

계명대학교

통계전산실습

# 기말 프로젝트

통계학과 5463312 안효준  
통계학과 5526369 안유나



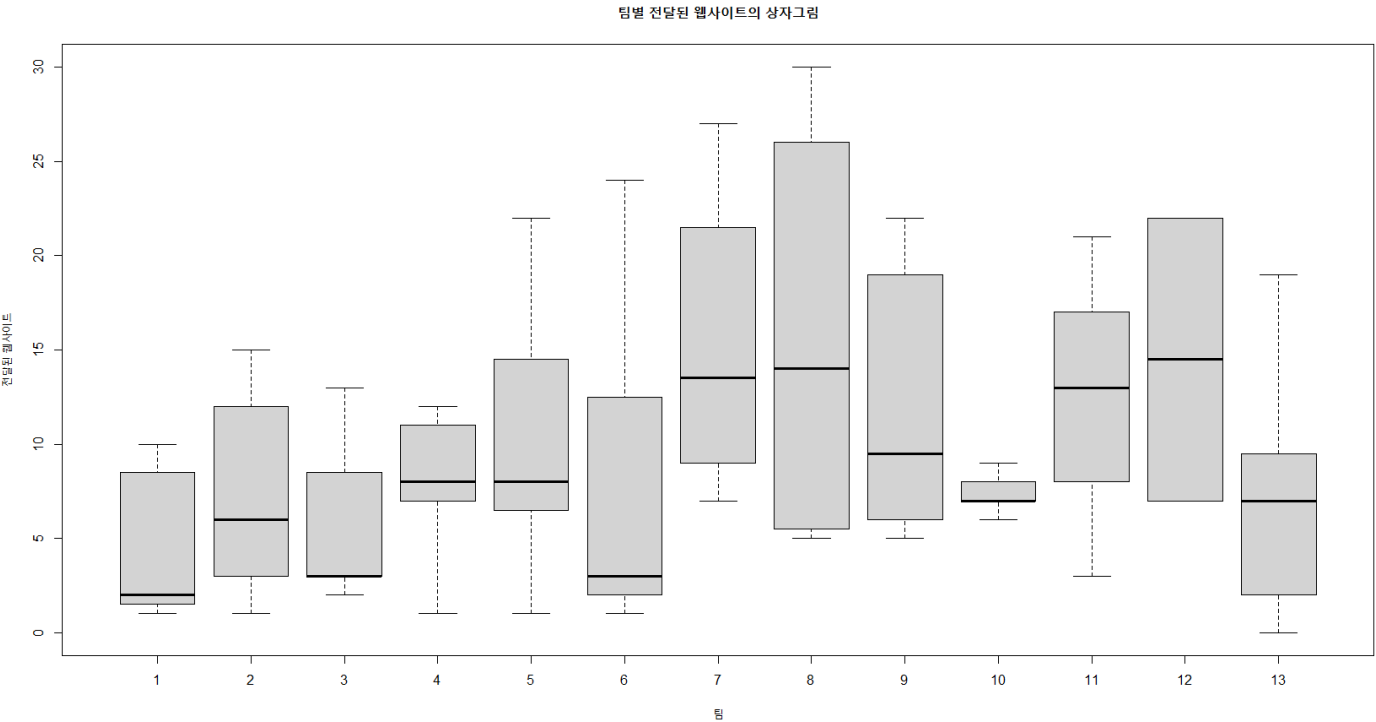
# 00. 목차

01	02	03	04
분석 목적	R을 활용한 분석	Python을 활용한 분석	회귀분석
Purpose	R program	Python	Regression

## 01) 분석 목적

- 2001년1월부터 2002년 8월까지 각 3명으로 구성된 13개 웹사이트 개발팀 데이터
- 고객에게 전달된 웹사이트 수에 가장 큰 영향을 미치는 변수파악 목적
- 통계적 분석을 통해 전달된 웹사이트 수에 영향을 미치는 최적의 변수 파악

