



Produce 101

연습생들의 데뷔 요인

박주원 이민경 이세윤

I 서론

- 01 분석목적
- 02 자료설명

II 본론

- 01 매출액 제외 분석
- 02 매출액 포함 분석

III 결론

- 01 최종 11인 예측
- 02 한계점

I 서론

- 01 분석목적
- 02 자료설명

II 본론

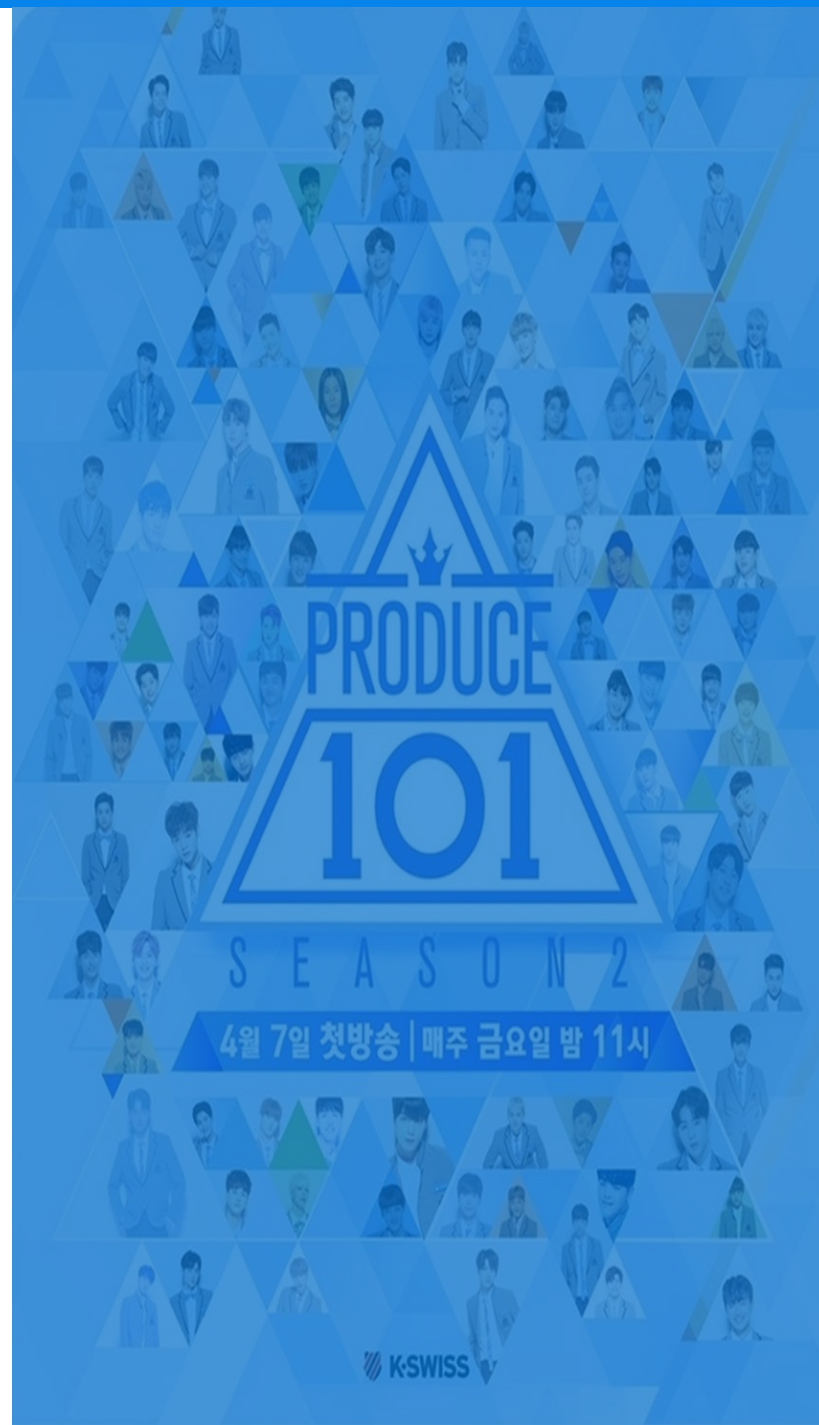
- 01 매출액 제외 분석
- 02 매출액 포함 분석

III 결론

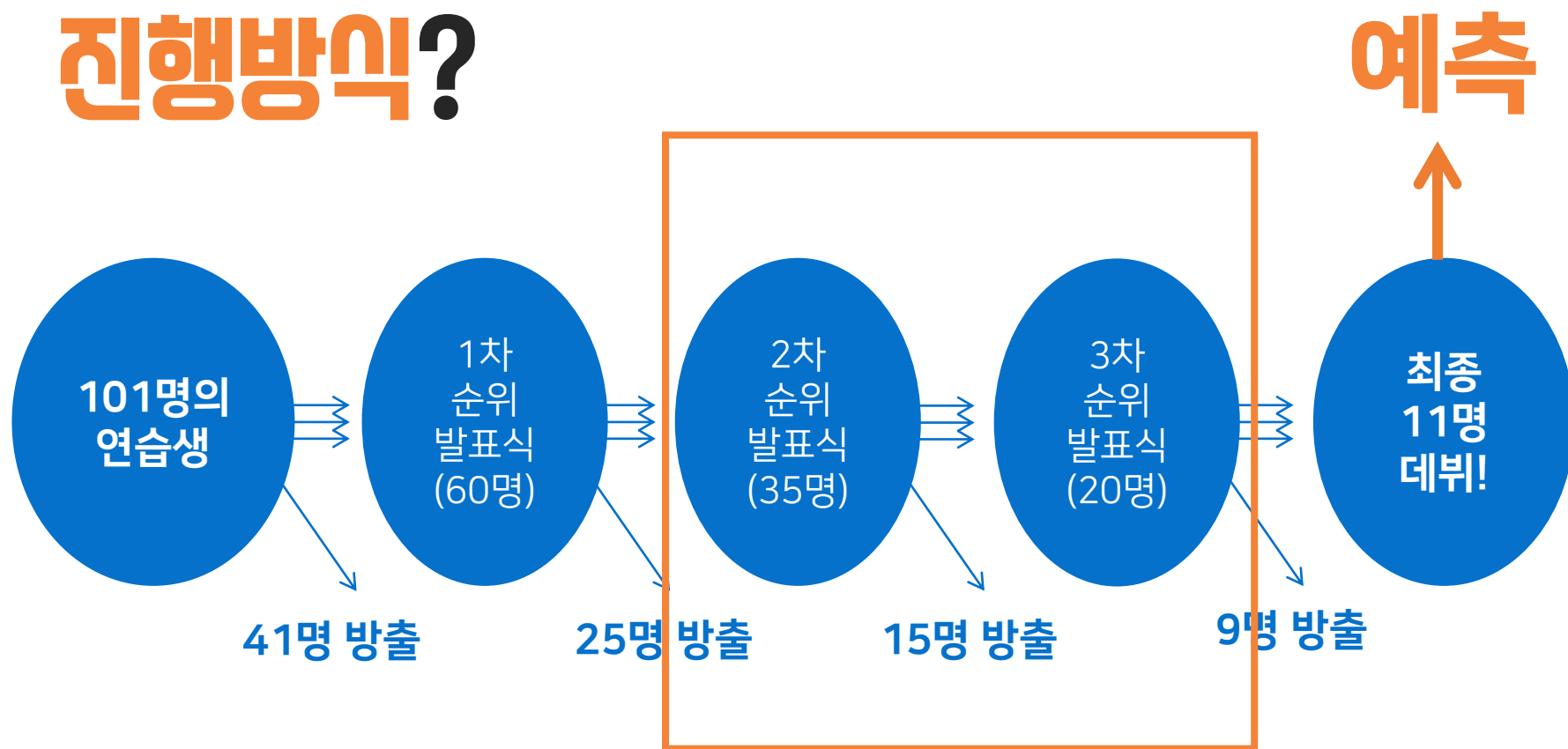
- 01 최종 11인 예측
- 02 한계점

프로그램 프로듀스 101?

