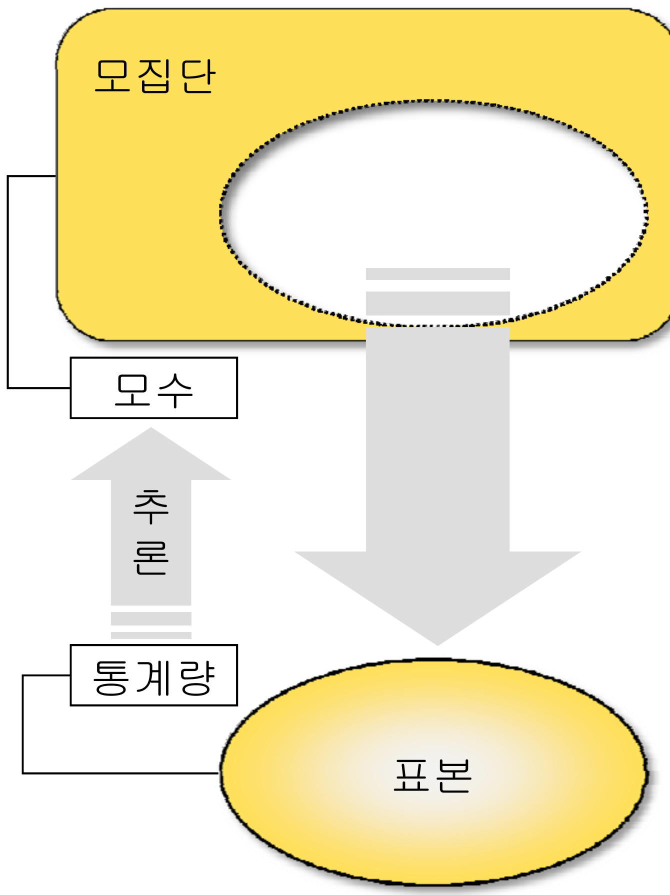
통계학
– 제 6 장 반복표본에서의 변동:표본분포

0011 0010 1010 1101 0001 0100 1011

모집단과 표본

- 모집단 (population)
  - 통계적인 관심의 대상이 되는 모든 개체들의 집합
    - 얻고자 하는 정보를 가지고 있는 하나의 연구 대상
    - 모수 :
      모집단의 수치적 특성치
- 표본 (sample)
  - 모집단에서 추출된 관측값들의 집합
    - 표본에 포함된 정보 분석 ⇒ 모집단의 특성 추론
    - 통계량 :
      표본관측값들의 실수값 함수



표본분포

- 통계량의 확률분포
  - 표본분포 (sampling distribution)
    - 통계량의 확률분포
    - 표본은 모집단의 일부분이므로 통계량의 값은 모수의 참값과 정확히 일치할 것이라고 기대할 수 없음
    - 통계량의 관측값은 많은 가능한 표본들 중에서 실제 추출된 특정한 하나의 표본에 의해서만 결정됨
    - 표본을 반복적으로 추출한다면 그로부터 계산된 통계량의 값들은 달라지므로 그 값들은 변동을 가지게 됨
    - 반복되는 표본추출에서의 변동은 표본분포로 설명될 수 있음
    - 모집단의 분포와 표본의 크기에 영향을 받음

- 예 – [주사위 던지기]

x	1	2	3	4	5	6
$f(x)$	1/6	1/6	1/6	1/6	1/6	1/6

  - 주사위를 던졌을 때 나오는 눈 : X
  - $E\left[X\right] = 3.5 = \mu$
  - $Var\left[X\right] = \frac{105}{36} = \sigma^2$

- 예 – [주사위 던지기]
  - 크기 2인 확률표본의 경우 (조합 수=36)

$x_1 + x_2$	2	3	4	5	6	7
$\bar{x}$	1	1.5	2	2.5	3	3.5
$h(\bar{x})$	1/36	2/36	3/36	4/36	5/36	6/36

