Journal of the Korean Data & Information Science Society
2017, **28**(5), 1055–1067

http://dx.doi.org/10.7465/jkdi.2017.28.5.1055
한국데이터정보과학회지

반복 측정 자료를 이용한 장애인 우울에 대한 분석†

홍혜선[1] · 허집[2]

[12]덕성여자대학교 정보통계학과

접수 2017년 8월 1일, 수정 2017년 8월 25일, 게재확정 2017년 9월 5일

요 약

지금까지 연구되어진 한국의 장애인 우울에 미치는 요인들을 찾는 연구들은 조사된 장애인별로 반복 측정된 자료임에도 대부분 그들 자료들이 독립이라는 가정 하에 통계적 분석을 하였다. 본 연구에서는 3년마다 반복 측정된 한국복지패널의 장애인 관련 부가조사 자료를 이용하여 장애인 우울에 미치는 요인들이 무엇인지를 반복 측정의 효과를 포함하는 선형혼합모형으로 분석한다. 각 장애인별로 얻어진 반복 측정 자료들의 종속성에 적절한 상관행렬을 고려하고, 회귀계수인 고정효과 뿐만 아니라 각 장애인별 특성을 반영하는 임의효과를 포함시켜 장애인 우울에 유의한 변수들을 찾고 최종모형을 구한다. 고정효과만 고려한 모형으로부터 얻어진 잔차들은 각 장애인별 잔차들의 평균이 0 근처에 있지 않다는 것을 보여주고, 임의효과를 포함하고 있는 최종모형의 잔차들의 잔차도와 분위수대조도를 제시하여 최종모형은 각 장애인별 잔차들의 평균의 문제점을 해결한다는 것을 보여주고자 한다.

주요용어: 상관행렬, 선형혼합모형, 임의효과, 정보판단기준, 한국복지패널자료.

1. 서론

Kim 등 (2013)에 의하면 우울은 한국에서 대표적인 정신건강 상의 문제로 대두되고 있으며, 우울증으로 발전하게 될 경우 정서적, 신체적 문제를 초래할 수 있다고 하였다. Bhattacharjee와 Chhetri (2014)에 의하면 장애인은 비장애인 보다 우울에 더 취약하다고 하였고 Shin 등 (2011)은 장애인이 우울의 부정적 영향에 더 많이 노출 되어 있다고 하였다. Choi 등 (2015)이 장애인의 우울이 음주에 부정적인 영향을 끼침을 밝혔으며 이는 알코올 중독을 야기할 수 있다고 하였고, Im과 Hwang (2014)은 장애인의 높은 우울 수준이 자살 충동에 취약함을 보였고 자살시도 위험에 빠지지 않도록 자살 및 우울의 고위험군의 선별이 필요함을 제시하였다.

횡단면 자료 (cross-sectional data)를 이용한 우리나라 장애인의 우울에 대한 다양한 연구가 이루어져 왔다. Lee와 Kahng (2009)은 구조방정식을 통해 스트레스, 심리·사회적 자원 및 우울의 관계를 파악하였고, Jeon (2010)은 회귀분석을 통한 매개효과 분석법과 조절효과 분석법을 사용하여 차별경험 횟수가 많을수록 장애인의 우울 수준이 높게 나타난다고 하였다. Park (2013)은 다중회귀분석을 사용하여 원 직장복귀자와 타 직장복귀자의 비교 중심으로 산재장애인의 우울에 미치는 요인에 대해 분석하였다. Kim 등 (2014)은 다중회귀분석을 통해 성별, 장애등급, 장애발병연령, 일상생활수행능력이 영향을

† 이 논문은 2016년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임. (NRF-2016R1D1A1B03931595)

[1] (01369) 서울시 도봉구 삼양로 144길 33, 덕성여자대학교 정보통계학과, 연구원.
[2] 교신저자: (01369) 서울시 도봉구 삼양로 144길 33, 덕성여자대학교 정보통계학과, 교수.
E-mail: jhuh@duksung.ac.kr

미친다고 분석한 바 있다. Jeon과 Kahng (2013)은 경시적 자료 (longitudinal data)인 한국복지패널 자료를 이용하여 잠재성장모형과 다중집단분석을 이용하여 차별경험, 일상생활수행능력, 도움필요정도, 자아존중감이 우울 수준과 관련이 있고, 2008년부터 2012년도 우울 수준의 표본평균이 감소하고, 65세 미만과 65세 이상의 집단 간에 우울 수준의 차이가 있음을 보였다. Jeong과 Park (2016)은 한국복지패널 자료를 이산시간위험모형으로 성별, 배우자 유무, 주거지역, 자아존중감 수준과 여가생활만족도에 따라 우울탈출 가능성에 대한 차이가 있는지를 분석하였다.

본 연구에서는 한국복지패널 자료의 경시적 자료를 이용하여 고정효과 (fixed effect)와 임의효과 (random effect)를 고려한 선형혼합모형 (linear mixed model)으로 장애인 우울에 미치는 요인들을 분석하고자 한다. 즉, 장애인의 우울 수준에 어떠한 변수가 영향을 끼치는지 확인하고 개인의 특성에 따라 우울 수준에 영향을 주는 변수들을 찾고자 한다. 이때 고려한 설명변수들은 장애인 우울에 대한 선행 연구들의 결과를 토대로 선택하였다. 즉, 반응변수는 장애인의 우울이며, 양적 설명변수로는 소득, 연령, 자아존중감, 일상생활수행능력을 선택하였고, 질적 설명변수로는는 성별, 교육수준, 장애차별경험과 장애등급을 고려하였다.

한국복지패널 자료의 조사는 매년 이루어지며 1차 조사는 2006년도에 7,072가구로 시작되었고, 자료가 공개된 2016년 현재 11차 조사로 4,560가구가 조사되었다. 장애인 관련 조사는 3년마다 조사가 이루어지는 부가조사로서 3차년도인 2008년, 6차년도인 2011년 그리고 9차년도인 2014에 조사되었다. 장애인 우울 수준에 관한 항목 조사가 이루어진 대상자의 수는 각각 3차년도의 706명, 6차년도의 535명, 9차년도의 432명으로 중도탈락자가 발생하였다. 2017년 장애인의 부가조사 자료는 아직 공개되지 않았기에 이번 연구에서 제외하였다.

