

다범주 자료의 다항로지트 모형과 로지스틱 회귀모형 비교-장애연금 특성분석 중심으로

김미정¹

¹연세대학교 수리과학연구소

(2008년 6월 접수, 2008년 6월 채택)

요약

순위형 다범주 자료에 있어서 범주값의 증감에 대한 설명변수의 특성분석을 위하여 다항로지트모형을 적합하여 분석하고 로지스틱 회귀모형을 적합하여 분석한 결과와 비교하였다. 이를 통하여 장애연금 수급자자료의 재정추계를 위해 필요한 일곱 가지 요인인 성별, 수급나이, 가입기간, 가입종별, 소득활동여부, 소득수준, 장애원인이 장애등급에 미치는 영향을 파악하였다. 일곱 요인 모두 장애등급에 대한 연관성이 있음을 확인하였고 이 가운데 다섯 요인은 장애등급의 증감에 있어서도 일정한 추세를 보였으나, 장애원인과 소득수준은 장애등급의 증감에는 일정한 추세를 보이지 않음을 확인하였다. 본 연구의 결과는 장애연금 관리방안을 모색하는데 있어서 장애등급에 따른 설명 요인의 특성을 반영하는데 필요한 가이드라인을 제공할 수 있을 것으로 기대한다. 장애등급 분류에 있어서 다중분류의 정분류율은 각각 42.56%와 42.43%로 로지스틱 회귀모형의 경우 다중로지트 모형의 경우보다 다소 높았지만 거의 비슷한 정확도를 보였다.

주요용어: 장애연금, 장애등급, 다항로지트, 로지스틱 회귀모형, 다중분류.

1. 서론

소득보장제도는 주로 단기 또는 장기 실업과 관련한 위험을 완화시키고 사회불안을 사전에 예방하며 소득 재분배를 통해 공평한 사회로 가기 위한 제도이다 (Moroney, 1991). 즉, 각종 사회적 위험에 의하여 소득원이 상실되었을 때 상실된 소득의 일부 혹은 전부를 보상하여 주는 것이라고 할 수 있다.

소득상실의 위험으로는 노령, 장애, 사망 등으로 인한 근로능력의 영구적 상실, 질병이나 사고에 의한 근로능력의 일시적인 상실로 구분될 수 있다.

이러한 관점에서 국민연금제도는 위에서 언급한 노령, 장애, 사망으로 발생하는 소득상실에 대해서 기본적인 보장이 가능하도록 설계되어 있다. 이 가운데 장애에 대한 소득보장 체계는 실업, 빈곤, 노령, 질병 등 일반적인 소득상실 위험에 대응하는 보편적인 소득보장체계의 일부로서 장애라는 독립적인 소득상실 위험 요인에 직면해 있는 장애인을 대상으로 한 특수한 소득보장체계라고 할 수 있다 (변용찬 외, 2004).

국민 연금 가운데 장애연금은 장애로 인하여 소득활동이 어려워진 가입자들을 대상으로 장애 정도에 따라 장애등급을 차등화하여 연금을 지급하고 있다. 하지만 장애연금에 있어서 지급기준(장애등급)의 조

본 연구의 자료는 국민연금관리공단의 공동연구 자료임.

¹(120-749) 서울시 서대문구 신촌동 134, 연세대학교, 수리 과학 연구소, 연구교수.

E-mail: mjkim@yonsei.ac.kr

표 2.1. 장애연금 지급현황(2006)

구분	계	장애연금				장애일시금 보상 (4급)
		소개	1급	2급	3급	
수급자수 (%)	66,660 (100%)	61,762 (93.3)	9,199 (13.8)	21,888 (32.84)	30,675 (46.02)	4,898 (7.35)
평균 지급액(원)	394,017	347,450	466,090	374,299	292,714	11,744,518

자료: 국민연금관리공단, 2006.

정이 필요하며 이를 반영한 장애연금 정책이 요구된다는 보고가 있다 (김성숙, 2002; 오근식, 2001; 원종욱, 2002).

본 연구는 장애연금 관리를 위한 장애연금의 특성을 파악함에 있어서 장애연금의 수급자 자료를 근거로 장애연금 수급자들의 장애등급에 따른 요인별 특성에 대한 객관적인 근거를 파악하기 위하여 자료분석을 실시하였다. 분석 자료는 1988년 1월부터 2007년 5월까지의 국민연금 가입자가운데 장애연금 수급자들의 자료로, 반응 변수는 장애등급이며 연금 재정추계 모형구축을 위해 필요한 수급나이, 성별, 소득활동 여부, 가입기간, 소득수준, 가입종별, 장애원인의 7개 요인가운데 장애등급에 영향을 미치는 요인들을 파악하고 이들 요인이 장애등급에 미치는 효과를 파악하였다. 다항로지 분석을 통하여 위에서 언급한 7가지 요인이 장애등급에 미치는 영향을 파악하였으며, 세 개의 장애등급 각각에 대하여 로지스틱 회귀모형을 적합하여 분석한 결과와 비교하였다. 이러한 분석의 결과는 장애연금관리방안을 고려함에 있어서 장애등급에 따른 설명요인들의 특성을 반영함으로써 보다 실질적인 방안을 제고하는데 필요한 자료로 사용될 수 있을 것으로 기대한다.

논문의 구성은 다음과 같다. 2장에서는 연구의 목적과 방법에 대하여 논의하였고 3장에서는 장애등급에 따른 수급자들의 특성을 파악하고 두 방법을 비교하였으며, 4장에서는 결론을 서술하였다.

2. 연구의 목적 및 방법

2.1. 장애연금의 현황

국민연금 관리공단에서 제공하는 2006년 통계연보에 따르면 장애연금을 수급하는 대상자는 총 61,762명이고, 장애일시보상금 수령자는 4,898명으로서 총 66,660명이다. 표 2.1은 장애등급에 따른 장애연금의 지급현황이다.

각 등급에 따른 비율을 살펴보면 1급은 13.8%, 2급은 32.84%, 3급은 46.02%로 3급이 가장 많은 비율을 차지하고 있다.

국민 연금의 장애연금 수급권은 국민연금 가입자 중, 국민연금 가입기간에 발생한 질병 또는 부상으로 인한 장애에 한하고 있다. 장애 4급인 경우에는 일시금을 지급받게 되며 등급에 따라 지급되는 금액이 다른데, 장애등급이 높을수록 많은 금액을 받으며 2006년 지급된 평균의 경우 1급은 466,090원을 2급은 374,299원을 그리고 3급은 292,714원을 지급받았다.

