

일단, 문과범위에서 “통계적추정” 단원에서 배우는 용어는 이게

전부입니다. (전수조사 같은 건 제외.)

그런데 사실, 과외해오면서 느끼는 건, 저렇게 써놓으면

다들 납득은 하지만,

정말 저 용어들 사이의 관계를 단박에 이해하는 학생이 드물어요.

그래서 문제 살짝만 비틀어도 어찌해야 할지 모르고 있습니다.

개념이 제대로 안되어 있다는 거죠.

물론 문제 많~이 풀어서 어렵פות한 기억에 의존해서 푸는 학생도 있지만,

(물론 다 맞긴함.)

적어도 교과 내 개념정돈 완벽해야 합니다.

지금 볼 것은,

이미 교과서에 다 서술 되어있는 것이지만,

완벽히 짚고 넘어가지 않았을 여러분 들이란걸 제가 잘 알기에,

다시 한번 이해하기 쉽게 풀어 설명해볼 것입니다.

자, 용어에 대한 개념은 다 알고 있던 전제하에 설명해볼게요.

우리가 어떤 전체 대상의 평균을 알고 싶다 칩시다.

그건 “모평균” 이겠죠??

만약 전체 대상의 숫자가 적다면, 그냥 전체 대상을 하나 하나 구한 뒤

그걸 전체 대상의 수로 나누면 되겠죠?

근데, 전체 대상의 수가 너무 크면 어찌죠?

하나하나 다 구할 수가 없겠죠? 물론 한다면야 구할 수가 있게지만.

그렇다고 안 구할 순 없고...

그래서 선택한게 바로 “표본평균의 평균”입니다.

왜냐면, “표본 평균의 평균” 은 “모평균” 과 같거든요.

(이유는 고등교과내에선 배우진 않습니다. ㅇㅇ 그렇구만 하시면 되요)

다시,

“모평균”을 구하기 어려워서 “표본평균의 평균”을 구한다.

자, 이해갔죠?

그런데, **표본평균의 평균** 은 무슨 말일까요?

첫 장에서도 다뤘지만, **“표본”은 “모집단의 일부”**입니다.

하나 짚고 가야할건, 표본은, 무한히 만들 수 있습니다.

그냥 계속 만들면 되는거 거든요. 방법이 어찌됐든.

10개든, 100개든, 1000개든.

(물론, 몇 개를 뽑을진 정해야 합니다. 가령, 모집단(전체개수)이 1000개라

치고, “내가 그 중 81개를 무작위로 뽑아 그 81개를 표본으로 삼겠다.”

하는 그 “81개” 는 정해야 합니다. 숫자가 클수록 좋겠죠?)

아무튼, 그럼 81개씩 뽑아서 그러한 표본 100개를 만들었다 칩시다.

그럼 100개의 표본 이 나한테 있는 거겠죠?? (표본간 중복은 가능합니다.)

그 중 1개의 표본의 평균을 “표본평균” 이라 합니다.

다시다시, 제가 전체 1000개 중 81개를 뽑아서 1개의 표본을 만들었죠?

그럼 그 1개의 표본내의 81개의 평균. 그걸 표본평균이라 한다 이겁니다.

(여기서 81개를 **표본의 크기**라 합니다.)

여기서 이해안가면 ... 흠...ㅏㅏ 가시죠?

그래도 안되시는 분들을 위해 교과서 예를 한번보죠.

*공 1,2,3,4,5 가 하나의 주머니에 들어있다. 철수와 영희가 무작위로 3개를

뽑아 표본으로 삼을 때, 표본평균이 더 높은 사람은 ?

- 철수의 표본 : 1, 1, 3

- 영희의 표본 : 2, 3, 5

철수와 영희는 전체 수가 5개인 모집단에서 3개씩 뽑아 하나의 표본을

만든겁니다.

철수의 표본평균은 $(1+1+3)/3 = 5/3$.

영희의 표본평균은 $(2+3+5)/3 = 10/3$.

