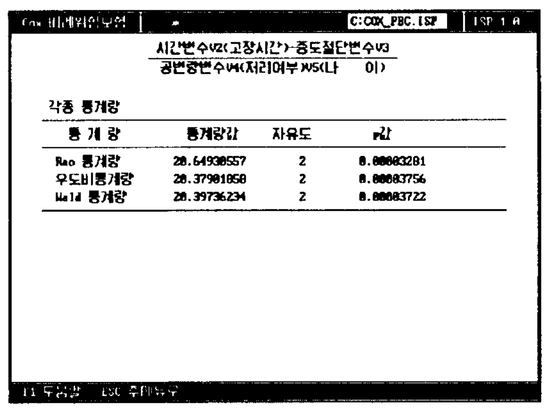
중도절단자료에 관한 회귀분석 모형으로 가장 널리 쓰이는 것이 Cox의 비례위험모형이다. 비례위험모형이란 공변량 $\underline{x}$를 가진 실험대상의 고장률 $\lambda(t;\underline{x})$에 대한 모형으로

$$\lambda(t;\underline{x}) = e^{\underline{\beta}'\underline{x}}\lambda_0(t)$$

을 가정한다. 즉, $\lambda(t;\underline{x})$는 공변량의 값에 상관없는 공통적인 고장률 $\lambda_0(t)$에 비례하며, 이 때 비례상수는 공변량의 값에 의해 결정된다. Cox(1972)는 위와 같은 비례위험모형을 가정하였을 때 모수 $\underline{\beta}$에 대한 추정량으로서, 이른바 부분우도함수(partial likelihood)를 최대화하는 추정량을 쓸 것을 제안하였다. 본 패키지에서는 <그림 3.3>에서와 같이, $\underline{\beta}$에 대한 Cox의 추정치와 함께, 각 공변량의 유의성 검정을 위해 p값을 출력하였다. p값 계산을 위해서, 각 회귀계수의 최우추정량의 성질을 이용한 Wald 통계량을 사용하였다. 한편 귀무가설 $H_0:\underline{\beta}=\underline{0}$ 도 검정할 수 있도록 하였다. 검정통계량으로는 Wald 통계량 외에, Score 통계량과 관련된 Rao 통계량, 우도비 통계량 등 세 통계량을 모두 구하여 p값과 함께 출력하였다. (<그림 3.4> 참조.)



<그림 3.3>



<그림 3.4>

$\underline{\beta}$에 대한 최우추정치를 구하기 위해 반복적 계산방법인 Newton-Raphson 방법을 적용하였으며, 수렴과정을 선택적으로 볼 수 있도록 하였다. 수렴성 점검을 위해 로그우도함수 값의 변화(자동설정값 10^{-6})와 회귀계수 값의 상대적 변화(자동설정값 10^{-4})를 동시에 고려하였으며, 이 값들을 최대반복횟수(자동설정값 10)와 함께 선택적으로 변경할 수 있도록 하였다.

Cox의 비례위험모형과 관련하여, 본 프로그램에서는 공변량의 $\underline{x}$의 값이 시간에 무관한 경우만 고려하였으나, 공변량 값이 시간에 따라 달라지는 모형(time dependent covariates)도 구현하여야

할 것이다. 그리고 비례위험모형의 적합성에 대한 진단 또는 검정도 앞으로 개발되어야 할 부분이다.

