

# 모든 분야의 데이터 분석

왜 맥주와 기저귀는 같이 팔리는가??

여러분이 마트 진열대에 갔다고 상상해 봅시다

비슷한 물건끼리 진열되 있는거 기억나시죠?

맥주랑 기저귀는 각각 어느 품목에 있나요?

적어도 같은 품목은 아니죠??

근데 같이 팔린다고??? 뭐지 이거?? 전혀 관련이 없는데 두 물건이??



기저귀 + 맥주 = .....?

혹시라도 이런거 상상하셨으면 데이터분석이고 나발이고 병원부터 가세요...



# 우주를 모래알로

- 우주를 모래알로 모두 채우려면 필요한 모래알 갯수는 몇개인  
가??

위 질문에 답하기 위해 필요한걸 정리해 봅시다.

1. 모래알의 크기: 음...모래알이라...어떻게든 정의내릴수 있을거  
같은데..

2. 우주의 크기: ㄷㄷㄷㄷ

모래알의 크기는 대충 정하면 될거같은데..

우주의 크기를 어떻게 정할까요?

# 우주를 모래알로

<모래알을 세는 사람> 논문에서는 이렇게 접근 했습니다.

지구의 반지름은 아무리 커도 300 미리아드 스타디아(myriad stadia, 고대 그리스의 측량단위)(약 5105 km)를 넘지 않을 것이다.

달은 지구보다 크지 않을 것이고, 태양은 달의 30배 이상 크지는 않을 것이다.

따라서 지구와 태양 사이의 거리는 달과 태양사이의 거리에 200배 정도 일 것이다.

또한 별의 연주 시차가 없으므로 별들은 지구와 같은 속도로 태양 주위를 돈다고 생각할 수 있고, 별들 사이의 거리는 태양과 지구 사이의 거리에 두 배 정도일 것이다.

# 우주를 모래알로

맞고 틀리고를 떠나서, 이러한 논리적인 접근이 필요합니다.

데이터 분석시도 마찬가지로입니다.

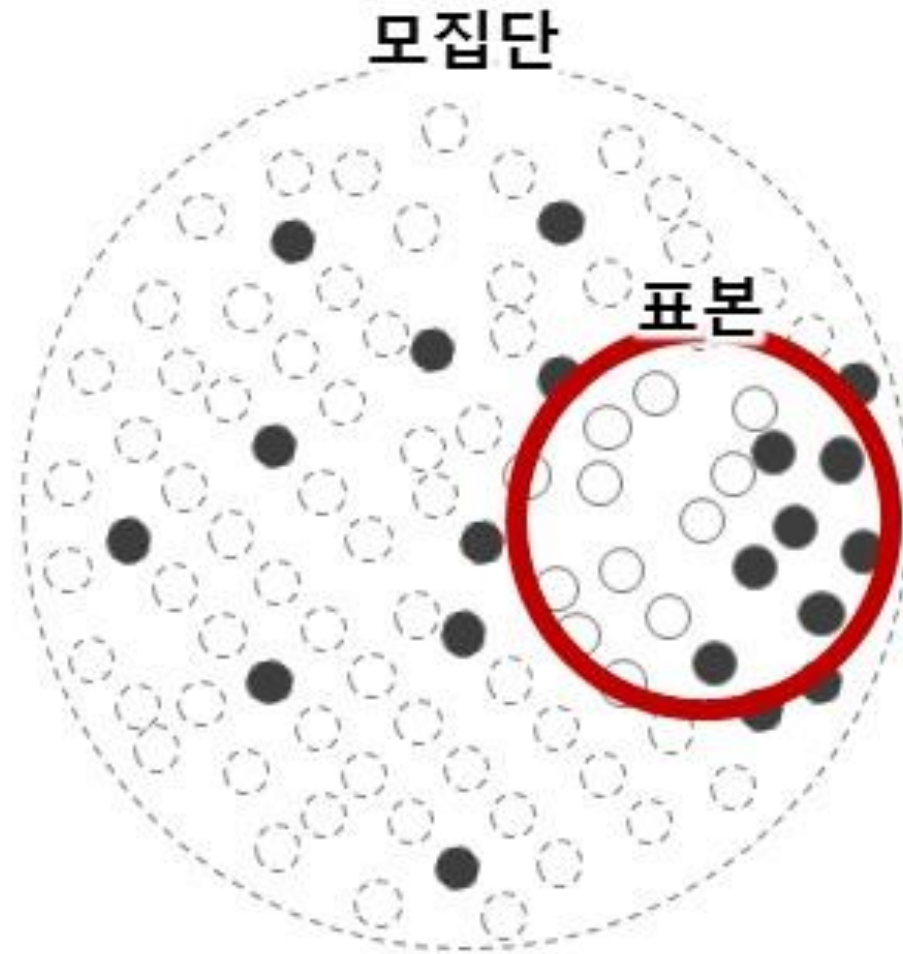
여러분이 데이터분석을 하다보면 이런 논리적인 접근을 해야되는 경우가 수없이 옵니다.

