

도요지 추정을 위한 로지스틱회귀분석 및 GIS 활용 예비연구 Preliminary study on application and design of Cultural Heritage Site Management System

이진영^{1)*}, 양동윤¹⁾, 김주용¹⁾, 정계옥²⁾, 나중화³⁾

Jin-young Lee^{1**}, Dong-yun Yang¹⁾, Ju-yong Kim¹⁾, Kye-ok Jung²⁾, Jong-hwa Na³⁾

¹⁾한국지질자원연구원 지질환경재해연구부, ²⁾문화재청 발굴조사과, ³⁾충북대학교 응용통계학과
*연락처 e-mail : boks99@hanmail.net, phone : 042-868-3066

서론

자연 및 인문환경으로 구성된 지리적 공간은 인간의 활동에 따라 그리고 시간의 흐름에 따라 여러 가지 공간적 특성을 나타낸다. 특히 지리공간에서 나타나는 위치, 분포 그리고 공간적 양상은 공간문제로 인식되어 연구의 대상이 되고 있다. 이러한 공간에 대한 연구는 공간 속에 존재하는 규칙과 원리에 대한 탐구과정에서 해결 가능하다. 특히 공간을 구성하는 구조적 해석이 필요하며 이를 위해서는 공간 속에 존재하는 규칙과 원리에 대한 탐구를 통해 공간구조를 이해해야 한다. 공간구조란 어떤 분포를 나타내고 있는 공간현상들이 체계적인 일련의 관계를 가지고 배열된 입지적 형태를 말한다.

때문에 공간에 대한 일반적인 기초자료로부터 분석적이고 예측적인 자료를 산출하는 다양한 분석이 적용된다. 결국 결과로서 얻어지는 2차적인 정보와 지식을 토대로 다양한 공간계획을 수립하고 공간의사결정, 공간변화예측 등을 포함하는 과학적이고 합리적이며 적정한 공간문제에 대한 의사결정이 이루어진다. 예측 모델링은 주로 통계적 방법을 이용하여 앞으로 일어날 수 있는 문제 (What if)에 대한 가능한 결과에 대한

답을 모색하는 것이다. 주로 통계적 확률론이 적용되며 공간적 회귀분석, 로지스틱 모델이 사용된다.

연구자료 및 방법

공간분석기법에 사용될 자료는 도요지가 분포하는 지역을 분석한 자료가 기초가 되었다. 이는 현재 도요지가 분포하는 입지를 통계적으로 처리한 자료로서 현재 지형에 분포하는 과거의 도요지 위치를 의미하며, 일반적인 도요지의 특성을 분석에 활용하는 것을 의미한다.

회귀분석이란 두 연속형(서열, 등간, 비율척도 등)변수의 상관관계를 살펴보는 차원이 아니라, 관찰된 연속형 변수에 대해 독립변수와 종속변수 사이의 선형식을 구하고 그 식을 이용하여 독립변수가 주어졌을 때 종속변수를 예측하는 분석방법이다. 1개의 종속변수와 1개의 독립변수 사이의 관계를 분석할 경우를 단순회귀분석(Simple Regression Analysis), 1개의 종속변수와 여러 개의 독립변수 사이의 관계를 규명하고자 할 경우를 다중회귀분석(Multiple Regression Analysis)이라고 한다(노형진, 2004).