4. 베이지안분석

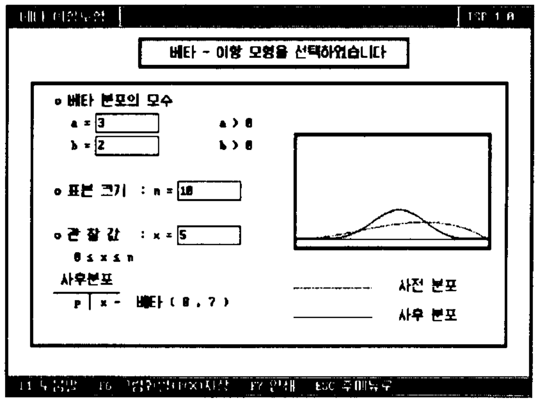
베이지안 분석 모듈은 아직까지 SAS나 SPSS 등에 없는, 본 패키지에서 독창적으로 설계를 한 것으로서 사전분포와 우도함수를 이용하여 사후분포를 계산하는 것이다. 베이지안 사후분포는 베이즈 정리를 사용하여 모수 θ에 대한 사전분포함수 $P(\theta)$와 우도함수 $P(x \mid \theta)$에 대해서 사후분포함수 $P(\theta \mid x)$를 계산한다. 사후분포는 부여된 사전분포가 conjugate prior인지 일반적인 사전분포인지에 따라 두 경우로 나누어서 모듈을 개발하였는데, 계산 알고리즘 개발에 참고한 문헌은 DeGroot [7]와 Press [13]이다.

4.1. Conjugate Prior의 이용

Conjugate Prior를 사전분포로 사용하는 경우에는 사후분포도 사전분포와 같은 family의 분포로서 직접적인 적분/합산의 필요없이 분포모수의 수정만으로 사후분포를 구할 수 있다. Conjugate Prior는 우도모형에 따라 다르고 사후분포의 유도에 필요한 충분통계량도 이 우도모형에 의해 결정되므로, 각 우도모형별로 Conjugate Prior의 모수값과 충분통계량의 자료를 입력받아서 사후분포의 모수값을 계산한다. 'Conjugate Prior' 명령을 선택하여 수행할 수 있는 우도모형은 다음과 같다.

(사전 - 우도)
[1] 이항모형 : 베타 - 이항
[2] 포아송모형 : 감마 - 포아송
[3] 음이항모형 : 베타 - 음이항
[4] 지수모형 : 감마 - 지수
[5] 정규모형(1) : 정규 - 정규 (σ를 아는 경우, μ의 분포)
[6] 정규모형(2) : 감마 - 정규 (μ를 아는 경우 , $1/\sigma^2$의 분포)
[7] 다항모형 : Dirichlet - 다항.

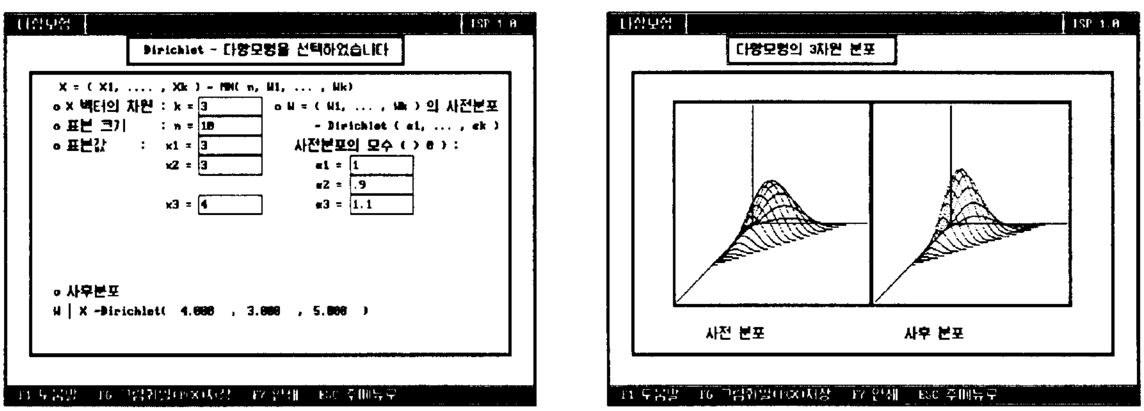
각 경우 모두 Conjugate Prior를 제시하고 그 모수값과 필요한 통계량의 값의 입력을 요구하여 사후분포를 구하고, 사전과 사후 분포의 그래프를 같은 축에 보여준다. <그림 4.1>에는 '베타-이항'모형에 대한 화면이 나타나 있다.