국외의 장애인 우울에 대한 연구로는 Geerlings 등 (2001)과 Geerlings 등 (2002)이 일반화추정방정식 (generalized estimating equation)을 이용하여 장애인의 우울에 대한 경시적 자료 분석 실시하였다. 선형혼합모형을 이용한 경시적 자료 분석으로는 Jo와 Chang (2013), Lee 등 (2014), Choi와 Huh (2014)와 Lee와 Huh (2017)가 있다.

한국복지패널 자료를 이용한 장애인 우울에 영향을 주는 변수들의 고정효과와 임의효과를 찾기 위하여, 2절에서는 각 설명변수들과 반응변수의 기술통계량들과 기본적인 자료 분석의 결과를 제시하였다. 고정효과와 임의효과를 포함한 선형혼합모형에 대하여 3절에서 소개하고 모형의 적합과 진단의 결과를 설명하고 4절에서는 적합된 모형을 통하여 장애인 우울에 영향을 주는 설명변수의 고정효과와 임의효과가 있는지 알아보고자 한다.

2. 변수의 설명 및 자료의 탐색

2.1. 변수의 설명

본 연구의 관심의 대상인 반응변수로 사용되는 장애인 우울 (depression)은 한국복지패널 부가 조사에서 사용된 관련 11개 문항을 사용하여 점수화하였다. 이는 9차 한국복지패널 유저가이드북을 참고한 것으로 16점 이상이면 우울증을 의심해봐야 하는 단계이다. 설명변수들 중 질적변수인 장애인의 교육수준 (educational background)은 한국복지패널 자료 조사의 결과에 의하면 일반적으로 비장애인들에 비해 낮은 편이다. 고등학교 졸업 이상의 학력의 도수가 많지 않기에, 중졸 이하의 저학력과 고졸 이상의 고학력으로 나누었다. 교육수준을 두 범주로 나눈 이유와 같은 이유로, 학창 시절과 사회 생활에서 장애차별경험 (experience of discrimination)은 있음과 없음으로 구분하였고 장애등급 (degree of disability)은 중증과 경증으로 나누었다.

양적 설명변수인 자아존중감 (self-esteem)은 10개의 관련 문항들을 사용하여 점수화하였고, 일상생활수행능력 (activities of day living, ADL)은 기본적 일상생활수행능력과 수단적 일상생활수행능력을

모두 고려하였고 22개의 조사된 관련 문항들을 고려하여 점수화하였다.

Table 2.1은 질적 설명변수들의 각 범주에 대한 연도별 도수들을 보여주고 있다. 괄호 안의 수치들은 각 해당 도수들의 백분율들이다. 이 표에 의하면 성별 (gender)에서는 여성 보다는 남성의 경우가, 고학력에 비해 저학력의 경우가, 장애차별경험이 있는 경우 보다 장애차별경험이 없는 경우가 도수가 많은 경향이 있음을 보여준다. 서론에 언급하였듯이 반복 측정 시에 결측치들이 생기고 있지만 전반적으로 각 설명변수들의 범주에 해당하는 백분율들은 연도별로 차이가 크지 않아 보인다.

Table 2.1 Frequencies with relative frequencies in the parentheses of the qualitative variables for each year.

	gender		educational background		experience of discrimination		degree of disability		
year	male	female	low	well	no	yes	grave	slight	total
2008	410	296	488	218	556	150	390	316	2824
	(58.1)	(41.9)	(69.1)	(30.9)	(78.8)	(21.3)	(55.2)	(44.8)	(42.2)
2011	313	222	371	164	429	106	306	229	2140
	(58.5)	(41.5)	(69.3)	(30.7)	(80.2)	(19.8)	(57.2)	(42.8)	(32.0)
2014	259	173	293	139	381	51	241	191	1728
	(59.9)	(40.1)	(67.8)	(32.2)	(88.2)	(11.8)	(55.8)	(44.2)	(25.8)
total	982	691	1152	521	1366	307	937	736	6692
	(58.7)	(41.3)	(68.9)	(31.1)	(81.6)	(18.4)	(56.0)	(44.0)	(100.0)

한편, Table 2.2는 양적 설명변수들과 반응변수인 장애인 우울에 대한 각 연도별 표본평균들과 표본표준편차들을 보여주고 있다. 표본표준편차들은 괄호 안에 제시하였다. 조사대상 장애인들의 소득 (income)은 점차 나아지는 경향이 있지만 물가 상승을 감안한다면 큰 변화가 없는 것처럼 보인다. 자아존중감과 ADL 또한 연도별 변화가 보이지 않고 있다. 장애인의 우울 점수는 1차 조사 시기가 그 외의 조사 시기보다 조금 높아보인다.

Table 2.2 Sample means with their standard deviations in the parentheses of the quantitative variables for each year.

year	income	age	self-esteem	ADL	depression
2008	2094.37	61.18	2.77	23.87	13.70
	(1885.85)	(14.74)	(0.47)	(5.37)	(11.43)
2011	2552.20	63.00	2.80	24.28	10.48
	(2500.75)	(13.98)	(0.45)	(6.17)	(10.21)
2014	2651.37	64.97	2.79	23.88	11.02
	(2350.52)	(13.02)	(0.45)	(5.60)	(9.85)
total	2232.61	62.73	2.78	24.00	11.98
	(2232.49)	(14.14)	(0.46)	(5.69)	(10.75)

2.2. 자료의 탐색

연도별 질적 설명변수들의 각 범주에 따른 장애인 우울 점수의 변화를 보기 위하여 Figure 2.1을 제

시하였다. Figure 2.1의 (a), (b), (c)와 (d)는 각각 성별, 교육수준별, 장애차별경험별과 장애등급별 세 차례의 조사에 대한 장애인 우울 점수의 변화를 보여주고 있다. 모든 질적 설명변수들에 대한 장애인 우울 점수의 변화 형태는 매우 유사하게 나타나고 있다. 한편, 조사된 모든 연도에서 남성 보다는 여성이, 고학력 보다는 저학력이, 차별경험이 없는 경우 보다는 있는 경우가 그리고 장애등급이 경증인 경우 보다는 중증인 경우가 장애인 우울의 점수의 평균이 더 컸다.