올해 4월에 방영한 101명의 연습생들이
11명 안에 진입하여 데뷔하기 위해 벌이는
경쟁을 담은 서바이벌 프로그램.



프로그램의 진행방식?



변수 소개

I(지시변수): 리더 및 센터 경험 유무(有:1, 無:0)

X1(등급): 연습생들의 실력을 나타내는 척도 (A:5, B:4, C:3, D:2, F:1)

X2(중간 누적 득표수): 1차, 2차 순위 발표식에서 받은 득표수

X3(동영상 조회수): 네이버 TV 캐스트의 공식 홈페이지에 올라오는 동영상 조회수

X4(동영상 하트수): 네이버 TV 캐스트의 공식 홈페이지에 올라오는 동영상 하트수

X5(분량): 프로그램 내에서 얼굴이 비춰진 시간 (단위: 초)

X6(매출액): 각 연습생들의 소속사의 매출액 (단위: 백만 원)

Y(반응변수): 3차 순위 발표식에서 받은 득표수

(데이터 기준) 동영상 조회수 및 하트수: 2017년 11월 9일 17:35 기준, 매출액: 2016년 1~4분기 총 매출액
(출처) 매출액: dart(전자공시시스템)

변수 소개

네이버 TV 캐스트 조회수 및 하트수(X3,X4)



구독 96,980명

✓ 구독

♡ 14,714

공유

N스토어

홈 재생목록 동영상

✓ 최신 ✓ 재생 ✓ 좋아요

[단독/직캠] 일대일아이컨택 | 강다니엘 - 슈퍼주니어 #Sorry Sorry_2조 @그룸베를

1,928,220 36,642

[단독/직캠] 일대일아이컨택 | 강다니엘 - 슈퍼주니어 #Sorry Sorry_2조 @그룸베를

2:43

[단독/직캠] 일대일아이컨택 | 박지훈 - BTS #상남자_1조 @그룸베를

2:48

[4회] *여기는 완성형* 이번 열차는 미안형~ | 슈퍼주니어 #Sorry_2조

2:46

[단독/직캠] 일대일아이컨택 | 김재현 - 슈퍼주니어 #Sorry Sorry_2조

2:42

[단독/직캠] 일대일아이컨택 | 김종현 - 슈퍼주니어 #Sorry Sorry_2조

2:43

[단독/직캠] 일대일아이컨택 | 강동호 - BTS #상남자_2조 @그룸베를

2:48

[단독/직캠] 일대일아이컨택 | 박우진 - 2PM #10점 만점에 10점_2조

2:47

[단독/직캠] 일대일아이컨택 | 황민현 - 슈퍼주니어 #Sorry Sorry_2조

2:43

[단독/직캠] 일대일아이컨택 | 홍성우 - 슈퍼주니어 #Sorry Sorry_2조

2:44

[단독/직캠] 일대일아이컨택 | 라이관린 - BTS #상남자_2조 @그룸베를

2:47

오앤요 | 장은복 | 힘통령의 귀환 @ 자기소개_1분 PR

1:53

자료 소개

자료1 :
매출액을 포함하지 않은
35개의 관측치 자료

이름	지시변수(I)	X1(등급)	X2(누적 득표수)	X3(조회수)	X4(하트수)	X5(분량)	반응변수(y)
강다니엘	1	5	2766092	24049235	221124	1164	828148
강동호	1	2	2057844	6438440	80042	1336	314807
권현빈	1	1	1054211	2349568	37944	1232	160693
김동한	1	4	804990	1162527	21947	248	120594
김동현	1	4	984643	981128	26488	454	125155
김사무엘	1	5	2170369	22650231	106225	1894	378491
김상균	1	2	983108	1006451	24954	809	140890
김예현	0	3	851187	280297	6988	393	53277
김용국	1	2	1386354	1511067	51537	692	184303
김재환	1	4	2392811	2774929	82310	799	259776
김종현	1	4	3547640	3681892	85210	1292	367052
김태동	1	5	1236939	1659199	29170	438	110091
노태현	1	5	1298357	2702459	40504	1183	143523
라이관린	1	1	2919940	11989642	62957	1257	188940
박우담	0	2	938913	473296	9212	881	31975
박우진	1	5	1800579	6760219	109336	1018	372493
박지훈	1	4	3226037	45480357	220069	1500	630198
배진영	0	1	2170203	12888503	65122	569	389982
서성혁	0	2	964971	465725	11233	659	54415
안형섭	0	5	1982078	2579545	49878	994	254984
웅성우	1	5	2818035	3536786	67648	1139	358656
유선호	0	1	2077540	1789701	45325	733	209168
윤지성	1	2	2217481	2359696	53062	1483	333974
이건희	1	2	1069762	694152	17620	943	50148
이대휘	1	5	2904798	4276024	80668	1485	325990
이우진	1	5	1347590	902664	22980	1150	47796
이의웅	1	2	1376467	1463977	28182	540	156572
임영민	1	5	2324781	2720153	64070	1076	197721
장문복	0	1	1342139	3210834	26131	1446	130324
정세운	1	4	1918401	2827844	71125	1175	196223
주학년	0	1	2332567	1930802	43553	892	197194
최민기	1	1	1578810	3152849	50867	614	217734
타가다 켄타	1	4	1071123	1845915	30783	384	146176
하성운	1	5	1298561	1742874	79831	741	413654
황민현	1	2	2684529	3896681	81313	692	315650