분석에 활용되는 변수는 단일 지점의 확

를계산을 위해 하나의 값으로 정의 되어야 한다. 하지만 분석결과자료는 특정 면적에 대한 통계자료를 기초로 하게 되므로 단일한 상수를 대푯값으로 선정해야만 사용이 가능한 상태이다. 때문에 활용되는 변수에 대한 대푯값은 각각의 변수 특성을 고려하여 선정하였다. 고도의 경우 최대고도, 최소고도, 평균고도 가운데 평균고도를 대푯값으로 선정하였다. 또한 사면의 경사는 최소경사, 최대경사, 평균경사 가운데 도요지의 입지에 대한 최소한의 경사가 의미상 입지조건에 맞는 것으로 하여 최소경사를 대푯값으로 선정하였다. 사면의 방향을 나타내는 사면향의 경우 20° 구간별 통계자료를 활용하였기 때문에 사면방향을 그대로 사용할 수 있었다. 때문에 분석에서는 0~360° 방향을 그대로 활용하여 구간을 선정하였다. 수계망 분석을 통해 얻어진 수계와의 최소거리, 최대거리, 평균거리는 의미상으로 최소거리가 입지에 활용되는 것이 타당할 것으로 판단된다. 또한 수계와의 비교의 경우도 평균 비교와 최대 비교의 경우 그 입지적 특성에 맞지 않으므로 최소비교를 입지조건을 대표변수로 활용하여 구간을 선정하였다. 도로망 분석결과에 경우는 현재 알려진 도요지 입지분포 특성을 감안할 때 현재의 도로에 대한 분석결과 보다는 과거의 교통로 관점에서 접근하는 것이 타당하다. 이는 현재의 교통로는 과거 도요지가 분포할 당시의 교통로와는 많은 차이가 있으므로 현재로서는 분석의 알고리즘과 현재 입지에 대한 통계적 특성을 파악하는데 주로 사용을 하고 본 연구에서는 제외하고자 하였다.

였다. 다만 과거 교통로의 구체적 자료가 확보되는 경우에는 이를 분석에 넣어 활용하는 것이 분석 결과의 신뢰도를 높이는 데 유용할 것으로 판단된다. 이와 같이 각각 선정된 변수는 '평균고도', '최소경사', '사면방향', '수계와의 최소거리', '수계와의 최소비교' 등으로 총 5개의 변수이다.

연구 결과

전체자료(N=1,280)에 대해 로지스틱 회귀분석을 실시한 결과는 다음과 같다. 먼저 모형에 포함되는 중요한 변수로는 6개의 변수(유적밀도:DENSITY, 고도:ALT, 도로와의 비교:R_ALT, 수계와의 비교:W_ALT, 수계와의 거리:W_DIST, 사면경사:SLP)가 선택되었으며, 이 가운데 변수 (DENSITY, ALT, R_ALT)가 변수 (W_ALT, W_DIST, SLP)들 보다는 좀 더 중요도가 높은 변수로 판단할 수 있다.

a. 모형에 대한 검정

아래의 결과는 위에서 언급한 6개의 변수를 포함하는 로지스틱 회귀모형이 통계적으로 유의한 모형임을 나타내고 있다. (p<.0001; 표 1)

b. 각 변수의 유의도

적합된 로지스틱 회귀모형의 각 변수별 유의도를 검정한 결과이다. 6개의 변수 모두가 유의한 것으로 나타나고 있다. 특히 (ALT, DENSITY, R_ALT) 변수의 유의도가 상대적으로 높게 나타난다(표 2).

Table 1. 모형에 대한 검정결과

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
-2 LOG L	4194.967	3497.303	697.664 with 30 DF (p<.0001)

Table 2. 각 변수의 유의도 분석결과

Effect	DF	Wald	Pr >
		Chi-Square	Chi-Square
ALT	5	61.3501	<.0001**
DENSITY	5	215.3825	<.0001**
R_ALT	5	38.1770	<.0001**
SLP	5	15.5511	0.0082**
W_ALT	5	17.5343	0.0036**
W_DIST	5	18.1258	0.0028**

Table 3. 각 변수의 중요도

Step	Effect	DF	Number	Score	Pr >
	Entered		In	Chi-Square	Chi-Square
1	DENSITY	5	1	552.8	<.0001**
2	ALT	5	2	51.1441	<.0001**
3	R_ALT	5	3	27.3539	<.0001**
4	W_ALT	5	4	26.0155	<.0001**
5	W_DIST	5	5	21.7027	0.0006**
6	SLP	5	6	15.8318	0.0073**

c. 각 변수의 중요도

아래의 결과는 단계별(stepwise) 변수선택법을 사용하여 모형을 적합할 경우, 모형에 포함되는 변수의 순서를 나타낸 것으로, 먼저 선택되는 변수일수록 중요도가 높은 변수라 말할 수 있다.