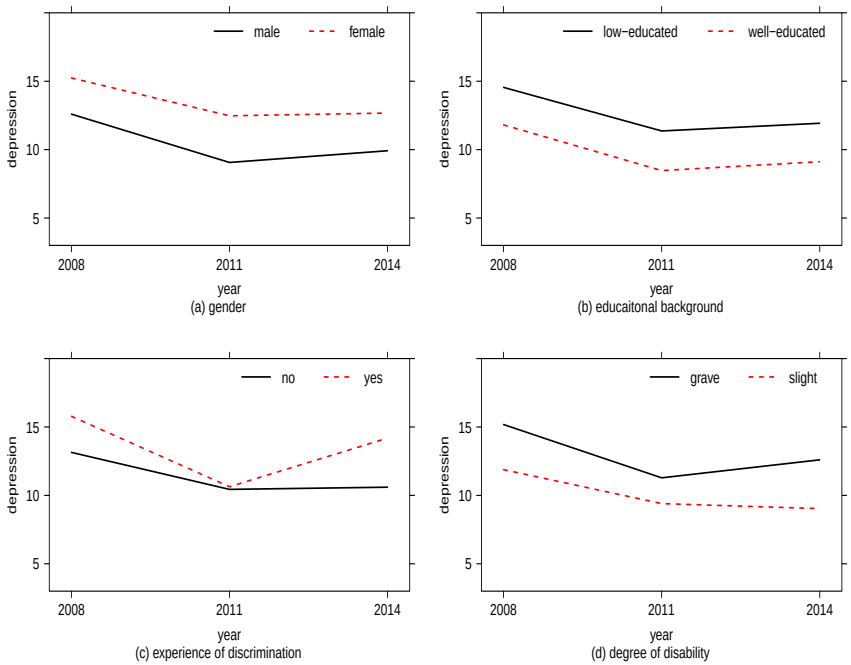

Figure 2.1 The flowcharts of sample means of the depression for each qualitative variable.

Figure 2.2는 각 양적 설명변수와 장애인 우울 점수에 대한 커널추정량의 국소선형추정량 (local linear estimator)을 이용한 추정된 회귀함수들을 보여주고 있다. 이때 사용된 커널함수는 Epanechnikov 커널이고, 띠폭은 두 변수간의 전체적인 관계를 파악하는 것이 주목적이기에 추정된 회귀함수의 부드러움 정도가 적절하다고 생각되는 띠폭으로 선택하였다. 실선은 2008년도, 굵은 점선은 2011년도, 가는 점선은 2014년도의 추정된 회귀함수를 보여주고 있다. Fan과 Gijbels (1996)에서 커널추정량에 대한 자세한 내용을 참조할 수 있다.

이 결과에 의하면, 소득이 높을수록 장애인 우울 점수가 낮아지는 경향이 있으나 아주 높은 소득에서는 오히려 장애인 우울 점수가 올라가는 것처럼 보인다. 이는 본 자료의 소득 자료도 대부분의 소득 자료들 처럼 양의 왜도를 가지고 있고 소수의 큰 소득의 장애인 우울 점수에 의한 현상일 뿐이다. 나이가 들면 장애인 우울 점수가 조금 상승하는 것처럼 보이나 크게 영향을 받는 것처럼 보이지는 않는다. 추측할 수 있듯이, 자아존중감이 커지면 장애인 우울 점수는 빠르게 감소하고 일상생활수행능력인 ADL의 점수가 커짐에 따라 장애인 우울의 정도가 올라가는 경향이 있다는 것을 알 수 있다. 하지만, ADL 또한 소득 자료 처럼 ADL이 높은 경우에는 오히려 장애인 우울 점수가 낮아지는 경향이 있는데 이는 소수의 큰 ADL의 장애인 우울 점수에 의한 현상이다. 이러한 양적 설명변수들에 대한 장애인 우울 점수의 회

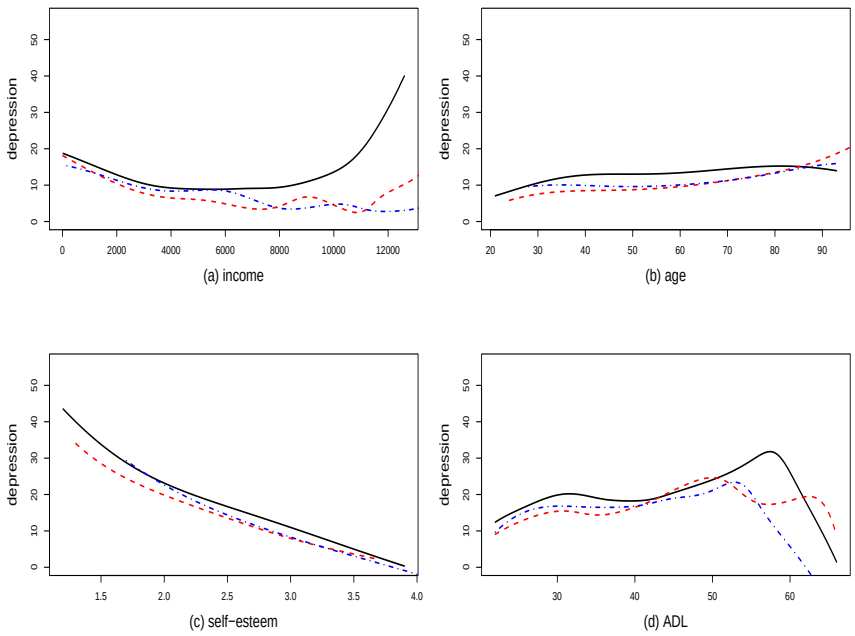

Figure 2.2 The kernel regression estimates of the depression for each quantitative variable and year. The solid, dashed and dotted lines are represented by the estimated regression functions for 2008, 2011 and 2014 respectively.

귀함수의 커널추정의 경향들은 조사된 모든 연도에서 비슷하게 나타났다.

