

제 7 장 표본조사와 표본분포

제1절 표본조사

1. 표본조사의 의의

유한모집단과 무한모집단

표본조사의 이유

1. 경제성
2. 신속성
3. 무한모집단의 경우
4. 조사 자체가 파괴 등 대상물의 성질이나 모양을 변경하는 경우

2. 오차의 종류

[1] 표본오차(Sampling error)

모집단의 일부를 선택하여 모수(population parameter)를 추정하기 때문에 생기는 오차

→ 모수(population parameter)와 통계량(statistic)의 불일치

예) 평균: μ 와 \bar{X} 의 불일치, 표준편차: σ 와 S 의 불일치

Statistics are used to make inferences(estimates or decisions) about **unknown population parameters**.

카드 세 장{10, 20, 30}에서 두 장의 카드를 뽑았을 때의 표본평균 ($\mu = 20$)

방법	표본	표본평균
1	10, 20	15
2	10, 30	20
3	20, 30	25

표본오차는 표본의 크기를 크게 하면 감소한다.

[2] 비표본오차(Nonsampling error)

표본의 성격을 관찰하는 방법이 부정확해 생기는 오차

표본으로 추출된 대상을 실제 관찰, 측정할 때 생기는 오차

→ 측정대상과 측정결과의 불일치

비표본오차는 표본의 크기를 크게 해도 반드시 감소하는 것은 아니다. 대상 모집단을 충분히 대표하는 표본을 선택하여 줄일 수 있다.

제2절 표본추출방법

1. 확률추출

확률추출이 효과적이라면 표본의 크기가 적정 규모의 표본을 가져야 하고 무작위성(randomness)이 확보되어야 한다.

[1] 단순확률추출(simple random sampling)

가장 기본적인 확률표본추출 방법

모집단을 구성하는 개체들이 선택될 기회가 동등하게 주어진다.

[2] 체계적 추출(systematic sampling)

모집단을 구성하는 개체들이 무작위로 배열된 상태에서 표본 추출

예: 100명 학생 중 10명 추출하는 방법 - 6, 16, 26, ..., 96번의 학생 추출

[3] 층별 추출(stratified sampling)

모집단을 일정한 수준에 의해 두 개 이상의 동질적인 소집단으로 분류, 각 소집단으로부터 단순무작위추출

예: 각 학년별 100명씩 무작위추출

[4] 군집추출(cluster sampling)

① 모집단의 구성 개체를 우선 자연적으로 또는 인위적으로 몇 개 군집으로 구별

② 무작위로 필요한 군집 선택

③ 선택된 군집에 대해 일부 또는 전수조사

예: 도시 전체 동 중에서 몇 동을 무작위 선정, 선정된 동의 주민 일부 또는 전수조사

2. 비확률추출

추출된 표본에 대해 오차를 명백히 추정하는 통계처리가 불가능한 추출이며, 무작위로 표본을 추출할 수 없음을 의미한다.

비확률추출을 사용하는 이유

① 연구목적이 모수추정에 있는 것이 아니고 특수한 경우를 분석하고자 하는 제한된 목적을 가지고 연구하는 경우

② 시간과 경비를 줄이기 위해

③ 확률추출의 효용성이 제한되어 있는 경우

[1] 편의추출(convenience sampling)

연구자가 자유롭게 연구대상을 표본에 포함시키는 것

예) 새로운 아이디어를 시험하거나 어떤 주제에 대해 아이디어를 얻을 때 유용

[2] 판단추출(judgement sampling)

연구자가 연구목적에 맞는 사항을 충족시킬 수 있다는 판단에 따라 모집단을 대표하는 표본을 선택

예) 자사 종업원을 대상으로 신제품 테스트 → 일반인보다 자사 종업원이 우호적이므로, 여기서 부정적이면 시장에서 성공 불가

탐색연구의 초기 단계에서 편의한 집단(biased group)을 선택하면 유용

[3] 할당추출(quota sampling)

표본이 모집단을 대표하도록 하는 경우에 사용한다.

모집단이 가지는 모든 속성이 선택되는 표본에 비례적으로 적절하게 포함되도록 하는 방법이다.

예) 어느 대학이 남녀 비율이 7:3일 때, 표본에 할당되는 표본비율을 7:3으로 하는 것

제3절 표본분포 (Sampling Distribution)**1. 표본분포의 의미****표본분포(Sampling Distribution)**

모집단에서 **일정한 크기(n)**로 뽑을 수 있는 표본을 모두 뽑았을 때, 그 모든 표본의 통계량의 확률분포를 말한다. → **각 표본의 크기는 모두 n 으로 일정하다.**

2. 평균의 표본분포**평균의 표본분포**

평균의 표본분포는 모집단에서 **일정한 크기(n)로 가능한 모든 표본**을 뽑아서 각 표본의 평균을 계산했을 때 표본평균 $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_k$ 의 확률분포를 말한다. (여기서 k 는 뽑은 표본의 수를 의미한다.)