영희가 더 높군요.

영희가 문제가 아니라, 표본평균자체를 이해하셨어야 합니다.

오키. 간단했고. 계속 보죠.

그럼 표본평균의 평균은 뭐냐??

쉽습니다.

우리가 표본을 100개 만들었죠?

1000개의 모집단에서 81개씩 취해서 1개의 표본을 만드는데,

그러한 표본을 100개 만들었습니다.(표본의 크기가 81)

그 중 1개의 표본의 평균을 표본평균이라 한다 했는데,

그럼 표본이 100개 있으니깐 표본평균 은 100개 있겠죠?

그럼 그 **“100개의 표본평균”**의 평균을

“ 표본평균의 평균”

이라 합니다.

뭐 사실 “ 표본평균들의 평균 ” 이 더 와닿겠지만요 ㅎㅎ

아무튼, 이해하셨죠? (**참고로 표본평균들은 다 다른값입니다. 물론 같을 수도 있구요**)

그런데 처음에 뭐라했었냐면,

“ 모평균 ”을 구하기 어려워서 **“ 표본평균의 평균”**을 구한다.

라 했었죠?

즉, 모평균을 구하기 어려워서 표본평균의 평균을 구해야하는데

그 표본평균의 평균은 표본평균들을 모아서 그것들을 평균내면 된다.

이겁니다.

자, 이제 총정리를 해보면,

모평균을 구하기 쉬우면? 강 구한다. 전체값을 구하고 전체 수로 나눠줌.

근데 모평균을 구하기 어려우면? 즉 그 전체 수가 너무 크면?

모평균과 같은 값인 표본평균의 평균을 구한다.

그러려면 표본평균들의 값과 그 개수를 구해야 하고, 그 값들을 개수로 나눠

표본평균의 평균을 구해준다. 그럼 그게 모평균이다. 끝.

아.. 쉽죠 ?? ㅎㅎ

자.. 근데 한 개가 더 남았죠..

만약, 모평균 구하기 어려워서 표본평균의 평균을 구하려 했고,

그러려고 표본평균 값들을 구하고 있었는데,

만약 표본 평균값 하나 밖에 모르면 어떨까요?

달리 방도가 없죠?

그래서 우린,

추정을 합니다.

모평균을요.

그게 바로 ‘모평균의 추정’입니다.

왜 이걸 배운다고요??? 말했죠? 모평균구하려고 표본평균의 평균구하려 했는데

그거조차 못 구하니, 그나마 있는 표본평균 값으로 추정이라도 하는겁니다.

연결이 되죠???? 다 연결되는 개념들입니다.

이쯤에서 개념한번 다시보죠.

모평균 m , 모표준편차 σ 인 모집단에서 임의 추출한 크기 n 인 표본의 표본평균을 \bar{X} 라 하면 n 이 충분히 클 때, 모평균 m 은

(1) 신뢰도 95%인 신뢰구간은 $[\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}]$

(2) 신뢰도 99%인 신뢰구간은 $[\bar{X} - 2.58 \frac{\sigma}{\sqrt{n}}, \bar{X} + 2.58 \frac{\sigma}{\sqrt{n}}]$

참고로, 이제 저 기호 \bar{X} 가 익숙해야합니다. $E(\bar{X})$ 란도 헛갈리면 안되구요.

결국, 전체적인 그림을 보면, 초점은

“모평균 ” 구하기

에 있음을 알 수 있습니다. 이해하시죠 ?? 저것을 중심으로 이러한 개념들이

등장한 겁니다. 이제 문제 볼 시간이군요.