<그림 4.1> 베타-이항 모형

즉, 이항분포를 따르는 우도모형에서 성공확률 p에 대한 사전분포로 베타분포를 부여하고 n개의 표본으로부터 x개의 성공을 관찰했을 때 p의 사후확률은 역시 베타분포가 되는데 그 베타분포의 모수값을 출력으로 보여준다. 그래프에는 p의 사전 및 사후분포의 밀도함수를 같은 축에 보여줌으로써 표본정보의 증가에 의한 사후분포 곡선의 첨예함을 보여준다.

<그림 4.2>는 'Dirichlet-다항' 모형에 대한 화면을 보여준다. 역시 $\underline{w}=(w_1,\cdot\cdot,w_k)$ 의 Dirichlet 사전분포에 대한 정보와 표본으로부터의 충분통계량을 입력하면 사후분포의 모수값을 보여준다. k=3의 경우에 한해서 $\underline{w}$백타의 사전, 사후분포의 그래프를 나타내 주도록 하였다.



<그림 4.2> Dirichlet-다항 모형

4.2 임의 모형

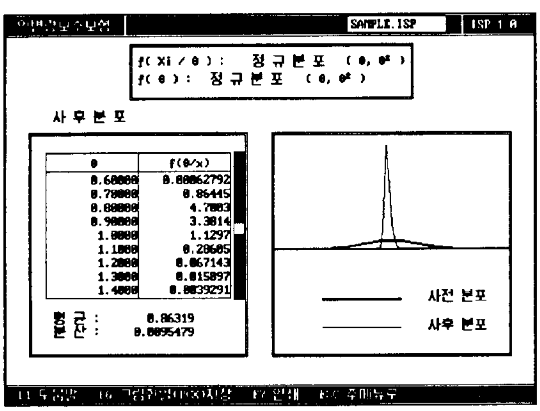
사전분포가 Conjugate Prior가 아닌 경우에는 간단한 분포모수의 수정만으로 사후분포를 구할 수 없다. 일반적인 경우의 사후분포는 θ가 연속형인지 이산형인지에 따라 다음의 베이즈 정리를

이용한 식에 의해서 구해진다.

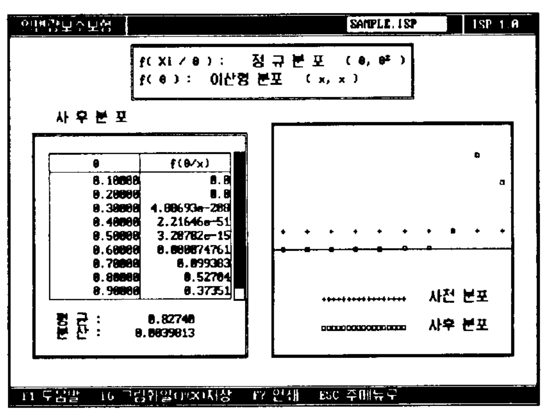
$$f(\theta \mid x) = \frac{f(\theta)\prod_{i=1}^{n} f(x_i \mid \theta)}{\int f(\theta)\prod_{i=1}^{n} f(x_i \mid \theta)d\theta}, \qquad \theta\text{연속}$$

$$f(\theta \mid x) = \frac{f(\theta)\prod_{i=1}^{n} f(x_i \mid \theta)}{\sum_{\theta} f(\theta)\prod_{i=1}^{n} f(x_i \mid \theta)}, \qquad \theta\text{이산}$$