자료 소개

자료2 :
매출액을 포함한
24개의 관측치 자료

이름	지시변수(I)	X1(등급)	X2(누적 득표수)	X3(조회수)	X4(하트수)	X5(분량)	X6(매출액)	반응변수(y)
강다니엘	1	5	2766092	24049235	221124	1164	1841	828148
강동호	1	2	2057844	6438440	80042	1336	23100	314807
권현빈	1	1	1054211	2349568	37944	1232	8352	160693
김동현	1	4	984643	981128	26488	454	3961	125155
김사무엘	1	5	2170369	22650231	106225	1894	1120	378491
김상균	1	2	983108	1006451	24954	809	9709	140890
김예현	0	3	851187	280297	6988	393	3073	53277
김용국	1	2	1386354	1511067	51537	692	500	184303
김재환	1	4	2392811	2774929	82310	799	1	259776
김종현	1	4	3547640	3681892	85210	1292	23100	367052
라이관린	1	1	2919940	11989642	62957	1257	19027	188940
박우진	1	5	1800579	6760219	109336	1018	3961	372493
박지훈	1	4	3226037	45480357	220069	1500	18243	630198
배진영	0	1	2170203	12888503	65122	569	1554	389982
옹성우	1	5	2818035	3536786	67648	1139	7029	358656
유선호	0	1	2077540	1789701	45325	733	19027	209168
윤지성	1	2	2217481	2359696	53062	1483	1841	333974
이건희	1	2	1069762	694152	17620	943	5248	50148
이대휘	1	5	2904798	4276024	80668	1485	3961	325990
임영민	1	5	2324781	2720153	64070	1076	3961	197721
정세운	1	4	1918401	2827844	71125	1175	22766	196223
주학년	0	1	2332567	1930802	43553	892	2573	197194
최민기	1	1	1578810	3152849	50867	614	23100	217734
황민현	1	2	2684529	3896681	81313	692	23100	315650

예상 결과

X1(등급)

X2(중간 누적 득표수)

X3(동영상 조회수)

X4(동영상 하트수)

X5(분량)

X6(매출액)

이 증가함에 따라

Y(3차 순위발표식 득표수)

가 증가 할 것이라고 예상

I(리더 및 센터 경험 여부)에 따른 모형의 기울기 변화 예상

I 서론

- 01 분석목적
- 02 자료설명

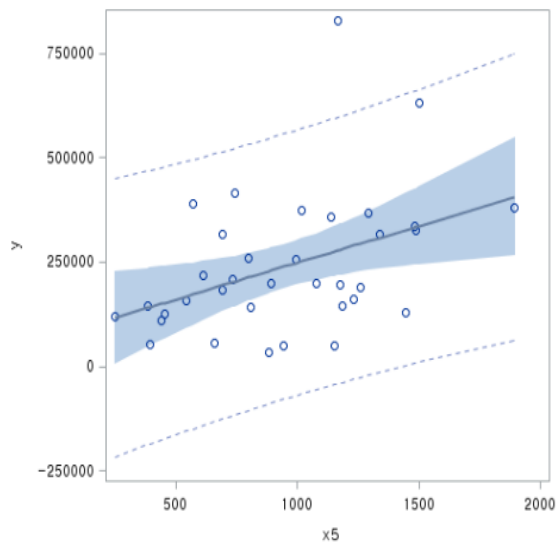
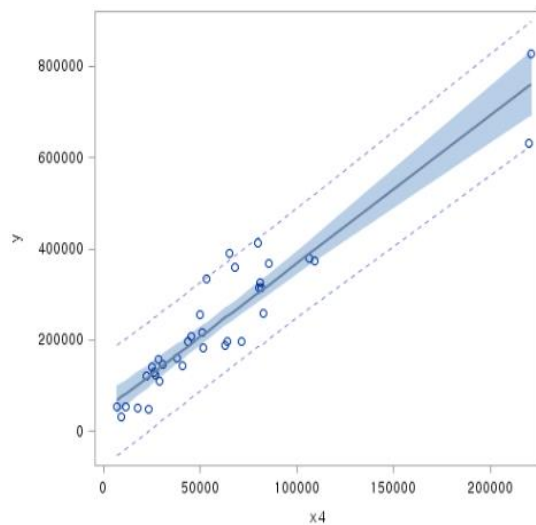
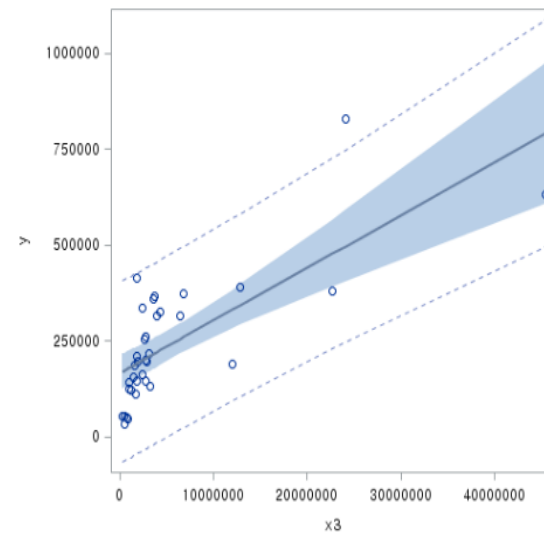
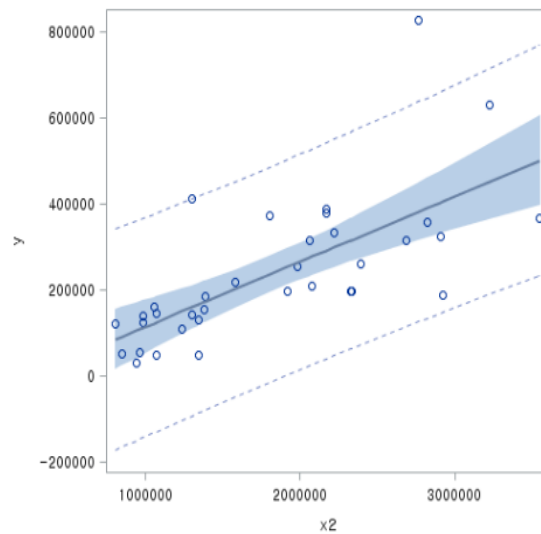
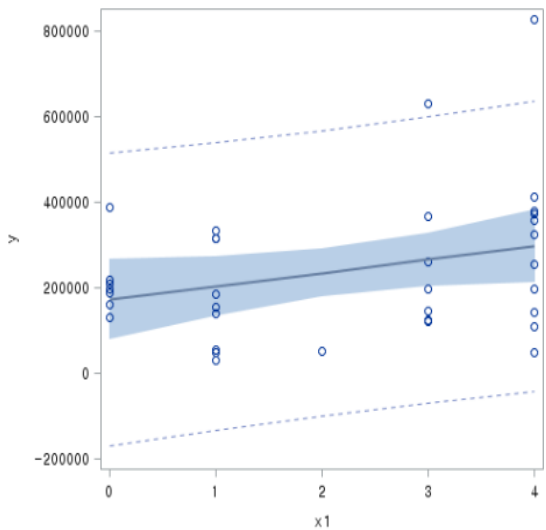
II 본론

- 01 매출액 제외 분석
- 02 매출액 포함 분석

III 결론

- 01 최종 11인 예측
- 02 한계점

단순 회귀 적합



단순 회귀 적합

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	141872	61240	2.32	0.0269
x1	1	31002	17292	1.79	0.0822

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-36935	53931	-0.68	0.4982
x2	1	0,15179	0,02732	5,56	<.0001