그때마다 방향을 정하는건 여러분들입니다.

정답이 없어요. 최대한 맞는걸 추측해서 해야합니다.



# 데이터 분석을 하려면 필요한 것들



# 데이터 분석을 하려면 필요한 것들

-야. 이거 데이터 다 모으려면 너무많고 시간도 오래걸리고 비용도 많이들고... 한번에 다 분석하지 못하겠는데?

엄청난 데이터를 수집하고 다룰수 있는 요즘에야 좀 덜해졌지만  
아직도 실생활에선 원하는 모든 데이터를 수집하지 못하는 경우가 많습니다.  
비용도 그렇고 시간 문제도 있죠.  
그래서 확률을 이용해서 이러한 방법을 썼습니다.



어차피 다 못하니까 일부만 랜덤하게 뽑자.  
랜덤하게 뽑으면 뽑힌 애들과 원래 애들의 비율이나 속성은 비슷하겠지.



# 데이터 분석을 하려면 필요한 것들

이 접근방법을 보고 무선표집(random sampling)이라고 부르며,

- 원래 분석하고 싶은 전체 집단을 모집단(population)
- 다 못하니까 일부만 뽑힌 녀석들을 표본집단(sample)

이라고 부릅니다.

랜덤하게 뽑으면 확률적으로 뽑힐 확률이 모두 동일하므로, 표본집단의 속성 = 모집단의 속성이라는 접근입니다.

그래서 전체 모집단을 알고있을때의 통계를 기술통계(technical이 아니고, descriptive. 즉, 말그대로 현상을 설명하는 것입니다)

표본집단만 알고있을경우, 표본집단의 특성을 이용해서 원래 알고싶은 모집단을 추정하는것을 추론통계 라고 부릅니다

부분으로 전체를 유추하겠다!! 라는게 핵심입니다

# 데이터 분석을 하려면 필요한 것들

## 모집단과 표본집단

먼저 모집단과 표본집단에 대해서 알아볼 필요가 있습니다. **모집단**이란 우리가 **관심있는 대상 전체**를 말합니다. **표본집단**이란 **모집단으로부터 추출된 모집단의 부분 집합**이라고 할 수 있습니다.

예를 들어 생각해볼까요? 우리가 대전사람들의 평균 키를 구하고 싶다고 해 봅시다. 여기서 모집단은 대전에 주민등록이 되어있는 모든 사람들입니다. 하지만 대전 지역에 거주하는 모든 사람들의 평균 키를 구할 수 없기 때문에 그 중의 일부를 뽑아서 구해야겠죠? 주민등록 상 대전이 거주지로 되어있는 사람들의 리스트를 모아서 일부를 추출한 집단이 표본집단입니다.

# 데이터 분석을 하려면 필요한 것들

표본집단의 평균키가 모집단의 평균키와 동일할까요?

표본집단의 평균키가 모집단의 평균키라고 추정할 수는 있겠지만 모집단의 평균키라고는 할 수 없습니다.

(뒤에 나오겠지만 표본평균이 모집단의 평균키라고 추정할 수 있는 근거가 중심극한정리에 담겨 있습니다.)

중심극한정리가 표본집단의 평균이 표본집단의 크기가 커지면 모집단의 평균과 같아진다고 생각한다면 큰일납니다.

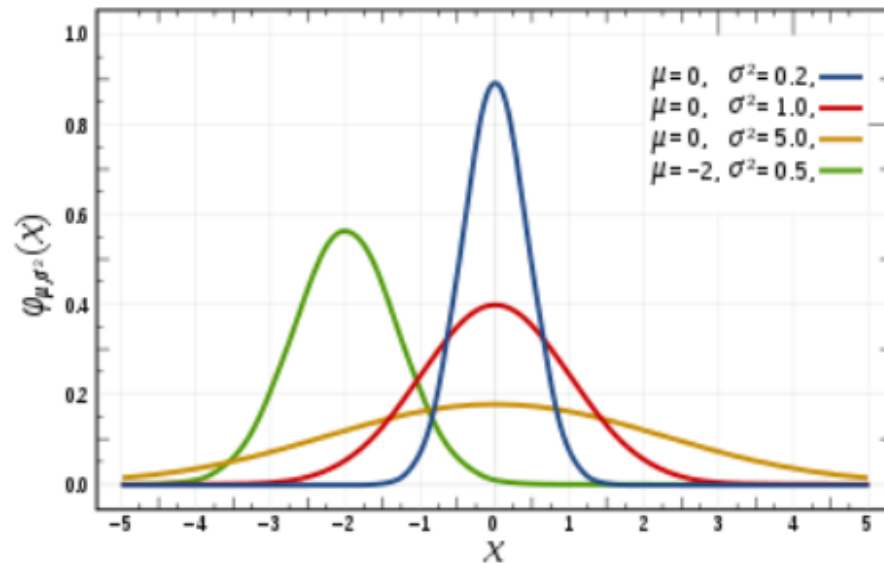
# 데이터 분석을 하려면 필요한 것들

## 분포의 모양은 어떻게 생겼는가??

또하나 분석을 시작하기 전에 중요한 것은, 분포의 모양입니다.