현재는 일변량 θ의 경우만 다루는데, θ의 사전분포가 연속형인 경우와 이산형인 경우로 나누어진다. 두 경우 모두 선택된 우도모형의 확률함수가 표시되고, 그 모형의 모수(들)을 θ의 식 또는 상수로 입력해야 하며 모수들 중 최소한 한개는 θ의 식으로 표현되어야 한다. 이 기능은 우도함수의 여러 개의 모수가 공통적으로 θ의 영향을 받거나, 또는 우도함수의 모수가 θ의 일정한 식으로 표시될 경우 등에도 적용될 수 있도록 한 것이다. <그림 4.3>은 X_i가 Normal(θ,θ^2) 의 분포를 따르고 θ의 사전분포가 Normal(1,1)인 경우에, n=10개의 표본으로부터 0.7, 1.4, 1.1, 2.1, 1.5, 0.2, -0.1, 0.9, 0.6, 1.2 의 자료를 얻었을 때 θ의 사후분포를 구하는 화면이다. 사후분포의 계산에 필요한 수치적분에는 Simpson의 공식을 사용하는데, 적분구간은 $\mu \pm 4\sigma$로 자동적으로 계산되고 - 단, $\theta > 0$인 경우에는 $(0, \mu+4\sigma)$ 의 범위 - 적분 소구간의 갯수는 사용자가 결정하도록 하였다. 계산된 사후확률분포의 밀도함수 값은 표와 그래프로 화면에 보여주고 또한 시스템에 저장되도록 하여 차후의 분석에서 사전분포의 입력 등으로 사용할 수 있도록 하였다. <그림 4.4>는 이산형 θ에 대한 화면이다.



<그림 4.3> 임의모형 - 연속형 모수



<그림 4.4> 임의모형 - 이산형 모수

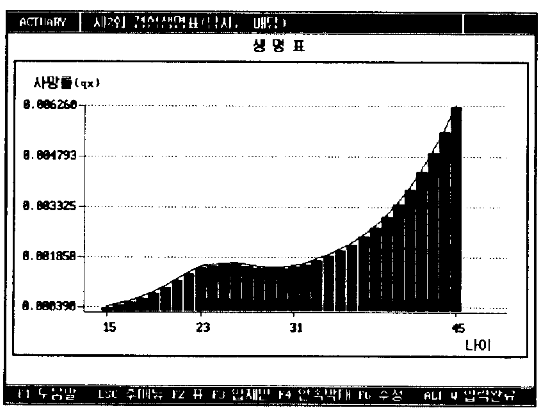
5. 보험통계

보험통계 모듈도 아직까지 SAS나 SPSS 등에 없는, 본 패키지에서 독창적으로 설계를 한 것으로서 경험생명표, 순보험료, 책임준비금 등을 표와 그래프를 이용하여 효율적으로 보여 준다. 계산 알고리즘 개발에 참고한 문헌은 Bowers [4]와 Kellison [9]이다.

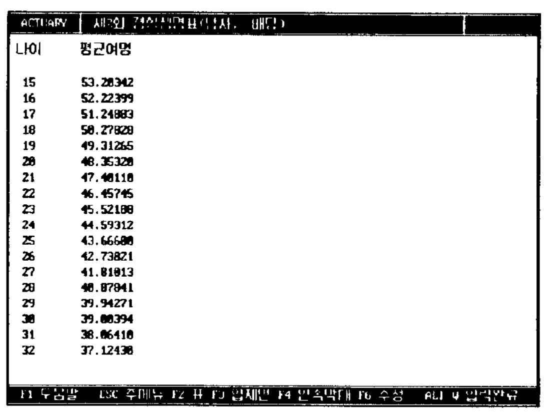
5.1 경험생명표

계약자 생존기간의 확률분포를 표현하기 위해 일반적으로 사용되는 것이 생명표이다. 생명표는 과거의 경험자료를 이용하여 각 나이구간별로 계약자들이 나이구간 내에 사망할 확률과 생존할 확률을 추정한 표이다. 생명표는 일반적으로 인구조사를 통해 얻어지는 인구통계 자료와 각 보험사들의 계약자통계 자료를 이용하여 작성된다. 국민통계 자료를 이용하여 작성되는 생명표를 국민생명표라 하고 계약자통계 자료를 이용하여 작성되는 생명표를 경험생명표라 한다. 수명의 확률분포가 다를 것으로 판단되는 집단들에 대해서는 정확성을 위해 별도의 생명표가 작성된다. 예를 들어 남자와 여자의 수명 분포에는 확연한 차이가 있으므로 남·녀의 생명표는 별도로 작성된다.