01) 팀 번호에 따른 전달된 웹사이트 수의 중앙값 비교



- 정규성검정: 전달된 웹사이트 수 Shapiro-Wilk

검정통계량	유의확률
0.9026	3.467e-05

Anderson-darling

검정통계량	유의확률
2.2302	1.034e-05

Kruskal-Wallis 검정

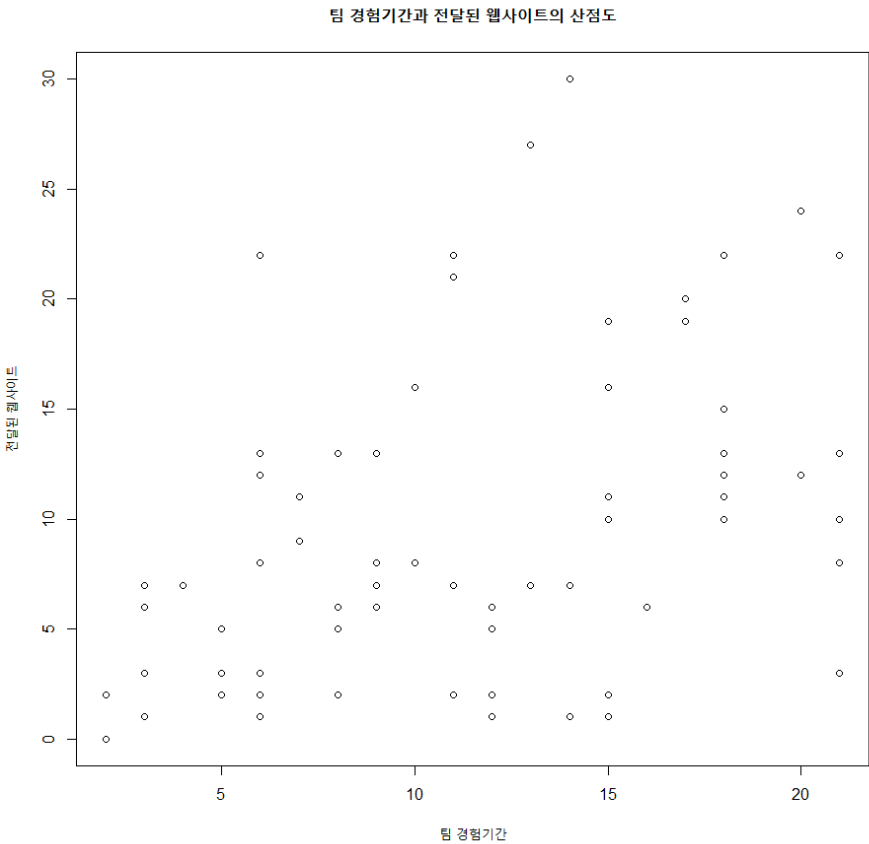
- 팀에 따른 역량의 차이 파악
- 비모수적 검정 방법

$H_0 : median_1 = median_2 = \dots = median_{13}$   
 $H_1 : \text{중앙값이 다른팀이 적어도 한 팀 있다.}$

Kruskal-Wallis

검정통계량	유의확률
12.905	0.376

02) 팀 경험기간과 전달된 웹사이트의 상관 분석



- 정규성검정: 팀 경험기간  
Shapiro-Wilk

검정통계량	유의확률
0.9483	0.0047

Anderson-darling

검정통계량	유의확률
0.92807	0.0175

Spearman의 순위 상관 분석

- 팀 경험기간과 전달된 웹사이트의 상관분석
- 비모수적 검정 방법

$H_0 : \rho = 0$  (상관관계가 없다.)