국민연금의 장애연금에 있어서 장애등급은 장애인복지법상의 장애의 기준과는 다르게 등급이 결정되며(부록 참조) 장애연금액은 기본금액과 가급연금액으로 구성되는데, 그 수준은 평균소득 및 장애등급에 따라 결정 된다 (권선진과 정경배 2005).

2.2. 연구 목적 및 방법

본 연구에서는 장애연금의 특성을 파악하기 위하여 국민연금이 도입된 시점인 1988년 1월부터 2007년 5월 현재까지의 자료를 근거로 장애등급에 따른 장애연금 수급자들의 특성을 파악하되, 국민연금의 재정추계를 위해 관찰하는 7개의 요인과 장애등급의 관계를 파악하고 이에 따라 장애등급에 따른 수급자의 특성을 파악함으로써 장애연금관리에 있어서 보다 현실적인 방안을 제시하는데 있어서 필요한 근거를 제시하고자 한다. 분석 자료는 국민연금의 장애연금 수급자 80,699명 중 결측값을 제거한 79,621명의 자료를 분석대상으로 하였다. 반응변수는 장애등급이며, 장애연금 수급대상이 되는 1등급, 2등급, 3등급의 세 범주로 구성되어 있다. 3개 이상의 순위형 반응변수의 경우 비례오즈모형을 적용할 수 있는데 비례오즈란 각 범주에 속하는 odds들이 순위가 변함에 따라 동일하다는 뜻이다. 그러나 비례오즈의 가정에 대한 P -value가 < 0.001 로 비례오즈가설이 유의하지 않아 비례오즈모형을 적용시킬 수 없었다.

따라서 셋 이상의 범주를 갖는 명목형 반응변수에 대한 다항로짓 모형을 적용하기 위하여 SAS 9.1의 CATMOD 프로시저를 활용하였다. 다항로짓 모형에서 반응변수의 기준범주 대비 각 범주에 속하는 확률을 구할 수 있고 이를 통하여 각 범주의 특성을 파악할 수 있다 (Maddala, 1983). 본 분석에서는 장애등급의 중앙값인 등급 2를 기준범주로 두어 장애등급의 증감에 따른 수급자들의 특성을 파악하였는데, CATMOD 프로시저의 경우, 기준범주의 값을 다른 범주들의 값보다 큰 값으로 할당함으로써 이러한 분석이 가능했다. 따라서 1등급은 0으로 2등급은 3으로 3등급은 1로 주었다.

장애등급의 증감에 따른 설명요인의 특성을 파악하기 위한 방법으로 다항로짓 모형을 적용하는 방법 외에 장애등급과 7개 요인과의 관계를 각 장애등급에 대하여 파악하기 위하여 로지스틱 회귀모형을 적용한 결과도 비교하였다. 전자의 경우 설명요인과 장애등급의 연관성뿐만 아니라 설명요인의 ‘일정한’ 특성이 장애등급의 증감에 반영되고 있는지 여부를 파악하는 것이 가능하고 후자의 경우 타 범주를 대비한 각 범주 고유의 특성을 파악하는 것이 가능할 뿐만 아니라 결과에 대한 해석이 용이하며 분석을 위해 요구되는 시간이 다항로짓 모형의 경우보다 적게 소요될 뿐만 아니라 각 범주에 대한 사후확률을 이용한 범주의 판별이 비교적 용이하다. 본 자료의 경우 로지스틱 모형을 통한 등급의 판별은 다항로짓 모형을 적용한 경우와 유사한 값을 보였다.

분석에 사용된 설명요인들은 수급나이, 성별, 소득활동 여부, 가입기간, 소득수준, 가입종별, 장애인이며 수급나이, 가입기간 그리고 소득수준은 연속형 변수이고, 소득활동여부, 가입종별, 성별 그리고 나이는 범주형 변수이다. 가입기간은 가입자가 연금가입 후 연금이 수급될 때까지의 기간을 개월 수로 나타냈으며, 소득수준은 소득이 있는 경우의 표준보수액을 기준으로 하여 1부터 45등급의 소득수준을 나타낸다. 1988년 1월 이후부터 95년 4월까지는 53등급으로 나누던 것을 95년 4월 이후에는 45등급으로 조정되어 분석에서는 95년 4월까지의 등급을 45로 보정하여 평균 소득수준을 계산하였다. 보정할 때 사용된 방법은 53등급으로 나누었을 때 53등급과 45등으로 나누었을 때의 45등급이 동일한 등급임을 가정하였으며 linear interpolation을 하여 45등급으로 맞추었다. 수급나이의 경우 수급이 시작된 나이를 의미한다.

7개의 설명변수를 적용한 다항로짓 모형의 일반식은 다음과 같다.

$$\log \left(\frac{\pi_i}{\pi_2} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7, \quad i = 1, 3,$$

여기서 π_i 는 i 범주에 속하는 확률을 나타내며 $\log(\pi_i/\pi_2)$ 는 반응이 i 범주나 기준범주에서 일어났다는 조건하에 반응이 i 범주일 로그 오즈(log odds)이다.

로지스틱 회귀모형의 경우 장애등급에 따라 세 개의 로지스틱 회귀모형을 세웠는데, 예컨대, 장애등급이 1등급인 경우 반응변수의 값을 1로 타범주에 속한 경우 0으로 두었다. 7개의 설명변수를 적용한 로지스

표 3.1. 연속형 독립변수의 기술통계

변수	N	평균값	표준편차	최소값	최대값
가입기간	79621	72.22	48.36	1	232
수급나이	79621	47.09	10.08	18	73
소득수준	79621	23.04	8.46	1	45

표 3.2. 명목변수의 분포

변수	범주	빈도	백분율(%)
가입종별	사업장가입= 0	32215	40.46
	지역가입= 1	43351	54.45
	임의가입= 2	4055	5.09
성별	남성= 1	70279	88.27
	여성= 2	9342	11.73
소득활동여부	취득= 0	43009	54.02
	납부예외= 1	19216	24.13
	대기= 2	17396	21.85
장애원인	질병= 1	53421	67.09
	부상= 0	26200	32.91
장애등급	1등급= 1	13854	17.40
	2등급= 3	29616	37.20
	3등급= 2	36151	45.40

틱 회귀모형식은 다음과 같다.