경험생명표의 주요 구성항목으로는 현재나이 x세인 계약자가 앞으로 1년 사이에 사망할 확률 (q_x)와 생존할 확률(p_x), $l_x, d_x, \mathring{e}_x$ 등이다. l_x는 신생아 l_0명의 생존여부를 계속 관찰한다고 할 때 x세까지의 생존자 수의 기대값을 나타낸다. l_0명의 신생아중 x세에서 x+n세 사이의 사망자수의 기대값을 $_nd_x$로 표기하며 특히 n가 1인 경우 생략할 수 있으며 d_x = $_1d_x$ 로 표기한다. 또한 x세 계약자의 생존기간 T(x)의 기대값을 $\mathring{e}_x$로 표기하고 완전평균여명(Complete-Expectation-of-Lifetime)이라 부른다. 보험통계 모듈에서는 이러한 값들을 주어진 경험생명표를 이용하여 계산한 뒤 메뉴를 통해 선택된 나이구간에 대해 표나 그림의 형태로 보여준다.



<그림 5.1> 나이구간 [15, 45]세에 대한 q_x 값의 그림(제2회 경험생명표 이용)



나이	평균여명
15	53.20342
16	52.22399
17	51.24883
18	50.27820
19	49.31265
20	48.35320
21	47.40110
22	46.45745
23	45.52180
24	44.59312
25	43.66680
26	42.73821
27	41.81013
28	40.87841
29	39.94271
30	39.00394
31	38.06410
32	37.12430

<표 5.1> 나이구간 [15, 45]세에 대한 $\mathring{e}_x$ 값의 표(* 제2회 경험생명표 이용)

5.2 순보험료

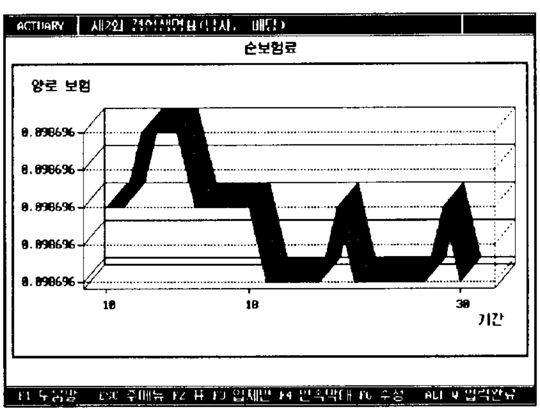
보험료는 순보험료(pure premium), 안전할증(safety loading), 각종 경비, 수수료(commission), 보험회사의 이익(profit)등의 요소로 구성되며 크게는 순보험료와 부가보험료로 나뉜다. 즉

$$(\text{영업})\text{보험료} = \text{순보험료} + \text{부가보험료}.$$

보험의 급부내용, 예정이율과 예정사망율만으로 보험회사가 계약자에게 지급할 것으로 기대되는 금액을 추정한 부분을 순보험료라 하며 나머지 부분을 부가보험료라 부른다. 보험회사의 입장에서 생각하면 납입되는 보험료는 수입이며 지급되는 보험금은 지출이다. 보험계약 시점을 기준으로 평가된 보험회사의 손실을 L로 표기하면

$$L = (\text{보험금의 현가}) - (\text{보험료의 현가})$$

이며 L은 보험금 지급시기와 보험료의 납입기간이 확률적으로 결정되므로 하나의 확률변수로 간주될 수 있다. 기대손실이 0이 되는 조건 즉, $E[L]=0$을 기준으로 순보험료를 결정하는 것을 수지상등의 원칙(Equivalence Principle)이라 하며 일반적으로 이 원칙에 따라 순보험료가 결정된다. 보험계약자의 입장에서는 납입하는 보험료가 지출이며 지급되는 보험금이 수입이므로 기대손실이 0이 되면 보험계약자의 입장에서도 수지상등의 원칙이 만족된다고 할 수 있다. 보험통계 모듈의 순보험료를 선택하면 여러가지 기본적인 보험종목에 대한 계약 조건을 입력할 수 있고 그러한 조건하에서 계약자의 연령이나 계약기간 등의 변화에 따라 순보험료를 계산하여 표나 그림의 형태로 보여준다.



