The Intuition Behind ANOVA

초 간략 판 예시, 각 배너를 타고 들어오는 일 방문객 수

일	기존 배너	New 1안	New 2안
1	49	56	51
2	47	54	55
3	46	61	57
4	50	57	53
평균	48	57	54

원본 데이터 전체 평균 53

The Intuition Behind ANOVA

초 간략 판 예시, 각 배너를 타고 들어오는 일 방문객 수

일	기존 배너	New 1안	New 2안
1	49	56	51
2	47	54	55
3	46	61	57
4	50	57	53
평균	48	57	54

원본 데이터 전체 평균 53

일	기존 배너	New 1안	New 2안
1	-4	3	-2
2	-6	1	2
3	-7	8	4
4	-3	4	0

데이터 – 전체 평균

데이터를 전체 평균으로만 설명하려고 할 때 생기는 편차 (오차, noise) 총 편차

일	기존 배너	New 1안	New 2안
1	1	-1	-3
2	-1	-3	1
3	-2	4	3
4	2	0	-1

데이터 – 그룹 평균

데이터를 **그룹별 평균으로** 설명하려고 할 때 생기는 편차 (오차, noise) 그룹 내 편차



그룹평균 - 전체평균

데이터를 전체 평균으로 설명할 때는 설명이 안되던 오차 중, 그룹별 평균으로 설명하면 설명이 되는 부분 그룹 간 편차

The Intuition Behind ANOVA

초 간략 판 예시, 각 배너를 타고 들어오는 일 방문객 수

일	기존 배너	New 1안	New 2안
1	49	56	51
2	47	54	55
3	46	61	57
4	50	57	53
평균	48	57	54

원본 데이터 전체 평균 53

일	기존 배너	New 1안	New 2안
1	-4	3	-2
2	-6	1	2
3	-7	8	4
4	-3	4	0

데이터 – 전체 평균

데이터 – 그룹 평균

New

1안

-3

4

0

New

2안

-3

-1

일 기존 New New HIH 1안 2안 1 -5 4 1 2 -5 4 1 3 -5 4 1 4 -5 4 1

그룹평균 - 전체평균

전체 평균으로는 이해 못할 오차

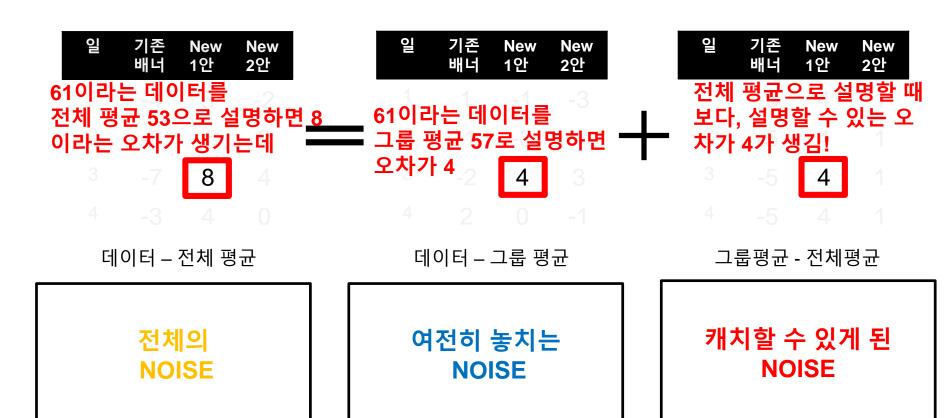
그룹별 평균으로는 이해 못할 오차 전체 평균 대비, 그룹별 평균을 사용하면 설명이 되는 오차

The Intuition Behind ANOVA

초 간략 판 예시, 각 배너를 타고 들어오는 일 방문객 수



원본 데이터 전체 평균 53



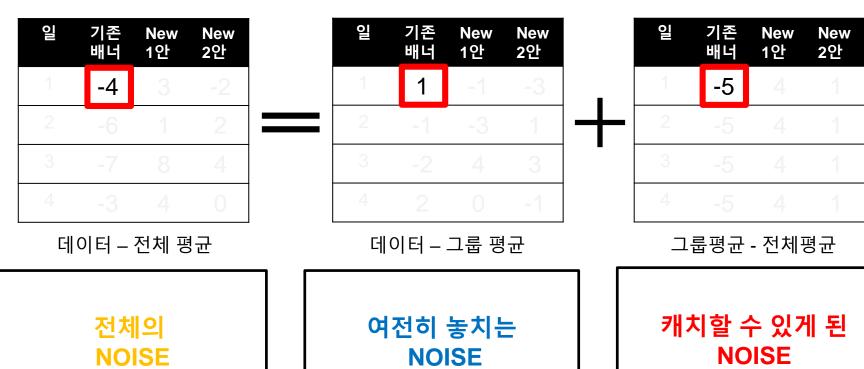
The Intuition Behind ANOVA

초 간략 판 예시, 각 배너를 타고 들어오는 일 방문객 수

단순히 편차로만 계산해서 하나하나 따 지고 들면 어색한 부분은 있음.

