

A Study of Underweight Adolescents Characteristics and Health Promotion by Applying Data Mining Techniques

Sun Mi Shin, In Sook Kim¹, Young Moon Chae², Joo Hyung Kim¹, Hee Woo Lee³

Graduate School of Public Health, Yonsei University
College of Nursing, Yonsei University¹
Graduate School of Health Science and Management, Yonsei University²
Seoul School Health Center³

Abstract

The purpose of this paper is to describe general characteristics of underweight adolescents and to search for ways to promote the health of underweight adolescents through assessing health related factors by using data mining techniques. The study sampled(n=4352) 1,180 underweight(BMI<18.5) and 3,172 average weight (18.5<=BMI<23) adolescents, 10th grade students in Seoul, 2000, and investigated the differences between two groups. Related variables were input in a decision tree and an association rule of SAS E-Miner. The most predictable model was CART. In frequency, the proportion of underweight adolescents was higher on the south of the Han-river than on the northern side; but in association rule, associated variables with high support rate and confidence rate were females, north of Han-river, and scoliosis. Therefore, approaches for health promotion of underweight adolescents are not only intervention of physical health, but also the education of proper weight perception to prevent low birth weight and underweight adolescents because mother's education and child's low birth weight are related to underweight adolescents. In conclusion, the following sample groups in Seoul are suggested: female adolescents with scoliosis on the north of Han-river in Seoul. (*Journal of Korean Society of Medical Informatics* 8-3,61~69, 2002)

Keyword : Underweight, Data Mining, Health Promotion

I.

가

1.

가

1)

가

. 1999

7,342

2

2.7%, 0.5%

18.8%,

3.4%

가

2)

97

3

8,100

97

27.5%

98 265%, 99 14.8%

97 11.8%

98 12.1%

99 30.3%

Chi-square test

가

5.0%

97 1.4%

98 1.5%

3

4 가

3)

100

, CHAID^A, C4.5^B,

0.9

CART^C

가

90%

80%

가

가

II.

4)

1.

5)

6)

가

A. Chi-squared Automatic interaction Detection :

B. C45 : ID3(Iterative Dichotomizer 3)

C. Classification and Regression trees : 가

2

24
 1 10,300 3)
 14 , 5 fold Cross-validation 가
 , CHAID, C4.5, CART
 가
 1 5,188
 BMI (Body Mass Index kg/m²) 185
 1,180 BMI 18.5 23 4) 가 CART
 3,172 , 4,352
 (Train set) 3,264
 (75%), 가 (Validation set) 1,088 (25%) 5)
 가

III.

3

SAS(version 8.1) SAS Enterprise Miner(version
 3.0) 1) 5,188
 (BMI 18.5) 22.74%
 27.94%, 21.05%
 가
 2) (BMI 25) 6.96%

Table 1. Characteristics of study subjects

		(%)	(%)	X ² value	P-value
	356	(33.74)	699	(66.26)	
	824	(24.99)	2,473	(75.01)	30.97 <.001
	1,180	(27.11)	3,172	(72.89)	
	36	(32.73)	74	(67.27)	
	402	(25.69)	1,163	(74.31)	
	9	(37.50)	15	(62.50)	9.80 0.020
	70	(20.29)	275	(79.71)	
	517	(25.29)	1,527	(74.71)	
	1,154	(26.94)	3,130	(73.06)	
	26	(38.24)	42	(61.76)	4.32 0.037
	1,180	(27.11)	3,172	(72.89)	
	1,123	(26.69)	3,084	(73.31)	
	41	(38.32)	66	(61.68)	7.15 0.007
	1,164	(26.98)	3,150	(73.02)	

*

9.89% 가 6.0% 가 1.7 , 1.15 가 가 (Table 2).

2) 가 2

1) 가 가

26.94% 38.24% 가 가 (Table 3).
26.69% 38.32% 가 가 (Table 1).