세 차례 반복 측정에 의해 조사된 자료이기에 조사 대상인 각 장애인별 장애인 우울 점수들은 독립이 아닐 가능성이 높다. Table 2.3은 이들 반복 측정된 연도별 장애인 우울 점수들의 표본상관계수를 보여주고 있다. 괄호 안은 각 표본상관계수들의 표준오차들이다. 2008년 1차 조사 시기와 2011년 2차 조사시기인 경우와 2차 조사시기와 2014년 3차 조가 시기 사이의 장애인 우울 점수간의 표본상관계수인 0.42와 0.43 보다 1차 조사시기와 3차 조사시기 간의 장애인 우울 점수간의 표본상관계수인 0.39가 조금 작아진듯하게 보인다. 이는 조사 시기의 차가 클수록 상관관계가 약해지는 일반적인 경향에 부합한다고 할 수 있다. 이들 모상관계수들이 0인지 아닌지를 검정하기 위하여 구한 유의확률은 Table 2.3에 제시되어 있듯이 매우 작아 장애인 우울 점수들이 정규분포를 따른다고 한다면 반복 측정 자료들은 독립이 아닐 가능성이 높다. 이러한 조사대상인 각 장애인별 반복 측정된 반응변수의 상관관계에 대한 정보는 다음 3절의 모형의 적합에 이용될 것이다.

Table 2.3 Sample correlations with standard errors in the parentheses and p-values of the depression for among years.

years	2008 and 2011	2011 and 2014	2008 and 2014
sample correlation coefficient	0.42 (0.039)	0.43 (0.043)	0.39 (0.044)
p-value	<0.001	<0.001	<0.001

3절의 모형의 적합 시에 연속인 설명변수들간의 다중공선성의 여부를 파악하기 위하여 분산팽창계수 (variance inflation factor, VIF)를 구하여 Table 2.4에 제시하였다. 분산팽창계수는 $VIF_j = 1/(1 - R_j^2)$이고, 여기서 R_j^2은 j번째 설명변수를 반응변수로 간주하였을 때, 다른 설명변수들과의 다중선형회귀모형의 적합에 의한 결정계수이다. 즉, j번째 설명변수와 그 외의 설명변수간의 선형관계가 클수록 분산팽창계수는 커져 다중공선성의 존재 가능성이 커지게 된다. 일반적으로 분산팽창계수가 10 이상인 경우에 다중공선성 존재 가능성이 있다고 본다. 분산팽창계수에 대한 자세한 내용은 Montgomery 등 (2012)를 참조하기 바란다. Table 2.4의 결과에 의하면 분산팽창계수가 1보다 조금 크기에 양적 설명변수들간의 다중공선성의 존재 가능성이 낮아보인다.

Table 2.4 Variance inflation factors (VIF) of the quantitative variables.

variable	income	age	self-esteem	ADL
VIF	1.14	1.07	1.45	1.07

3. 자료 분석

3.1. 모형

한국복지패널 장애인 부가 조사 자료는 반복 측정된 것이기에, 설명변수들이 반응변수인 장애인 우울에 어떻게 영향을 주는 지를 파악할 고정효과 뿐만 아니라 장애인 개개인의 특성이 반영된 임의효과를 포함하는 선형혼합모형이 고려되어야 더욱 적절한 자료 분석이 될 것이다. 다음의 선형혼합모형과 추정량들에 대한 서술은 Verbeke와 Molenberghs (2009)를 참고한 것이다.

선형혼합모형의 서술을 위하여 조사 대상자의 수를 m이라 하고, 결측치가 존재할 수 있기에 i번째 조사 대상자의 반복 측정 횟수를 n_i라 하고 설명변수의 갯수는 p개라 하자. i번째 조사 대상자의 j번째 반복 측정된 반응변수의 관측치를 y_{ij}라 하고 그 때의 k번째 설명변수의 관측치를 x_{kij}라 하자. 모형을 벡터와 행렬로 표현하고자 i번째 조사 대상자의 관측치들에 대하여 $\boldsymbol{y}_i = (y_{i1}, \cdots, y_{in_i})^T$, $\boldsymbol{x}_{ij} = (1, x_{1ij}, \cdots, x_{pij})^T$ 그리고 $X_i = (\boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{in_i})^T$라 두자. 그러면 선형혼합모형은 다음과 같이 표현된다.

$$\boldsymbol{y} = X\boldsymbol{\beta} + Z\boldsymbol{b} + \boldsymbol{\varepsilon}, \tag{3.1}$$

여기서 $\boldsymbol{y} = (\boldsymbol{y}_1^T, \cdots, \boldsymbol{y}_m^T)$, $X = (X_1^T, \cdots, X_m^T)^T$, $Z = \mathrm{diag}(Z_1, \cdots, Z_m)$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_p)^T$, $\boldsymbol{b} = (\boldsymbol{b}_1^T, \cdots, \boldsymbol{b}_m^T)^T$ 그리고 $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^T, \cdots, \boldsymbol{\varepsilon}_m^T)^T$이다. 벡터 $\boldsymbol{\beta}$는 $(p+1)$ 차원이며 고정효과의 회귀계수이고, $(q+1)$차 확률벡터 $\boldsymbol{b}_i = (b_{i0}, b_{i1}, \cdots, b_{iq})^T$는 임의효과이며, 임의효과의 보조벡터는 $\boldsymbol{z}_{ij} = (1, z_{1ij}, \cdots, z_{qij})^T$이고 이들 보조벡터들을 이용하여 $Z_i = (\boldsymbol{z}_{i1}, \cdots, \boldsymbol{z}_{in_i})^T$를 구성하였다. 여기서 보조벡터 $\boldsymbol{z}_{ij}$는 벡터 $\boldsymbol{x}_{ij}$의 일부이기에 양의 정수 q는 설명변수의 갯수 p보다 같거나 작다. 한편, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \cdots, \varepsilon_{in_i})^T$는 i번째 조사 대상자의 j번째 반복 측정 자료들의 오차들이고 $\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_m$과 $\boldsymbol{b}_1, \cdots, \boldsymbol{b}_m$은 서로 독립이라 가정한다. 자연수 k에 대하여, N_k는 k 차원 정규분포라 한다면, 오차들과 임의효과들에 대한 분포는 다음과 같다.

$$\boldsymbol{\varepsilon}_i \sim N_{n_i}(\mathbf{0}_{n_i}, \Sigma_{n_i}), \quad \boldsymbol{b}_i \sim N_{q+1}(\mathbf{0}_{q+1}, D), \tag{3.2}$$

여기서 $\mathbf{0}_k$는 k차원 0벡터이고 Σ_{n_i}과 D는 각각 $\boldsymbol{\varepsilon}_i$와 $\boldsymbol{b}_i$의 공분산행렬 (covariance matrix)로 각각 n_i차와 $(q+1)$차 정방행렬이다. 오차 $\boldsymbol{\varepsilon}_i$의 성분들은 i번째 조사 대상자의 반복 측정 자료들의 오차들이기에 일반적으로 독립이 아니다. 이를 고려한 상관행렬 (correlation matrix)을 자료분석 시에 고려하여야 한다.