<그림 5.2> 계약기간 [10, 30]년에 대한 30세 계약자, 10년 보험료납입, 예정이율 7.5%, 거치기간 10년인 양로보험의 연납순보험료의 그림 (* 제2회 경험생명표 이용)

ACTUARY

(나이 : 30)

기간	양로보험
10	0.09879
11	0.09879
12	0.09879
13	0.09879
14	0.09879
15	0.09879
16	0.09879
17	0.09879
18	0.09879
19	0.09879
20	0.09879
21	0.09879
22	0.09879
23	0.09879
24	0.09879
25	0.09879
26	0.09879
27	0.09879

<표 5.2> 계약기간 [10, 30]년에 대한 30세 계약자, 10년 보험료납입, 예정이율 7.5%, 거치기간 10년인 양로보험의 월납순보험료 값의 표 (* 제2회 경험생명표 이용)

5.3 책임준비금

수지상등의 원칙에 따라, 계약시점에서 평가된 피보험자가 납입할 순보험료와 피보험자에게 지급될 보험금의 현가의 기대값은 일치한다. 그러나 계약 후 시간이 경과한 후에는 일반적으로 이러한 수지상등의 상태가 더 이상 성립하지 않는다. 책임준비금은 계약 후 일정 시간이 경과한 후의 시점에서 앞으로 지급될 보험금의 현가와 납입될 보험료의 현가의 차이를 평가한 금액이다. 일반적인 보험상품은 피보험자가 계약을 해지하는 것 보다 계속 유지하는 것이 유리하도록, 책임준비금이 항상 양의 값이 되도록 설계된다. 보험회사는 미래의 수지불균형을 보전하기 위해 책임준비금만큼의 금액을 준비하여야 한다.

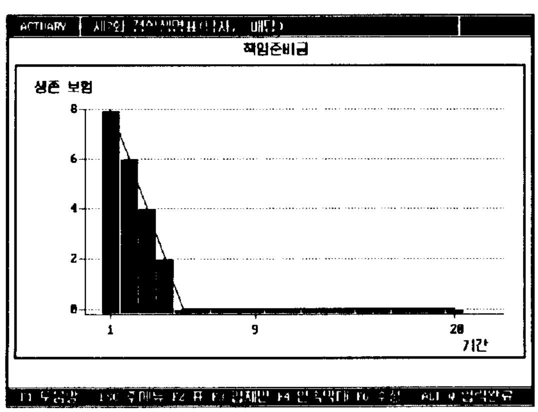
일반적으로 계약 후 t+U시기에 사망할 경우 지급되는 보험금을 B_{t+U}로, 계약 후 t+y시점에서의 납입되는 보험료의 연액을 π_{t+y}로 표기할 때 계약 후 t년이 경과한 후의 시점에서 그 후 U년 뒤에 피보험자가 사망하는 경우에 보험회사의 수지차 ${}_tL$은 아래와 같은 식으로 표현될 수 있을 것이다.

$$ {}_tL = B_{t+U} * v^U - \int_0^U \pi_{t+y} * v^y \, dy $$

따라서 계약 후 t년이 경과한 후의 시점에서 책임준비금 ${}_t\overline{V}$는 아래와 같이 구할 수 있다.

$$ \begin{aligned} {}_t\overline{V} &= E[\ {}_tL\ ] \\ &= \int_0^\infty \left(B_{t+u} * v^u - \int_0^u \pi_{t+y} * v^y \, dy \right) {}_u p_{t+y} \mu_{x+t+u} \, du \\ &= \int_0^\infty B_{t+u} * v^u * {}_u p_{t+y} \mu_{x+t+u} \, du - \int_0^\infty \pi_{t+y} * v^y \ {}_y p_{x+t} \, dy \end{aligned} $$

보험통계 모듈의 순보험료를 선택하면 여러가지 기본적인 보험종목에 대한 계약 조건을 입력할 수 있고 그러한 조건하에서 계약자의 연령이나 계약기간, 계약 후 경과기간 등의 변화에 따라 책임준비금을 계산하여 표나 그림의 형태로 보여준다.



