

박주원 이민경 이세윤



# 01 분석 목적 02 분석 과정 03 결론 해석자료 설명



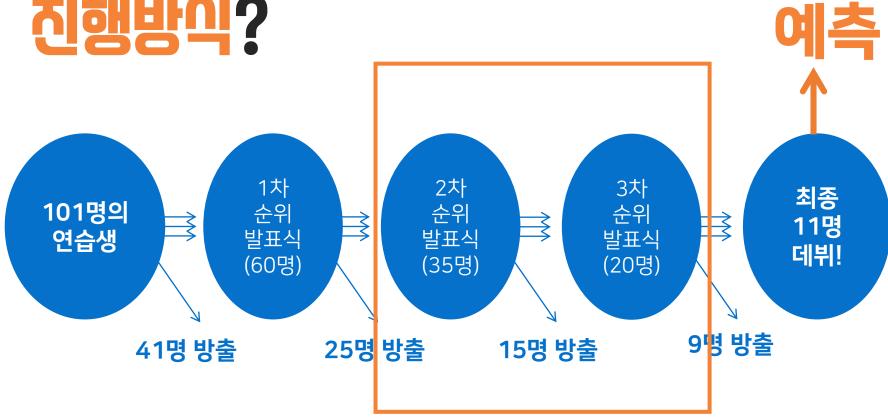
# 01 분석 목적 02 분석 과정 03 결론 해석 자료 설명 결과 예측

## 프로그램 프로듀<u></u> 101?

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



## 프로그램의 진행방익?



#### ▲ 변우 오개

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 기준, 매출액: 2016년 1~4분기 총 매출액 (출처) 매출액: dart(전자공시시스템), 사람인



#### 네이버 TV 캐스트 조회수 및 하트수(X3,X4)



재생목록 동영상

[ 공유 **14,714** N스토어

✓ 최신 ✓ 재생 ✓ 좋아요





아이컨택ㅣ강다 [단독/직캠] 일대일아이컨택ㅣ박지 [4회] "여기는 완성형" 이번 열차 Sorry Sorry\_2 훈 - BTS ♬상남자\_1조 @그룹배틀



는 미안행~ | 슈퍼주니어 #Sorry, ▷804,726 ♡ 19,991 | 1주전



[단독/직캠] 일대일아이컨택 | 김재 환 - 슈퍼주니어 #Sorry Sorry\_2조 ▷ 321.561 ♡ 15.959 | 2주 전



[단독/직캠] 일대일아이컨택 | 김종 현 - 슈퍼주니어 #Sorry Sorry\_2조 호 - BTS #상남자\_2조 @그룹배틀



▷ 427.960 ♡ 14.515 | 2주전

▷ 660.524 ♡ 13.106 | 2주전



[단독/직캠] 일대일아이컨택 | 옹성 우 - 슈퍼주니어 #Sorry Sorry\_2조 관린 - BTS #상남자\_2조 @그룹배



▷ 430,831 ♡ 9,967 2주전



[단독/직캠] 일대일아이컨택 | 라이 오앤오 | 장문복 | 힙통령의 귀환 @ 자기소개\_1분 PR

▷ 1,308,622 ♡ 9,767 | 1달전

#### 사람 고개

#### 자료1:

매출액을 포함하지 않은 35개의 관측치 자료

이름	지시변수(I)	X1(등급)	X2(누적 득표 수)	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		280297	6988	393	53277
김용국	1	2	1386354	1511067	51537	692	184303
김재환	1	4		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		45480357	220069	1500	630198
배진영	0	1	2170203	12888503	65122	569	389982
서성혁	0	2	964971	465725	11233	659	54415
안형섭	0	5		2579545	49878	994	254984
옹성우	1	5	2818035	3536786	67648	1139	358656
유선호	0	1	2077540	1789701	45325	733	209168
윤지성	1	2		2359696	53062	1483	333974
이건희	1	2	1069762	694152	17620	943	50148
이대휘	1	5		4276024	80668	1485	325990
이우진	1	5		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		2827844	71125	1175	196223
주학년	0	1	2332567	1930802	43553	892	197194
최민기	1	1	1578810	3152849	50867	614	217734
타가다 켄타	1	4		1845915	30783	384	146176
하성운	1	5		1742874	79831	741	413654
황민현	1	2	2684529	3896681	81313	692	315650