Example. 세 장의 카드 {2, 3, 4}. 한 장씩 2회 추출(복원추출)할 때 표본분포와 표본평균은?

바구니에 2, 3 및 4가 적힌 카드 3장이 있다고 하자. 바구니에서 카드 한 장을 꺼내 숫자를 확인 후 카드를 다시 바구니에 넣고 다시 한 장을 꺼내 숫자를 확인한다고 하자. 이 실험에서 바구니에 있는 카드의 수는 항상 3이므로, 복원추출에 해당한다. 이 복원추출을 충분히 많이 반복할 때, 추출되는 두 장의 카드(표본) 종류는 총 9이다.

① 한 장씩 2회 추출시 가능한 모든 표본, 각각 표본평균 및 표본이 추출될 확률

번호	가능한 표본	표본 평균	표본 추출 확률	번호	가능한 표본	표본 평균	표본 추출 확률
1	2, 2	2.0	1/9	6	3, 4	3.5	1/9
2	2, 3	2.5	1/9	7	4, 2	3.0	1/9
3	2, 4	3.0	1/9	8	4, 3	3.5	1/9
4	3, 2	2.5	1/9	9	4, 4	4.0	1/9
5	3, 3	3.0	1/9				

② 평균의 표본분포

위의 표에서 표본평균 \bar{X} 가 취할 수 있는 경우는 {2.0, 2.5, 3.0, 3.5, 4.0, 4.5}이며, 각 평균이 추출될 확률은 다음과 같다.

번호	표본평균, \bar{X}	표본평균 확률, $P(\bar{X})$	비고
1	2.0	1/9	(2,2)
2	2.5	2/9	(2,3), (3,2)
3	3.0	3/9	(2,4), (3,3), (4,2)
4	3.5	2/9	(3,4), (4,3)
5	4.0	1/9	(4,4)

③ 평균의 표본분포의 기댓값과 분산

$$\text{기댓값} = \mu_{\bar{X}}$$

$$= E(\bar{X})$$

$$= \sum_{i=1}^5 \bar{X}_i \times P(\bar{X}_i)$$

$$= 2.0\left(\frac{1}{9}\right) + 2.5\left(\frac{2}{9}\right) + 3.0\left(\frac{3}{9}\right) + 3.5\left(\frac{2}{9}\right) + 4.0\left(\frac{1}{9}\right) = 3.0$$

$$\text{분산} = \sigma_{\bar{X}}^2$$

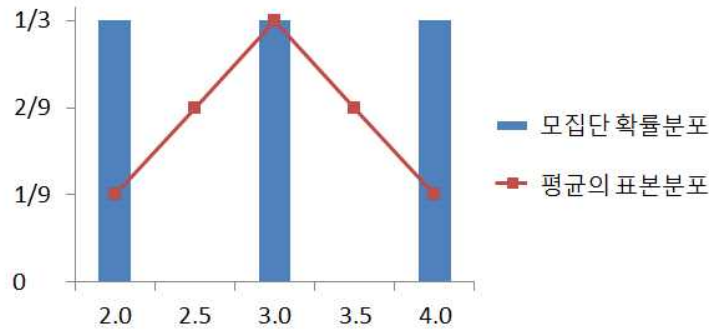
$$= V(\bar{X})$$

$$= E(\bar{X}_i - E(\bar{X}))^2$$

$$= \sum_{i=1}^5 (\bar{X}_i - E(\bar{X}))^2 \times P(\bar{X}_i)$$

$$= (2.0 - 3.0)^2 \left(\frac{1}{9}\right) + (2.5 - 3.0)^2 \left(\frac{2}{9}\right) + (3.0 - 3.0)^2 \left(\frac{3}{9}\right) + (3.5 - 3.0)^2 \left(\frac{2}{9}\right) + (4.0 - 3.0)^2 \left(\frac{1}{9}\right)$$

$$= \frac{1}{3}$$



모의실험 1.

