제 12 장 기대값과 표준오차

1. 기 대 값
2. 표 준 오 차
3. 정규분포곡선 이용하기
4. 분류하고 횟수세기

1. 기대값

기대값

기대값

- 하나의 확률 과정에 의해 결정되는 숫자는 하나의 값 주위로 분포한다. 이때 기대값(expected value)은 분포의 무게 중심에 해당되는 값이다.

표준오차

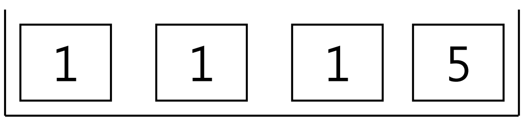
- 실현된 값이 기대값과 차이가 나는 정도는 표준오차(standard error)에 의해 측정된다.

→확률과정에 의해 결정되는 숫자는 기대값 주위로 분포하며, 기대값과 표준오차 정도 차이가 나는 경향이 있다.

1. 기대값

예1

5가 한 장, 1 이 세 장 들어 있는 상자로부터 100회 무작위 복원추출을 한다. 이 때, 추출된 숫자들의 합의 기대값은?



$\boxed{5}$ 가 나올 확률은 1/4

$\boxed{1}$ 이 나올 확률은 3/4

총 100회 추출에서 나온 숫자들의 합은 25 × 5 + 75 × 1 = 200 으로 예상됨.

이 값이 바로 **기대값(expected value)**임

합의 기대값을 구하는 다른 방법으로 아래 공식을 이용할 수도 있다.

합의 기대값 = 추출횟수 × 상자의 평균

1. 기대값

기대값

앞의 공식에 대해 좀더 살펴보자.

- 상자의 평균 = (1+1+1+5) / 4 = 1 X (3/4) + 5 X (1/4) =2
- 추출 횟수 = 100
- 합의 기대값 = 100 X 2 = 200

1. 기대값

기대값

잘 섞인 한 벌의 카드에서 한 장의 카드를 뽑는다. 이 카드가 스페이드인지, 다이아몬드인지, 하트인지, 아니면 클럽인지를 맞추면 2천원을 따고, 틀리면 천원을 잃는 게임이 있다. 이 게임을 100번 반복할 때, 총 얼마를 딸 것으로 기대하는가?

☞ 상자모형을 만들자. 순이득이 2000원 증가할 확률은 1/4, 천원 감소할 확률은 3/4 이다.

2000	-1000	-1000	-1000

상자의 평균은 2000X(1/4)+(-1000)X(3/4)=-250이고, 게임을 100회 반복한다면 100X250=25,000원 정도 손해를 볼 것으로 기대된다.

2. 표준오차

표준오차

(추출한 숫자들의 합) = (합의 기대값) + (합에 든 확률오차)

전반적인 크기=**표준오차(SE)**

제곱근법칙(square root law):

상자에서 무작위 복원 추출할 경우 합의 표준오차 계산

(합의 표준오차) = (상자의 표준편차) X (추출회수)$^{1/2}$

→ 추출회수가 1인 경우 합의 표준오차는 바로 상자의 표준편차와 같다.

2. 표준오차

표준오차

(합의 표준오차) = (상자의 표준편차) X $(\text{추출회수})^{1/2}$

- **상자의 표준편차**: 상자 안의 든 숫자들이 서로 다른 정로를 나타냄. 상자 안에 든 숫자들이 서로 다를수록 예측하기 어렵다. → 표준편차가 커지면 표준오차도 커진다.

- **추출회수**: 상자로부터 카드를 뽑을 때 어떤 숫자를 뽑을지 알 수 없기 때문에, 추출회수를 늘리면 그 합을 예측하기는 점점 더 어려워진다. 그러나 그 난이도가 추출회수에 비례하여 증가하지는 않는다. 난이도를 측정하는 표준오차는 추출회수의 제곱근에 비례하여 증가한다.

2. 표준오차

제곱근 법칙

케리히(John Kerrich, 1903–1985, 남아공)의 동전던지기 실험에서

(앞면이 나오는 횟수) = (던진 횟수의 절반) + (확률오차)

- (앞면이 나온 횟수의 표준오차) = ½ x (던진 횟수)$^{1/2}$
- 던진 횟수가 증가하면서 표준오차의 절대적인 크기는 증가하지만 표준오차의 상대적인 크기는 감소

(앞면이 나오는 비율) = (앞면이 나오는 횟수)/던진 횟수

- (앞면이 나오는 비율의 표준오차) = (앞면이 나오는 횟수의 표준오차)/던진 횟수
  = 1/2 / (던진 횟수)$^{1/2}$

제곱근 법칙

- 오차의 절대적 크기는 던진 횟수의 제곱근으로 곱해져 증가함
- 오차의 상대적 크기는 던진 횟수의 제곱근으로 나누어져 감소함

2. 표준오차

표준오차

0, 2, 3, 4, 6의 카드가 한 장씩 들어 있는 상자로부터 25회 무작위 복원 추출하는 경우 합의 표준오차를 구해보자.