## 자료 오개

자료2:

매출액을 포함한 24개의 관측치 자료

이름	지시변수(I)	X1(등급)	X2(누적 득 표수)	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

#### ▲ 예앙 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

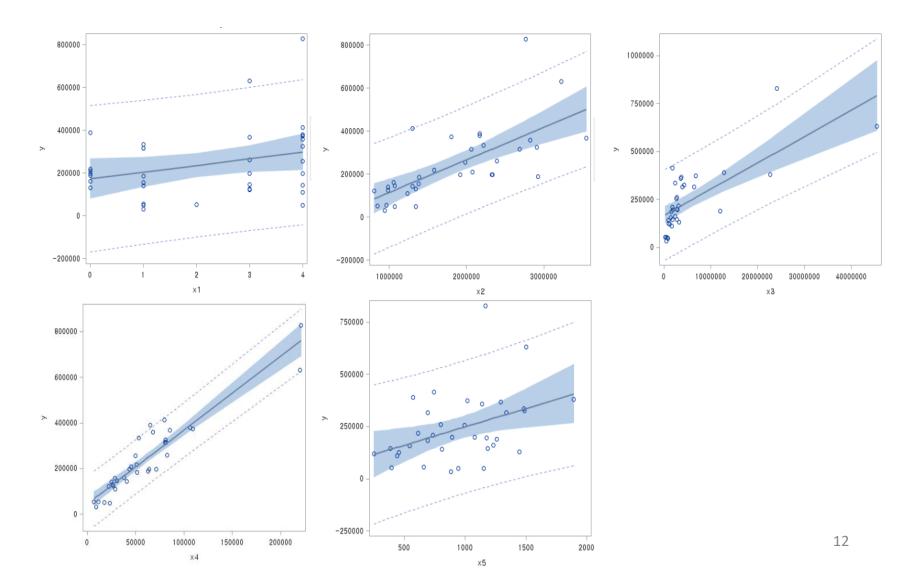


# 01 분석 목적 02 분석 과정 03 결론 해석자료 설명

#### Contents

01 매출액 포함 02 매출액 포함 하지 않은 자료 한 자료

## ▲ 단운 회귀 적합



## ▲ 단운 회귀 적합

Parameter Estimates									
Variable	t Value	Pr >  t							
Intercept	1	141872	61240	2,32	0,0269				
х1	1	31002	17292	1,79	0,0822				

Parameter Estimates									
Variable	DF	t Value	Pr >  t						
Intercept	1	-36935	53931	-0,68	0,4982				
x2	1	0,15179	0,02732	5,58	<,0001				

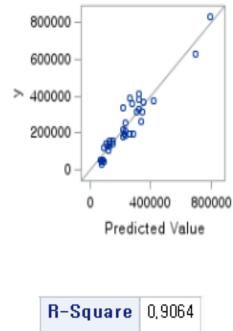
	Parameter Estimates									
Variable	DF	Parameter Estimate		t Value	Pr >  t					
Intercept	1	166198	22653	7,34	<,0001					
хЗ	1	0,01376	0,00220	6,25	<,0001					

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				

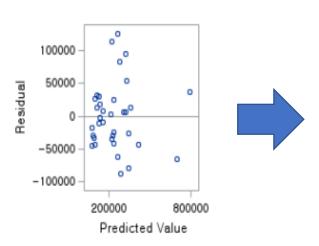
	Parameter Estimates										
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t						
Intercept	1	72683	69850	1,04	0,3056						
х5	1	176,03591	68,11316	2,58	0,0144						