식 (3.1)의 적합으로 다음의 고정효과의 추정인 최량선형불편추정량 (best linear unbiased estimator)과 임의효과의 예측인 최량선형불편예측 (best linear unbiased predictor)이 이용된다.

$$\widehat{\boldsymbol{\beta}} = (X^T V^{-1} X)^{-1} X^T V^{-1} \boldsymbol{y}, \quad \widehat{\boldsymbol{b}}_i = D Z_i^T V_i^{-1} (\boldsymbol{y}_i - X_i \widehat{\boldsymbol{\beta}})^T. \tag{3.3}$$

여기서 $V_i = Z_i D Z_i^T + \Sigma_i$, $i = 1, \cdots, m$, $V = \mathrm{diag}(V_1, \cdots, V_m)$이다. 선형혼합모형과 이들의 추정량들에 대한 성질의 자세한 내용은 Verbeke와 Molenberghs (2009)를 참고하기 바란다.

3.2. 결측치와 상관행렬

반복 측정된 자료들에서 결측치는 자연적인 현상이다. Verbeke와 Molenberghs (2009)의 15장에 서술되었듯이 결측치들은 완전임의결측 (missing at completely random)과 임의결측 (missing at random)인 무시할 수 있는 결측치와 비임의결측 (missing not at random)인 무시할 수 없는 결측치로 나눌 수 있다. Diggle 등 (2001)의 13장에 의하면 같은 관측 연도 내의 결측치 갯수별 반응변수의 모평균들의 신뢰구간들이 공통 구간을 가지게 되면 무시할 수 있는 결측인 임의결측으로 판단할 수 있다고 하였다. 본 자료의 각 연도별 결측치 갯수별 모평균의 99% 신뢰구간을 Table 3.1에 제시하였다. 이 결과에 의하면 각 연도별 모평균의 신뢰구간은 공통 구간을 가지고 있기에 본 자료의 결측치는 임의결측인 무시할 수 있는 결측치로 간주하고 자료 분석을 진행하였다.

Table 3.1 Confidence intervals of the means of the depression for each number of missing value and year.

	year	2008	2011	2014
	2	(13.49,18.32)		
number of missing values	1	(10.73,16.42)	(10.84,16.95)	
	0	(11.47,14.25)	(8.46,10.86)	(9.80,12.25)

식 (3.2)에서 정방행렬인 i번째 조사 대상자의 오차들의 공분산행렬 Σ_{n_i}는 결측치의 갯수에 따라 차원이 달라진다. 각 조사 대상자들의 반복 측정 자료들 내에서 측정 시차가 커질수록 상관계수가 작아지는 것이 일반적이고 상식적인 현상이다. 2절의 Table 2.3에서 본 자료의 장애인 우울 점수의 조사 연도간 표본상관계수를 보면 조사 연도의 차이에 따라 큰 차이를 보이고 있지는 않다. 흔히 쓰이는 상관행렬 중 본 연구에서는 단위행렬, 교환가능 (exchangeable)행렬, 일차자기상관행렬 (frist-order autocorrelation)행렬, 지수 (exponential)행렬을 자료 분석 시에 고려하였다.

단위행렬은 각 조사 대상자 내의 반복 측정 자료들이 독립인 경우이며, 교환가능행렬은 조사 대상자 내의 모든 서로 다른 오차항들의 상관계수가 ρ로 동일한 경우이다. 일차자기상관행렬은 시계열모형에서 1차 자기회귀모형 AR(1)모형의 자기상관함수의 특징을 가진 것으로, 임의의 서로 다른 j와 j'에 대하여 $corr(\varepsilon_{ij}, \varepsilon_{ij'}) = \rho^{|j-j'|}$가 된다. 지수상관행렬은 $corr(\varepsilon_{ij}, \varepsilon_{ij'}) = \exp(-\phi|j-j'|)$, $\phi > 0$ 관계가 성립한다. 즉, 시차의 차이가 커질수록 지수적으로 상관계수가 작아진다. 상관행렬에 대한 자세한 내용은 Diggle 등 (2001)을 참고하기 바란다.

3.3. 모형 적합

양적 설명변수들인 소득, 나이, 자아존중감과 ADL을 각각 x_1, x_2, x_3와 x_4라 하고 질적 설명변수인 성별, 교육수준, 장애차별경험과 장애등급을 가변수 (dummy variable)화 한 변수들을 각각 x_5, x_6, x_7과 x_8이라 하자. 모형의 적합을 위하여 식(3.3)의 계산법으로 제한최대가능도추정량 (restricted maximum likelihood estimator)과 R 소프트웨어의 `lme()`를 사용하였다.

먼저, 식 (3.1)에서 고정효과만을 고려한 경우, 즉 일반선형모형 (general linear model)으로 본 자료를 분석해 보았다. 모형의 선택 기준으로 널리 쓰이는 AIC (Akaike Information Criterion)을 고려하였다. AIC의 자세한 내용은 Fitzmaurice 등 (2011)을 참고하기 바란다. 앞의 3.2절의 네 가지 상관행렬을 적용하였을 때 교환가능행렬인 경우의 AIC가 가장 작았다. 네 가지 상관행렬의 결과들이 유사하여 AIC가 가장 작은 교환가능행렬인 경우의 식 (3.3)의 $\widehat{\boldsymbol{\beta}}$의 추정치와 그들의 표본오차를 Table 3.2에 제시하였다. 각 설명변수들의 고정효과인 회귀계수가 0인지를 검정하기 위한 제시된 유의확률에 의하면 양적 설명변수인 소득, 자아존중감과 ADL 그리고 질적 설명변수인 성별이 장애인 우울에 영향을 준다고 볼 수 있다. 소득과 자아존중감은 기울기가 음이고 ADL은 양의 기울기를 가지며, 여성인 경우가 남성에 비해 장애인 우울 점수가 큰 경향을 나타내고 있다. 이러한 결과는 2.2절의 Figure 2.1과 2.2의 결과와 유사하다. 또한 상관행렬이 교환가능행렬일 때 AIC가 가장 작게 나온 것은 Table 2.3에서 조사연도 간의 장애인 우울 점수의 표본상관계수들이 큰 차이를 보이지 않는 것과 유사한 결과라 할 수 있다.