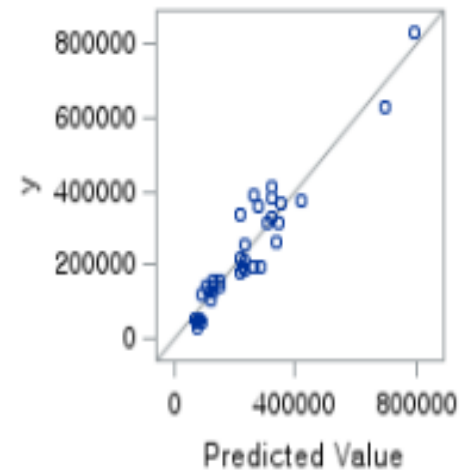
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	166198	22653	7.34	<.0001
x3	1	0,01376	0,00220	6,25	<.0001

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	45619	15670	2.91	0.0064
x4	1	3,23459	0,20415	15,84	<.0001

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	72683	69850	1,04	0,3056
x5	1	176,03591	68,11316	2,58	0,0144

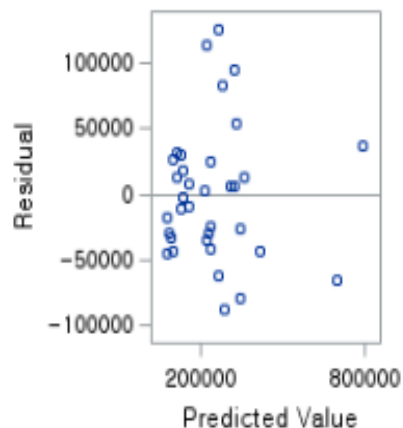
▲ 다중 회귀 적합

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	48137	35655	1,35	0,1878
i	1	-29600	25854	-1,14	0,2619
x1	1	-3095,14731	6996,76510	-0,44	0,6616
x2	1	0,01221	0,01971	0,62	0,5404
x3	1	-0,00453	0,00226	-2,01	0,0545
x4	1	3,97488	0,51249	7,76	<,0001
x5	1	-12,98578	31,05079	-0,42	0,6790



R-Square 0,9064

모형 진단



∴ 선형성 만족

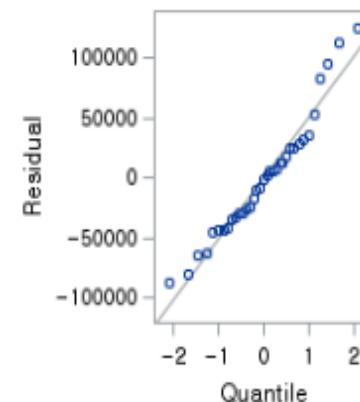
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2,93290	2,93290	1,31	0,2598
Error	33	76,54109	2,23055		
Corrected Total	34	76,54109			

$$SS_{reg} = 2,93290 / 2 = 1,46645 < \chi^2(1) = 3,841$$

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	4,72023	0,78670	0,31	0,9281
Error	28	71,82086	2,56503		
Corrected Total	34	76,54109			

$$SS_{reg} = 4,72023 / 2 = 2,360115 < \chi^2(6) = 12,592$$

∴ 등분산성 만족



정규성 검정		
통계량		p 값
W	0,967006	Pr < W

∴ 정규성 만족

변수 변환

등분안생, 언형생, 정규생
모두 만족

변수 변환 진행 X

다중 공선성 확인

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Variance Inflation
Intercept	1	48137	35655	1.35	0.1878	0
i	1	-29600	25854	-1.14	0.2619	1.31083
x1	1	-3095.14731	6996.76510	-0.44	0.6616	1.35290
x2	1	0.01221	0.01971	0.62	0.5404	2.43817
x3	1	-0.00453	0.00226	-2.01	0.0545	4.35973
x4	1	3.97488	0.51249	7.76	<.0001	6.63918
x5	1	-12.98578	31.05079	-0.42	0.6790	1.56726

✓ VIF 문제 있음

Collinearity Diagnostics (intercept adjusted)								
Number	Eigenvalue	Condition Index	Proportion of Variation					
			i	x1	x2	x3	x4	x5
1	3.00976	1.00000	0.01541	0.01463	0.02854	0.01726	0.01420	0.03393
2	1.21388	1.57463	0.26364	0.26502	0.01812	0.01007	0.00038843	0.01852
3	0.66832	2.12214	0.13585	0.07193	0.02313	0.06456	0.02600	0.45331
4	0.57997	2.27804	0.51510	0.54323	0.00167	0.01402	0.00556	0.13525
5	0.43719	2.62381	0.00168	0.00209	0.58171	0.09435	0.00039273	0.28715
6	0.09087	5.75499	0.06833	0.10310	0.34684	0.79974	0.95346	0.07184

피어슨 상관 계수, N = 35 H0: Rho=0 가정하에서 Prob > r						
	i	x1	x2	x3	x4	x5
i	1.00000	0.40451 0.0159	0.17624 0.3112	0.15039 0.3885	0.31940 0.0614	0.18608 0.2845
x1	0.40451 0.0159	1.00000	0.15096 0.3867	0.16661 0.3388	0.34175 0.0445	0.15103 0.3865
x2	0.17624 0.3112	0.15096 0.3867	1.00000	0.52696 0.0011	0.69368 <.0001	0.54777 0.0007
x3	0.15039 0.3885	0.16661 0.3388	0.52696 0.0011	1.00000	0.84797 <.0001	0.46462 0.0049
x4	0.31940 0.0614	0.34175 0.0445	0.69368 <.0001	0.84797 <.0001	1.00000	0.46834 0.0045
x5	0.18608 0.2845	0.15103 0.3865	0.54777 0.0007	0.46462 0.0049	0.46834 0.0045	1.00000

✓ 상태수 = 5.75499 문제 없음

✓ 분산비율에 문제 있어 보임



변수 선택

후진 제거법, 단계적 회귀에 의한 변수 선택

Summary of Backward Elimination							
Step	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	x5	5	0.0006	0.9058	5.1749	0.17	0.6790
2	x1	4	0.0007	0.9051	3.3975	0.23	0.6357
3	x2	3	0.0011	0.9040	1.7236	0.34	0.5616
4	i	2	0.0071	0.8969	1.8380	2.28	0.1410

x_3 & x_4

Summary of Stepwise Selection								
Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	x4		1	0.8838	0.8838	3.7677	251.03	<.0001
2	x3		2	0.0131	0.8969	1.8380	4.08	0.0519
3	i		3	0.0071	0.9040	1.7236	2.28	0.1410
4		i	2	0.0071	0.8969	1.8380	2.28	0.1410