분포 라는것은 모집단이든 표본집단이든 간에, 집단에 속하는 놈들을 일렬로 죽 나열하였을때 찍히는 y값들로 이루어진 그래프를 의미합니다.

이런겁니다.



빨간색 녀석이 가장 정상분포에 가깝습니다

# 데이터 분석을 하려면 필요한 것들

## 확률표본

이제 표본집단이란 용어를 말고 확률표본이란 단어를 사용해보겠습니다.

**확률표본**이란 모집단에 속한 표본 하나하나가 표본으로 추출될 확률이 동일하다는 조건 하에서 추출된 표본입니다.

여기서 유의해야할 부분은 확률표본은 구체적인 관측치가 아닌 **확률변수**(정의역이 표본공간이고 치역이 실수인 함수)이기 때문에 표본에 속할 하나하나가 무엇일지는 사건이 일어나기 전까지 미리 알 수 없다는 것입니다.

여기서 우리가 관심있는 것은 **확률표본의 수(크기)**입니다. 확률표본을 여러 개( $n$ 개) 뽑아서 우리는  $X_1, X_2, \dots, X_n$ 과 같이 나타낼 수 있습니다.

# 데이터 분석을 하려면 필요한 것들

한쪽으로 쏠리거나, 너무 뾰족하거나, 너무 납작하면 '편포'라고 부릅니다. 그럼 왜 그렇게 정상분포 정상분포 노래를 부르는거냐? 라고 물어보실겁니다. 그이유는 분석의 용이함 + 다음의 이녀석 때문입니다.

## 중심극한정리

([https://ko.wikipedia.org/wiki/%EC%A4%91%EC%8B%AC\\_%EA%B7%B9%ED%95%9C\\_%EC%A0%95%EB%A6%AC](https://ko.wikipedia.org/wiki/%EC%A4%91%EC%8B%AC_%EA%B7%B9%ED%95%9C_%EC%A0%95%EB%A6%AC))

"N(표집수)가 충분히 많으면, 모든 확률분포는 정규분포로 수렴하며, 평균은 실제 평균에 점점 가까워진다 " 가 중심극한정리의 요지입니다.



# 데이터 분석을 하려면 필요한 것들

중심극한정리는 표본평균의 표본분포에 관한 것으로 이해하면 좋습니다.

모집단은 어떤 분포를 따를텐데(이항분포, 초기하분포, 균일분포, 정규분포 등등) 어떤 분포를 따르든지 상관없습니다.

이때 모집단은 평균이  $\mu(\text{mu})$ 이고 표준편차가  $\sigma(\text{sigma})$ 입니다.

이때, 모집단에서 확률표본을  $n$ 개 만큼 뽑는다고 해봅시다.

그렇다면 확률표본의 표본평균은 통계량이므로 분포를 따를것입니다.

이 표본평균의 표본분포는  $n$ 이 커지면(일반적으로  $n \geq 30$ ) 평균이  $\mu$ 이고 표준편차가  $\sigma/n$ 인 정규분포를 따른다는 것입니다.

# 데이터 분석을 하려면 필요한 것들



동전의 앞면과 뒷면이 나올확률은 누구나 아다시피 각각  $0.5(1/2)$ 다  
하지만 실제로 동전을 던져봤을때, 30번을 한거랑 1000번을 한것 중 어느것이 더  $1/2$ 에 가깝게 나올까??

# 데이터 분석을 하려면 필요한 것들

## 중심극한정리가 왜 중요할까요?

통계학에서 중요한 부분 중 하나가 모집단의 특성(모수)를 추정하는 것입니다. 왜냐하면 모집단 전체를 조사하여 모집단의 특성을 알아내는 것은 거의 불가능에 가깝고, 시간과 비용이 많이 들기 때문이지요. 그래서 통계학의 여러 기법을 통해 모집단의 표본들을 통해서 모집단의 특성을 추정하는 것입니다.