Table 3.2 Estimated regression coefficients with their standard errors in the parentheses and p-values using the exchangeable correlation matrix in the fixed effect model.

exploratory variable		regression coefficient	p-value
intercept		33.57 (2.37)	0.0000
income (x_1)		-0.001 (0.0001)	0.0000
age (x_2)		0.04 (0.02)	0.0836
self-esteem (x_3)		-9.99 (0.53)	0.0000
ADL (x_4)		0.17 (0.04)	0.0001
gender (x_5)	male	·	·
	female	2.22 (0.52)	0.0000
educational background (x_6)	low	·	·
	well	0.84 (0.64)	0.1879
experience of discrimination (x_7)	no	·	·
	yes	0.99 (0.59)	0.0894
degree of disability (x_8)	grave	·	·
	slight	-0.10 (0.53)	0.8449

Table 3.2에 의한 유의한 설명변수들만으로 고정효과와 임의효과의 유의성을 파악하고 적절한 모형을 찾기 위하여 다음과 같은 모형들을 고려하였다.

- 모형 1 : $y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_3 x_{3ij} + \beta_4 x_{4ij} + \varepsilon_{ij}$.
- 모형 2 : $y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_3 x_{3ij} + \beta_4 x_{4ij} + b_{i0} + \varepsilon_{ij}$.
- 모형 3 : $y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_3 x_{3ij} + \beta_4 x_{4ij} + b_{i0} + b_{i1} x_{1ij} + \varepsilon_{ij}$.

- 모형 4 : $y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_3 x_{3ij} + \beta_4 x_{4ij} + b_{i0} + b_{i3} x_{3ij} + \varepsilon_{ij}$.

- 모형 5 : $y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_3 x_{3ij} + \beta_4 x_{4ij} + b_{i0} + b_{i4} x_{4ij} + \varepsilon_{ij}$.

모형 1은 고정효과만 있는 모형이고, 모형 2는 고정효과와 임의절편 (random intercept)이 있는 모형이다. 모형 3, 4, 5는 고정효과와 임의절편을 포함하고 유의한 설명변수 x_1, x_3, x_4에 대한 각각의 임의기울기를 포함하고 있는 모형들이다. 위 5가지의 모형과 4가지의 상관행렬에 대하여 각각 본 자료를 적합하여 구한 AIC를 Table 3.3에 제시하였다. 유의한 설명변수들의 고정효과와 임의절편 및 자아존중감의 임의기울기가 있는 교환가능행렬인 경우의 모형 4가 가장 작은 AIC를 가지기에 본 자료의 적절한 모형으로 최종적으로 선택하고자 한다.

Table 3.3 AICs of each model

model	1	2	3	4	5
identity	12117.77	12071.51	12073.16	11998.04	12048.26
exchangeable	12071.51	12073.51	12055.39	11985.50	12050.26
AR(1)	12081.14	12073.42	12074.46	11999.71	12050.20
exponential	12081.14	12073.42	12125.77	12001.82	12058.71

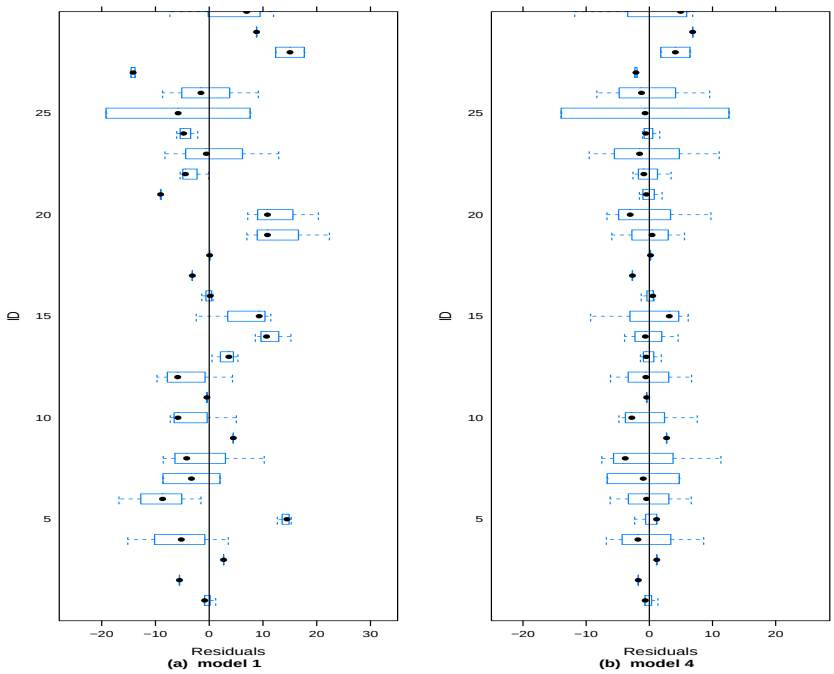

Figure 3.1 The residual plots of model 1 and 4 with the exchangeable correlation matrix

교환가능행렬의 고정효과만 있는 모형 1과 최종적으로 선택한 모형 4을 비교해보기 위하여 다음의 잔차도인 Figure 3.1을 제시하였다. 모든 조사 대상자의 잔차도에 대한 상자 그림을 보여주기에 시각적

효율성이 없어 처음 30명의 조사 대상자의 잔차에 대한 상자 그림만 제시하였다. Figure 3.1의 (a)는 고정효과만 있는 모형 1의 잔차도이다. 모형 1의 잔차도에서는 각 조사 대상자의 잔차들이 모두 양수이거나 모두 음수인 경우가 많아 보인다. 즉, 고정효과만으로 각 조사 대상자들의 특성을 반영하지 못함을 알 수 있다. 반면, 최종적으로 선택되어진 모형 4의 잔차도 Figure 3.1의 (b)는 조사 대상자들의 잔차들이 대부분 0 근처에 분포 되어 있어 모형 1보다 자료들을 더 잘 설명해준다고 볼 수 있다. 추가적으로 전체 자료의 스튜던트화 잔차들에 대한 분위수대조도 (Q-Q plot)를 보여주고 있는 Figure 3.2에 의하면 전체적으로 정규분포의 가정이 적절하다고 할 수 있다. 한편, 모형 1과 모형 4의 차이가 유의한 지를 보기 위하여 가능도비검정 (likelihood ratio test)를 실시하였고, 이때의 검정통계량은 92.01이었으며 유의확률은 0.0001보다 작아 두 모형이 유의한 차이가 있었다.

