

Problem Set 3

Data Analysis and Visualization using R

RStudy2020

Due by 2020년 12월 7일 월요일 오전 12시 (한국시간)

Information & Instructions

출제자: 박상훈

데이터 파일

첫 번째 문제는 여기에서 2012년 미국 총선 시계열 설문조사 자료를 사용.

두 번째 문제는 여기의 'Quality of Governance'의 시계열 자료를 사용.

제출

정해진 시간까지 정해진 문제에 대한 코드를 작성한 R 스크립트를 sp23@email.sc.edu 혹은 Dropbox의 4_QnA or Discussion 폴더에 업로드.

모든 코드는 다른 작업환경에서 열더라도 재생산가능하게 디렉토리 설정 등을 모두 고려한 결과물이어야 함.

Here starts the actual test

문제 1

관심을 가지고 있는 데이터를 가지고 연속형 변수 하나, 이항변수 하나, (최소 둘 이상의 카테고리를 가진) 순위형 변수 하나의 설명변수를 가지는 함수로 종속변수를 설명하는 회귀모델을 수립하라. 선형회귀모델을 추정하라. 회귀모델을 추정할 결과로부터 얻을 수 있는 모든 정보를 서술 및 설명하라.

문제 2

x, z, y, w 의 네 연속형 변수를 가진 함수 $y = \gamma + \mu x + \alpha z + \mu x z + \beta w$ 를 관심을 가지고 있는 데이터를 이용해 추정하라. x 와 z 변수의 낮은 수준의 값을 각각 \underline{x} 와 \underline{z} 로, 높은 수준의 값을 각각 \bar{x} 와 \bar{z} 로 나타내라. 이때, 아래의 함수는 수립한 모델과 데이터에 있어 실질적으로 어떠한 함의를 전달하는가? 1

$$E(y|x = \underline{x}, z = \underline{z}) - E(y|x = \bar{x}, z = \underline{z})$$
$$E(y|x = \underline{x}, z = \bar{z}) - E(y|x = \bar{x}, z = \bar{z}).$$

문제 3

Brambor, Golder, and Clark의 논문을 읽고 정리하라. 가능한 한 자세히 논문의 함의를 정리하여 설명하라.

문제 4

King, Tomz, and Wittenberg의 논문을 읽고 논문이 주장하는 바와 제안하는 바를 정리하라. 나아가 1) 이 논문의 제목이 함의하는 바를 설명하고, 2) 예측값(predicted values)와 기대값(expected values)의 차이를 설명하라. 1차 차분(first difference)란 무엇인가? 3) 논문에서 언급한 베이지안 접근법이란 무엇인가?

문제 5

관심을 가지고 있는 데이터셋을 가지고 모델을 수립하여 King, Tomz, and Wittenberg의 Figure 1의 양식을 참조하여 그래프 두 개를 만들어라. 즉, 하나의 회귀모델, $y = \beta_0 + \beta_1x + \beta_2z + \beta_3xz + \beta_4w + \beta_5m + \beta_6mw$ 을 수립하고 추정한 뒤, x 와 z 의 효과에 초점을 맞춘 그래프를 하나, m 과 w 의 효과에 초점을 맞춘 그래프를 또 하나 작성하라. 논문의 일부를 작성하듯, 그래프가 가지는 함의를 정연하게 서술하고 설명하라.