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_5x_5 + \beta_6x_6 + \beta_7x_7, \quad i = 1, 2, 3,$$

여기서 π_i 는 i 범주에 속하는 확률을 나타내며 $\pi_i/(1-\pi_i)$ 는 설명변수 x 값들이 주어졌을 때 타범주에 속할 확률 대비 각 범주에 속할 확률이다.

3. 장애연금 수급자 분석

3.1. 기술통계

본 연구에 사용된 연속형 변수의 평균, 표준편차, 최저, 최대는 표 3.1과 같다. 가입기간의 평균은 72.2개월이었으며, 수급나이는 47세였다. 또한 소득수준의 경우는 23등급이었다.

한편, 범주형 변수의 분포는 표 3.2에 제시하였다. 가입종별은 사업장가입은 근로소득이 있는 직장 가입자로 40.46%를 차지하고 있으며, 지역가입이 54.45%를 차지하고, 임의가입은 사업장가입과 지역가입을 제외한 경우로 5.09%의 가장 낮은 분포를 보이고 있다. 또한 성별의 경우 남성이 88.27%이고 여성의 경우 11.73%를 차지하고 있어 남성이 여성에 비해 장애연금을 수급하는 사람이 많은 것을 알 수 있다.

또한 소득활동여부의 경우 취득은 소득활동이 있는 국민연금 가입자의 경우로 전체 54.02%를 차지하고 있어 절반이상이 장애연금 수급 전 소득이 있었음을 알 수 있으며, 납부 예외는 현재 국민연금 가입자 범주에 속하지만 소득활동이 없어서 국민연금 납부가 일정기간 유예된 자로서 24.13%를 대기는 국민연금 가입이력을 가지고 있으나 현재 소득활동이 없는 관계로 현 가입자 범주에 해당되지 않는 자이면서 수급 발생이 일어나지 않은 자로 21.85%를 차지하는 것을 볼 수 있다. 마지막으로 장애원인이 질병인

표 3.3. 장애 2급을 기준으로 각 장애등급에 대한 추정치

		장애 2급 관계					
		장애 1급			장애 3급		
		계수1	표준오차	p-value	계수2	표준오차	p-value
상수항		-1.05460	0.06700	< .0001	0.12130	0.051600	-
가입기간		-0.00023	0.00025	0.3656	0.00311	0.000189	< .0001
수급나이		0.01080	0.00119	< .0001	-0.00028	0.000926	0.7601
소득수준		0.00378	0.00144	0.0087	0.00436	0.001120	0.0001
성별	0	0.01520	0.01680	0.3656	-0.12650	0.012800	< .0001
소득활동여부	0	-0.34560	0.01430	< .0001	0.12860	0.011100	< .0001
	1	0.11940	0.01760	< .0001	-0.00371	0.014500	0.7982
가입종별	0	-0.10230	0.02190	< .0001	0.05840	0.017800	0.0010
	1	-0.03830	0.01930	0.0476	0.00934	0.015800	0.5540
장애원인	1	-0.38540	0.01260	< .0001	-0.55280	0.009790	< .0001

경우가 전체의 67.09%를 부상이 32.91%를 차지하고 있음을 확인할 수 있다. 종속변수인 장애등급에서는 1등급이 17.40%이었으며, 2등급은 47.20% 그리고 3등급은 45.40%로 앞에서 살펴보았던 2006년의 장애연금 수급현황과 마찬가지로 3등급에서 빈도가 가장 높았다.

본 연구에서는 다항로짓 분석을 이용하여 2등급을 기준으로 장애등급이 높을수록(혹은 낮을수록) 요인들이 각 모형에 미치는 영향을 파악하고 기준범주(2등급)를 대비하여 그 특징을 파악하였다.

3.2. 다항로짓 모형

다항로짓에서는 범주의 수에 따라 모형이 결정된다. 본 자료의 장애등급과 같이 범주가 3개인 경우, 장애 2급을 기준범주로 하여 장애1급/장애2급, 장애3급/장애2급에 대한 두 모델을 세울 수 있다. 이렇게 세워진 모델을 통하여 설명요인이 장애등급에 영향을 미치는 정도를 파악할 수 있으며, 이를 통하여 장애등급의 증감에 따른 요인의 특성을 살펴보는 것이 가능하다.

분석 결과는 표 3.3에 제시하였으며 의사 결정은 유의수준 0.05에서 이루어졌다.

장애 2급 대비 장애 1급의 비교 결과는 첫 번째 열의 계수1에 제시되어 있으며 장애 2급을 대비한 장애 1급에 영향을 미치는 유의한 요인으로는 수급나이, 소득수준, 소득활동여부, 가입종별 그리고 장애원인으로 파악되었다. 이를 바탕으로 추정된 회귀식은 다음과 같다.

$$\log\left(\frac{\pi_{1\text{등급}}}{\pi_{2\text{등급}}}\right) = -1.0546 - 0.00023\text{가입기간} + 0.0108\text{수급나이} + 0.00378\text{소득수준} \\ + 0.0152S - 0.3456C1 + 0.1194C2 - 0.1023J1 - 0.0383J2 - 0.3854R,$$

여기서 J1은 가입종별이 사업장 가입일 경우 1 그렇지 않을 경우 0을 J2는 가입종별이 지역가입일 경우 1 그렇지 않을 경우 0을 나타낸다. C1은 소득활동여부가 취업일 경우 1 그렇지 않을 경우 0을 그리고 C2는 납부예외 일 경우 1을 그렇지 않을 경우 0을 나타낸다. S는 성별이 남성일 경우 1을 여성인 경우 0을 그리고 R은 장애원인이 질병일 경우 1을 부상일 경우 0을 나타낸다.