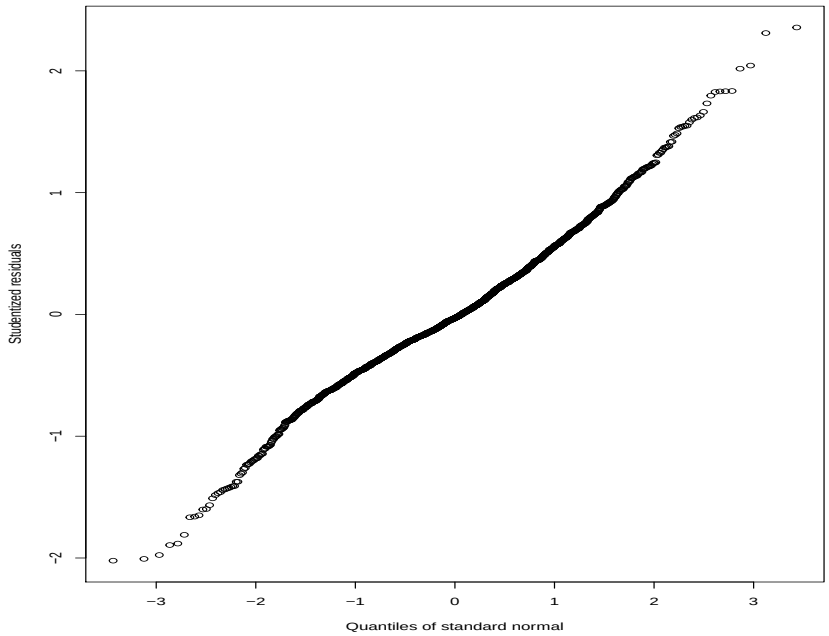

Figure 3.2 The Q-Q plot of the studentized residuals

4. 적합된 모형과 맺음말

서론에서 언급하였던 장애인 우울에 영향을 주는 설명변수들을 대상으로 고정효과와 임의효과를 반영한 선형혼합모형으로 한국복지패널 자료를 분석한 앞 3.3절의 결과를 보면, 장애인 우울에 영향을 주는 변수로는 소득, 자아존중감, ADL과 성별이 유의하였다. 소득과 자아존중감은 낮을수록, ADL은 높을수록 장애인 우울 점수가 높아지는 경향이 있었다. 또한 남성 보다는 여성들이 장애인 우울에 취약하였다. 그에 비하면 나이, 교육수준, 장애차별경험과 장애등급은 장애인 우울에 크게 영향을 주지 않는 것으로 파악되었다. 한편, 임의절편과 자아존중감의 임의기울기가 유의한 것으로 나타났기에 적합된 모형은 개인의 특성이 반영된 것으로 각 조사 대상자들의 모형의 절편과 자아존중감의 장애인 우울에 대한

영향이 다르다고 할 수 있다.

유의한 설명변수들이 장애인 우울에 미치는 정도를 파악하기 위하여, 선택된 모형 4에 포함된 변수들을 표준화하여 적합하면 다음과 같다.

$$\widehat{y}_{ij} = -0.090x_{1ij} - 0.422x_{3ij} + 0.100x_{4ij} + 0.189x_{5ij} + \widehat{b}_{i0} + \widehat{b}_{i3}x_{3ij}. \tag{4.1}$$

이때 i번째 조사 대상자의 임의효과들의 예측은 $\widehat{\boldsymbol{b}}_i = (\widehat{b}_{i0}, \widehat{b}_{i3})^T$으로 $\widehat{b}_{i0}$는 임의기울기의 예측이고, $\widehat{b}_{i3}$는 설명변수인 자아존중감 x_3의 임의기울기의 예측이다. 추정된 모형으로부터 얻어진 식 (3.2)의 교환가능행렬인 오차의 공분산행렬의 추정치 $\widehat{\Sigma}_i$과 임의효과의 공분산행렬의 추정치 $\widehat{D}$는 다음과 같다.

$$\widehat{\Sigma}_i = \begin{pmatrix} 0.358 & -0.178 & -0.178 \\ -0.178 & 0.358 & -0.178 \\ -0.178 & -0.178 & 0.358 \end{pmatrix}, \quad \widehat{D} = \begin{pmatrix} 2.79 & -0.68 \\ -0.68 & 0.16 \end{pmatrix}.$$

오차의 공분산행렬의 추정치는 결측치가 없는 경우의 예를 든 것이다. 위 추정된 오차의 공분산행렬은 결측치가 하나인 경우는 2차 정방행렬이 되며, 결측치가 두 개인 경우는 스칼라가 된다. 위 임의효과의 공분산행렬의 추정치를 보면 임의절편의 분산의 추정치가 자아존중감의 임의기울기의 분산의 추정치보다 크기에 임의절편이 자아존중감의 임의기울기 보다 각 개인의 특성이 더 강하다고 할 수 있다. 한편, 오차의 공분산행렬의 분산 추정치와 상관계수 추정치의 표준오차들은 각각 0.0002와 0.765이다. 그리고 임의기울기와 임의절편의 분산의 추정치들의 표준오차들은 각각 0.138과 0.050이다.

표준화 후 적합된 선형혼합모형의 결과인 (4.1)에 의하면 자아존중감이 장애인 우울에 매우 큰 영향을 주는 것처럼 보인다. 이러한 결과는 2.2절의 양적 설명변수들에 대한 장애인 우울 점수의 회귀함수의 커널추정 결과들과 유사하다. Figure 2.2의 네 그림 중 (c)가 자아존중감을 설명변수로 한 회귀함수의 추정 결과이다. 다른 세 결과에 비해 추정된 회귀함수의 기울기가 매우 가파르게 나타난다. 본 분석의 결과로 보면 장애인 우울에 대한 정책으로 장애인들의 자아존중감을 가지게 하고 키워주는 방향을 모색해야 할 것으로 보인다. 또한, 장애인 여성들이 장애인 우울에 더 취약한 경향이 있기에 여성을 배려한 정책이 필요해 보인다.