# 데이터 분석을 하려면 필요한 것들

각각의 표본은 모집단의 특성을 나타내기에는 부족할 것입니다. 하지만 (표본들의 더하여 그 개수만큼 나눈) 표본평균의 분포가  $n$ 이 커지기만 한다면 모집단의 특성을 나타낼 수 있게되는 것이죠.

즉, 통계량인 **표본평균을 통해서** 모집단의 모수(모집단의 특성을 나타내는 값)인 **모평균**과 **모표준편차**를 **추정할 수 있는 확률적 근거를 제시**해주는 것입니다.

더군다나 정규분포는 표준화하여 표준정규분포로 나타낼 수 있기 때문에 표준정규분포표 하나가 주어지면 확률값을 구하기도 쉽습니다.

정리하자면, 중심극한정리는 모집단의 모수를 추정하는데 큰 기여를 하고 있습니다.

# 데이터 분석을 하려면 필요한 것들

## 분석이 용이하다는 이유는

추론통계등 대부분의 통계적 분석방법이 모두 다양한 분포의 형태를 고려하여 만들지 않고, **모집단이 정규분포임을 가정한** 상태에서 만들어졌기 때문입니다. 즉, 정규분포가 아니면 대부분의 분석방법은 다 말짱깡....

## 표본의 확률분포는 어떠한가?

- 어떤 분포이냐에 따라 추정을 위한 기법은 달라지기 때문에 중요하다. 다만 표본의 수가 많아질 수록 정규분포에 근사하게 된다.

# 데이터 분석을 하려면 필요한 것들

분포 모양이 정해졌다 칩시다.

또다른 문제가 있습니다.

[1,1,2,2,3,3,4,4,5,6,6,6,6,6,7,7,8,8,8,8]

위와 같은 데이터가 있고 데이터를 옆사람에게 전달해준다고 해봅시다.

가장좋은 방법은 "첫번째는 1, 그다음은 1...." 일일이 전달해주는겁니다.

하지만 이게 수백개 수천개라면? 데이터를 어떻게 상대방에게 이해시킬수 있을까요?

전부 다 일일이 줄순 없으니까. 일단 대표를 뽑읍시다.

대표를 할만한녀석은 다음 둘중 하나입니다.

1. 가장 '가운데' 있는 녀석
2. 가장 빈도가 높은 녀석



# 데이터 분석을 하려면 필요한 것들

**가운데있는 녀석**은 사람마다 2가지로 정의할 수 있습니다.

1. 그냥 순서대로 작은거부터 높은거까지 주욱 나열했을때, 가운데 '순서'에 있는 놈 (=중앙치, median 이라고 부릅니다)
2. 값들을 수학적으로 계산해서, 그 수학적인 값이 가운데 있는놈 (= 평균, mean 이라고 부릅니다)

**빈도가 제일 높은 녀석**

말그대로 제일많이 출현한 녀석 입니다. 최빈치(mode)라고 부릅니다.