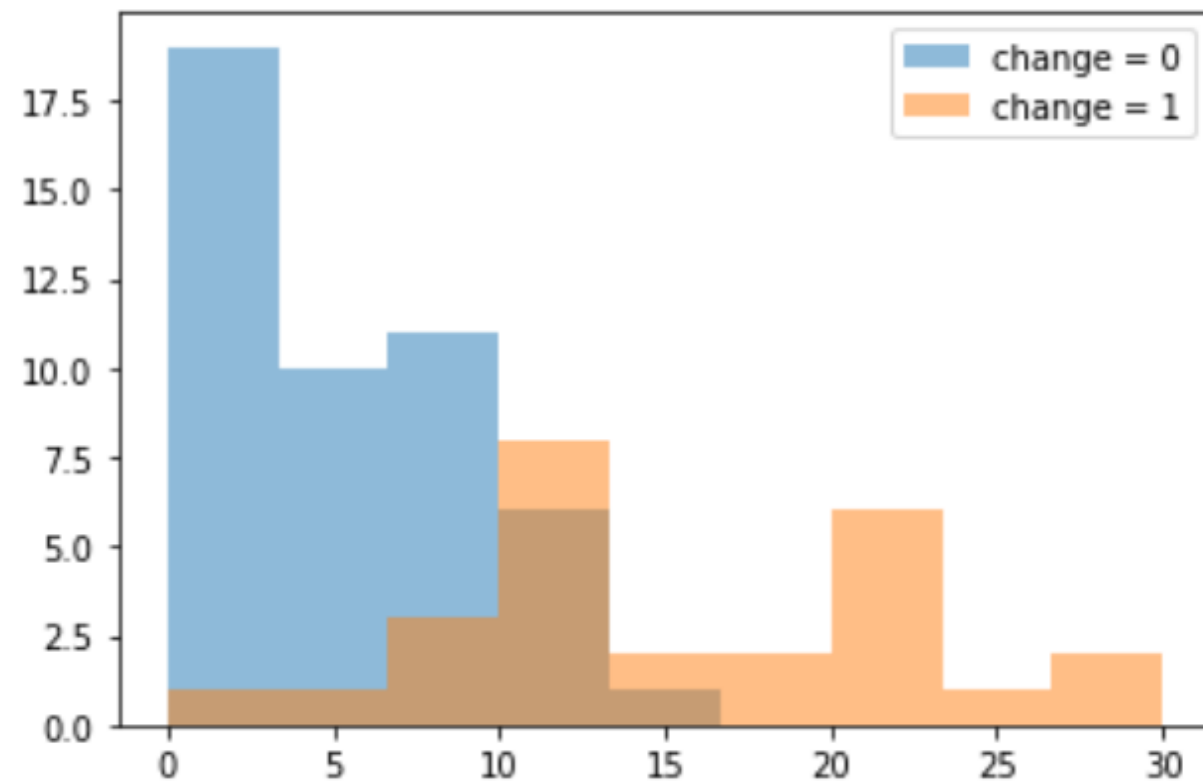
$H_1 : \rho \neq 0$  (상관관계가 있다.)

Spearman's rank correlation

검정통계량	유의확률
34354	2.719e-05

$\hat{\rho} = 0.4700$

## 01) 절차변경에 따른 전달된 웹사이트 수의 중앙값 비교



- 정규성검정: 절차변경 여부 별 전달된 웹사이트 수
- |             | Shapiro-Wilk | Anderson-darling |
|-------------|--------------|------------------|
| 전달된 웹사이트 수  | 0.9786       | 0.2111           |
| 비전달된 웹사이트 수 | 0.9999       | 0.1411           |

집단	검정통계량	유의확률	집단	검정통계량	유의확률
절차변경=0	0.9288	0.0069	절차변경=0	1.0320	0.733
절차변경=1	0.9684	0.5819	절차변경=1	0.3741	0.705