<2012 수능 가형 9번 문항>

어느 회사에서 생산하는 음료수 1병에 들어 있는 칼슘 함유량은 모평균이 m , 모표준편차가 σ 인 정규분포를 따른다고 한다. 이 회사에서 생산한 음료수 16병을 임의추출하여 칼슘 함유량을 측정한 결과 표본평균이 12.34이었다. 이 회사에서 생산한 음료수 1병에 들어 있는 칼슘 함유량의 모평균 m 에 대한 신뢰도 95%의 신뢰구간이 $11.36 \leq m \leq a$ 일 때, $a + \sigma$ 의 값은? (단, Z 가 표준정규분포를 따를 때 $P(0 \leq Z \leq 1.96) = 0.4750$ 이고, 칼슘 함유량의 단위는 mg이다.) [3점]

- ① 14.32 ② 14.82 ③ 15.32
 ④ 15.82 ⑤ 16.32

한번 풀어보죠. 풀이는 옆에서.

사실 개념이 완벽하다 해서 바로 문제가 술술 풀리진않아요.

다만 문제를 통해 개념적용을 할 때 수월할 뿐이죠.

근데 그게 엄청 중요한겁니다 ㅎㅎ

자 시작해봅시다.

첫줄 보면, 음료수 1병의 모평균과 모표준편차가 등장하죠?

그런데, 굳이 모평균, 모표준편차라 하지 않고

그냥 1병의 평균과 표준편차라 하면 . 알아서 그게 모평균, 모표준편차라고

알아야 해요.

아무튼, 계속 보면.

음료수 16병을 임의추출했다네요. 그러면 표본의 크기가 16인거죠?

$n = 16$ 이란겁니다.

표본평균이 12.34 이네요. 이건 모평균이 아니죠???모평균은 표본평균의 평균.

표본평균을 하나 밖에 모르니 추정을 했군요.

신뢰구간이 주어져 있고, 95% 일때의 z 를 아래의 주었네요.

그럼 신뢰구간을 직접 구해줘서 신뢰구간의 왼쪽값이 11.36, 오른쪽 값이 a 라

하고 풀면 됩니다.

풀이는 생략! 다음문제 풀어보세요.

어느 약품 회사가 생산하는 약품 1병의 용량은 평균이 m , 표준편차가 10인 정규분포를 따른다고 한다. 이 회사가 생산한 약품 중에서 임의로 추출한 25병의 용량의 표본평균이 2000 이상일 확률이 0.9772일 때, m 의 값을 오른쪽 표준정규분포표를 이용하여 구한 것은? (단, 용량의 단위는 mL이다.) [3점]

z	$P(0 \leq Z \leq z)$
1.5	0.4332
2.0	0.4772
2.5	0.4938
3.0	0.4987

- ① 2003 ② 2004 ③ 2005 ④ 2006 ⑤ 2007

풀이는 다음장에.

어느 약품 회사가 생산하는 약품 1병의 용량은 평균이 m , 표준편차가 10인 정규분포를 따른다고 한다. 이 회사가 생산한 약품 중에서 임의로 추출한 25병의 용량의 표본평균이 2000 이상일 확률이 0.9772일 때, m 의 값을 오른쪽 표준정규분포표를 이용하여 구한 것은? (단, 용량의 단위는 mL이다.) [3점]

z	$P(0 \leq Z \leq z)$
1.5	0.4332
2.0	0.4772
2.5	0.4938
3.0	0.4987

- ① 2003 ② 2004 ③ 2005 ④ 2006 ⑤ 2007

어느 공장에서 생산되는 제품의 길이 X 는 평균이 m 이고, 표준편차가 4인 정규분포를 따른다고 한다. $P(m \leq X \leq a) = 0.3413$ 일 때, 이 공장에서 생산된 제품 중에서 임의추출한 제품 16개의 길이의 표본평균이 $a-2$ 이상일 확률을 오른쪽 표준정규분포표를 이용하여 구한 것은? (단, a 는 상수이고, 길이의 단위는 cm이다.) [4점]

z	$P(0 \leq Z \leq z)$
1.0	0.3413
1.5	0.4332
2.0	0.4772

- ① 0.0228 ② 0.0668 ③ 0.0919
 ④ 0.1359 ⑤ 0.1587

자, 하나씩 뜯어볼까요?

아까도 말했지만,

약품1병의 용량의 평균이 m 이니깐, 그게 그냥 모평균입니다.