d. T-점수

아래의 그림은 각 변수들의 T-점수(T-score)를 시각적으로 표현한 것이다. T-score는 Wald-카이제곱 통계량의 절대값에 제곱근을 취한 것으로 각 변수의 중요도를 반응변수의 범주와 연관지어 나타내고 있다. 예를 들어, 중요도가 가장 높은 변수는 DENSITY이며, 반응변수인 가마터의 3번째 범주(GAMA=3)에 영향을 크게 미치는 것으로 해석될 수 있다. 아래 그림에서 중요도가 높은 변수들을 중복을 제하고 나열해보면 변수 (DENSITY, ALT, R_ALT,

W_ALT, W_DIST)들이 선택되며, 이 결과는 단계별 변수선택법에서의 변수중요도와 일치함을 알 수 있다.

위의 로지스틱 분석의 결과로부터 다음의 식 (1)~(6)을 얻을 수 있다. 이 식에서 Y =반응변수(가마터 유형)를 나타내며, 그 값에 따라 1(토기가마), 2(기와가마), 3(청자가마), 4(분청사기), 5(백자가마), 6(기타가마)를 의미한다. 예를 들어, 아래의 식 (1)은

$\log(\{\text{기타가마(6)에 속할 확률}\}/\{\text{토기가마(1)에 속할 확률}\})$ 의 추정식을 나타낸 결과이다.(이를 "토기가마에 대한 기타가마의 오즈(Odds)의 로그값"으로 표현하기로 한다.)

동일한 방법으로, 토기가마($Y=1$)에 대한 다른가마($Y=2, 3, 4, 5$)들의 $\log(\text{오즈})$ 값을 추정하면 다음의 식 2, 3, 4, 5와 같다.

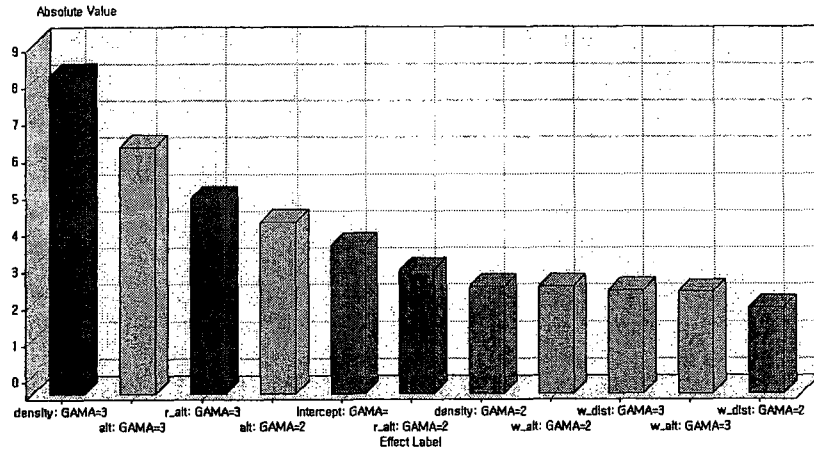


Figure 1. 각 변수의 반응변수에 대한 영향력 (T-score)

$$\log \frac{\hat{P}(Y=6)}{\hat{P}(Y=1)} = -0.7208 - 0.0023 \cdot Alt + 0.0043 \cdot Density - 0.0053 \cdot R_{alt} \quad (1)$$

$$+ 0.0506 \cdot Slp + 0.0002 \cdot W_{alt} + 0.0008 \cdot W_{dist} \equiv A$$

$$\log \frac{\hat{P}(Y=5)}{\hat{P}(Y=1)} = 0.4672 + 0.0006 \cdot Alt + 0.0175 \cdot Density + 0.0008 \cdot R_{alt} \quad (2)$$