# Wilcoxon Ranksum 검정

- 절차변경에 따른 웹사이트 수 비교
- 비모수적 검정 방법

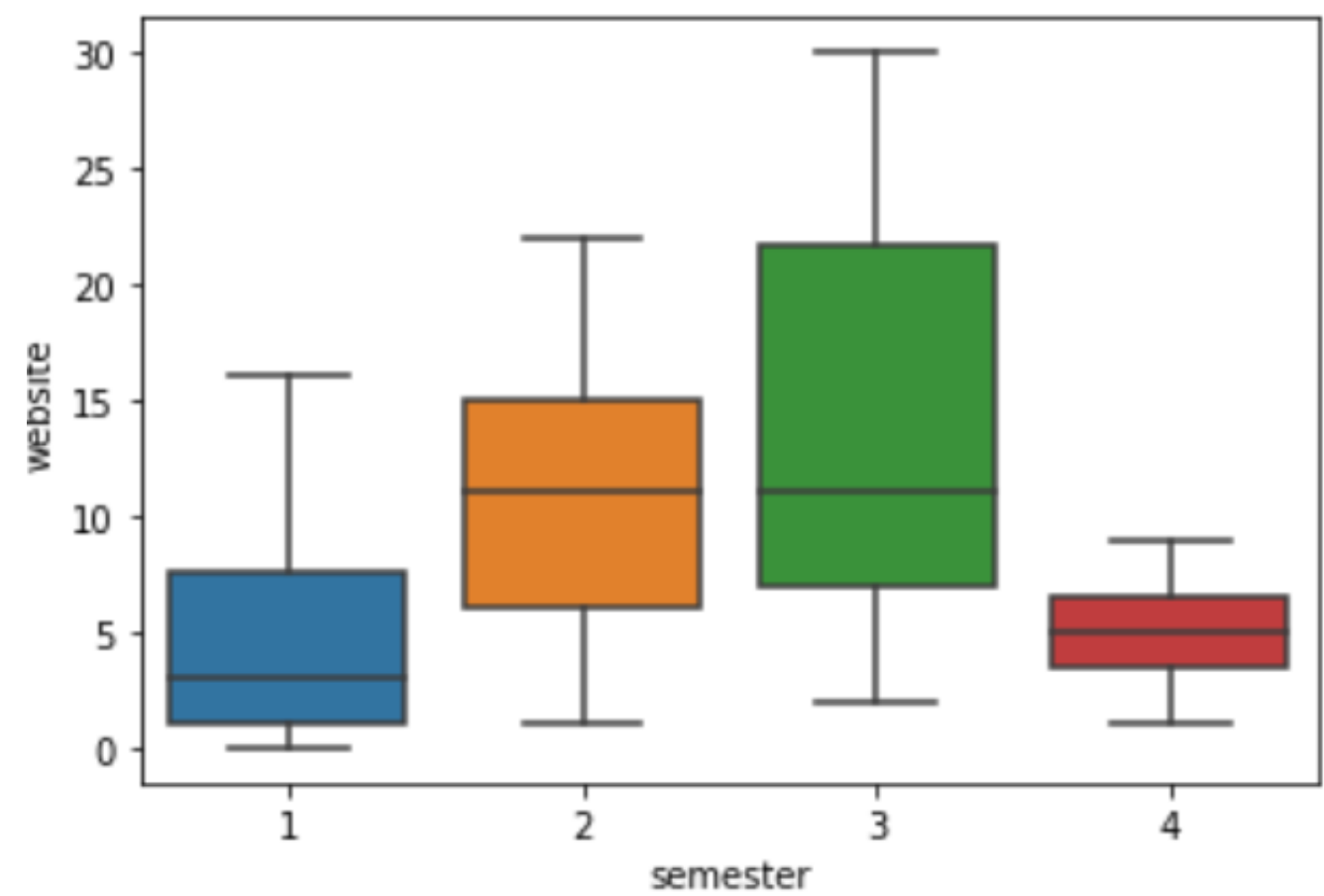
$$H_0 : median_1 = median_2$$

$$H_1 : median_1 \neq median_2$$

## Wilcoxon Ranksum

검정통계량	유의확률
-7.7646	1.9446e-11

02) 분기 에 따른 전달된 웹사이트 수의 중앙값 비교



- 정규성검정: 절차변경 여부 별 전달된 웹사이트 수
- Shapiro-WilkAnderson-darling

집단	검정통계량	유의확률
semester=1	0.8849	0.0262
semester=2	0.9508	0.3525
semester=3	0.9027	0.0336
semester=4	0.9327	0.4389

집단	검정통계량	유의확률
semester=1	0.8072	0.69
semester=2	0.3065	0.694
semester=3	0.9245	0.696
semester=4	0.4165	0.68

Kruskal-Wallis 검정

- 분기에 따른 전달된 웹사이트 수 차이 파악
- 비모수적 검정 방법

$H_0 : median_1 = median_2 = median_3 = median_4$   
 $H_1 : not H_0$

Kruskal-Wallis

검정통계량	유의확률
18.6325	0.0003

사후검정 : Bonferroni(유의수준=0.05/6 =0.008)

- 1분기와 2분기, 1분기와 3분기, 3분기와 4분기에서 차이가 존재

### 01) 회귀분석

- 각 변수가 전달된 웹사이트 수에 미치는 영향 파악하고자 회귀분석 진행
- 팀 번호 변수 제외
- 범주형 변수 factor화
- 수치형 변수 Min-Max Scale 방법으로 스케일링

- Full-model

*website*

$$= \text{intercept} + \text{balance} + \text{exper} + \text{change1} + \text{year2002} + \text{semester2} + \text{semester3} + \text{semester4}$$
$$+ \text{balance} : \text{exper} + \text{balance} : \text{change1} + \text{balance} : \text{year2002} + \text{balance} : \text{semester2}$$
$$+ \text{balance} : \text{semester3} + \text{balance} : \text{semester4} + \text{exper} : \text{change1} + \text{exper} : \text{year2002}$$
$$+ \text{exper} : \text{semester2} + \text{exper} : \text{semester3} + \text{exper} : \text{semester4} + \epsilon$$

- 수정된 결정계수가 낮고, VIF값이 매우 높게 나타남  
-> 변수제거 진행



02) 변수선택

- 선택지 생성 : 전진선택법, 후진소거법, 단계선택법
- 성능 평가: Adj-R2, RMSE , AIC, BIC

전진선택법, 단계선택법

$website = intercept + change1 + semester3 + year2002 + \epsilon$

성능지표

성능지표	검정통계량
Adj-R2	0.4936
RMSE	0.168
AIC	-47.3601
BIC	-35.9085

VIF 표

변수	VIF
Chane1	2.3561
Semester3	1.1656
Year2002	2.1301

후진소거법

$website = intercept + balance + exper + change1 + semester3 + semester4 + balance : exper + balance : semester3 + exper : semester4 + \epsilon$