<그림 5.3> 계약 후 경과기간 [0, 20]년에 대한 30세 계약자, 10년 보험료납입, 예정이율 7.5%, 거치기간 10년인 20년양로보험의 순보험료식 책임준비금의 그림 (* 제2회 경험생명표 이용)

6. 결론 및 향후과제

일반인을 대상으로 하는 통계패키지를 개발할 때 얻은 경험과 기술축적은 전문가용 통계패키지를 개발하는데 많은 시간과 노력을 단축시켜 주었다. 이제는 전문가들의 좋은 아이디어만 있으면 그것을 패키지로 구체화하는 작업은 그리 어려운 작업이 아닌 것 같다. 본 연구는 비슷한 분야의 통계전문가들이 합심하여 시스템이나 결과출력 화면에 대한 좋은 설계만 할 수 있다면 외국산 패키지보다 훨씬 우수한 통계패키지를 개발할 수 있는 가능성을 보여 주었다고 생각한다.

앞으로 하여야 할 일은 향후 PC 시스템의 발전에 맞게 WINDOWS용 소프트웨어로의 전환이 될 것이다. 이 시스템을 이용하면 다작업기능, 그래픽 사용자 환경(GUI: Graphical User Interface) 등 다양한 형태의 바람직한 통계패키지를 개발할 수 있을 것이다.

참 고 문 헌

[1] 안현순 (1992). 「터보 C로 구현한 과학기술계산 프로그래밍」, 가남사.

[2] 이정진, 강근석, 이윤오 (1992). 통계학 교육용 한글소프트웨어 개발연구, 「응용통계연구」, 제5권, 제1호, 81-92.

[3] 이정진, 강근석 (1994). 한국형 통계패키지 개발연구, 「응용통계연구」, 제7권, 제2호, 279-288.

[4] Bowers, N.L. et. al. (1986). *Actuarial Mathematics*, Society of Actuaries.

[5] Cooke, D., Craven, A.H., Clarke, G.M. (1989). *Basic Statistical Computing*, Arnold.

[6] Cox, D.R. (1972). Regression models and life tables (with discussion). *Journal of Royal Statistical Society B*, 34, 187-220.

[7] DeGroot, M.H. (1970). *Optimal Statistical Decisions*, McGraw-Hill.

[8] Hall, W.J. and Wellner, J.A. (1980). Confidence Bands for a Survival Curve from Censored Data. *Biometrika*, Vol. 67, 133-143.

[9] Kellison, S.G. (1991). *The Theory of Interest*, IRWIN.

[10] Kennedy, W.J. and Gentle, J.E. (1980). *Statistical Computing*, Dekker.

[11] Kim, J. (1993). Two Sample Inference for Quantiles based on Bootstrap for Censored Survival Data, *Journal of the Korean Statistical Society*, Vol. 22, 159-169.

[12] Maindonald, J.H. (1984). *Statistical Computation*, John Wiley & Sons.

[13] Press, S.J. (1989). *Bayesian Statistics: Principles, Models, and Applications,* John Wiley & Sons.

[14] SAS Institute Inc. (1985). *SAS User's Guide: Statistics*, 5th Edition.

[15] SPSS Inc. (1988). *SPSS-X User's Guide*, 3rd Edition.

[16] Wang, J. and Hettmansperger, T.P. (1990). Two-Sample Inference for Median Survival Times Based on One-Sample Procedures for Censored Survival Data, *Journal of the American Statistical Association*, Vol. 85, 529-536.