

[REPORT-hw2]

**■ 과 목 명 : 시계열 자료 분석과 예측**

**■ 담당교수 : 김명석 교수님**

**■ 제 출 일 : 2022.10.30**

**■ 학과: 데이터 사이언스. 인공지능**

**■ 학번: A64037**

**■ 성명: 서승덕**

**1.** 다음 물음에 답하시오(사이버캠퍼스에서home.txt 다운로드).

(Note: Y: price, X: SqFt, LotSize, Baths)

**(1)** ANOVA table을보고하시오.



**(2)** R^2와 adjusted R^2를구하고, 모델의 적합성 평가를 위해서 둘 중 어느값을 쓸 것인지를

정하고, 이유를 설명하시오.



**R^2 값은 설명변수의 개수가 늘어날수록 그 값이 높아진다. 따라서 지금과 같이 x값이 여러 개인 다중회귀분석을 진행할 때에는 k를 사용하여 단순히 변수의 개수가 늘어남에 따라 설명계수가 늘어나는 것(종속변수를 설명하지 못하는 독립변수를 넣어도 높아지는 경향)을 보정해주는 adjusted R^2를 사용하는 것이 적절하다.**

**(3)** 다중공선성(multicollinearity) 검사를 위해서 VIF를 구하시오

.

**(4)** 새로운 home이 아래의 정보를 가지고 있을 때 가격을 예측하시오.

SqFt=2955, LotSize=20.2, Baths=3

**Price = 659.6985**

**2.** 다음 물음에 답하시오(사이버캠퍼스에서election1.txt 다운로드).

Choice=b0+b1Age65+b2Urban+b3ColGrad+b4Union+b5Neast+b6West+b7Seast+e

**(1)** Ols 방법으로 b5, b6, b7을 추정하여 보고하고, 각 계수가 유의수준 10%에서 통계적으로

유의한지(0과 다른지)를 p-value를 이용하여 검정하시오.

**b5 = -7.27780, p-value = 0.000474 🡺 유의함,**

**b6 = -1.92319, p-value = 0.314025 🡺 유의하지 않음.**

**b7 = -5.67146, p-value = 0.007928 🡺 유의함**.

**(2)** 추정한계수b5, b6, b7의의미를설명하시오(힌트: 기울기의 개념을 그대로 적용할지 아니면 다른 방식으로 해석할지 고려)

**기울기의 개념을 그대로 적용한다면 b5, b6, b7는 Neast, West, Seast의 기울기를 의미하게 됨.**

**각각의 변수가 독립이라면 b5+b0, b6+b0, b7+b0는 각각 집단의 Choice 평균이 됨.. 즉 b5= 집단 5의 Choice평균 – b0 , b6= 집단 6의 Choice평균 – b0, b7= 집단 7의 Choice평균 – b0이 됨.**

**일단 위의 3개의 변수는 이항형태의 독립변수이기 때문에 기울기로 계산하기 보다는 평균으로 계산하는 것이 옳다고 생각함.**

**(3)** AIC 평가 기준을 기반으로 Backward variable selection을 실시하여 최종적으로 포함하여야 할

X변수를 모두 나열하고, 그때의 AIC 값을 보고하시오.



**X변수: Age65 + Urban + ColGrad + Union + Neast + Seast, AIC=152.05**

**(4)** 새로운 state가 아래의 정보를 가지고 있을때, 위의 (3)에서 얻은 결과를 이용한 모델로 choice를 예측하시오.

Age65=17, Urban=79, ColGrad=68, Union=21 이고 area는 MidWest.

**Choice = 15.7486**