# 데이터 분석을 하려면 필요한 것들

국어

수학

영어

사회

과학

95점

100점

90점

50점

97점

데이터 분석을 하려면 필요한 것들

평균

$$\frac{95 + 100 + 90 + 50 + 97}{5} = 86.4$$

# 데이터 분석을 하려면 필요한 것들

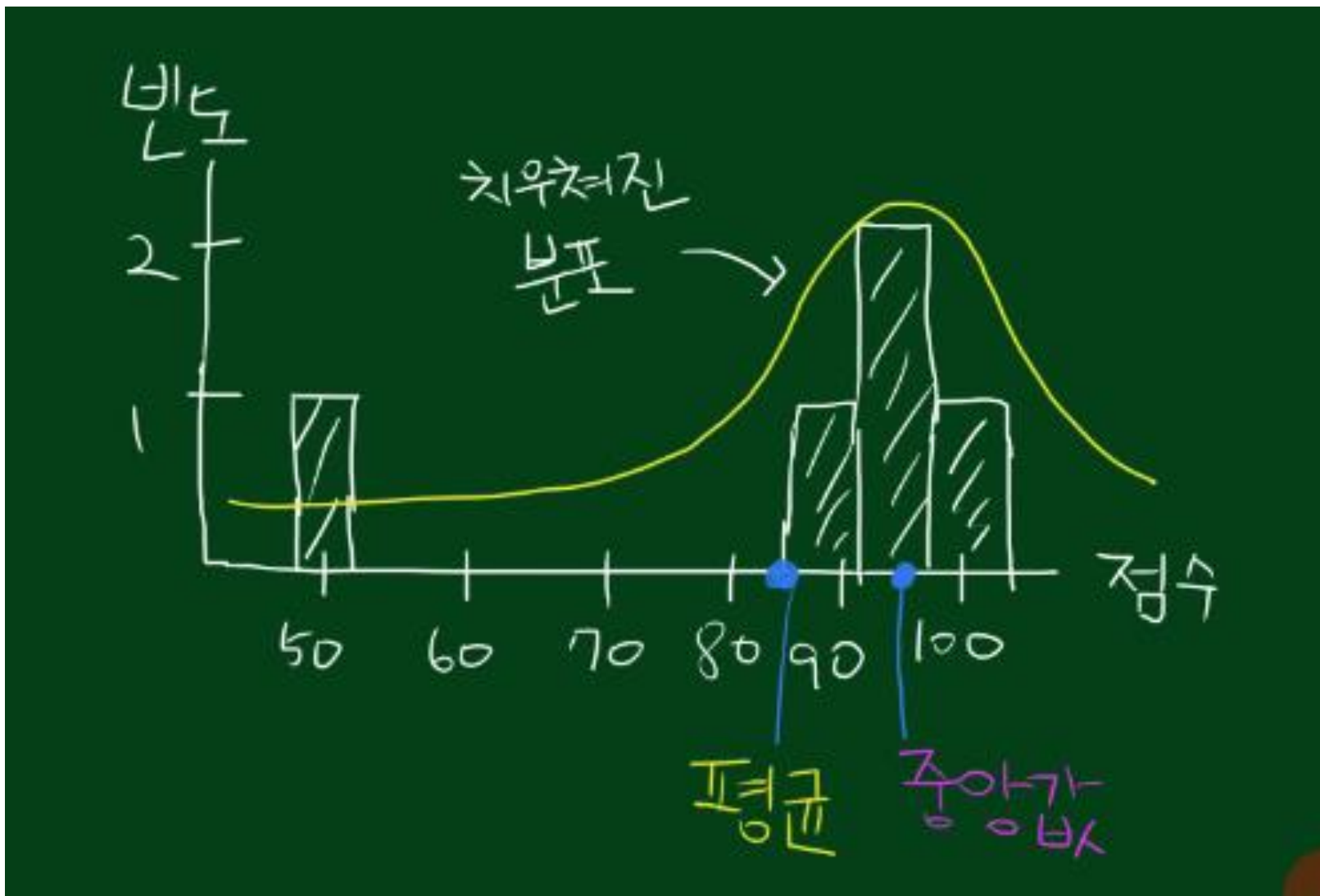
중앙값

↳ 작은 순서대로 정렬했을때 가운데

50, 90, 95, 97, 100  
↑

중앙값 95

# 데이터 분석을 하려면 필요한 것들



# 데이터 분석을 하려면 필요한 것들

새치기 ~~방지~~ 줄 서기



정확하게 한사람씩 떨어져 있으면 좋으련만...현사창



# 데이터 분석을 하려면 필요한 것들

## 뭐가 이렇게 들쭉날쭉하냐?

데이터를 설명하려고 짱은 구해봤습니다

어, 근데 하나더 생각해야될게 있음요...

각 애들이 짱을 기준으로 어느정도 간격으로 떨어져있냐....

데이터마다 이 간격의 길이가 다 다르고 들쭉 날쭉합니다.

# 데이터 분석을 하려면 필요한 것들

예를들어.

$a = [1, 2, 3, 4, 5]$

$b = [1, 3, 5]$

$a$ 와  $b$ 의 짱(평균으로 하겠습니다)는 모두 3입니다. 같죠?

하지만 평균으로부터 떨어진 거리가 다릅니다.

평균만 줘서는 데이터를 설명하기 부족하다는게 이해되시죠?

떨어진 거리도 알려줘야 합니다!

# 데이터 분석을 하려면 필요한 것들

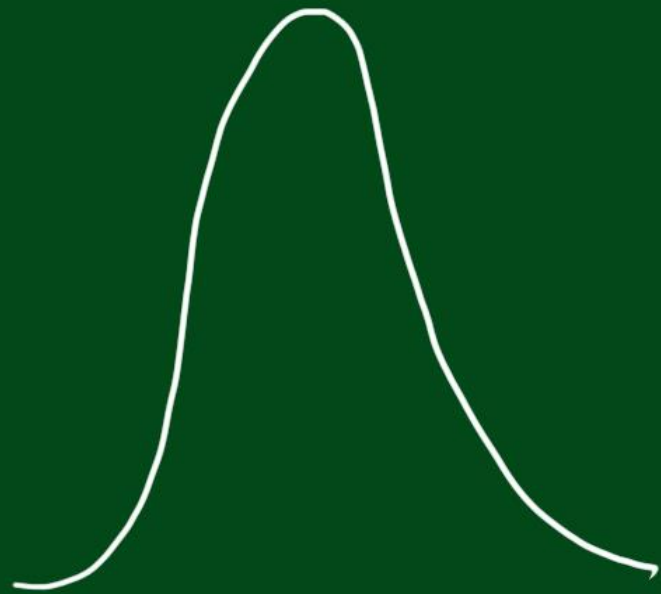
이렇게,

평균으로부터 각 수치들이 평균적으로 얼마나 떨어져있는가?