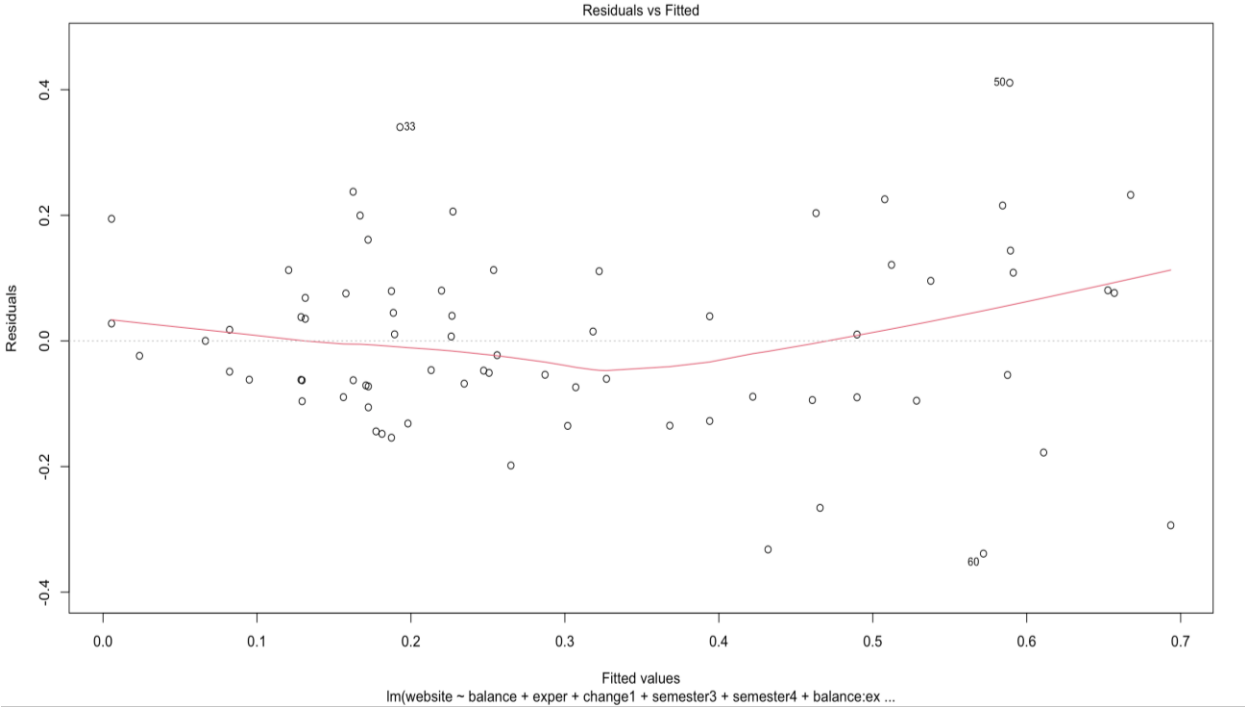
성능지표

성능지표	검정통계량
Adj-R2	0.5723
RMSE	0.1544
AIC	-55.1857
BIC	-32.2811

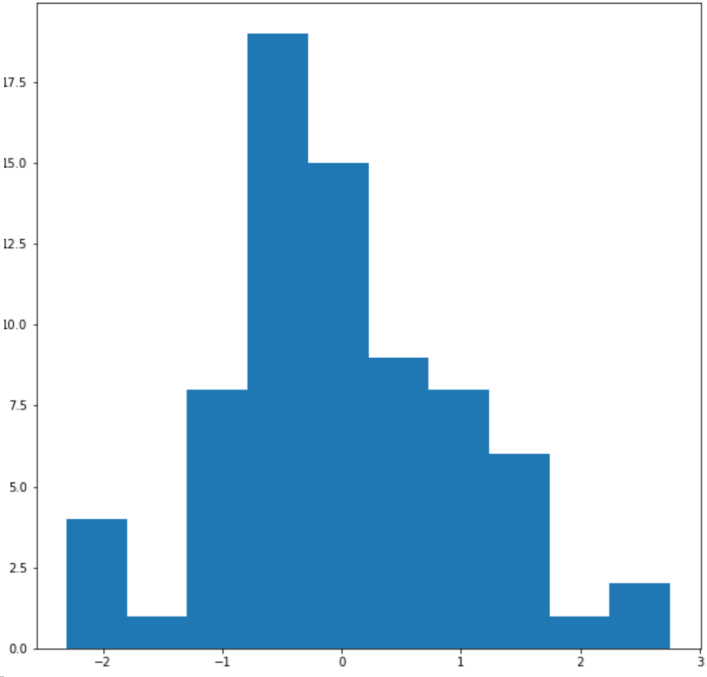
VIF 표

변수	VIF