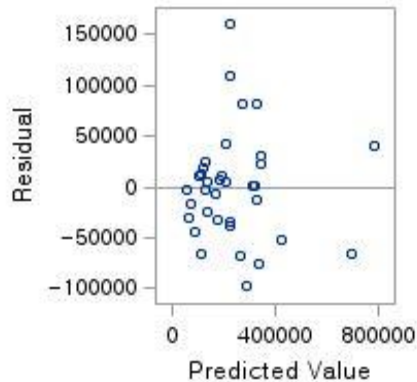
▲ 변수 선택

Cp통계량에 의한 변수 선택

1	0.0888	0.0612	241.6802	26004284296	x1
1	0.0620	0.0336	249.6920	26768335112	i
2	0.8969	0.8905	1.8380	3032787648	x3 x4
2	0.8874	0.8804	4.6955	3313804916	x2 x4
2	0.8867	0.8797	4.8910	3333035644	i x4
2	0.8850	0.8778	5.4226	3385314433	x4 x5
2	0.8844	0.8772	5.5829	3401080131	x1 x4

x_3 & x_4

모형 진단



∴ 선형성 만족

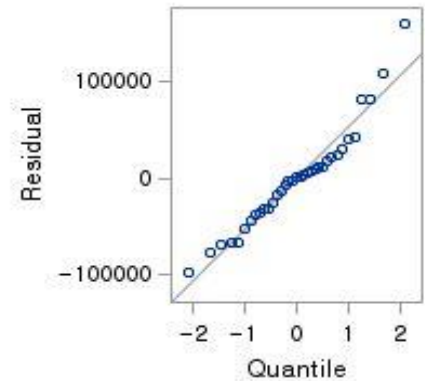
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	2.96596	1.48298	0.44	0.6448

$$SS_{reg} = 2.96596 / 2 = 1.48298 < \chi^2(2) = 5.991$$

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2.12001	2.12001	0.65	0.4256

$$SS_{reg} = 2.12001 / 2 = 1.06 < \chi^2(1) = 3.841$$

∴ 등분산성 만족



정규성 검정			
통계량		p 값	
W	0.961819	Pr < W	0.2591

∴ 정규성 만족

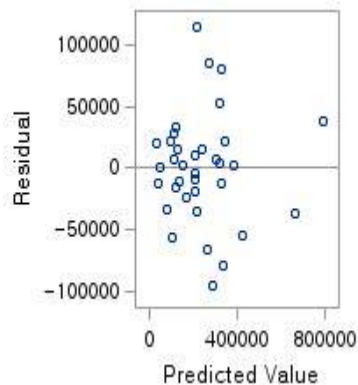
다중 공선성 확인

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Variance Inflation
Intercept	1	29398	17004	1.73	0.0935	0
x3	1	-0.00404	0.00200	-2.02	0.0519	3.55937
x4	1	3.86536	0.36837	10.49	<.0001	3.55937

✓ 다중공선성 줄어듦 !

▲ 교호 작용 추가

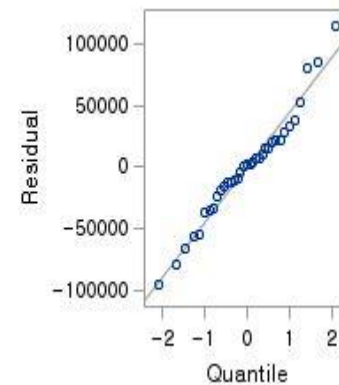
(I, Ix_3, Ix_4) 모형진단



Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	15.38507	3.07701	1.22	0.3262

$$SS_{reg} = 15.38507 / 2 = 7.69029 < \chi^2(5) = 11.070$$

∴ 선형성, 등분산성 만족



정규성 검정				
검정	통계량		p 값	
Shapiro-Wilk	W	0.977308	Pr < W	0.6701

∴ 정규성 만족

▲ 교호 작용 추가

(I, Ix_3, Ix_4) 모형진단

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Variance Inflation
Intercept	1	713.74411	33008	0.02	0.9829	0
i	1	14375	37731	0.38	0.7060	3.86876
x3	1	0.00793	0.00642	1.23	0.2268	49.04274
x4	1	4.38240	1.22220	3.59	0.0012	52.32658
ix3	1	-0.01351	0.00669	-2.02	0.0528	54.08386
ix4	1	-0.27240	1.27334	-0.21	0.8321	69.26082

Number	Eigenvalue	Condition Index	Proportion of Variation				
			i	x3	x4	ix3	ix4
1	3.77636	1.00000	0.00363	0.00120	0.00122	0.00118	0.00093464
2	0.95675	1.98672	0.22036	0.00241	0.00013915	0.00068677	0.00043469
3	0.22796	4.07008	0.08958	0.01656	0.02904	0.01746	0.01103
4	0.03374	10.57882	0.34949	0.16478	0.09622	0.15191	0.10035
5	0.00517	27.01458	0.33694	0.81505	0.87339	0.82877	0.88724

✓ 다중공선성과 분산비율에 문제 있어 보임

>>> 변수 선택 필요



교호 작용 변수 선택

Summary of Forward Selection							
Step	Variable Entered	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	x4	1	0.8838	0.8838	18.5935	251.03	<.0001
2	ix3	2	0.0345	0.9184	5.8507	13.54	0.0009
3	x3	3	0.0113	0.9297	3.0128	5.00	0.0327

Summary of Backward Elimination							
Step	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	ix4	3	0.0000	0.9297	3.0128	0.01	0.9107

Summary of Stepwise Selection								
Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	x4		1	0.8838	0.8838	18.5935	251.03	<.0001
2	ix3		2	0.0345	0.9184	5.8507	13.54	0.0009
3	x3		3	0.0113	0.9297	3.0128	5.00	0.0327

전진선택법,
후진 제거법,
단계적 회귀에 의한
변수 선택

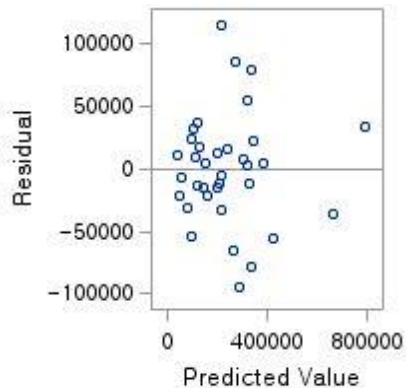
x_3 & x_4

& ix_3

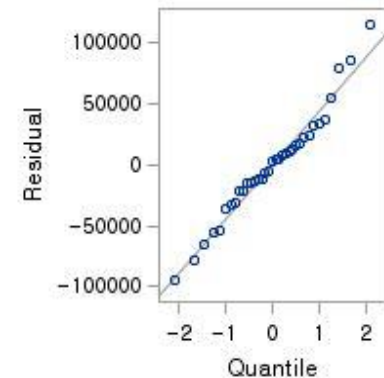
▲ 교호 작용 모형 진단

x_3, x_4, Ix_3

선택된 모형



Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	10.80717	3.60239	1.42	0.2550



$$SS_{reg} = 10.80717 / 2 = 5.4036 < \chi^2(3) = 7.815$$

∴ 선형성 만족

∴ 등분산성 만족

정규성 검정			
통계량		p 값	
W	0.976621	Pr < W	0.6472

∴ 정규성 만족

▲ 최종 모형

$$\hat{y} = 11283 + 0.00803x_3 + 4.16389x_4 - 0.01377Ix_3$$

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	11283	15045	0.75	0.4590
x3	1	0.00803	0.00359	2.24	0.0327
x4	1	4.16389	0.31897	13.05	<.0001
ix3	1	-0.01377	0.00362	-3.80	0.0006

리더, 멘터를 한 경험이 **있는** 연습생 (I=1)