일	기존 배너	New 1안	New 2안
1	49	56	51
2	47	54	55
3	46	61	57
4	50	57	53
평균	48	57	54

원본 데이터 전체 평균 53 |



The Intuition Behind ANOVA

이 편차들을, 한번에 요약해서 비교할 필요가 있다.

일	기존 배너	New 1안	New 2안
1	49	56	51
2	47	54	55
3	46	61	57
4	50	57	53
평균	48	57	54

원본 데이터 전체 평균 53

일	기존 배너	New 1안	New 2안
1	-4	3	-2
2	-6	1	2
3	-7	8	4
4	-3	4	0

데이터 – 전체 평균



데이터 – 그룹 평균



그룹평균 - 전체평균

전체의 NOISE (어쨌든 편차) 여전히 놓치는 NOISE (얘도 편차) 캐치할 수 있게 된 NOISE (결국 이것도 편차)

The Intuition Behind ANOVA

이 편차들을, 한번에 요약해서 비교할 필요가 있다.

일	기존 배너	New 1안	New 2안
1	49	56	51
2	47	54	55
3	46	61	57
4	50	57	53
평균	48	57	54

원본	데이	터
전체	평균	53

일	기존 배너	New 1안	New 2안
1	-4	3	-2
2	-6	1	2
3	-7	8	4
4	-3	4	0

데이터 – 전체 평균

	일	기존 배너	New 1안	New 2안				
	1	1	-1	-3				
•	2	-1	-3	1	╅			
	3	-2	4	3				
	4	2	0	-1				

데이터 – 그룹 평균



그룹평균 - 전체평균

어쨌든 편차니까 제곱해 서 평균내면 분산! 얘도 결국 제곱해서 평 균내면 분산!

The Intuition Behind ANOVA

Case2

일	기존 배너	New 1안	New 2안
1	49	56	51
2	47	54	55
3	46	61	57
4	50	57	53
평균	48	57	54

원본 데이터 전체 평균 53

귀무 가설 기준의 총 오차 결국 놓친 오차 잡아낸 오차

결국 놓친 오차 Case1

차

잡아낸 오 차

결국 놓친 오

잡아낸 오차

놓친 오차 대비 별로 잡아낸 오차가 없음. 별로 배너간 차이 없는 것 같음

놓친 오차 대비 잡아낸 오차가 상당 함! 배너간 차이가 있는 듯!

오차의 크기들을 비교해야 하는데 그 요약 수단으로 분산을 사용했음.

그리고 구한 분산을 비교해서 잡아낸 오 차가 어느 정도로 큰지 관찰함

그래서 분산분석 이라고 함!

즉, 기준(전체 평균 : 귀무 가설 / 그룹 별 평균 : 대립 가설)을 무엇으로 하느냐에 따라서 생기는 오차의 크기를 분산을 이용해서 비교한 것!

귀무 가설 : 모든 그룹 간 평균 차이가 없다 ; 배너에 따른 방문객 수 차이는 없을 것이다.

대립 가설 : 최소한 한 그룹은 평균 차이가 있을 것이다 ; 배너에 따른 방문객 수 차이는 있을 것이다.

SS = Sum of Squares : 제곱해서 더한 거.

MS = Mean of Squares : 제곱해서 더한 거를

자유도로 나누어 평균낸거

변동 요인	변동합(SS)	자유도	분산(MS)	F ratio	P-value	5%
그룹 간 (Between Groups)	SSB : 168	2 (그룹 수 – 1)	MSB : 84	13.5	0.002	
그룹 내 (Within Groups)	SSW : 56	9 (데이터 수 – 그룹 수)	MSW : 6.22			
Total	SST : 224	11 (데이터 수 -1)	(보통 안 구함)	분산 비교 부분!	ni di m	
자으도느 익다 찬자					말하면,	

자규노는 걸린, 심자.

놓친 오차보다 잡아낸 오차가 13.5배

Intution에 의한 모델링이라기 보다. 수학적인 보정의 성격이 훨씬 크다.

는 크다! 라고 설명해 볼 수 있음.