$$+ 0.0266 \cdot Slp + 0.0005 \cdot W_{alt} + 0.0002 \cdot W_{dist} \equiv B$$

$$\log \frac{\hat{P}(Y=4)}{\hat{P}(Y=1)} = -1.0597 - 0.0014 \cdot Alt + 0.0267 \cdot Density - 0.0068 \cdot R_{alt} \quad (3)$$

$$+ 0.0674 \cdot Slp + 0.0086 \cdot W_{alt} - 0.0013 \cdot W_{dist} \equiv C$$

$$\log \frac{\hat{P}(Y=3)}{\hat{P}(Y=1)} = 0.1250 - 0.0147 \cdot Alt + 0.1252 \cdot Density + 0.0319 \cdot R_{alt} \quad (4)$$

$$+ 0.0240 \cdot Slp - 0.0272 \cdot W_{alt} - 0.0039 \cdot W_{dist} \equiv D$$

$$\log \frac{\hat{P}(Y=2)}{\hat{P}(Y=1)} = -0.2255 - 0.0094 \cdot Alt + 0.0425 \cdot Density + 0.0188 \cdot R_{alt} \quad (5)$$

$$+ 0.0357 \cdot Slp - 0.0335 \cdot W_{alt} + 0.0028 \cdot W_{dist} \equiv E$$

통계 분석을 통해 얻어진 각 요업형태별 회귀 모형은 야외에서 확인된 요업유적이 분포하는 지역을 대상으로 실험적으로 적용하였다. 현재 만들어진 회귀식이 의미하는 바와 같이 실제 요업유적이 위치하는 지역에서 어떠한 형태의 요업유적이 속하는지를 확률적으로 검토해 보는데 의의가 있다.

대상으로 선정된 지역은 서울 관악산 주변의 요업유적 분포 가능지역 2개소와 대전

성북동 일대의 요업유적 분포가능지역이다. 각각의 지역은 야외조사 및 자료조사를 통해 개략적인 위치가 확인된 지역이며, 요업의 형태도 확인된 지역이므로 확률식의 적용과 검증에 유용한 지역이다. 식 P1, P2, P3, P4, P5, P6은 각각 값에 따라 토기가마, 기와가마, 청자가마, 분청사기, 백자가마, 기타가마에 속할 확률을 의미하며 개발된 매장문화재 분석프로그램을 활용하여 결과가 도출되었다.