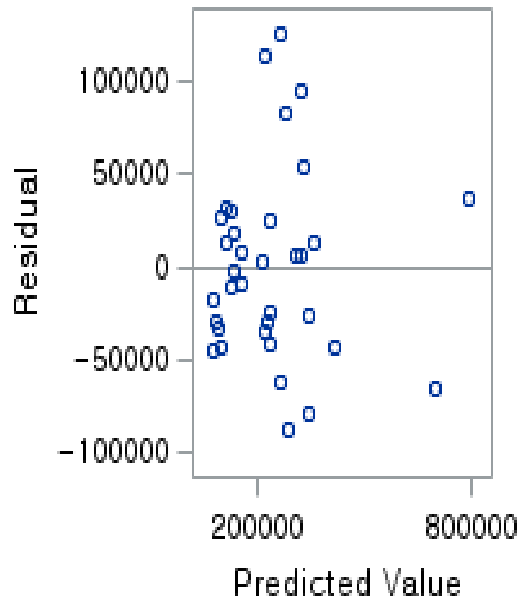
$$\hat{y} = 11283 - 0.00574x_3 + 4.16389x_4$$

리더, 멘터를 한 경험이 **없는** 연습생 (I=0)

$$\hat{y} = 11283 + 0.00803x_3 + 4.16389x_4$$

▲ 이상점 & 영향력 관측값 예상

[Produce101 자료]



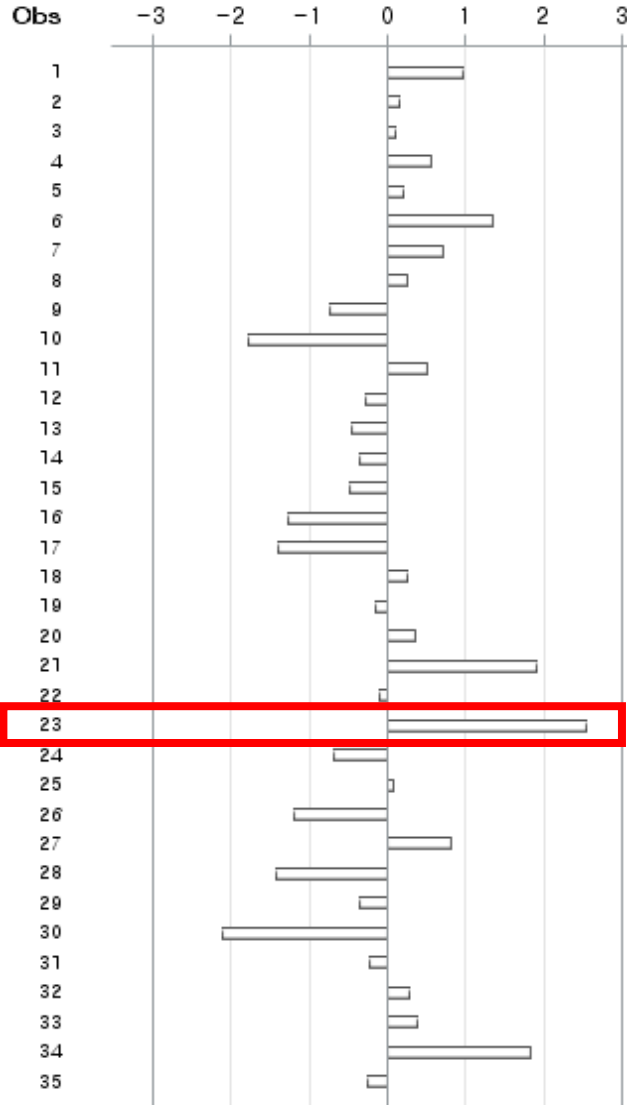
	순발식 점수합 (x2)	조회수 (x3)	하트수 (x4)	분량합 (x5)
17 (박지훈)	3,226,037	45,480,357	220,069	1,500
RANK(순위)	2	1	2	2
평균	1,825,710	5,378,162	60,155	952
표준편차	762,292	8,908,263	48,371	388

※ 표준편차 : 소수점 첫번째 자리에서 반올림



이상점 검정

Studentized Residuals



1	0.9608
2	0.1560
3	0.1063
4	0.5417
5	0.2011
6	1.3527
7	0.6932
8	0.2368
9	-0.7205
10	-1.8363
11	0.4917
12	-0.2867
13	-0.4551
14	-0.3512
15	-0.4773
16	-1.2817
17	-1.4188
18	0.2541
19	-0.1631
20	0.3359
21	1.9892
22	-0.1136
23	2.8075
24	-0.6761
25	0.0735
26	-1.2098
27	0.8012
28	-1.4578
29	-0.3464
30	-2.2472
31	-0.2385
32	0.2763
33	0.3780
34	1.8907
35	-0.2611

H_0 : 23번째 관측값(윤지성)이상점이 아니다

H_1 : 23번째 관측값(윤지성)이상점이다
 $|t_{23}| = 2.8075 < t_{Bonf}(0.05; 35, 4) = 3.51$

∴ 귀무가설 기각을 못하므로, 이상점이 아니다.

여하러 과츠가



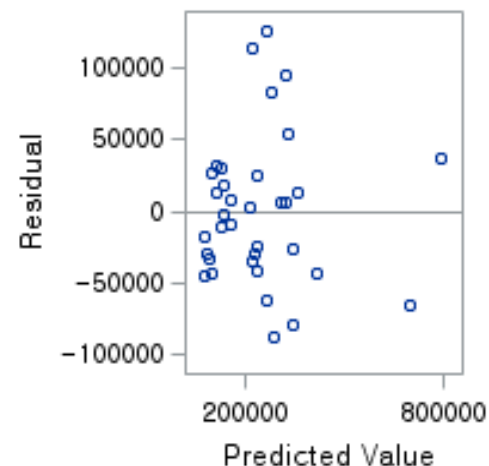
17. 박지훈

*Cook*의 통계량
 $D_{17} = 1.038 > 1$

▲ 영향력 관측값 제거한 회귀 모형

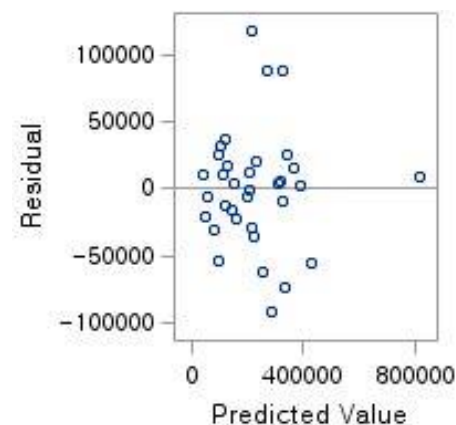
영향력 관측값 포함 회귀 모형

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	11283	15045	0.75	0.4590
x3	1	0.00803	0.00359	2.24	0.0327
x4	1	4.16389	0.31897	13.05	<.0001
ix3	1	-0.01377	0.00362	-3.80	0.0006



영향력 관측값 제거 회귀 모형

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	11880	14811	0.80	0.4288
x3	1	0.00899	0.00360	2.50	0.0183
x4	1	4.00475	0.33332	12.01	<.0001
ix3	1	-0.01226	0.00372	-3.29	0.0025