$x_1 + x_2$	8	9	10	11	12
$\bar{x}$	4	4.5	5	5.5	6
$h(\bar{x})$	5/36	4/36	3/36	2/36	1/36

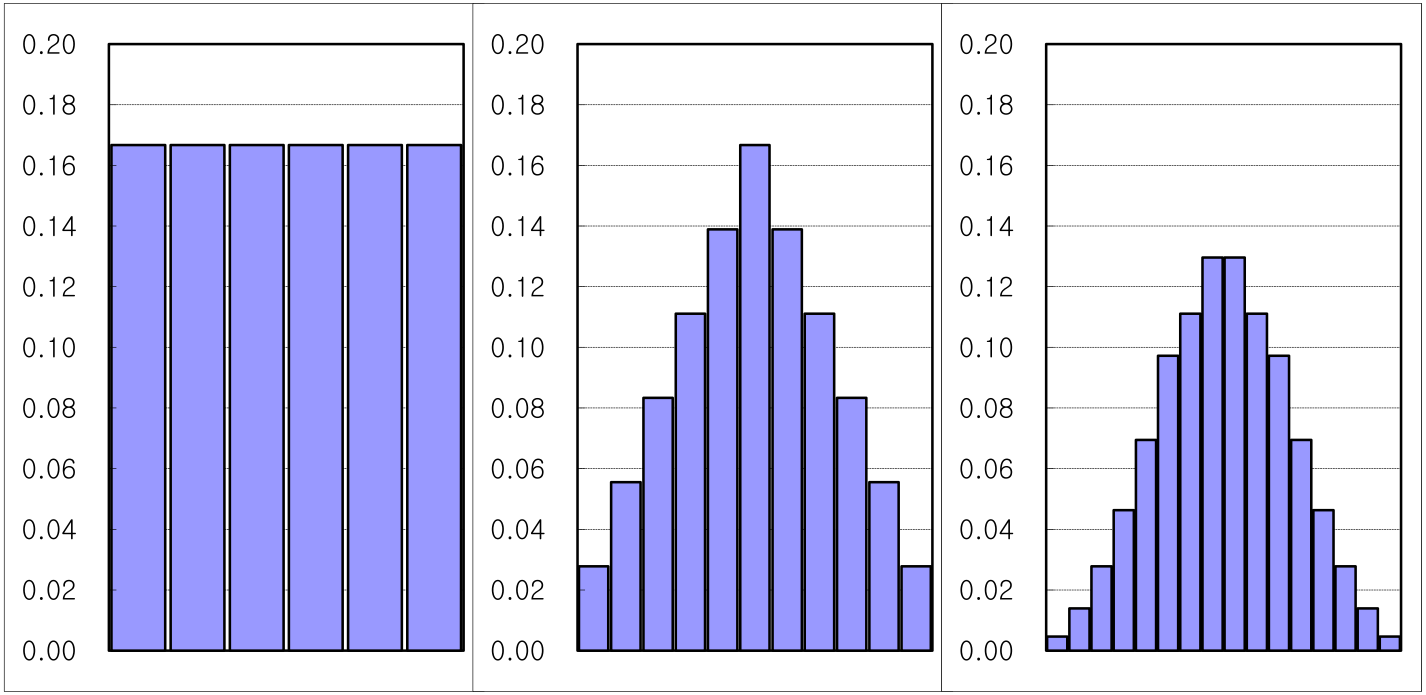
  - 주사위를 던져 나온 눈의 평균 : $\bar{X}$
  - 표본평균의 평균 : $E\left(\bar{X}\right) = 3.5$
  - 표본평균의 분산 : $Var\left(\bar{X}\right) = \dfrac{105}{72}$

- 예 – [주사위 던지기]
  - 크기 3인 확률표본의 경우 (조합 수=216)

$x_1 + x_2 + x_3$	3	4	$\cdots$	17	18
$\bar{x}$	1	4/3	$\cdots$	17/3	6
$h(\bar{x})$	1/216	3/216	$\cdots$	3/216	1/216