Table 3. Input variables in model assessment

	BMI	X-Ray
	/	/
	/	/
	/	/
	/	/
	/	/
	/	/
	/	/
	/	/
	/	/
	/	/
	/	/
	/	/
	/	/
	/	/

가

3) 1.52 0.52

가 1.14 1.68

Table 2. Results of logistic regression analysis

		Parameter Estimate	Odds ratio	Pr>Chisq	95% wald confidence limits
	()	0.21	1.52	<.001	1.31-1.77
(母)	()	-0.45	0.52	0.003	0.32-0.84
	()	0.14	1.14	0.050	0.91-1.43
	()	0.25	1.68	0.030	0.36-0.97
	()	0.26	1.70	0.000	1.14-2.53
	()	0.07	1.15	0.030	1.00-1.31
가	()	-0.02	0.98	0.830	0.73-1.31
	()	0.06	1.22	0.460	0.92-1.60
	()	-0.08	0.74	0.390	0.50-1.09

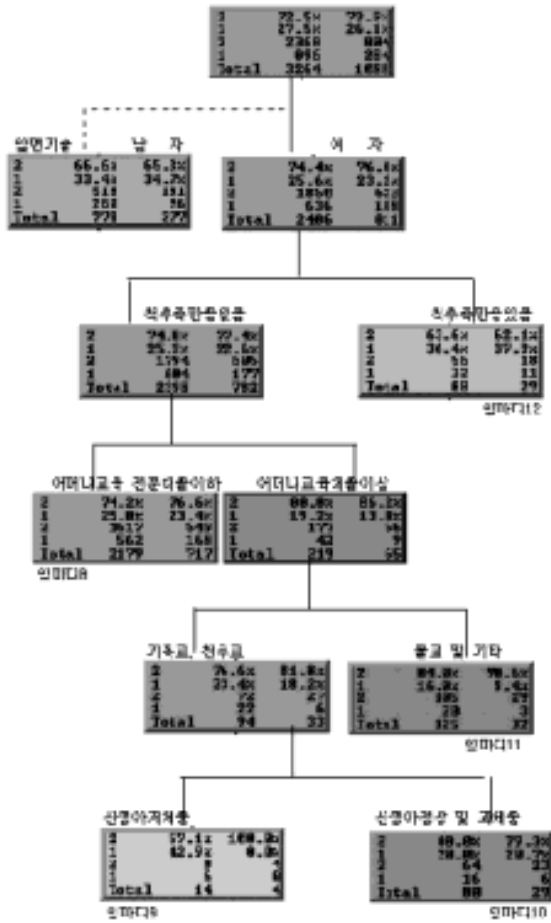
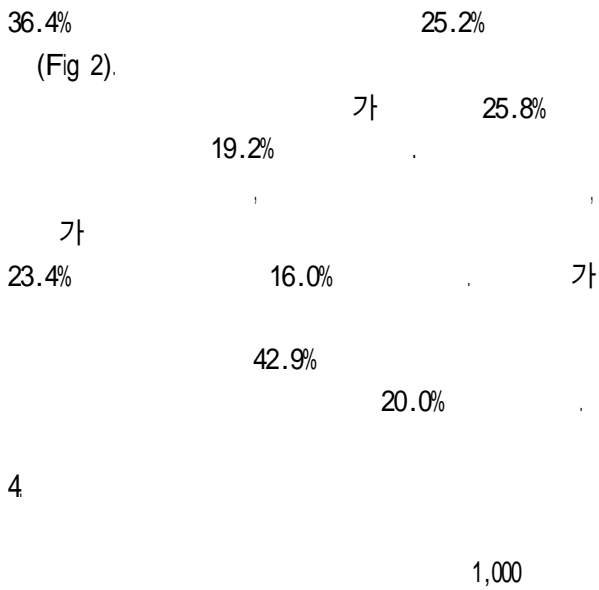


Fig 2. Decision tree in CART model(Female)



가 (Pr(A B)/N)

8)

(Support)

가

(Pr(A B)/P(A)) (Confidence) 8)

(Lift

: Pr(A B)/Pr(A)Pr(B) 1 가

가

1

가

8)

가

가 25

(Table 5).