balance	4.8194
exper	31.6299
Change1	2.1242
Semester3	34.5208
semester4	7.7108

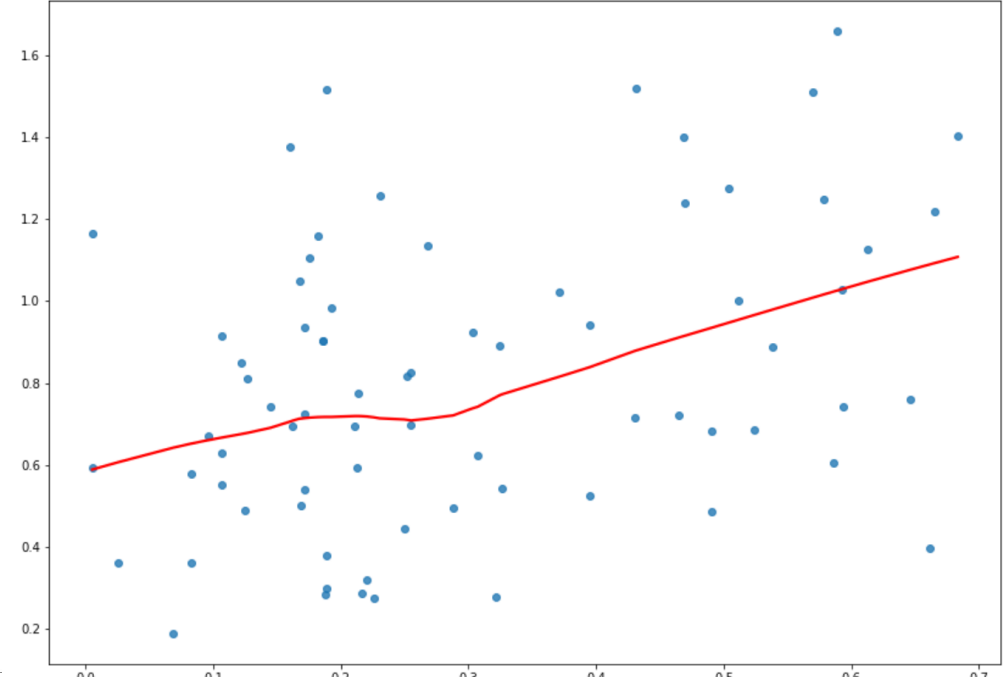
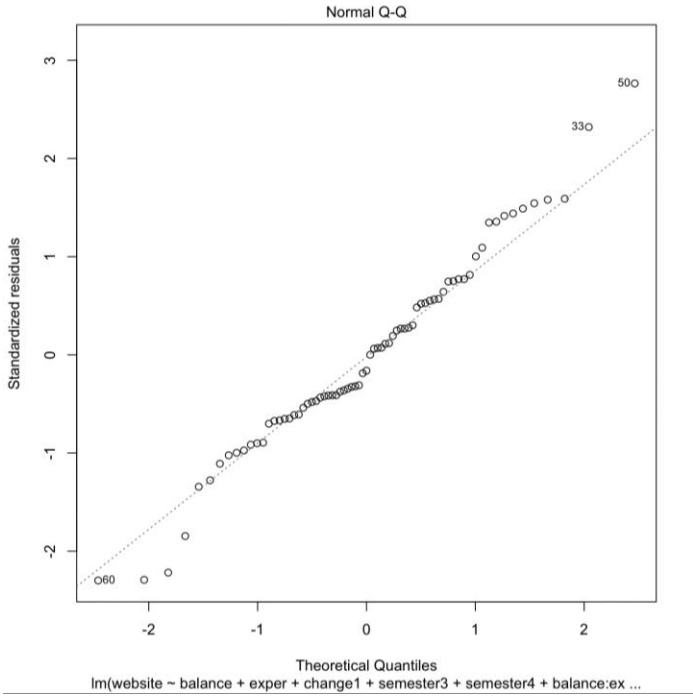
03) 잔차분석