	X_1	X_2	\bar{X}	빈도					누적빈도				
				2.0	2.5	3.0	3.5	4.0	2.0	2.5	3.0	3.5	4.0
1	3	3	3.0	0	0	1	0	0	-	-	1	-	-
2	2	4	3.0	0	0	1	0	0	-	-	2	-	-
3	2	4	3.0	0	0	1	0	0	-	-	3	-	-
4	3	4	3.5	0	0	0	1	0	-	-	3	1	-
5	3	3	3.0	0	0	1	0	0	-	-	4	1	-
6	4	4	4.0	0	0	0	0	1	-	-	4	1	1
7	4	3	3.5	0	0	0	1	0	-	-	4	2	1
8	4	2	3.0	0	0	1	0	0	-	-	5	2	1
9	4	2	3.0	0	0	1	0	0	-	-	6	2	1
10	3	2	2.5	0	1	0	0	0	-	1	6	2	1
...													
100,000	4	4	4.0	0	0	0	0	1	11,100	22,351	33,311	22,204	11,034
상대빈도									0.1110	0.2235	0.3331	0.2220	0.1103

$$\begin{aligned}
 E(\bar{X}) &= \frac{2.0 \cdot 11,100 + 2.5 \cdot 22,351 + 3.0 \cdot 33,311 + 3.5 \cdot 22,204 + 4.0 \cdot 11,034}{100,000} \\
 &= \frac{2.0 \cdot 11,100}{100,000} + \frac{2.5 \cdot 22,351}{100,000} + \dots + \frac{4.0 \cdot 11,034}{100,000} \\
 &= 2.0 \times 0.1110 + 2.5 \times 0.2235 + \dots + 4.0 \times 0.1103 \\
 &= 2.999955 = 3.0 \\
 &= \bar{X}_1 \cdot P(\bar{X}_1) + \bar{X}_2 \cdot P(\bar{X}_2) + \dots + \bar{X}_n \cdot P(\bar{X}_n) \\
 &= 2.0 \times 1/9 + 2.5 \times 2/9 + \dots + 4.0 \times 1/9 \\
 &= \sum_{i=1}^n \bar{X}_i \cdot P(\bar{X}_i) \quad \rightarrow \sum () \cdot p() \text{ 형식이므로 } \bar{X} \text{의 평균을 의미한다.}
 \end{aligned}$$

$$\begin{aligned}
 V(\bar{X}) &= \frac{(2.0 - 3.0)^2 \cdot 11,100 + (2.5 - 3.0)^2 \cdot 22,351 + \dots + (4.0 - 3.0)^2 \cdot 11,034}{100,000} \\
 &= \frac{1.0 \cdot 11,100}{100,000} + \frac{0.250 \cdot 22,351}{100,000} + \dots + \frac{1.0 \cdot 11,034}{100,000}
 \end{aligned}$$

$$\begin{aligned}
&= 1.0 \times 0.1110 + 0.250 \times 0.2235 + \dots + 1.0 \times 0.1103 = 0.3324 \\
&= (\bar{X}_1 - E(\bar{X}))^2 \cdot P(\bar{X}_1) + (\bar{X}_2 - E(\bar{X}))^2 \cdot P(\bar{X}_2) + \dots + (\bar{X}_n - E(\bar{X}))^2 \cdot P(\bar{X}_n) \\
&= 1.0 \times 1/9 + 0.250 \times 2/9 + \dots + 1.0 \times 1/9 \\
&= \sum_{i=1}^n (\bar{X}_i - E(\bar{X}))^2 \cdot P(\bar{X}_i) = 0.3333 \\
&\rightarrow \sum () \cdot p() \text{ 형식이므로 } \bar{X} \text{ 편차 제곱의 평균, 즉 분산을 의미한다.}
\end{aligned}$$

평균의 표본분포(\bar{X})의 기댓값과 분산, 표준편차

기댓값 : $\mu_{\bar{X}} = \mu_X$ 1)

분산 : $\sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n}$ 2) 조건: X 는 서로 독립적이다.