가

10.53%

100%가

1.55 가

가

가

가 5.04%

가

가 100%가

가

가 11.42%

가 99.67%가

가 309

(99.67%)가

가

가 15.41%

가 99.04%

가

가

가

가 1,000

가

가

Table 5. Association rule in underweight adolescents

Confidence	Support	Lift	Count	Rule	
100	10.53	1.55	285	&	& ==>
100	7.57	1.55	205	&	& ==>
100	5.73	1.55	155	&	& ==>
100	5.04	1.55	137	& 가	& ==>
99.67	11.42	1.54	309	&	& ==>
99.54	8.17	1.54	221	&	& ==>
99.41	6.28	1.54	170	&	& 1 ==>
99.31	5.36	1.53	145	&	& ==>
99.27	5.02	1.53	136	&	& ==>
99.19	9.13	1.53	247	&	& ==>
99.09	16.26	1.53	440	&	==>
99.04	15.41	1.53	417	&	& ==>
98.58	7.72	1.52	209	&	& ==>
98.56	10.12	1.52	274	&	& ==>
98.45	7.06	1.52	191	&	& ==>
98.32	8.68	1.52	235	&	& ==>
98.15	5.91	1.45	160	&	& ==>
98.14	5.87	1.52	159	&	& ==>
97.53	5.84	1.44	158	& 1	& ==>
97.42	6.98	1.44	189	&	& ==>
97.40	5.54	1.44	150	&	& ==>
97.38	6.87	1.44	186	& 1	& ==>
97.27	9.24	1.44	250	&	==>
97.00	5.98	1.43	162	&	& ==>
96.96	8.28	1.43	224	&	& ==>

CART

가 가 가 가 가

IV. 가

가 0.52

가 Paricio¹⁰⁾ 가 가 가

9) 1.13

(p-value 0.88).

(miss classification)

가 25

가
12

가

가

가 가

가

가

(nonlinearity) 가
가

가

가

(interaction)

¹²⁾

가

¹¹⁾

²⁾

가

-

(-BMI 185)
23)

(-BMI 185

1. Report on national health and nutrition survey, Korea lipid and Atherosclerosis. 2001;11(3):408-410
2. Lee. H. W, Shin. S. M. A study of health-related behavior in 10th grade students of 12 high schools located in Seoul, yearbook of School Health, 1999;29:
3. Hankyoreh newspaper. 2000.125
4. Huh K. B. Pathogenetic Heterogeneity of Type 2 Diabetes Mellitus in Korea The Journal of Korean Diabetes Association. 1999;23(1):62-69
5. the Reuters News. 1999.12.10
6. Behrman RE, Kliegman RM, Arvin AM, Nelson textbook of pediatrics, 15th ed. W.B. Saunders Company, Philadelphia, 1995:169-172
7. Kim, Y.D. Growth of Tree Model by one side purity. Korean intelligent information system society, 2000;7(1):17-25
8. Choi K. R. Theory and Practice of Datamining, Chonggu , Seoul, 2000:116-143
9. Rikimaru T. Risk factors for the prevalence of malnutrition among urban children in Ghana. J Nutr Sci Vitamino(Tokyo), 1998;44(3):391-407
10. Paricio JM. Health examination of children from the democratic Sahara Republic(North West Africa) on Vacation in Spain. An Esp pediater, 1998;49(1):33-38
11. Lee T. H, Shin. S. M. Health Promotion Behavior and Related Psychosocial variables among High School Students in Seoul. The journal of Korean Community Nursing, 1998;11(1):459-467
12. Kang H. H. Datamining Free Academy, Seoul, 2000;1:153-16