분포가 평균을 중심으로 얼마나 넓게 퍼져있는가를 따지는 겁니다.

이러한 것을 나타내는 수치가 분산과 표준편차이며 구하는 공식은 다음과 같습니다.

# 데이터 분석을 하려면 필요한 것들



분산이 작음



분산이 큼

# 데이터 분석을 하려면 필요한 것들

$$(x - \bar{x})$$



편차

$$(x - \bar{x})^2$$



편차 제곱

$$\sum (x - \bar{x})^2$$



편차 제곱합

$$\frac{\sum (x - \bar{x})^2}{n}$$



편차 제곱의 평균(분산)

갑자기 이게 다 뭐여...

# 데이터 분석을 하려면 필요한 것들

아 복잡해!!! 라고 생각하지말고 차근차근 보면 어렵지 않습니다  
우리의 목적은 평균으로부터 각 수치들이 얼마나 떨어졌나를 계산 하려는 겁니다.

우선 위에 x바 라는것은 평균을 나타냅니다. 아하 그럼...

편차 = 각 수치에서 평균을 뺀것 = 각 수치에서 평균까지의 거리(양수든 음수든)



# 데이터 분석을 하려면 필요한 것들

아 ㅇㅋ! 그럼 편차가 거리니까, 이걸다 더해서 샘플갯수(N)개 만큼 나누면 '평균'적인 거리가 나오겠네?

하지만 안됨...

왜냐!!! 편차의합은 항상 0!!!!

# 데이터 분석을 하려면 필요한 것들

당연하지. 평균은 수학적으로 가장 가운데 있는놈인데 그녀석을 기준으로 거리를 -, + 한걸 다 더해버리면...당근 0 아니겠음??

어떻게 해버릴까... 에라 모르겠다. 0만 안되면 되니까 그냥 거리를 다 제곱해서 양수로 만들고 더해버려!!!! = 편차제곱합(sum of square = ss)

그리고 그걸 평균내자.

즉, (거리)<sup>2</sup> 한걸 다 더해서 이걸 N으로 나누자! = 분산

# 데이터 분석을 하려면 필요한 것들

분산의 단점은 거리를 다 제공해서 더했기 때문에 거리랑 상관없는 무슨 이상한 어마무시한 수치가 뱅뛰기 되서 나온다는게 문제입니다. 그쵸?

아 제공이 문제라고? 분산에다가 그럼 root 씹어버려!!! = 표준편차

즉, 분산과 표준편차는 같은 개념이고 root를 씹었냐 안씹었냐의 차이일 뿐입니다. 간단하죠?

$$\text{root}(\text{분산}) = \text{표준편차} \circ \circ$$

# 데이터 분석을 하려면 필요한 것들

## 평균, 분산, 표준편차

Start	1번	2번	3번	4번	5번
수량	175	177	179	181	183
<div>평균</div> <div>(수량의 합 / 인원수)</div>	$= \frac{175+177+179+181+183}{5} = 179$				
<div>편차</div> <div>(수량 - 평균)</div>	$175 - 179$ $= -4$	$177 - 179$ $= -2$	$179 - 179$ $= 0$	$181 - 179$ $= 2$	$183 - 179$ $= 4$
<div>편차제곱</div> <div>(편차 * 편차)</div>	$(-4) * (-4)$ $= 16$	$(-2) * (-2)$ $= 4$	$(0) * (0)$ $= 0$	$(2) * (2)$ $= 4$	$(4) * (4)$ $= 16$
<div>분산</div> <div>(편차제곱의 평균)</div>	$= \frac{16+4+0+4+16}{5} = \frac{40}{5} = 8$				
<div>표준편차</div> <div>(<math>\sqrt{\text{분산}}</math>) </div>	$= \sqrt{8} \approx 2.828$				
결론	인원수=5, 평균=179, 분산=8, 표준편차=2.828				

# 데이터 분석을 하려면 필요한 것들

평균, 분산, 표준편차

Start	1번	2번	3번	4번	5번
변량	175	175	175	175	175
<div>평균</div> <div>(<math>\frac{\text{변량의 합}}{\text{인원수}}</math>)</div>	$= \frac{175+175+175+175+175}{5} = 175$				
<div>편차</div> <div>(<math>\text{변량} - \text{평균}</math>)</div>	$175 - 175$ $= 0$	$175 - 175$ $= 0$	$175 - 175$ $= 0$	$175 - 175$ $= 0$	$175 - 175$ $= 0$
<div>편차제곱</div> <div>(<math>\text{편차} * \text{편차}</math>)</div>	$(0) * (0)$ $= 0$	$(0) * (0)$ $= 0$	$(0) * (0)$ $= 0$	$(0) * (0)$ $= 0$	$(0) * (0)$ $= 0$
<div>분산</div> <div>(<math>\text{편차제곱의 평균}</math>)</div>	$= \frac{0+0+0+0+0}{5} = \frac{0}{5} = 0$				
<div>표준편차</div> <div>(<math>\sqrt{\text{분산}}</math>)</div>	$= \sqrt{0} = 0$				
결론	인원수=5, 평균=175, 분산=0, 표준편차=0				