  - 표본평균의 평균 : $E\left(\bar{X}\right) = 3.5$
  - 표본평균의 분산 : $Var\left(\bar{X}\right) = \dfrac{105}{108}$

- 표본평균의 분포 변화



표본평균의 분포와 중심극한정리

- 표본평균의 분포
  - 표본평균의 평균과 분산
    - $X_1,\ X_2,\ \cdots,\ X_n$을 평균 μ , 분산 σ^2 인 모집단으로부터의 크기 n 인 확률표본이라 할 때
      - 표본평균의 평균은 모평균과 같다

$$E\left(\bar{X}\right) = \mu$$

      - 표본평균의 분산은 (모분산/ n)과 같다

$$Var\left(\bar{X}\right) = \frac{\sigma^2}{n}$$

    - 모집단이 정규분포를 따르면 표본평균도 정규분포를 따른다
      - 모평균이 μ 이고 모분산이 σ^2인 정규모집단에서 n 개의 표본을 임의로 추출할 때, 그 표본의 평균 $\bar{X}$ 의 분포는 $N\left(\mu,\ \sigma^2/n\right)$ 임

- 중심극한정리 (Central Limit Theorem)
  - $X_1,\ X_2,\ \cdots,\ X_n$을
    평균 μ, 분산 σ^2인 임의의 모집단에서 추출한
    크기 n의 확률표본이라 할 때,
    표본의 크기 n이 커짐에 따라 표본평균 $\overline{X}$의 확률분포는
    근사적으로 정규분포 $N\left(\mu,\ \sigma^2/n\right)$를 따르게 된다

as $n \to \infty$

$$\overline{X} \sim N\left(\mu,\ \sigma^2/n\right)$$

$$Z = \frac{\overline{X} - \mu_{\overline{X}}}{\sqrt{\sigma^2_{\overline{X}}}} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

- 예 – [감자 칩의 실제 무게]
  - 어떤 제과회사에서 감자 칩을 생산
  - 생산된 감자 칩의 포장지에 무게가 455g이라고 표시
  - 공장의 포장 기계가 봉지 당 채워 넣는 감자 칩의 무게는 표준편차가 3.5g인 정규분포에 가까움
  - 봉지 당 무게의 평균이 462g이 되도록 포장기계를 설정하였다면 임의로 추출된 9봉지의 평균 무게가 455g 이하일 확률은?

  - X : 감자 칩 한 봉지의 무게
  - $X \sim N\left(462,\ 3.5^2\right)$
  - $\overline{X}$: 임의로 추출된 감자 칩 9봉지의 평균 무게
  - $\overline{X} \sim N\left(462,\ {3.5^2}/{9}\right)$
  - $P\left(\overline{X} < 455\right) = P\left(\frac{\overline{X} - 462}{3.5/3} < \frac{455 - 462}{3.5/3}\right) = P\left(Z < -6.0\right) = 0$

- 예 – [감자 칩의 실제 무게]
  - 봉지 당 무게의 평균이 462g이 되도록 포장기계를 설정하였다면 포장되는 감자 칩 봉지 중 무게가 455g 이하일 확률은?
  - $P\left(X < 455\right) = P\left(\frac{X - 462}{3.5} < \frac{455 - 462}{3.5}\right) = P\left(Z < -2.0\right) = 0.0228$
  - 임의로 추출된 9봉지의 평균 무게가 455g 이하일 확률이 0.0228이 되기 위해서는 봉지 당 무게의 평균이 얼마가 되도록 포장기계를 설정하여야 하는가?
  - $0.0228 = P\left(\overline{X} < 455\right) = P\left(Z < -2.0\right) = P\left(\frac{\overline{X} - \mu}{3.5 / 3} < \frac{455 - \mu}{3.5 / 3}\right)$
  - $\mu = 455 + 2.0 \times \frac{3.5}{3} = 457.33$