

# LGE - SNU DS 고급 과정

## 확률통계 및 공통 알고리즘

### 과제5 -회귀분석

2022. 01

※필요하다면 다음의 값을 이용하시오.

:  $F_{0.95}(1, 10) = 4.96$ ,  $t_{0.975}(10) = 2.23$

**문제 1.** 설명변수 X(기계의 사용연수(단위:년)), 반응변수Y(기계가 마모된 정도)로부터 다음과 같은 결과를 얻었다.

$$n = 12, \bar{x} = 10, \bar{y} = 20$$

$$S_{xx} = 100, S_{xy} = 90, S_{yy} = 97$$

이를 바탕으로 한 단순선형회귀모형이 선형성, 독립성, 등분산성, 정규성의 가정을 만족한다고 하자.

(a) 회귀 직선을 구하고 아래 분산분석표의 빈칸을 이 모형에 대해서 채워넣어라.

요인	자유도	제곱합	평균제곱	F값
회귀	( )	( )	( )	( )
잔차	( )	( )	( )	
계		( )		

(b) F통계량을 이용해 회귀직선의 유의성을 유의수준 5%에서 검정하여라.

**문제 2.** 위 문제1에 적합된 회귀모형에서, 연 평균 마모도의 변화율에 대한 95% 신뢰구간을 구하고 그 의미를 설명하여라.

**문제 3.** 아래의 표는 당뇨병 진행도 (target)를 반응변수, 나이(age), 성별(sex; 남:1, 여:0), bmi지수 (bmi), 혈압(bp)을 설명 변수로 하여 다중회귀모형을 적합한 결과이다. 이를 바탕으로 아래의 물음에 답하여라.

OLS Regression Results						
Dep. Variable:	target	R-squared:	0.400			
Model:	OLS	Adj. R-squared:	0.395			
Method:	Least Squares	F-statistic:	72.91			
Date:	Tue, 29 Dec 2020	Prob (F-statistic):	2.70e-47			
Time:	21:57:24	Log-Likelihood:	-2434.2			
No. Observations:	442	AIC:	4878.			
Df Residuals:	437	BIC:	4899.			
Df Model:	4					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	156.8913	3.979	39.433	0.000	149.072	164.711
age	37.2412	64.117	0.581	0.562	-88.776	163.258
sex	-10.1592	5.922	-1.716	0.087	-21.798	1.480
bmi	787.1793	65.424	12.032	0.000	658.594	915.764
bp	416.6738	69.495	5.996	0.000	280.088	553.259
Omnibus:	9.858	Durbin-Watson:	1.933			
Prob(Omnibus):	0.007	Jarque-Bera (JB):	6.464			
Skew:	0.146	Prob(JB):	0.0395			
Kurtosis:	2.485	Cond. No.	31.8			

- 데이터로 적합한 회귀직선의 식을 나타내고 그 의미를 설명하여라.
- F통계량을 이용해 회귀직선의 유의성을  $\alpha = 0.05$ 에서 검정하여라.
- 각각의 회귀계수 및 상수항의 유의성을  $\alpha = 0.05$ 에서 검정하여라.
- 회귀모형의 결정계수를 구하고 그 의미를 설명하여라.