

오늘은 지난 주와 마찬가지로 뉴질랜드 오클랜드의 인구센서스 데이터를 사용하여 실습을 진행합니다. 우선 실습에 필요한 패키지들을 `library()` 함수를 사용해 불러옵니다.

```
library(rgdal)
library(classInt)
library(RColorBrewer)
library(spdep)
```

다음으로 오클랜드 인구센서스 데이터가 담긴 CSV 파일과 CAU 경계를 나타내는 셰이프파일을 각각 다음과 같이 워크스페이스로 가져올 수 있습니다.

```
auck.df <- read.csv("auck_pop.csv", header = TRUE)
auck.sp <- readOGR("auck.shp")
```

오늘 실습의 목표는 공간적 자기상관 정도를 나타내는 모란 I 통계를 R에서 직접 계산해보는 것입니다. 모란 I의 계산에는 월요일 수업시간에 살펴본 바와 같이 `spdep` 패키지의 `moran.mc()` 함수를 사용하게 됩니다. `moran.mc()`에는 지역(여기서는 CAU) 간의 인접성이 저장되어 있는 `listw` 객체가 필요합니다. `listw` 객체는 다음과 같은 순서로 생성할 수 있습니다.

```
auck.xy <- coordinates(auck.sp)
auck.nb <- poly2nb(auck.sp, queen = TRUE)
auck.nblast <- nb2listw(auck.nb, zero.policy = TRUE)
plot(auck.sp, col = "Grey 90", border = "Grey 50")
plot(auck.nb, auck.xy, add = TRUE)
```



**Q1.** 위의 `nb2listw()` 함수에서 `zero.policy` 인자의 역할은 무엇인가요? 만약 위의 예에서 `zero.policy = FALSE`로 설정 되어있었다면 어떤 결과가 나왔을까요? [2 점]

listw 객체가 생성되었다면, 이제 이를 이용해 모란 I 통계를 계산할 수 있습니다. 아래에서는 2006 년 뉴질랜드 원주민(마오리) 인구의 공간적 자기상관 정도를 확인하고 있습니다. 여기서 nsim 인자는 통계적 유의성(p-value) 계산을 위한 시뮬레이션 횟수를 지정합니다.

```
moran.mc(auck.df$maori06, auck.nblast, nsim = 999, zero.policy = TRUE)

      Monte-Carlo simulation of Moran I

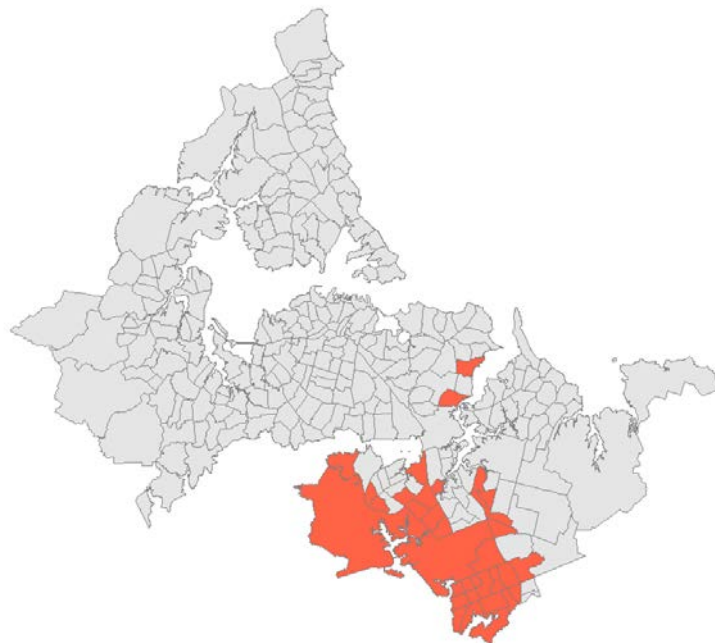
data:  auck.df$maori06
weights: auck.nblast
number of simulations + 1: 1000

statistic = 0.63344, observed rank = 1000, p-value = 0.001
alternative hypothesis: greater
```

**Q2. 위의 모란 I 통계는 퀸(Queen) 방식의 인접성 정의를 바탕으로 한 결과입니다. 지역 간의 인접성을 다른 방식으로 정의한다면 결과도 달라질 수 있을 것입니다. 록(Rook) 방식과 최근린 이웃( $k=3$ ) 방식으로 지역 간의 인접성을 정의하고, 그 결과를 사용하여 모란 I 통계를 계산하는 R 코드를 적어주세요. [3 점]**

위의 moran.mc() 함수는 전체 지역에서 전반적으로 나타나는 공간적 자기상관 정도를, 평균과 같이 하나의 숫자로 요약하여 보여줍니다. 만약 각각의 지역에서 나타나는 공간적 자기상관 정도를 확인하고 싶다면 localG()와 같은 함수를 사용하여 지역별로 공간적 자기상관을 측정해야 합니다. R 에서 localG() 함수는 moran.mc() 함수와 마찬가지로 spdep 패키지에 포함되어 있으며, 사용법도 moran.mc()와 매우 유사합니다.

```
auck.localG <- localG(auck.df$maori06, auck.nblast, zero.policy = TRUE)
sig <- which(auck.localG >= 1.96)
plot(auck.sp, col = "Grey 90", border = "Grey 50")
plot(auck.sp[sig,], col = "tomato", border = "Grey 50", add = TRUE)
```



위의 지도에서 붉은색으로 표시된 지역은 G 통계(모란 I와 유사한 또다른 공간적 자기상관 측도) 값이 통계적으로 유의미하게 높은 지역( $z \geq 1.96$ )을 의미합니다.

**Q3. 위 지도에 G 통계값이 통계적으로 유의미하게 낮은 지역( $z \leq -1.96$ )을 파란색으로 추가하여 나타내는 R 코드를 적어주세요. [5 점]**

#### 기타 평가항목

R 파일의 실행 가능 여부 및 작성 스타일(Google's R Style Guide 준수 여부): 5 점

---

각 문항에 대한 답안은 앞선 실습과 마찬가지로 R 콘솔에서 직접 실행이 가능한 형태의 스크립트 파일로 작성하여 제출합니다. 파일의 확장자는 \*.R 로 하며, source() 함수를 통해 오류 없이 실행되어야 합니다. 답안에서 R 코드 이외의 부분은 주석으로 처리하여, 오류가 발생하지 않도록 해주세요. 답안은 12 월 13 일 금요일까지 KLAS 를 통해 제출하면 됩니다.

2019 년 8 월 13 일

홍성연([syhong@khu.ac.kr](mailto:syhong@khu.ac.kr))