모델의 선형성



잔차의 정규성



잔차의 등분산성과 독립성

### 04) 결론

#### 최종모델

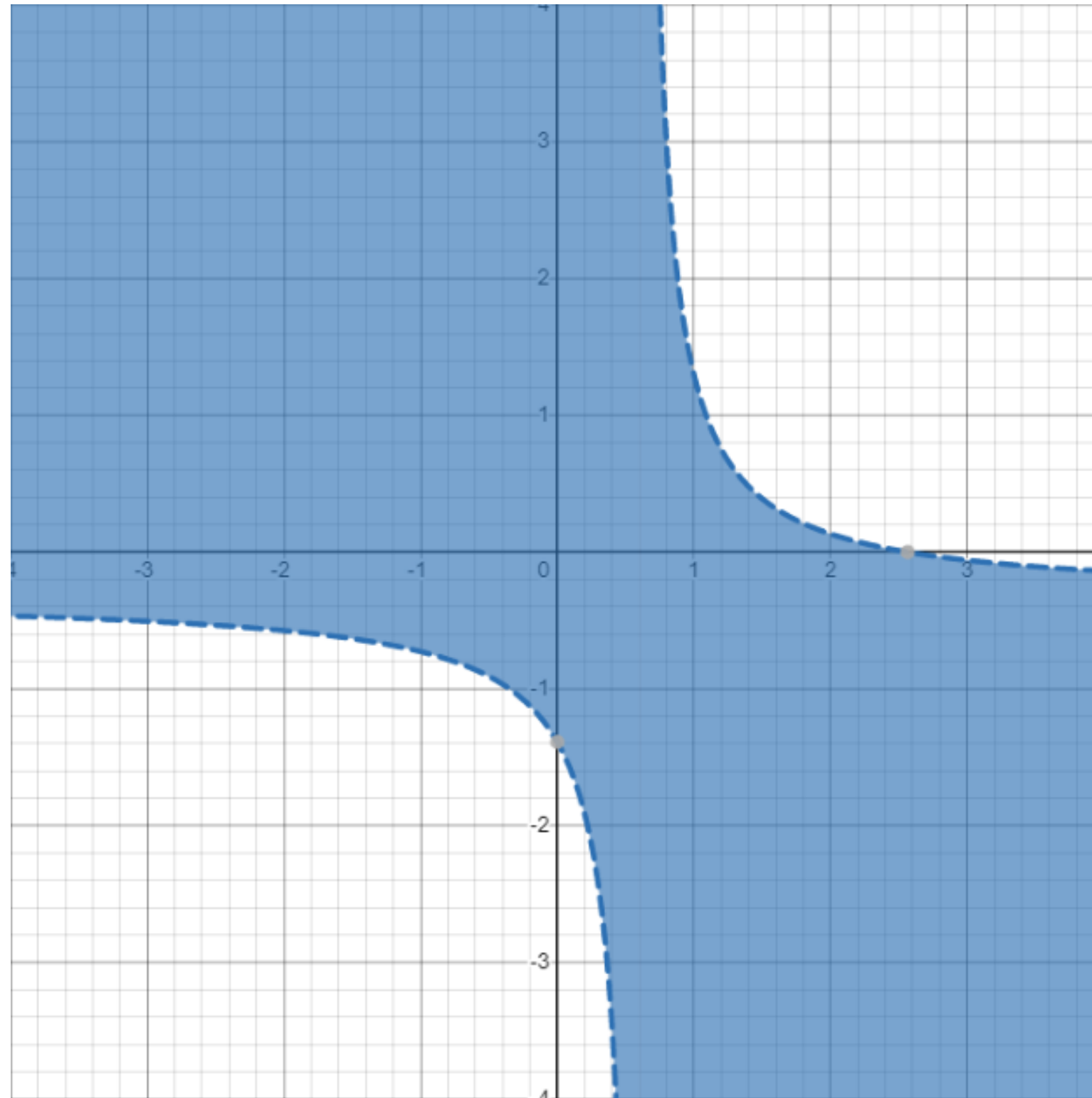
$$\begin{aligned} website = & -0.15 + 0.60 \times balance + 0.72 \times exper + 0.37 \times change1 \\ & + 0.63 \times semester3 + 0.21 \times semester4 - 1.19 \times balance : exper \\ & - 0.99 \times balance : semester3 - 0.57 \times exper : semester4 \end{aligned}$$