장애 2급 대비 장애 3급을 비교한 결과는 넷째 열의 계수 2에 제시되어 있으며 장애 2급 대비 장애 3급에 대해 영향을 미치는 유의한 요인으로는 가입기간, 소득수준, 성별, 소득활동여부, 가입종별 그리고

표 3.4. ODDS RATIO와 그 95% 신뢰구간

		장애 2급과의 관계					
		장애 1급			장애 3급		
		odd ratio	95% 하한	95% 상한	odd ratio	95% 하한	95% 상한
가입기간		0.9997	0.9990	1.00026	1.0031	1.0027	1.0034
수급나이		1.0108	1.0085	1.0132	0.9997	0.9979	1.0015
소득수준		1.0037	1.0009	1.0066	1.0043	1.0021	1.0065
성별	1 vs. 2	1.0153	0.9824	1.0493	0.8811	0.8593	0.9035
소득활동여부	0 vs. 2	0.7077	0.6882	0.7279	1.1372	1.1127	1.1622
	1 vs. 2	1.1268	1.0886	1.1663	0.9962	0.9683	1.0250
가입종별	0 vs. 2	0.9027	0.8648	0.9423	1.0601	1.0237	1.0977
	1 vs. 2	0.9624	0.9266	0.9995	1.0093	0.9786	1.0411
장애원인	1 vs. 2	0.6801	0.6635	0.6971	0.5753	0.5644	0.5864

장애원인으로 나타났다. 이를 바탕으로 추정된 회귀식은 다음과 같다.

$$\log\left(\frac{\pi_{3\text{등급}}}{\pi_{2\text{등급}}}\right) = 0.1213 + 0.00311\text{가입기간} - 0.00028\text{수급나이} + 0.00436\text{소득수준} \\ - 0.1265S + 0.1286C1 - 0.00371C2 + 0.0584J1 + 0.00934J2 - 0.5528R.$$

계수 1과 계수 2값이 서로 반대의 부호를 갖는 요인들은 기준범주인 장애 2급을 대비하였을 때 장애 1급과 장애 3급이 서로 상반되는 영향을 주고 있다는 의미이며, 이는 요인에 따라 장애등급이 증가 또는 감소의 일정한 반응을 보임을 뜻한다. 즉, 이러한 요인들은 장애등급의 증감에 일정한 추세를 보이는 요인들이다. 따라서 표에서 살펴보는 바와 같이 가입기간, 성별, 소득활동여부, 가입종별 그리고 수급나이는 각 요인의 범주에 따른 장애등급의 차이에 있어서 요인의 범주에 따른 장애등급의 차이가 증가 또는 감소의 일정한 경향을 보이는 요인들임을 알 수 있고 이는 각 요인의 범주가 장애등급에 적절하게 반영되고 있음을 의미한다.

로지모형의 경우 Odds ratio는 설명변수 단위가 한 단위 증가할 때 장애 2등급 대비 장애 1등급 혹은 장애 2등급 대비 장애 3등급에 속할 가능성을 나타내며, 이러한 Odds ratio값을 살펴봄으로써 장애 2등급에 대비한 1등급 혹은 2등급 대비 3등급의 특성을 파악할 수 있다. 표 3.4에 추정된 회귀식에 대한 Odds ratio와 그에 대한 95% 신뢰구간을 제시하였다.

유의한 경우는 진하게 표시하였다. Odds Ratio가 1인 경우 설명변수는 종속 변수와 연관성이 없음을 의미하며, Odds Ratio 값이 1보다 큰 경우, 기준범주 대비 각 범주에 속할 가능성이 높은 반면에 1보다 작은 경우, 기준범주 대비 각 범주에 속할 가능성이 낮음을 의미한다.

95% 신뢰도에서 유의한 요인들의 Odds Ratio를 살펴보면 기준범주인 장애 2급에 대비하여 높은 장애 등급(1급)의 경우, 수급개시 연령이 높고, 소득수준이 높으며, 소득활동여부가 대기이거나 납부예외인 경우가 많았으며, 가입종별은 임의가입인 경우가 많았다. 장애원인으로는 부상일 가능성이 높았다. 기준범주 대비 낮은 장애등급(3급)에 대한 수급자의 특징은, 가입기간이 길고, 소득수준이 높으며, 성별이 여성이 많았고, 소득활동여부는 취득인 경우가 많으며, 가입종별은 사업장가입이 많고, 장애원인이 부상일 경우가 많음을 확인하였다.

장애등급의 상대적 특성으로부터 내릴 수 있는 결론은, 장애등급이 낮을수록(즉, 3등급으로 갈수록) 가입기간이 길어지고, 수급나이가 적었으며, 성별이 여성인 경우가 많았다. 그리고 소득활동여부가 취득

이거나 대기인 경우가 많아지며, 납부예외인 경우가 줄어든다. 가입종별은 사업장가입과 지역가입인 경우가 많아진다. 반면, 소득수준은 장애등급에 영향을 미치는 요인이기는 하나, 장애등급의 증감은 이에 비례하여 반응하지 않았다. 즉 소득수준의 경우 장애 2급 대비 1급에 속할 확률과 장애 2급 대비 3급에 속할 확률이 모두 1보다 큰 값을 보임으로써 장애등급의 증감은 소득수준의 증감에 따라 일정한 추세를 보이지는 않음을 확인하였다. 장애원인 또한 장애등급에 영향을 미치는 요인이지만 장애등급의 증감에 일정한 추세를 보이지 않고 있음을 알 수 있었다.

결론적으로, 소득수준과 장애원인을 제외한 다섯 개의 요인에 따른 장애등급의 증감은 일정한 추세를 보였다. 따라서 이들 다섯 요인에 대하여 장애등급별 증감의 비율을 반영한 연금관리가 이루어진다면 장애연금의 현실적인 현상을 반영한 연금관리가 이루어 질 수 있음을 알 수 있다. 즉, 분석결과로부터 얻은 이들 다섯 요인에 대한 odds ratio는 각 등급의 상대적 비교를 통하여 장애연금 관리방안을 고려할 경우 등급에 따른 차등을 어느 정도로 두어 정책을 수립해야 하는지를 판단하는 정보로 사용될 수 있다. 또한, 장애연금의 수급여부에 대한 예측이 먼저 이루어진 후 예측된 장애연금 수급대상자에 대하여 장애등급의 특성분석을 위한 모형을 통하여 장애등급에 대한 설명요인의 특성 파악이 가능하므로 이들 설명요인에 따른 등급 간 상대적 차등을 반영하여 연금지급액을 예측, 관리할 수 있다. 이와 더불어 장애연금 수급대상자들의 관리에 있어서도 대상자들의 설명요인의 특성에 따라 장애등급의 분류 및 관리가 가능하다. 이러한 제안은 실제 자료의 분석으로부터 얻은 결과에 근거하므로 연금관리에 있어서 현실반영이 제대로 이루어 질 수 있다는 장점을 가지고 있다.