### ▲ 다중 회귀 적합

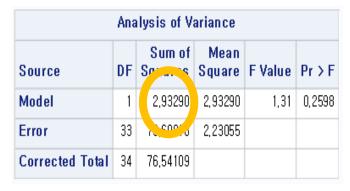
	Parameter Estimates										
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > [t]						
Intercept	1	48137	35655	1,35	0,1878						
i	1	-29600	25854	-1,14	0,2619						
x1	1	-3095,14731	6996, 76510	-0,44	0,6616						
x2	1	0,01221	0,01971	0,62	0,5404						
хЗ	1	-0,00453	0,00226	-2,01	0,0545						
x4	1	3,97488	0,51249	7,76	<,0001						
х5	1	-12,98578	31,05079	-0,42	0,6790						



#### ▲ 모형 진단



: 선형성 만족

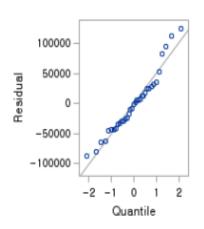


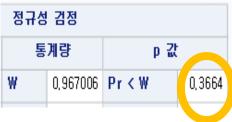


Analysis of Variance									
Source	Source DF Square					Pr > F			
Model	6		4,72023	0,78670	0,31	0,9281			
Error	28		71,02500	2,56503					
Corrected Total	34		76,54109						



∴ 등분산성 만족





::정규성 만족

#### ▲ 변우 변환

등분안영, 언형영, 정규영 모두 만족 변우 변환 진행 X

#### ▲ 다중 공언영 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			
х3	1	-0.00453	0.00226	-2.01	0.0545	4.35973			
x4	1	3,97488	0.51249	7.76	<.0001	6,63918			
х5	1	-12,98578	31.05079	-0.42	0.6790	1.56726			

#### ✓ VIF 문제 있음

Collinearity Diagnostics (intercept adjusted)												
	Condition			Proportion of Variation								
Number Eigenvalue	Index	i	x1	x2	хЗ	×4	х5					
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				

	i	x1	x2	х3	x4	x5
i	1,00000	0.40451 0.0159	0.17624 0.3112	0.15039 0.3885	0.31940	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
хЗ	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
х5	0.18608 0.2845	0.15103 0.3865	0.54777 0.0007	0.46462 0.0049	0.46834 0.0045	1.00000

✓ 상태수= 
$$\left(\frac{3.00976}{0.09087}\right)^{\frac{1}{2}}$$
 문제 없음 =5.7551

✓ 분산비율에 문제 있어 보임



#### 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	х5	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	is a	2	0.0071	0.8969	1.8380	2.28	0.1410					

 $x_3 & x_4$ 

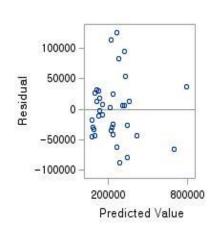
	Summary of Stepwise Selection											
Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F				
1	x4		1	0.8838	0.8838	3.7677	251.03	<.0001				
2	хЗ		2	0.0131	0,8969	1.8380	4.08	0.0519				
3	i		3	0.0071	0.9040	1.7236	2.28	0.1410				
4		i	2	0.0071	0.8969	1.8380	2.28	0.1410				

#### ▲ 변우 언택 Cp통계량에 의한 변수 선택

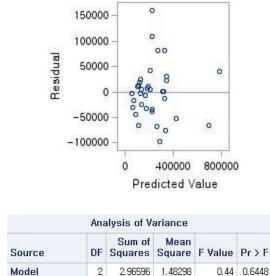
1	0.0888	0.0612	241.6802	26004284296	х1
1	0.0620	0.0336	249.6920	26768335112	i
2	0.8969	0.8905	1.8380	3032787648	x3 x4
2	0.8874	0.8804	4,6955	3313804916	x2 x4
2	0.8867	0.8797	4.8910	3333035644	i x4
2	0.8850	0.8778	5.4226	3385314433	x4 x5
2	0.8844	0.8772	5,5829	3401080131	x1 x4

$$x_3$$
 &  $x_4$ 

#### ▲ 모형 진단



: 선형성 만족

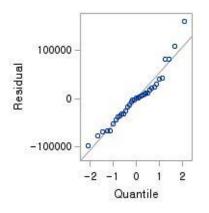


 $SS_{reg}$ =2.96596/2=1.48298  $< \chi^2(2)$ =5.991

Analysis of Variance									
Source	DF	Sum of Squares		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