- 상자의 평균 = 3. 합의 기대값 = 25 x 3 = 75
- 상자의 표준편차(모표준편차)

$$= \sqrt{\frac{(-3)^2 + (-1)^2 + 0^2 + 1^2 + 3^2}{5}} = \sqrt{\frac{9+1+0+1+9}{5}} = \sqrt{\frac{20}{5}} = 2$$

- 합의 표준오차 = $25^{1/2}$ x 2 = 10
- 관측된 합은 그 기대값인 75 주위에 그 표준오차인 10 정도의 차이를 두고 분포한다.

2. 표준오차

표준오차

25회씩의 추출을 100번 반복 시행한 결과

- 원칙적으로 합은 0 에서 150 까지 가능.
  그러나 관측치들은 그 대부분인
  99개가 50과 100 사이, 즉,
  기대값 ± 2.5SE 이내에 위치

- 하나의 관측치가 그 기대값으로부터
  ±2SE 혹은 ±3SE 이상 떨어져 실현될
  가능성은 매우 희박함

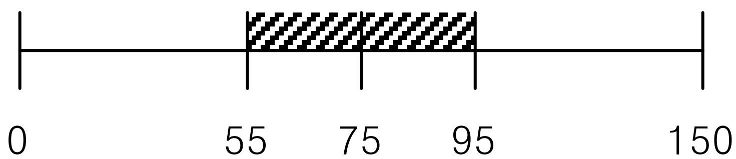
25씩의 추출을 100번 반복한 결과의 기록

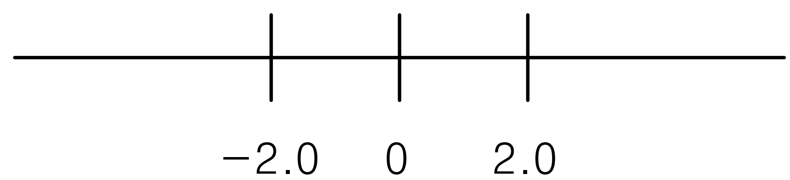
반복 시행	관측된 합	반복 시행	관측된 합	반복 시행	관측된 합	반복 시행	관측된 합	반복 시행	관측된 합
1	74	21	75	41	93	61	67	81	66
2	72	22	76	42	72	62	87	82	67
3	82	23	74	43	70	63	72	83	83
4	72	24	99	44	79	64	68	84	80
5	72	25	82	45	74	65	70	85	72
6	59	26	74	46	58	66	79	86	70
7	89	27	79	47	67	67	59	87	76
8	89	28	64	48	77	68	80	88	58
9	80	29	73	49	85	69	70	89	77
10	74	30	81	50	71	70	84	90	76
11	69	31	70	51	72	71	91	91	86
12	67	32	49	52	77	72	77	92	78
13	76	33	80	53	69	73	63	93	74
14	71	34	67	54	72	74	74	94	60
15	77	35	78	55	69	75	80	95	94
16	75	36	60	56	55	76	84	96	85
17	62	37	85	57	70	77	92	97	81
18	69	38	83	58	75	78	82	98	82
19	72	39	74	59	59	79	73	99	80
20	81	40	84	60	72	80	74	100	62

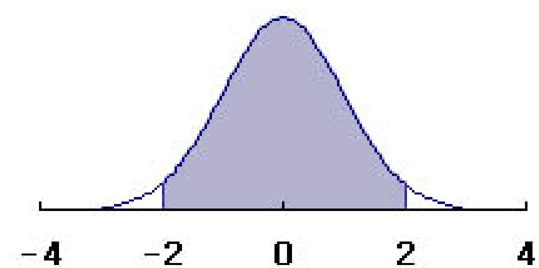
3. 정규분포곡선 이용하기

정규분포곡선 이용하기

앞의 예에서 합이 55에서 95사이에 있을 확률을 구해보자.







- 기대값이 75이고 표준오차가 10 이므로 95는 +2SE, 55는 –2SE로 나타낼 수 있다.

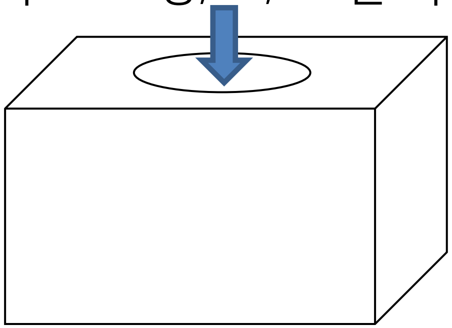
- (55, 95) 구간은 기대값 주위로 ±2SE이내에 해당되는 구간이다. 따라서 68-95 규칙을 이용하면 합의 관측치가 (55, 95)구간에 속할 확률은 95%로 짐작된다.