한편, 소득수준과 장애원인에 따른 장애등급의 증감에 있어서는 일정한 추세를 보이지 않음을 확인하였다. 장애등급간 소득수준의 차이에 대하여는, 장애연금에 있어서 장애등급을 결정하는 기준으로 근로능력이 반영되고 있지 않다는 보고가 있는데 (권선진과 정경배, 2005), 근로 능력이 소득 수준과 연관이 있다고 본다면 본 자료 분석의 결과로부터 얻은 장애등급간 소득의 차이를 장애등급에 따른 연금 지급율에 반영하는 것을 고려해 보는 것도 제기된 문제에 대한 해결의 실마리가 될 수 있으리라 사료된다. 가령, 현 국민연금의 장애연금에서는 장애등급에 따라 연금의 지급율이 다른데, 장애등급에 따른 연금지급율의 결정에 있어서 앞의 분석 결과가 보여주는 장애등급 간 소득수준 차이(odds ratio)를 반영하는 방법을 고려해 볼 수 있다. 이렇게 함으로써, 장애등급에 따른 연금지급액에 현실적인 소득수준의 차이가 반영되고, 이는 장애등급의 결정에 근로능력이 간접적으로 반영되는 효과를 얻을 수 있을 것이라 판단하기 때문이다.

3.3. 로지스틱 회귀모형

다항로지트 모형을 이용하여 범주의 특성을 파악하는 경우 하나의 범주를 기준으로 하여 기준 범주 대비 각 범주의 상대적인 특성을 살펴볼 수 있는데, 타 범주들에 대한 각 범주 고유의 특성을 파악하기 위하여 본 절에서는 로지스틱 회귀분석을 통하여 각 장애등급에 따른 설명요인의 특성을 살펴보고자 한다.

우선 각각의 장애등급에 대한 특성을 파악하기 위해 세 개의 로지스틱 회귀모형을 세우되 해당 장애등급의 경우 반응변수 값은 1을 그렇지 않은 경우 0의 값을 갖도록 하였다. 예컨대, 장애 1등급에 대한 로지스틱 회귀모형의 경우 1등급에 속하면 반응변수의 값을 1로 2등급 혹은 3등급에 속한 경우 0으로 두었다. 따라서 예측된 사후확률 값이 높을수록 각 장애등급에 속할 가능성이 높다고 예측할 수 있다. stepwise 변수 선택법을 사용하였으며, 각 장애등급에 대한 로지스틱 회귀모형을 1 모형, 2 모형, 3 모형으로 명명하기로 한다.

표 3.5은 세 모형에 대해 로지스틱 회귀분석을 실시한 결과 도출된 회귀계수의 값과 p -value들이고 표 3.6은 이들의 odds 비이다.

표 3.5. 장애등급에 따른 세 개의 로지스틱 회귀모형의 변수 선택 결과 및 추정치

		장애 1급			장애 2급			장애 3급		
		계수1	표준오차	p-value	계수2	표준오차	p-value	계수3	표준 오차	p-value
상수항		-1.591	0.063	< .0001	-0.992	0.047	< .0001	0.177	0.050	0.0000
가입기간		-0.002	0.000	< .0001	-0.002	0.000	< .0001	0.003	0.000	< .0001
수급나이		0.011	0.001	< .0001	-0.003	0.001	0.001	-0.004	0.001	< .0001
소득수준		-	-	-	-0.005	0.001	< .0001	0.004	0.001	0.0000
성별	0	0.172	0.030	< .0001	0.171	0.024	< .0001	-0.267	0.023	< .0001
소득활동 여부	0	-0.708	0.023	< .0001	-	-	-	0.445	0.019	< .0001
	1	-0.169	0.028	< .0001	-	-	-	0.160	0.024	< .0001
가입종별	0	-0.085	0.024	0.0000	-	-	-	0.066	0.019	0.0010
	1	0.220	0.043	< .0001	-	-	-	-0.136	0.036	0.0000
장애원인	1	-0.122	0.022	< .0001	1.019	0.018	< .0001	-0.838	0.017	< .0001

표 3.6. 세 모형의 로지스틱 회귀분석 결과 ODDS RATIO

		장애 1급			장애 2급			장애 3급		
		odd ratio	95% 하한	95% 상한	odd ratio	95% 하한	95% 상한	odd ratio	95% 하한	95% 상한
가입기간		0.998	0.998	0.999	0.998	0.997	0.998	1.003	1.003	1.003
수급나이		1.011	1.009	1.013	0.997	0.996	0.999	0.996	0.995	0.998
소득수준		-	-	-	0.995	0.993	0.997	1.004	1.002	1.006
성별	1 vs. 2	1.188	1.120	1.260	1.187	1.133	1.243	0.766	0.731	0.802
소득활동 여부	0 vs. 2	0.493	0.471	0.515	-	-	-	1.560	1.503	1.619
	1 vs. 2	0.844	0.800	0.891	-	-	-	1.174	1.120	1.230
가입종별	0 vs. 2	0.918	0.877	0.962	-	-	-	0.766	0.731	0.802
	1 vs. 2	1.246	1.146	1.354	-	-	-	1.068	1.029	1.109
장애원인	1 vs. 2	0.885	0.848	0.924	2.770	2.673	2.870	0.433	0.418	0.447

표 3.6에 제시되어 있는 요인들 가운데 odds ratio의 95% 신뢰구간이 1을 포함하지 않는 요인들은 각 장애등급에 대해 유의한 요인들이다. 장애 1등급은 가입기간이 짧을수록, 수급개시연령이 길수록, 성별이 남성인 경우, 소득활동여부가 대기이고, 가입종별이 임의가입이거나 지역가입인 경우 그리고 장애원인이 부상인 경우가 이에 속할 가능성이 높았다. 장애 2등급은 가입기간이 짧을수록, 수급개시연령이 짧을수록, 소득수준이 낮을수록, 성별이 남자인 경우 그리고 장애원인이 부상인 경우가 이에 속할 가능성이 높았다. 장애 3등급은, 가입기간이 길수록, 수급개시연령이 낮을수록, 소득수준이 높을수록, 성별이 여성인 경우, 소득활동여부가 취득이거나 납부예외인 경우, 가입종별이 사업장가입이거나 임의가입인 경우 그리고 장애원인이 부상인 경우가 이에 속할 가능성이 높았다.

3.4. 두 방법의 비교

3.4.1. 두 분석 방법에 따른 설명요인의 특성 비교 3.2절과 3.3절에서는 장애등급에 따른 설명요인들의 특성을 살펴보기 위하여 다항로짓 모형과 로지스틱 회귀모형을 적합하여 분석한 결과를 유의한 설명요인 중심으로 살펴보았다. 본 절에서는 두 분석의 차이점을 살펴보기 위하여 두 가지 방법 각각의 경우 반응변수에 대한 7개 설명변수의 특성을 살펴보았으며 이로부터 얻은 결과를 표 3.7에 요약하였다.