	통계량	p at			
W	0.961819	Pr < W	0.2591		

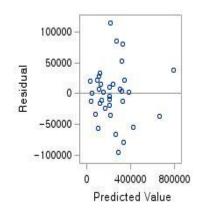
: 정규성 만족

#### ▲ 다중 공언영 확인

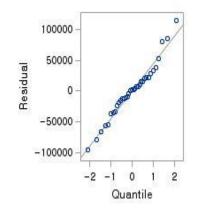
Parameter Estimates										
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > [t]	Variance Inflation				
Intercept	1	29398	17004	1.73	0.0935	0				
хЗ	1	-0.00404	0.00200	-2.02	0.0519	3,55937				
x4	1	3,86536	0.36837	10,49	<.0001	3,55937				

✓ 다중공선성 없어짐!

#### $\Lambda$ 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				



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

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

: 선형성, 등분산성 만족

::정규성 만족

#### $\Lambda$ 교호 작용 추가 $(I,Ix_3,Ix_4)$ 모형진단

Parameter Estimates									
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			
хЗ	1	0.00793	0.00642	1.23	0.2268	49.04274			
x4	1	4.38240	1.22220	3,59	0.0012	52,32658			
ix3	Ť	-0.01351	0.00669	-2.02	0.0528	54.08386			
ix4	1	-0.27240	1.27334	-0.21	0.8321	69,26082			

		Condition Index	Proportion of Variation							
Number	Eigenvalue		i	х3	×4	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



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

	Summary of Backward Elimination										
Step			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	200		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	х3		3	0.0113	0.9297	3.0128	5.00	0.0327				

### 전진선택법, 후진 제거법, 단계적 회귀에 의한 변수 선택

$$x_3$$
 &  $x_4$ 

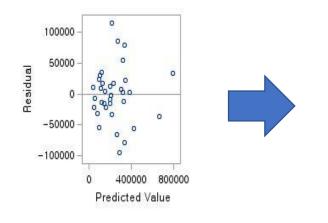
& 
$$Ix_3$$

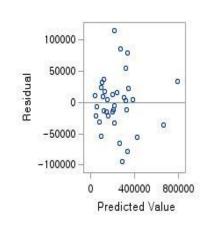


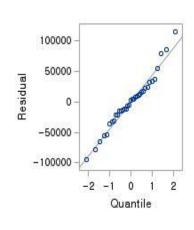
#### ▲ 교호 작용 모형 진단

#### $x_3, x_4, Ix_3$

#### 선택된 모형







••	서	형실	넌 다	ŀ족
		$\cdot$	∟ כ	- 1

	Ana	alysis of V	ariance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F			
Model	3	10.80717	3,60239	1.42	0.2550			

정규	성 검정			
- 5	통계량	p at		
W	0.976621	Pr < W	0.6472	

:: 등분산성 만족

#### ▲ 최종 모형

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	11283	15045	0.75	0.4590
х3	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