# 데이터 분석을 하려면 필요한 것들

## **안되겠다. 대충 때려맞추자 : 점추정과 구간 추정**

"표본집단만 알고있을경우, 표본집단의 특성을 이용해서 원래 알고싶은 모집단을 추정하는것을 추론통계 라고 부릅니다"

라고 위에서 언급했습니다.

근데 추정방법에는 이렇게 두가지가 있음요....

# 데이터 분석을 하려면 필요한 것들

•

점추정

|————|

구간추정

# 데이터 분석을 하려면 필요한 것들

**점추정:** 표본에서 얻은 값을 그냥 모집단 값이라고 우긴다  
가장 단순한 방법입니다. 예를들어 어떤 데이터에서 랜덤하게 뽑아낸 수치가 다음과 같다고 합시다.

[1,2,3,4,5]

이 표본 집단의 평균은 얼마죠? 3이죠?

점추정은 이 3이 곧 모집단의 평균이다!!! 라고 추정하는겁니다.  
아무도 모르니까 일단 우기자??



# 데이터 분석을 하려면 필요한 것들

**구간추정:** 표본에서 얻은 값에 확률적으로 상, 하한(구간)을 추가하여, 그 구간으로 모집단 수치를 예측한다

위 예를 들어, 3을 기준으로 95% 신뢰구간 값이  $\pm 1$ 이라고 하면,  $(3-1) \sim (3+1)$ . 즉, 2~4 사이에 실제모집단의 평균이 있을 확률이 95%이다. 라고 하는겁니다. 어쨌든 추정이죠?

# 데이터 분석을 하려면 필요한 것들

**또라이불변의 법칙:** 세상 살다보면 또라이는 하나씩 있다(**outlier**)  
학교나 사회생활 하시게 되면 누구나 공감하는 부분....data에도  
있습니다.

문제는 정작

# 데이터 분석을 하려면 필요한 것들



문제는 정작 자기가 또라이라는걸 모른다는거...

# 데이터 분석을 하려면 필요한 것들

data를 다루다보면 이런일들이 많이 발생합니다

갑자기 말도 안되는 수치가 튀어나올때가 있죠. 누가봐도 어디서 문제가 생긴게 분명합니다.

이게 왜 그렇게 분석에 문제가 되냐고요?? 아래를 보시면 한번에 이해됩니다

# 데이터 분석을 하려면 필요한 것들

In [6]:

```
import numpy as np
a = [1,2,3,4,5]
b = [1,2,3,4,999999] # 999999은 자기가 또라이인지 스스로 알방법이 없습니다...

print('a의 평균 :', np.mean(a))
print('b의 평균 :', np.mean(b))
```

a의 평균 : 3.0

b의 평균 : 200001.8

# 데이터 분석을 하려면 필요한 것들

**이 세상에 하얀 까마귀는 없는것인가??**

까마귀가 까맣다는것을 통계적으로 증명하는 방법

이번엔 좀 다른 애길 해보겠습니다

까마귀가 까맣다는 주장(명제)이 있다고 합시다.

이게 참이라고 통계적으로 증명하고 싶습니다.

어떻게 해야 할까요??

아래 두 접근 방법이 있습니다

# 데이터 분석을 하려면 필요한 것들

1. 이 세상 모든 까마귀를 다 잡아와서 모두 검다는것을 확인한다:  
단하나의 예외만 나와도 거짓인 주장이 됩니다
2. 까마귀중 일부를 잡아와서, '하얀' 까마귀가 나올 확률을 역으로 추정한다: 즉, 나의 주장과 반대의 주장이 나올 확률을 본다  
이 반대 주장이 나올 확률이 너무너무너무너무 낮으면, 반대주장은 확률적으로 틀린것이고 나의 주장이 맞다

# 데이터 분석을 하려면 필요한 것들

위 2의 접근 방법을 '**통계적 가설검증**' 이라고 부릅니다.