표 3.7. 다항로짓 모형과 로지스틱 회귀모형에 따른 설명요인의 특성

장애등급	다항로짓 모형	장애등급	로지스틱 회귀모형
1 vs. 2	가입기간이 짧고, 수급개시연령이 높고, 소득수준이 높으며, 성별이 남성, 소득활동여부가 대기이거나 남부예외, 가입종별이 임의가입 그리고 장애원인이 부상인 경우가 2급에 비해 1급에 속할 가능성이 높아진다.	1	가입기간이 짧고, 수급개시연령이 많고, 소득수준이 높고, 성별이 남성이고, 소득활동여부가 대기이고, 가입종별이 임의가입이거나 지역가입인 경우 그리고 장애원인이 부상인 경우가 이에 속할 가능성이 높았다.
3 vs. 2	가입기간이 길고, 수급개시연령이 적고, 소득수준이 높으며, 성별이 여성인 경우, 소득활동여부가 취득이거나 대기인 경우, 가입종별이 사업장가입이거나 지역가입인 경우, 장애원인이 부상인 경우가 2급에 비해 3급에 속할 가능성이 높아진다.	2	가입기간이 짧고, 수급개시연령이 적고, 소득수준이 낮고, 가입종별이 사업장가입이거나 임의가입, 소득활동여부가 대기이며, 성별이 남자이고 그리고 장애원인이 부상인 경우인 특징을 가지고 있다.
		3	가입기간이 길고, 수급개시연령이 적고, 소득수준이 높고, 성별이 여성이고, 소득활동여부가 취득이거나 남부예외인 경우, 가입종별이 사업장가입이거나 임의가입인 경우 그리고 장애원인이 부상인 특징을 가지고 있다.

표 3.7은 두 모형에 따른 분석의 결과로부터 얻는 각 장애등급에 대한 설명변수의 특성이 유사함을 보여준다. 그러나 다항로짓 모형과 로지스틱 회귀모형은 분석하는 방법과 성격에 있어서 차이가 있다. 본 연구에서 적용한 다항로짓 분석의 경우 범주값의 중앙값(장애 2급)을 기준범주로 하여 각 등급의 상대적 특징을 살펴보았는데 이를 통하여 ‘소득수준’과 ‘장애원인’을 제외한 모든 설명요인이 장애등급의 증감에 따라 ‘일정한’ 특성을 보임을 파악할 수 있었고 이들에 대한 통계적 유의성을 검정하는 것이 가능했으며 뿐만 아니라, 각 등급의 상대적 비교를 통하여 장애연금관리방안을 고려할 경우 장애등급에 따라 차등을 둔 관리가 가능함도 보일 수 있었는데 이와 같은 사항은 비례오즈가정을 만족하지 않는 다범주 자료 분석에 있어서 범주값의 증감에 대해서 일정한 반응을 보이는 설명요인을 찾기 위한 시도로 범주값들의 중앙값을 기준범주로 하여 범주간 상대비교를 함으로써 가능하였다.

반면, 로지스틱 회귀분석의 경우, 타 등급에 대한 장애등급 고유의 특성을 살펴보는 것이 가능했으며, 로지스틱 회귀모형이 갖는 일반적인 장점으로 각 등급에 대한 분류 모델의 설정과 해석이 용이하였다. 본 연구의 일차적인 목적이 장애등급에 대하여 주어진 7가지 설명요인의 특성분석에 있고 따라서 장애등급의 분류를 위한 모형으로 7개의 설명요인이 제한적일 수는 있지만 다중분류에 있어서 두 방법을 적용하여 정확도를 비교한다면, 다음의 3.4.2절에서 보여주는 바와 같이 두 모형에 의한 장애등급 판별에 있어서 로지스틱 모형을 통한 장애등급의 판별은 다항로짓 모형에 의한 판별과 유사한 정분류율을 보임을 확인하였다. 그러나 로지스틱 회귀모형을 통하여 장애등급의 증감에 대한 설명요인의 일정한 특성을 설명하는 것이 가능하더라도 이에 대한 통계적 유의성을 보여주는 것은 어렵다.

3.4.2. 두 분석 방법에 따른 정확도의 비교 본 절에서는 다범주 반응자료에 대하여 다중분류(multicategory classification)를 함에 있어서 다항로짓 모형을 적용한 분류와 로지스틱 회귀모형을 이용한 분류 방법에 대하여 기술하고 결과를 비교하여보고자 한다.

표 3.8. 로지스틱 회귀모형에 의한 다중분류표

장애등급		판별 등급			총합
		1	2	3	
자료값	1	5,898	5,117	2,839	13,854
		42.57	36.94	20.49	
	2	9,343	14,607	5,666	29,616
		31.55	49.32	19.13	
	3	9,777	12,992	13,382	36,151
		27.04	35.94	37.02	
총합		25,018	32,716	21,887	79,621

표 3.9. 다항로지트 모형에 의한 다중분류표

장애등급		판별 등급			총합
		1	2	3	
자료값	1	5,902	5,084	2,742	13,728
		42.99	37.03	19.97	
	2	9,410	14,585	5,486	29,481
		31.92	49.47	18.61	
	3	9,948	12,947	13,132	36,027
		27.61	35.94	36.45	
총합		25,260	32,616	21,360	79,236

다중분류를 함에 있어서 Dietherich와 Bakiri (1991, 1995)가 제안한 Error Correcting Output Coding 방법은 하나의 다중분류문제를 여러 개의 이진분류(binary classification)문제로 변환하여 해결한 후 그 결과를 결합하여 다중분류를 하는 방법이다. 이진분류 단계에서 가장 많이 사용하는 방법 가운데 하나로 One Versus All(OVA)이 있는데 이는 한 범주 대비 그 외의 모든 범주에 대한 이진분류를 시행하는 방법이다. 본 연구에서는 세 범주 각각에 대하여 하나의 범주 대비 나머지 범주에 대하여 세운 세 개의 로지스틱 회귀모형을 이진분류를 위한 방법으로 적용하였다. 또한 이진분류 후 다중분류를 위한 기준을 다음과 같이 정하였다.

$$\max \frac{\hat{\pi}_i - p_i}{p_i}, \quad i = 1, 2, 3,$$

여기서 $\hat{\pi}_i$ 는 i 범주에 대하여 예측된 사후확률을 말하며 p_i 는 해당 범주의 비율이다. 즉, 세 개의 로지스틱 회귀모형으로부터 각각의 장애등급에 대하여 예측한 사후확률에서 해당등급의 비율을 뺀 값을 다시 해당등급의 비율로 나눈 값들 가운데 maximum을 갖는 등급으로 대상자의 장애등급 판별을 하였다. 이 경우 1, 2, 3 등급 각각을 옳게 판별한 경우는 33,887명으로 전체 자료 79,621명에 대하여 로지스틱 모형에 의한 정분류율은 42.56%임을 확인하였다(표 3.8). 이는 세 범주에 대하여 임의로 분류할 경우 옳게 분류할 확률인 33.3%의 128%에 해당하는 값으로 28% 향상된 값이다.