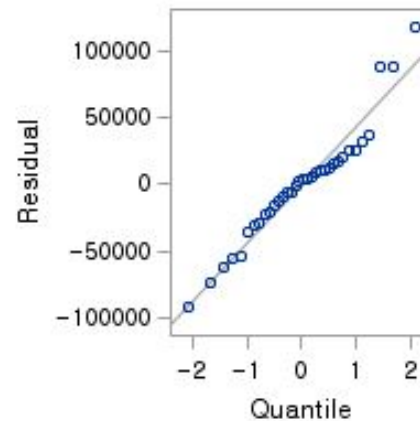


영향력 관측값 제거 회귀 모형_모형진단

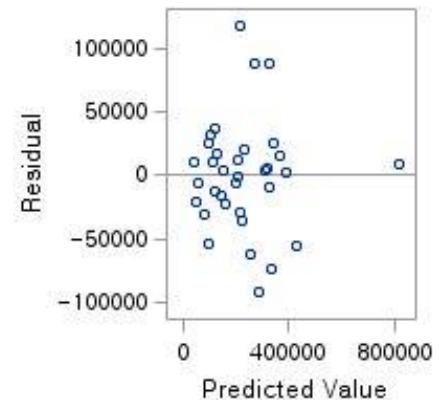
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1.27564	1.27564	0.42	0.5233
Error	32	98.02587	3.06331		
Corrected Total	33	99.30151			

정규성 검정				
검정	통계량		p 값	
Shapiro-Wilk	W	0.954133	Pr < W	0.1632

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	14.78526	4.92842	1.75	0.1781
Error	30	84.51625	2.81721		
Corrected Total	33	99.30151			



∴ 정규성 만족



∴ 선형성 만족

$$SS_{reg} = 1.27564 / 2 = 0.63782 < \chi^2(1) = 3.841$$

$$SS_{reg} = 14.78526 / 2 = 7.39263 < \chi^2(3) = 7.815$$

∴ 등분산성 만족

▲ 영향력 관측값 제거 회귀 모형_다중공선성

영향력 관측값 포함 회귀 모형

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Variance Inflation
Intercept	1	11283	15045	0.75	0.4590	0
x3	1	0.00803	0.00359	2.24	0.0327	16.31258
x4	1	4.16389	0.31897	13.05	<.0001	3.78930
ix3	1	-0.01377	0.00362	-3.80	0.0006	16.88730

Collinearity Diagnostics (intercept adjusted)					
Number	Eigenvalue	Condition Index	Proportion of Variation		
			x3	x4	ix3
1	2.78086	1.00000	0.00755	0.02974	0.00732
2	0.18718	3.85446	0.05526	0.96757	0.04587
3	0.03196	9.32764	0.93718	0.00269	0.94681

영향력 관측값 제거 회귀 모형

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Variance Inflation
Intercept	1	11880	14811	0.80	0.4288	0
x3	1	0.00899	0.00360	2.50	0.0183	6.53076
x4	1	4.00475	0.33332	12.01	<.0001	2.85907
ix3	1	-0.01226	0.00372	-3.29	0.0025	6.92836

Collinearity Diagnostics (intercept adjusted)					
Number	Eigenvalue	Condition Index	Proportion of Variation		
			x3	x4	ix3
1	2.66371	1.00000	0.01973	0.04071	0.01879
2	0.25288	3.24555	0.11095	0.95188	0.07338
3	0.08341	5.65096	0.86932	0.00741	0.90783

▲ 영향력 관측값 제거 회귀 모형_최종 회귀 모형 결과

$$\hat{y} = 11880 + 0.00899x_3 + 4.00475x_4 - 0.01226Ix_3$$

영향력 관측값 **제거** 회귀 모형

VS

$$\hat{y} = 11283 + 0.00803x_3 + 4.16389x_4 - 0.01377Ix_3$$

영향력 관측값 **포함** 회귀 모형

리더, 멘터를 한 경험이 **없는** 연습생 (I=0)

$$\hat{y} = 11880 + 0.00899x_3 + 4.00475x_4$$

I 서론

- 01 분석목적
- 02 자료설명

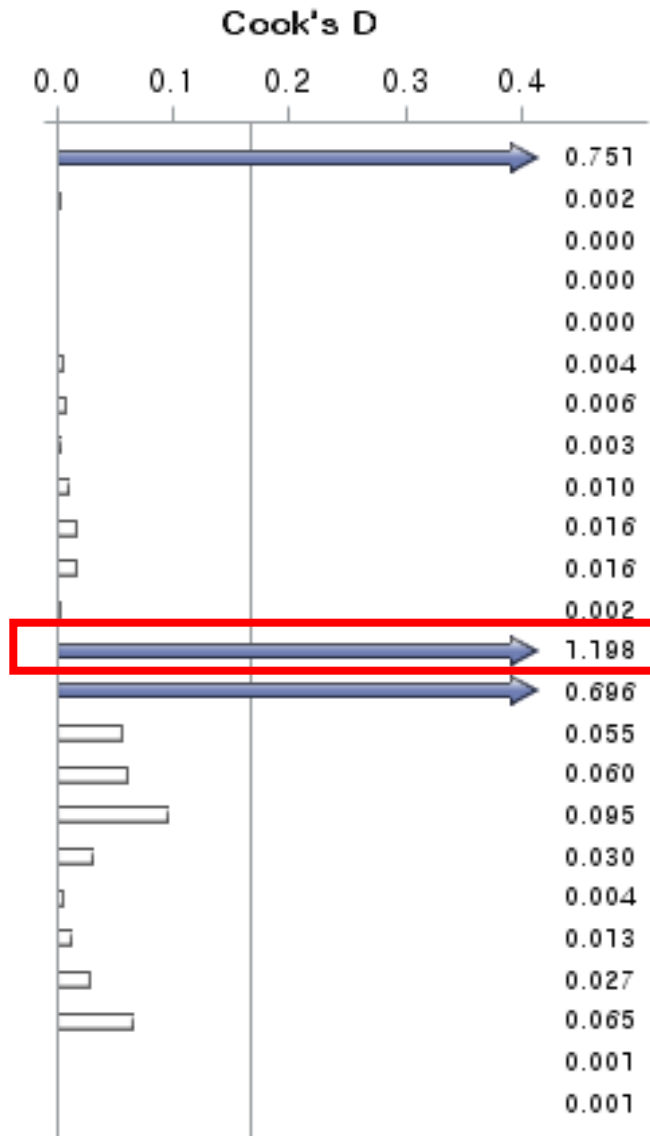
II 본론

- 01 매출액 제외 분석
- 02 매출액 포함 분석

III 결론

- 01 최종 11인 예측
- 02 한계점

▲ 매출액(x6) 포함 – 영향력 관측값 제거



13. 박지훈



*Cook*의 통계량

$$D_{13} = 1.198 > 1$$

▲ 매출액(x6) 포함 – 최종 회귀 모형_모형진단

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.17432	0.17432	0.08	0.7765
Error	21	44.25035	2.10716		
Corrected Total	22	44.42466			

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	0.65772	0.32886	0.15	0.8614
Error	20	43.76694	2.18835		
Corrected Total	22	44.42466			