현실적으로 1은 어렵기 때문에, 2의 방법으로 접근 합니다. 이때,

**내 주장:** 까마귀는 까맣다(애를 검증하지 않습니다)를 **대립가설**

**반대 주장:** 까마귀는 까맣지 않다(애를 실제로 검증해서 확률을  
봅니다) 를 **0가설** 이라고 부릅니다



# 데이터 분석을 하려면 필요한 것들

## 내말이 맞는 이유는 니가 틀려서야 : 0가설 검정

0가설을 실제로 검증하기 때문에 이것을 0가설 검정이라고도 부릅니다. 이해되시죠?

세상을 반으로 가른다고 보면 됩니다. 내가 맞는 세상과 내가 틀린 세상

내가 틀린 세상이 참임을 가정하여, 구간을 정해놓고 실제로 어떤값이 나오는지 봅니다.

이 확률이 낮으면?

내가 틀린 세상이 존재하지 않는다고 보고, 그럼 나머지는 내가 맞는 세상밖에 없습니다.

ㅇㅋ 복잡하지만 이해했어!!!

어?? 근데 잠깐만.. 이거 0가설 검정이 뭔지는 알겠는데...하나 더 필요한게 있는데????

# 데이터 분석을 하려면 필요한 것들

내 반대주장이 맞다/틀리다의 '기준'이 필요하잖아?

어떨때 틀렸다고 하지?: 알파값과 p - value

정확한 지적입니다. (뭐가?)

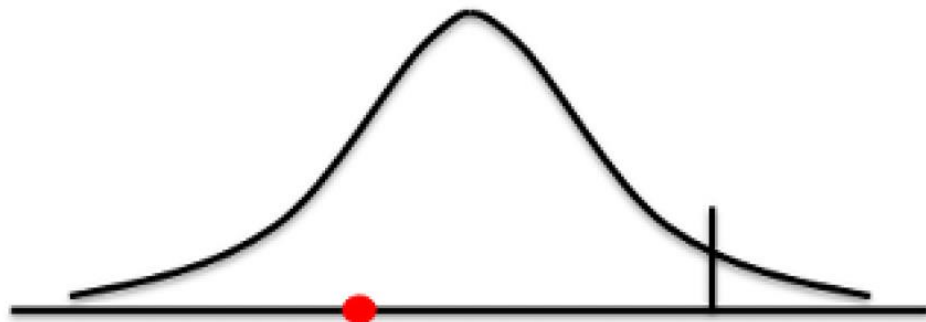
내 반대주장(0가설)을 검정한다는 의미는 알겠는데,

"그럼 어떤 기준에 따라서 검정하지??" 의 문제가 있습니다

이때의 기준이 알파값이고 보통 0.05, 즉 95% 신뢰구간으로 정합니다.

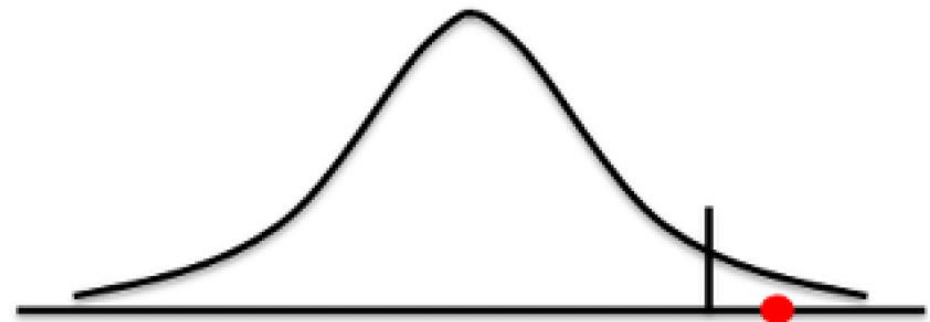
P-value는 실제로 관측된 확률입니다

# 데이터 분석을 하려면 필요한 것들



검정통계량

<귀무가설 채택>



검정통계량

<귀무가설 탈락>

귀무가설 = 내 반대주장 = 0가설  
검정통계량 : 실제로 관측한 값

# 데이터 분석을 하려면 필요한 것들

이야기는 이러한 의미를 내포하고 있습니다.

내 반대주장이 참인 세상을 만들고, 이때의 평균과 평균으로부터 상, 하한 구간을 (95%로) 정해 놓는다.

실제 실험이나 관측치가 이 신뢰구간 안에 포함되는지 본다. 즉, 우리는 경계선을 95%로 정하는 것이다. 5%는 인간적으로 너무 낮잖아???

95% 안에 포함되면 실제 생활에 있는 경우로 판단하여, 내 반대주장을 기각할수 없고, 아주 극단적으로 낮은(5%)이하면 반대주장을 기각한다

즉 상대방 말이 맞을 확률이 5% 이하이면, 그 주장을 기각하고 내가 맞다고 친다. 왜냐...5%는 너무낮잖아???

# 데이터 분석을 하려면 필요한 것들



뇌 주름이 모두다 꼬이는거같다