마찬가지로 표준편차가 10인데 그 표준편차가 모표준편차이구요.

임의로 25병 추출했는데, 표본의 크기가 25인거죠?

용량의 표본평균이 2000이상이다...이고 표가 주어져있으니,

표준화 하는 문항임을 알 수 있습니다.

식을 써보면,

$$P(\bar{X} \geq 2000) = P(Z \geq 2000 - m/2) = 0.9772$$

표를 참고하면,

$$0.9772 = 0.5 + 0.4772 = P(Z \geq -0.2)$$

아하. 그럼 $2000 - m/2$ 가 -0.2 겠네요. 끝.

쉽네요. 다음문제 바로 볼게요.

풀이는 다음 장에서.

어느 공장에서 생산되는 제품의 길이 X 는 평균이 m 이고, 표준편차가 4인 정규분포를 따른다고 한다.

$P(m \leq X \leq a) = 0.3413$ 일 때, 이 공장에서 생산된 제품 중에서 임의추출한 제품 16개의 길이의 표본평균이 $a-2$ 이상일 확률을 오른쪽 표준정규분포표를 이용하여 구한 것은? (단, a 는 상수이고, 길이의 단위는 cm이다.) [4점]

z	$P(0 \leq Z \leq z)$
1.0	0.3413
1.5	0.4332
2.0	0.4772

- ① 0.0228 ② 0.0668 ③ 0.0919
 ④ 0.1359 ⑤ 0.1587

마찬가지로, 처음에 m , 4 가 모평균과 모표준편차 임을 알아야하고.

(계속 반복되는게 보이죠? 이래서 기출풀이가 필요합니다.)

또 확률 하나 주어져 있고 표도 주어져있으니 표준화 하는 문항이겠네요.

이정도 보여야하고.

표본의 크기가 16인 표본을 하나 뽑았군요.

하나뽑았으니깐 모평균을 추정하는 문항일수도 있게지만

표줬으니 표준화 문항이겠죠.

길이가 $a-2$ 이상이다 .. 하는걸 보니 표준화 맞네요 ㅎㅎ

풀어볼까요?

3번째 줄을 표준화 해주면,

$$P(0 \leq Z \leq a-m/4) = 0.3413$$

즉, $a-m/4 = 1.0$ (by 표)

$a-m = 4$ 군요.

그 후, 표본평균이 $a-2$ 일 확률을 구해보면,

$$P(\bar{X} \geq a-2) = P(Z \geq a-2 - m/4) \quad (\because \text{표본평균의 표준화 공식..})$$

그런데 $m = a-4$ 이므로,

$$P(Z \geq a-2 - m/4) = P(Z \geq 2/1) = 0.0228.$$

1 번.

쉽죠 ?? ㅎㅎㅎ

이 문제풀이 자체가 안되는 분들은,

part 2 보단 앞으로 쓰여질 part 1 자체를 공부하는게 나올거예요.

추정부분에서의 문제보단

아직 표준화 자체에 대한 개념이 덜잡힌겁니다.

흠. 최근 3개년 수능에서의 문제를 찾아봤는데.

확실한 개념의 정립여부를 묻는 문항은 아직 출제되지 못한 것같네요.

현재 출판되고 있는 사설 실모에선 종종 보이긴 하다면.

이 칼럼보고 완전히 정립된 개념이 없어도,

기출의 무수한 반복으로 애시당초 풀이틀이 갖춰지면, 수능에서 크게 어려움은

없을 듯 싶습니다만.

칼럼 초반에도 말씀드렸다시피, 적어도 교과내 개념정돈, 완벽히 해서

나쁠 건 없습니다.

기출로 부족한 것 같다 !! 싶으면 사설문제집으로 공부해도 되는 부분입니다.

이 칼럼으로 각 용어들 사이 관계가 완벽히 잡혔다 하면

그 어떤 문항풀이습득도 어렵지 않을 겁니다 ㅎㅎ 열공하세요.