다항로지트 모형에 있어서도 로지스틱 모형의 경우와 동일한 기준으로 장애등급을 판별할 경우 결측치를 제외한 전체 79,236명의 대상자 중 33,621명에 대한 장애등급을 옳게 판별하여 42.43%의 정분류율을 보였다(표 3.9). 이는 임의확률의 127%에 해당하는 값으로 로지스틱 모형을 이용하여 이진분류를 시행한 후 다중분류를 할 때의 정분류율인 128%와 비슷한 값을 보였다. 본 자료의 일차적 분석 목적이 장애등급을 판별하는 데 있는 것이 아니라 장애등급에 대하여 주어진 7개 설명요인의 특성을 파악하는 데 있었기 때문에 주어진 7개 요인만으로는 모형의 적합도가 높지 않았음을 고려할 때 세 범주 분류에 있어

서 27%~28% 증가한 정확도를 보인 것은 다중분류에 있어서 로지스틱 모형이나 다항로짓 모형으로부터 합리적인 결과를 얻었다고 판단된다. 또한 두 방법으로부터 구한 정분류율이 거의 일치함을 확인함으로써 로지스틱 모형으로 이진분류를 수행한 후 다중분류를 하더라도 정확도에 있어서 다범주 모형으로 분류하는 것보다 열등하지 않음을 확인하였다. 장애등급을 결정하는 중요한 설명요인에 대한 자료가 추가된다면 적합도가 높은 모형을 구축할 수 있고 이러한 모형을 통하여 분류한 장애등급에 대한 정확도 또한 향상된 값을 보일 것으로 기대한다.

표 3.8과 3.9는 다중분류에 있어서 두 가지 방법에 의한 등급분류의 결과로, entry의 첫 행은 각 등급에 속한 대상자에 대하여 해당 등급으로 분류된 대상자 빈도이고 둘째 행은 이에 대한 분류율(%)이다. 1, 2, 3 등급의 민감도(sensitivity)는 로지스틱 모형으로는 각각 42.57%, 49.32%, 37.02%을 보였고 다항로짓 모형으로는 각각 42.99%, 49.47%, 36.45%를 보여 두 방법 모두 1, 2등급에서 임의 확률보다 28%~48% 증가한 민감도를 보였고 3등급에서는 9%~11% 증가한 값을 보였다.

4. 결론

본 연구는 장애연금의 특성을 파악하기 위하여, 장애연금 수급자 자료를 토대로 장애등급에 영향을 미치는 요인들의 특성을 파악하고 장애등급의 적정성에 관한 근거를 살펴보고자 하였다. 국민연금의 재정추계를 위해 살펴보는 7가지 요인들에 대한 장애등급과의 연관성 및 장애등급에 따른 요인들의 특성을 다항로짓 분석으로 확인하였다. 다항로짓 분석의 경우 기준범주 대비 각 범주에 속하는 확률을 파악할 수 있는데, 본 연구의 경우 중앙 등급인 장애 2급을 기준 범주로 하여 장애 2급 대비 장애 1급과 3급의 특징을 살펴봄으로써 장애등급의 증감에 따른 요인의 특성 파악이 가능하였다. 분석 모형에서 7개의 요인 간 교호 효과는 일부 설명요인에서 유의성을 보였으나 모형에서 고려하지 않은 이유는 교호 요인을 모형에 추가했을 때 정확도는 1~2% 증가하는데 머문 반면, 본 연구의 일차적인 목적인 장애연금의 관리 방안과 재정추계 등을 위하여 필요한, 수급자의 장애등급에 대한 설명요인의 특성 파악에 있어서는 교호 요인을 추가함으로써 등급에 영향을 미치는 정도를 파악하는 것이 크게 복잡해졌기 때문이다.

분석 결과, 기준 범주 대비 장애 1급과 장애 3급에 대한 요인들의 특성은 소득수준과 장애원인을 제외한 다섯 요인들의 특성에 따라 장애등급의 증감은 일정한 추세를 보이고 있음을 확인하였다. 이로부터, 장애연금의 장애등급은 7개의 요인들 가운데 가입기간, 성별, 소득활동여부, 수급나이 그리고 가입종별의 다섯 개 요인으로는 잘 분류되고 있음을 알 수 있었다. 이들에 대한 특성으로, 장애등급이 낮을수록 가입기간이 길어지고, 수급나이가 적었으며, 성별이 여성인 경우가 많았다. 그리고 소득활동여부가 취득이거나 대기인 경우가 많았으며, 납부예외인 경우가 적었다. 가입종별은 사업장가입과 지역가입인 경우가 많았다. 따라서 장애등급에 따른 이들 다섯 요인의 특성을 반영한 연금관리가 필요하다고 판단하였다. 자료 분석 결과로부터 파악한 장애등급별 요인의 특성을 반영한다면, 보다 현실적인 연금 정책이 가능한데, 가령, 장애등급이 높은 경우와 장애등급이 낮은 경우 등급 간 차등을 두되 실제의 자료 분석 결과가 보여주는 요인별 odds ratio에 따른 차등을 고려한다면 현실적인 상황에 맞는 연금 관리가 가능하리라 판단하기 때문이다.