3. 정규분포곡선 이용하기

정규분포곡선 이용하기

어느 카지노에 있는 하나의 룰렛 머신에서 한 달 동안 10,000 번의 룰렛 게임이 이루어진다고 하자. 한 번의 룰렛 게임에서 카지노가 돈을 딸 확률은 20/38이고 고객이 돈을 딸 확률은 18/38이다. 카지노를 이용하는 고객은 게임당 1,000원씩 걸며 매 게임은 독립이라고 하자. 카지노 측이 이 머신 하나로부터 한 달간 25만원 이상 딸 확률은?

+1,000원 카드 20 장, -1,000원 카드 18 장



3. 정규분포곡선 이용하기

정규분포곡선 이용하기

평균: $$(+1{,}000)\times\frac{20}{38}+(-1{,}000)\times\frac{18}{38}=50$$

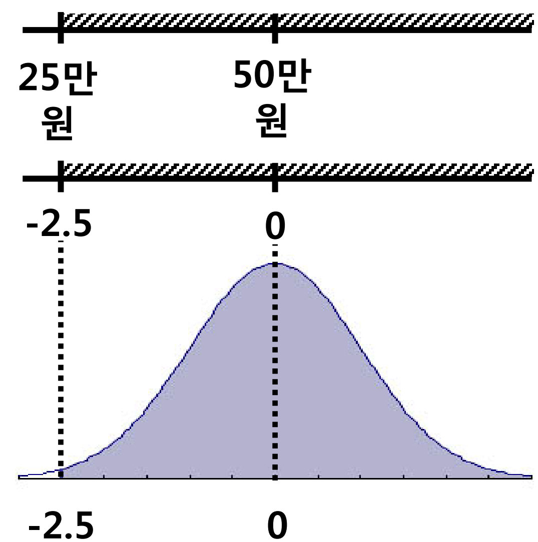
표준편차: $$\sqrt{(+1{,}000-50)^2\times\frac{20}{38}+(-1{,}000-50)^2\times\frac{18}{38}}\approx 1{,}000$$

- 이득의 기대값 = 10,000×50=50만원.
- 이득의 표준오차가 $10000^{1/2}$=100배로 커짐.

  이득의 표준오차 =1,000×100=10만원

카지노가 한 달 동안 특정 룰렛 머신 하나로부터

25만원 이상 딸 확률은 99%나 된다.



3. 정규분포곡선 이용하기

상자 안에 두 종류의 카드만 있는 경우

상자의 표준편차

= (큰수 – 작은수) x [(큰수의 비율) x (작은수의 비율)]$^{1/2}$

앞의 카지노 예에서 상자의 표준편차는

$$(1{,}000-(-1{,}000))\times\sqrt{\frac{20}{38}\times\frac{18}{38}}\approx 1{,}000$$

4. 분류하고 횟수세기

주사위를 60회 던지는 경우 나오는 눈의 합

주사위 하나를 60번 던진다.

1) 나오는 숫자들의 합은 대략 ( )이고, ( ) 정도 차이가 날 수 있다.

→ 1,2,3,4,5,6이 든 상자로부터 60회 무작위로 복원추출

- 상자의 평균 : 3.5
- 상자의 표준편차 = 1.71
- 합의 기대값 = 60×3.5 = 210
- 합의 표준오차 = $60^{1/2} \times 1.71 = 13.25$ 이다.

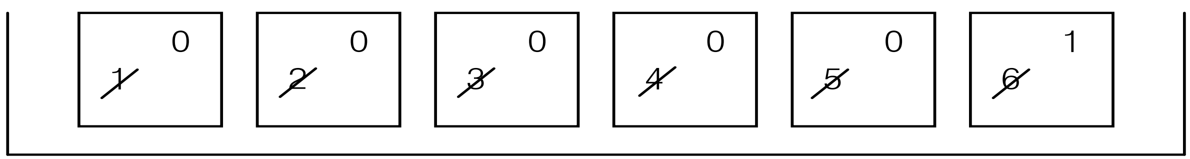
4. 분류하고 횟수세기

주사위를 60회 던지는 경우 6이 나오는 횟수

주사위 하나를 60번 던진다.

2) 6이 나오는 횟수는 대략 ( )이고, ( ) 정도 차이가 날 수 있다.

→ 결과가 성공과 실패로만 분류되는 새로운 상자 모형을 만들어야 한다.



- 6이 나오는 “성공” 횟수의 기대값은 60×(1/6) = 10. 대략 10번 정도 나온다.
- 6이 나오는 “성공” 횟수의 표준오차는 $\sqrt{60}\times\sqrt{1/6\times5/6} \approx 3$
- 기대값 10과 ±3 정도 차이가 날 수 있다.

4. 분류하고 횟수세기

동전을 100번 던져 나오는 앞면의 횟수

동전 하나를 백 번 던질 때, 앞면이 나오는 횟수

- 기대값 : $100 \times (1/2) = 50$
- 표준오차 : $100^{1/2} \times (1/2) = 5$
- 앞면이 나오는 횟수는 50주위로 ±5 정도 오차가 날 수 있다.
- 앞면이 40회에서 60회 사이로 나올 확률은? 이 구간은 기대값 주위로 ±2SE에 해당되므로 구하는 확률은 95%