- 주문 잔고, 팀 경험기간, 절차변경, 3분기 | 4분기
- 주효과의 회귀계수 모두 양수



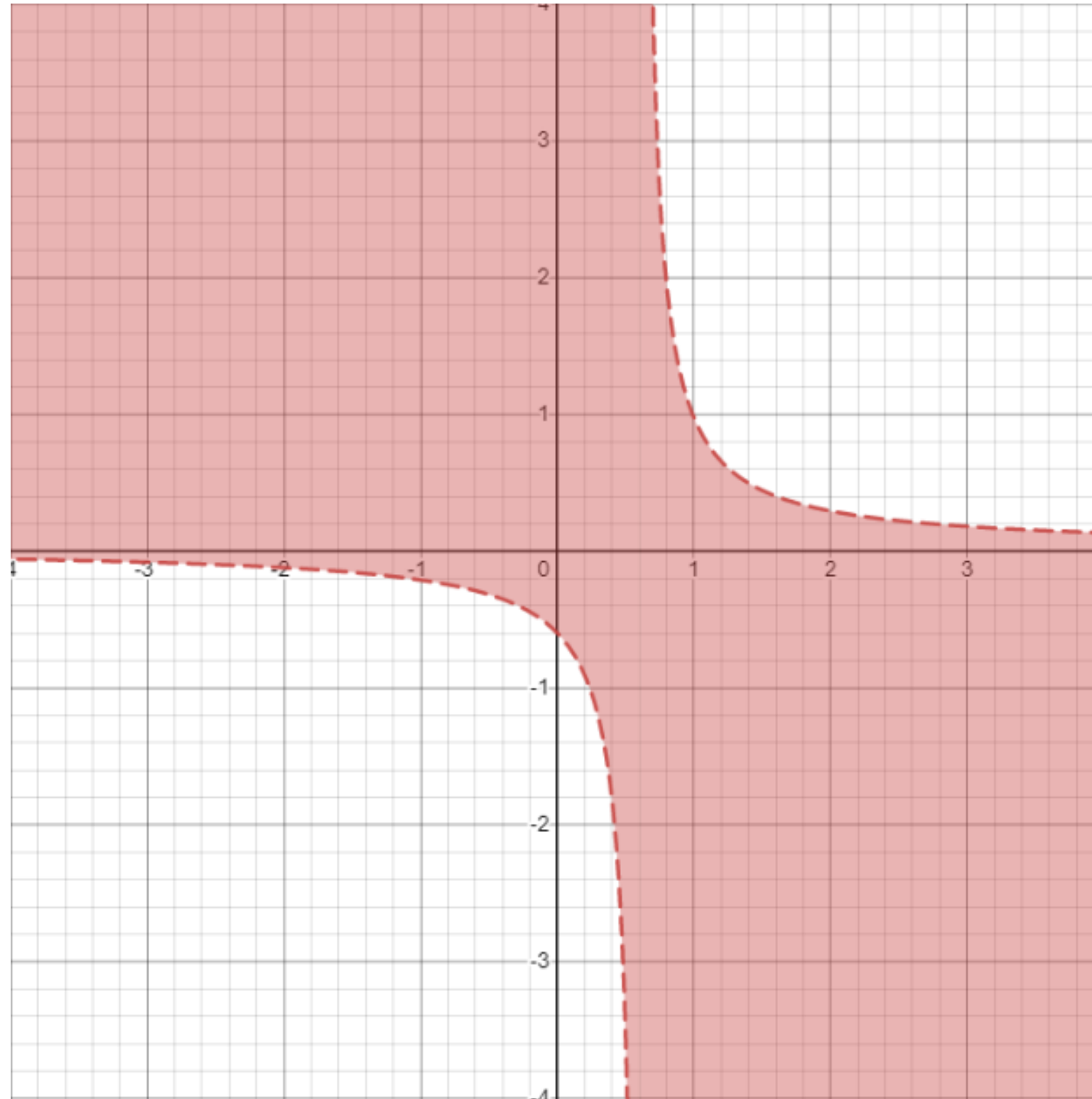
감사합니다.





X축: balance (주문잔고)  
Y축: exper (경험기간)

최종 모형식에서  $\text{change1} = 1$ ,  $\text{semester3} = 1$ ,  $\text{semester4} = 0$ 을 대입



X축: balance (주문잔고)  
Y축: exper (경험기간)

최종 모형식에서  $\text{change1} = 1$ ,  $\text{semester3} = 0$ ,  $\text{semester4} = 1$ 을 대입