자료분석의 방법에 있어서 다항로짓 모형 외에 로지스틱 회귀모형에 적합한 결과도 살펴보고았는데, 다항로짓의 경우 기준범주 대비 각 범주의 상대적 특성 파악을 통하여 장애등급의 증감에 따른 설명변수의 '일정한' 효과 및 이에 대한 통계적 유의성 검정이 가능한 반면 로지스틱 회귀모형을 통하여서는 타범주 대비 각 범주 고유 특성 파악이 가능하고 모형을 통하여 설명변수의 특성에 따라 장애등급을 분류할 경우 분류 모형에 대한 평가가 용이하고 로지스틱 모형을 이용하여 다중분류를 할 경우 정분류율에 있어서 다항로짓 모형의 경우와 동일한 값을 보여 정분류율에 있어서는 두 방법이 같은 정확도를 보임을 확

<참고> 국민연금과 장애인복지법의 장애등급 체계 비교

구분	국민연금법	장애인복지법
목적	장애로 인한 소득능력상실 또는 감소로 인한 직접적인 금전 보전	장애인의 자립 및 보호를 위한 세금감면, 복지 시설 및 편의시설 이용 등의 혜택 제공
장애인의 정의	국민연금 가입기간 중 발생한 질병 또는 부상으로 인하여 완치 후에도 신체 또는 정신상의 장애가 있는 자	신체적, 정신적 장애로 오랫동안 일상생활이나 사회생활에서 상당한 제약을 받는 자
수급 요건	국민연금 가입기간 중 발생한 질병 또는 부상으로 인하여 신체 또는 정신상의 장애가 있으면 인정	지체장애, 시각장애, 청각장애, 언어장애, 정신장애, 심장장애, 간장애 등 한정된 장애에 대하여 인정
장애 등급	질병, 부상의 완치 일을 기준으로 노동력 손실 또는 감소정도에 따라 1급 4급으로 결정하며 장애등급의 결정 및 심사는 국민연금공단이 행함	6등급 체계이며, 6등급이 장애인복지법 상에서 가장 경미한 장애
판정 시기	완치 또는 초진일로부터 1년 6개월 경과된 시점	장애유형에 따라 다르나 대체로 장애등록 신청 시점
판정 기준	노동의 제한정도에 따라 등급 결정	일상생활, 사회생활의 제약정도에 따라 등급 결정

인하였다. 장애등급에 영향을 주는 중요한 요인들이 더 주어진다면, 모형의 적합도를 높일 수 있고 따라서 모형에 의한 분류율에 있어서도 향상된 값을 얻을 수 있을 것이라 판단한다.

본 연구는 장애연금의 장애등급에 따른 주요 요인들의 특성과 장애등급의 적정성을 확인한 것으로, 1988년 1월부터 2007년 5월까지의 실제 자료로부터 도출한 결과인 만큼 실질적인 장애연금 관리방안을 마련하는데 기여할 수 있을 것으로 기대하며, 본 연구에서 파악한 장애등급에 따른 요인들의 특성을 반영함으로써, 타 연금에 비해 상대적으로 연구가 적은 장애연금에 대하여, 실제상황에 적합한 연금관리방안의 근거를 제시해줄 수 있을 것으로 기대한다.

부록

국민연금의 장애등급결정 및 이를 위한 장애심사는 국민연금공단에서 시행하고 있다. 장애를 구분하는 세 가지 개념인 기능손실, 직업적 불능, 일반장애 중 국민연금은 기능손실의 개념에 해당하는 정의를 사용하고 있으며 의학적 판단에 입각하여 장애여부 판정을 결정한다. 즉, 소득가득능력(earning capacity)이나 과거 직업, 구할 수 있는 직업 등을 감안하지 않는 것으로 되어 있다. 첫째, 기능손실은 신체의 일부분이나 어떤 신체적 또는 정신적 기능의 완전한 또는 불완전 상실을 의미하는 개념으로 이 개념에서의 장애는 그것이 경제적으로 또는 직업적으로 미칠 수 있는 결과와는 아무런 연계를 갖지 않는다. 둘째, 직업적 불능은 장애가 발생하기 전의 직업을 계속 수행할 수 없다거나 또는 그 직업을 수행하더라도 전과 같은 소득을 벌 수 없는 소득가득능력의 상실을 의미하는 개념이다. 셋째, 일반장애는 장애가 없었더라면 가질 수도 있는 노동시장에서의 일자리 기회를 더 이상 가질 수 없는데서 오는 소득가득능력의 영구적 상실로서의 개념을 뜻한다 (ILO. 1984, 73-81).

참고문헌

- 권선진, 정경배 (2005). 국민연금의 장애연금 개선방안에 관한 연구: 장애연금 수급확대를 중심으로, <사회복지정책> **22**, 99-122.
- 김성숙 (2002). <국민연금 급여제도 개선방안(II)>, 국민연금연구센터.

- 변용찬, 김경희, 윤상용, 배회옥, 박성민, 최미영 (2004). <장애인의 종합적 소득보장체계 구축방안>, 한국 보건사회연구원
- 오근식 (2001). <장애연금제도 개성방안>, 국민연금연구센터.
- 원종욱 (2002). <장애, 유족연금 심사업무 효율화 방안>, 한국보건사회연구원.
- Dietterich, T. G. and Bakiri, G. (1991). Error-correcting output codes: A general method for improving multi-class inductive learning programs, In *Proceedings of the Ninth National Conference on Artificial Intelligence (AAAI-91)*, 572-577.
- Dietterich, T. G. and Bakiri, G. (1995). Solving multi-class learning problems via error-correcting output codes, *Journal of Artificial Intelligence Research*, **2**, 263-286.
- Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Economics*, Cambridge university Press, Cambridge.
- Moroney, R. (1991). *Social Policy and Social Work: Critical Essays on the Welfare State (Modern Applications of Social Work)*, Aldine De Gruyter, Berlin.
- SAS (2000). *SAS System for Windows V.9.1*, SAS Institute Inc.

Comparison of Multinomial Logit and Logistic Regression on Disability Pensioners' Characteristic

Mijung Kim¹

¹Institute for Mathematical Sciences, Yonsei University

(Received June 2008; accepted June 2008)

Abstract

This article studies on disability pensioners' characteristic with multinomial logit and logistic regression model. Seven factors are examined on whether each factor is reflected in degree of disability in the disability pension. By incorporating multinomial logit and logistic regression model, effectiveness and characteristic of the seven factors are investigated on the degree of disability. Result shows all the seven factors are significant on the degree of disability, while among the seven, five factors, age, sex, type of coverage, type of category, insured duration show a trend in degree of disability and the other two, cause of disability and class of standard monthly income are not effective on trend in degree of disability. Results from analyses might be useful for disability pension management.

Keywords: Disability pension, multinomial logit, logistic regression model, degree of disability, multicategory classification.

¹Research Professor, Institute for Mathematical Sciences, Yonsei University, 134 Shinchon-Dong, Seodaemun-Gu, Seoul 120-752, Korea. E-mail: mjkim@yonsei.ac.kr