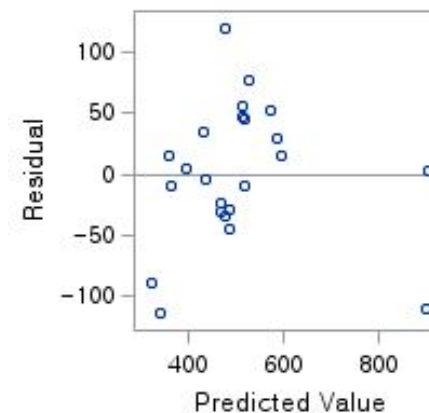
$$SS_{reg}=0.17432/2=0.08716 < \chi^2(1)=3.841$$

$$SS_{reg}=0.65772/2=0.32886 < \chi^2(2)=5.991$$

∴ 등분산성 만족

정규성 검정				
검정	통계량		p 값	
Shapiro-Wilk	W	0.974077	Pr < W	0.7852
Kolmogorov-Smirnov	D	0.104847	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.035602	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.260466	Pr > A-Sq	>0.2500

∴ 정규성 만족



∴ 선형성 만족

▲ 영향력 관측값 제거 회귀 모형_최종 회귀 모형 결과

$$\widehat{sy} = 289.57649 + 0.00434x_4 - 0.00154Ix_4$$

영향력 관측값 **제거** 회귀 모형

VS

$$\widehat{y} = 32676 + 4.68117x_4 - 1.46514Ix_4$$

영향력 관측값 **포함** 회귀 모형

▲ 영향력 관측값 제거 회귀 모형_최종 회귀 모형 결과

이름	y	\hat{sy}	$(\hat{sy})^2$	\hat{y}
강다니엘	1578837	1357.97	1844082	1259813
강동호	755436	742.3729	551117.5	552749.9
김사무엘	391529	801.6769	642685.8	620865.4
김재환	1051735	661.3213	437345.8	459655.4
김종현	704148	815.2093	664566.2	636408.4
라이관린	905875	608.2053	369913.7	398647.3
박우진	937379	827.9493	685500	651041.4
박지훈	1136014	1338.529	1791661	1237484
배진영	807749	785.8598	617575.7	567972.5
안형섭	609085	733.5411	538082.6	511541
옹성우	984756	681.8901	464974.1	483280.4
유선호	551745	684.8463	469014.5	459018.2
윤지성	902098	541.9293	293687.4	322523.9
이대휘	1102005	695.4869	483702	498897.4
임영민	654505	579.0881	335343	365203.9
정세운	769859	605.5873	366736	395640.4
주학년	349040	629.494	396262.7	399314.6
최민기	277108	613.7661	376708.8	405034.4
황민현	862719	757.6161	573982.1	570257.9

I 서론

- 01 분석목적
- 02 자료설명

II 본론

- 01 매출액 제외 분석
- 02 매출액 포함 분석

III 결론

- 01 최종 11인 예측
- 02 한계점

▲ 최종 데뷔 11인 예측해보기

※ \hat{y} : 소수점 첫번째 자리에서 반올림해서 기입

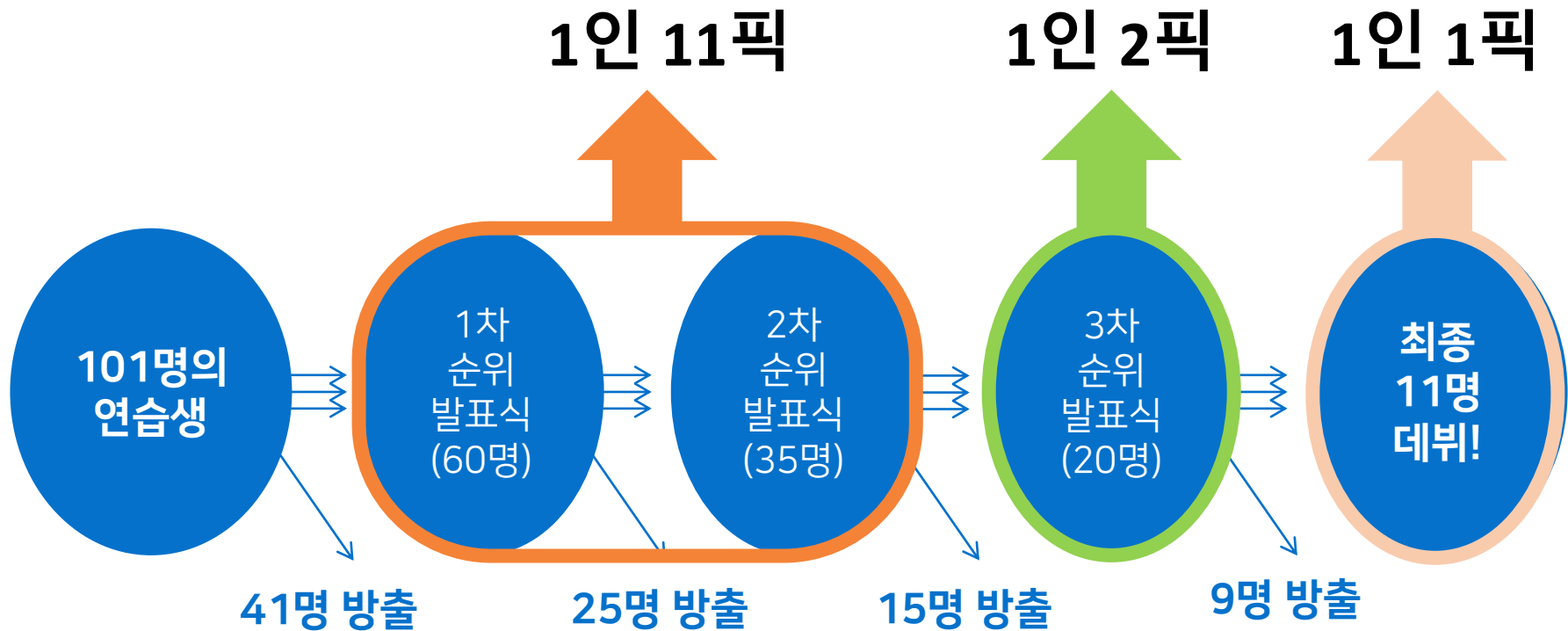
이름	\hat{y}	예측순위	실제순위	예측성공
강다니엘	1398424	1	1	○
강동호	629718	6	13	
김사무엘	568785	9	18	
김재환	532950	11	4	○
김종현	755726	3	14	
라이과린	402770	18	7	
박지민	724153	7	15	
배성영	175114	12	2	○
배성영	605445	7	10	○
안형섭	474297	13	16	
옹성우	560325	10	5	○
유선호	415226	17	17	
윤지성	367674	19	8	
이대휘	576647	8	3	○
임영민	415353	16	15	
정세운	454094	14	12	
주학년	357846	20	19	
최민기	454011	15	20	
하성운	498932	12	11	
황민현	670837	5	9	○

예측 성공률 약 73%

▲ 한계점 ①_4분할 컷



▲ 한계점 ②_투표방식의 변화



[illegible]

▲ 한계점 ④_자료수집의 한계

x_3

(조회우)

&

x_4

(하트우)

2017.11.09 17:35~45 기준자료

x_6

(매출액)

▼ Thank you ▲



Q&A