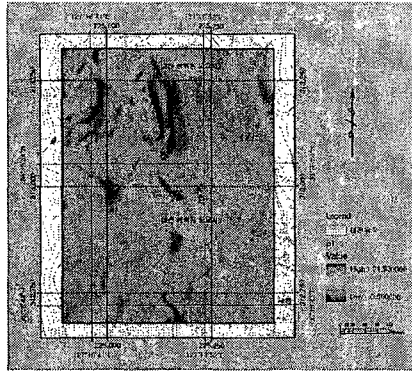


Figure 1. 토기가마유적 분포확률도

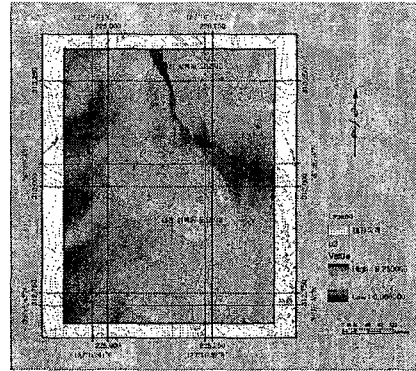


Figure 2. 기와가마유적 분포확률도

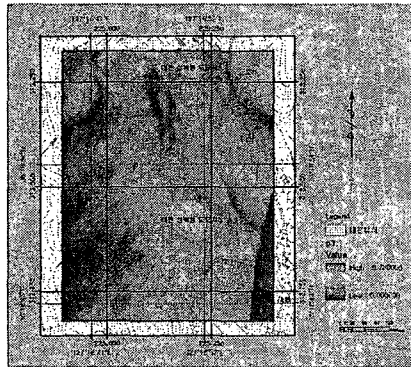


Figure 3. 청자가마유적 분포확률도

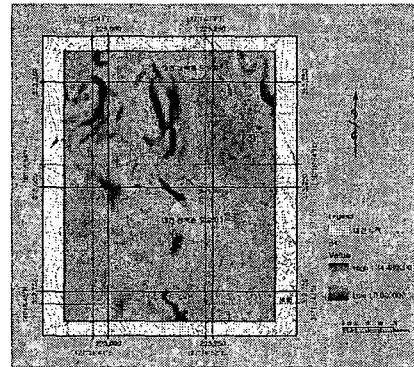


Figure 4. 분청가마유적 분포확률도

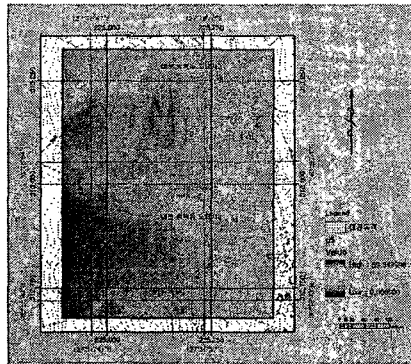


Figure 5. 백자가마유적 분포 확률도

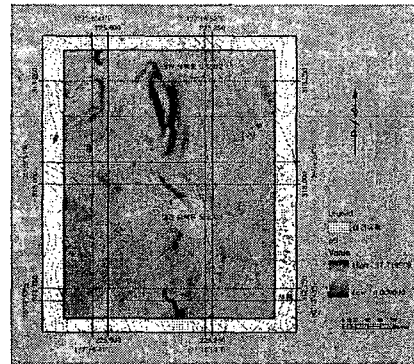


Figure 6. 기타가마유적 분포 확률도

결론

지리정보시스템(GIS)은 다양한 공간정보를 관리하고, 분석하여 문화유적의 보존과 국토의 효율적 이용을 위한 목적에 유용하게 사용될 수 있다. 본 연구는 지리정보시스템을 이용하여 전국적으로 분포하는 요지유적의 위치를 입력하고, 요지유적의 입지특성을 통계적으로 분석하여 로지스틱 회귀모형을 개발하고 이를 이용하여 도요지의

특성을 추정하고자 하였다. 이를 위하여 전국의 약 1,200개소 요지유적의 위치를 입력하였고, 각 위치별 입지분석(고도, 사면의 경사, 사면방향, 수계 및 도로와의 거리) 및 로지스틱 통계분석을 통해 요지의 입지특성을 추정할 수 있는 회귀모형을 제작하였다. 전체자료(N=1,280)에 대해 로지스틱 회귀분석을 실시한 결과 모형에 포함되는 중요한 변수는 유적의 밀도, 고도, 도로와의 거리, 수계와의 비교, 수계와의 비교, 경사 등

이 선택되었으며, 이 가운데 유적의 밀도, 고도, 도로와의 거리 변수가 수계와의 비고 및 수계와의 거리, 경사보다 중요한 변수로 판단되었다.

제작된 로지스틱 모형은 각각 값에 따라 토기가마, 기와가마, 청자가마, 분청사기, 백자가마, 기타가마에 속할 확률을 구할 수 있었으며, 요업유적으로 확인된 지역에서 어떠한 형태의 요업에 해당할 확률이 높은지를 판단하는데 유용하였다. 모형에 대한 통계적 검증결과 상위 1개 확률로 예측할 경우 49%의 정분류율을 나타냈으나, 상위 2개의 확률로 예측할 경우 63%의 정분류

율을 나타냈다. 도요지로 조사된 지역에 대해 지리정보시스템 상에 확인된 유적을 토대로 로지스틱 모형을 적용한 결과 활용 가능성이 높은 것으로 나타났다.

본 예비연구 결과를 통해 향후 로지스틱 모형과 같은 통계적 모형의 활용을 통해 문화유적 보존과 국토의 효율적 이용에 대한 목적에 충분하게 기여할 것으로 기대된다.

참고문헌

노형진, 2004, 한글 SPSS 10.0에 의한 조사방법 및 통계분석, 형설출판사.