References

Bhattacharjee, A. and Chhetri, K. (2014). Who is more depreesed?: A study among persons with disability. *Indian Journal of Health and Wellbeing*, **5**, 1339-1341.

Choi, N. and Huh, J. (2014). A longitudinal study for child aggression with Korea Welfare Panel Study data. *Journal of the Korean Data & Information Science Society*, **25**, 1439-1447.

Choi, B. I., Park, Y. S. and Song, J. Y. (2015). The influence of depression and self-esteem on the drinking level of the disabled: Focused on the mediating effect of life satisfaction. *Korean Journal of Social Welfare Education*, **29**, 141-163.

Diggle, P. J., Heagerty, P., Liang, K. Y. and Zeger, S. L. (2001). *Analysis of longitudinal data, 2nd edition*, Oxford, New York.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications*, Chapman and Halls, London.

Fitzmaurice, G. M., Laird, N. M. and Ware, J. H. (2011). *Applied longitudinal analysis, 2nd edition*, John Wiley & Sons, Hoboken, New Jersey.

Geerlings, S. W., Beekman, A. T. F., Deeg, D. J. H., Twisk, J. W. R. and van Tilburg, W. (2001). The longitudinal effect of depression on functional limitations and disability in older adults: An eightwave prospective community-based study. *Psychological Medicine*, **31**, 1361-1371.

Geerlings, S. W., Twisk, J. W. R., Beekman, A. T. F., Deeg, D. J. H., and van Tilburg, W. (2002). Longitudinal relationship between pain and depression in older adults: Sex, age and physical disability. *Social Psychiatry and Psychiatric Epidemiology*, **37**, 23-30.

Im, W. G. and Hwang, S. H. (2014). Investigation of the causes of suicidal ideation of persons with severe disabilities living in S-Gu-focusing on their daily activities, experience of committing suicide, and depression. *The Korean Society of Stress Medicine*, **22**, 35-42.

Joen, H. S. and Kahng, S. K. (2013). An exploratory study on the predictors of depression trajectory among individuals with disabilities: Multi-group comparisons between 65 and over and 65 less using korean welfare panel study. *Journal of Rehabilitation Research*, **17**, 41-67.

Jeon, J. H. (2010). Study for the mediators and relationship between disability discrimination experience and mental health of people with disabilities. *Mental Health & Social Work*, **35**, 51-80.

Jeong, J. S. and Park, M. E. (2016). A study on the factors influencing the depression-exit of disabled. *Korean Journal of Social Welfare Education*, **36**, 55-83.

Jo, J. and Chang, U. J. (2013). A statistical analysis of the fat mass repeated measures data using mixed model. *Journal of the Korean Data & Information Science Society*, **24**, 303-310.

Kim, H. and Park, K. (2013). The review of process and researches in acceptance commitment therapy for depression. *Journal of Psychotherapy*, **13**, 99-112.

Kim, Y. S., Kim, S. Y. and Kim, M. (2014). Factors related to depression and activities of daily living in persons with disability: Using the Korean welfare panel data. *Korean Public Health Research*, **40**, 51-62.

Lee, N. and Huh, J. (2017). A longitudinal data analysis for child academic achievement with Korea welfare panel study data. *Journal of the Korean Data & Information Science Society*, **28**, 1-10.

Lee, H. J. and Kahng, S. K. (2009). The relationships between stressors, psychosocial resources, and depression among individuals with disabilties. *Mental Health & Social Work*, **33**, 193-217.

Lee, J., Sung, J. and Won, S. (2014). Efficient strategy for the genetic analysis of related samples with a linear mixed model. *Journal of the Korean Data & Information Science Society*, **25**, 1025-1038.

Montgomery, D., C., Peck, E., A. and Vining, G. G. (2012). *Introduction to linear regression analysis, 5th edition*, John Wiley & Sons, Hoboken, New Jersey.

Park, S. K. (2013). Factors associated with depression among injured workers with disabilities returning to work in South Korea : Comparison of return to pre-injury job with work at a new firm. *Korean Journal of Social Welfare Research*, **37**, 149-174.

Shin, J. S., Lee, K. H., Kim, K. S. and Lee, Y. I. (2011). The impact of perceived social stigma on depression among people with disabilities living in Choongbuk. *Journal of Community Welfare*, **37**, 361-385.

Verbeke, G. and Molenberghs, G. (2009). *Linear mixed models for longitudinal data*, Springer Verlag, New York.

Journal of the Korean Data & Information Science Society
2017, **28**(5), 1055–1067

http://dx.doi.org/10.7465/jkdi.2017.28.5.1055
한국데이터정보과학회지

An analysis of depression of the individuals with disabilities using repeated measurement data†

Haesun Hong[1] · Jib Huh[2]

[12]Department of Statistics, Duksung Women's University

Received 1 August 2017, revised 25 August 2017, accepted 5 September 2017

Abstract

Most previous works to study for the depression of the disabilities in Korea have analyzed the repeated measured data of each individual under the mutually independent assumption. In this study, Korea Welfare Panel data of the disabilities surveyed additionally every three years are analyzed to detect the significant exploratory variables by the linear mixed models. A suitable correlation matrix is considered for the dependency of repeated measurement of each individual. The random effect to reflect the characteristics of the individuals as well as the fixed effect is included in the fitted linear mixed model. By the residual plot of the fixed effect model, the problem that the averages of residuals of each individual do not seem to be around zero is described. Further, the residual plot and the Q-Q plot coming from the selected final model are shown that the problem is modified well.

Keywords: Correlation matrix, information criterion, Korean Welfare Panel Data, linear mixed model, random effect.

† This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education. (No. 201602880001)

[1] Researcher, Department of Statistics, Duksung Women's University, Seoul 01369, Korea.
[2] Corresponding author: Professor, Department of Statistics, Duksung Women's University, Seoul 01369, Korea.