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

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

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

#### Contents

 01
 매출액 포함

 하지 않은 자료
 02

 한 자료

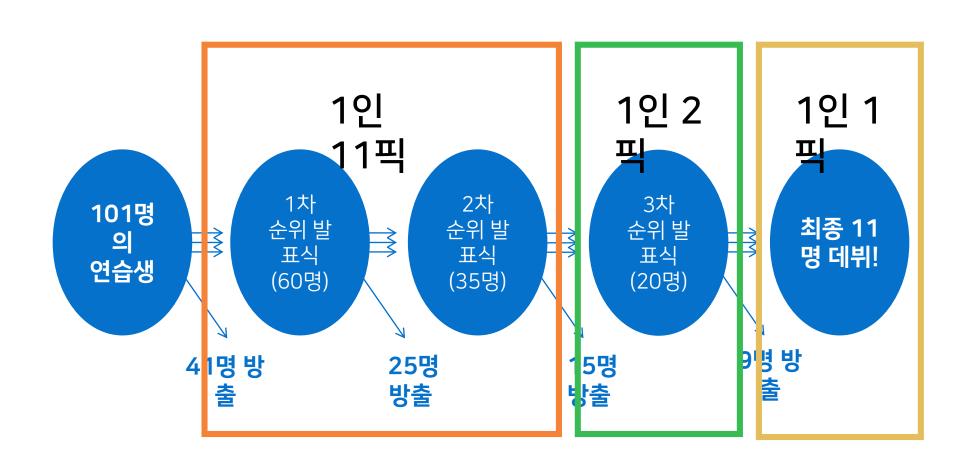


# 01 분석 목적 02 분석 과정 03 분석 결과 자료 설명

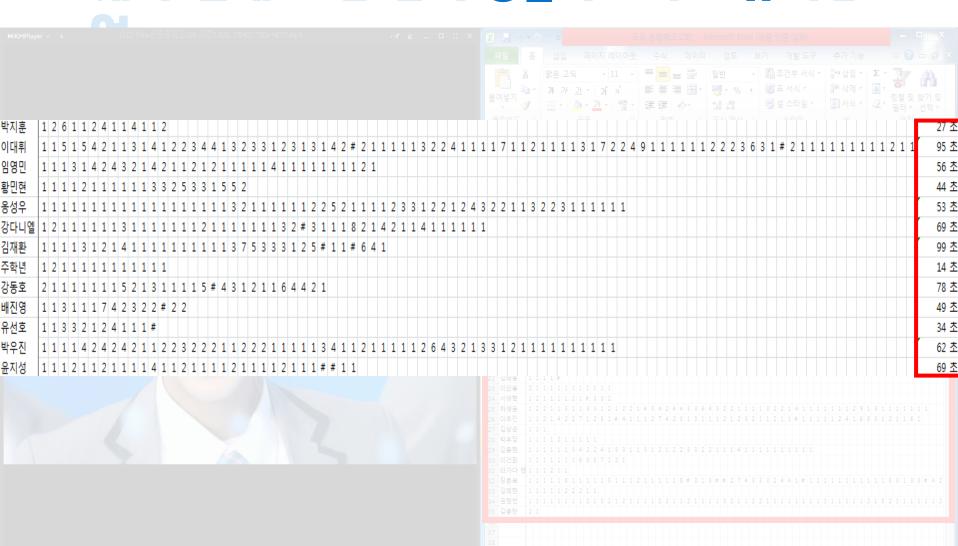
### ▲ 예측 일패 요인 분억 0\_4분할 컷



#### ▲ 예측 일패 요인 분억 ❷\_투표방익의 변화



## ▲ 예측 일패 요인 분억 ❸\_자료의 오류가능



## ▲ 예측 일패 요인 분억 ④\_방용 중 연습맹들

하민호, 결국 '프로듀스 101 시즌2' 하차... 네티즌 송" "착하게 살아라" "안타깝다"

2017-06-09 23:51:13



[뉴스엔 김예은 기자]

임영민이 부정행위 논란에 대해 직접 사과했다.

6월 9일 방송된 Mnet '프로듀스 101 시즌2'에서는 세 번째 순위 발표식이 진행됐다.

이날 17등에 이름을 올린 브랜뉴뮤직 임영민은 부정행위와 관련 사과의 말을 전했다. 그는 "부족한 저를 이 자리까지 올 수 있게 해주신 거에 대해 감사하다"며 "많은 이야기들이 있었 고 불편을 겪었을 국민 프로듀서들, 동료 연습생들에게 죄송하다는 말씀을 드리고 싶다. 더 발전하는 임영민 되겠다"고 말했다.



인력 2017-05-09 00:44















(사진=Mnet '프로듀스 101 시즌2' 영상 캡쳐)

# Thank you

# Q&A