$$\begin{aligned}
1) \quad \bar{X} &= \frac{X_1 + X_2 + \dots + X_n}{n} \\
E(\bar{X}) &= E\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{1}{n} E(X_1 + X_2 + \dots + X_n) \\
&= \frac{1}{n} (E(X_1) + E(X_2) + \dots + E(X_n)) = \frac{1}{n} (\mu + \mu + \dots + \mu) = \mu \\
2) \quad V(\bar{X}) &= V\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{1}{n^2} V(X_1 + X_2 + \dots + X_n) \\
&= \frac{1}{n^2} E[(X_1 + X_2 + \dots + X_n - E(X_1 + X_2 + \dots + X_n))^2] \\
&= \frac{1}{n^2} E[(X_1 + X_2 + \dots + X_n - E(X_1) - E(X_2) - \dots - E(X_n))^2] \\
&= \frac{1}{n^2} E[(X_1 - E(X_1) + X_2 - E(X_2) + \dots + X_n - E(X_n))^2] \\
&= \frac{1}{n^2} E[(X_1 - E(X_1))^2 + \dots + (X_n - E(X_n))^2 + 2 \sum_i \sum_{i < j} (X_i - E(X_i))(X_j - E(X_j))] \\
&= \frac{1}{n^2} \{E[(X_1 - E(X_1))^2] + E[(X_2 - E(X_2))^2] + \dots + E[(X_n - E(X_n))^2]\} \\
&\quad + \frac{1}{n^2} 2 \sum_i \sum_{i < j} E[(X_i - E(X_i))(X_j - E(X_j))] \\
&\quad X \text{가 서로 독립적이면 } E[(X_i - E(X_i))(X_j - E(X_j))] = 0 \text{이 된다.} \\
&= \frac{1}{n^2} \{E[(X_1 - E(X_1))^2] + E[(X_2 - E(X_2))^2] + \dots + E[(X_n - E(X_n))^2]\} \\
&= \frac{1}{n^2} [V(X_1) + V(X_2) + \dots + V(X_n)] = \frac{1}{n^2} [V(X) + V(X) + \dots + V(X)] \\
&= \frac{1}{n^2} [n V(X)] = \frac{V(X)}{n} = \frac{\sigma_X^2}{n}
\end{aligned}$$

(여기서 n 은 표본의 크기; $\frac{\sigma_x^2}{n}$ 는 $\frac{\sigma^2}{n}$ 로 표기할 수 있다.)

표준편차: $\sigma_{\bar{X}} = \frac{\sigma_x}{\sqrt{n}}$ 조건: X 는 서로 독립적이다. (여기서 $\frac{\sigma_x}{\sqrt{n}}$ 는 $\frac{\sigma}{\sqrt{n}}$ 로 표기할 수 있다.)

연습문제 1.

번호	X	$P(X)$
1	2.0	1/3
2	3.0	1/3
3	4.0	1/3

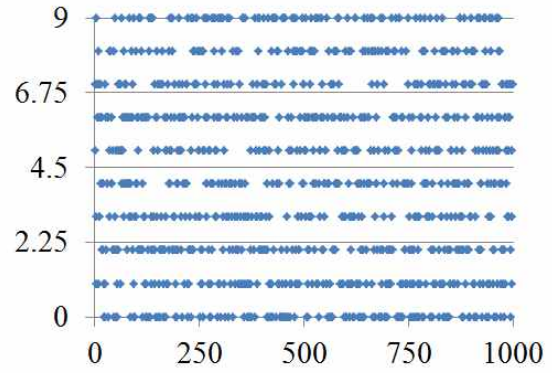
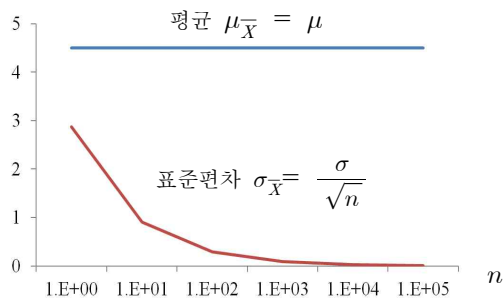
문제 1. $E(X)$, $V(X)$ 를 구하시오.

문제 2. $n=2$ 일 때, $E(\bar{X})$ 와 $V(\bar{X})$ 를 구하시오.

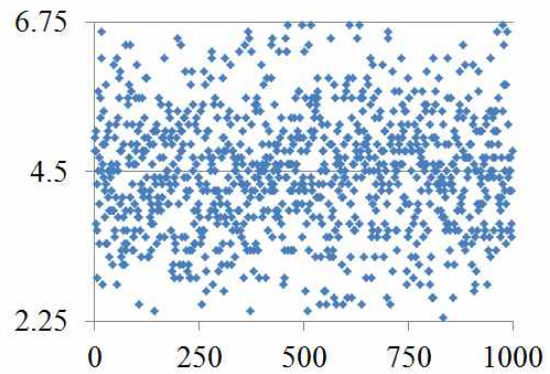
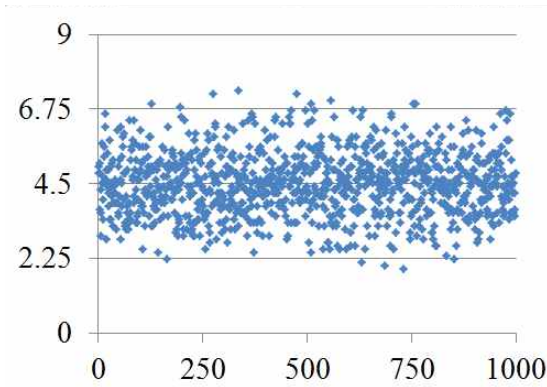
문제 3. $n=2$ 일 때, $E(\bar{X})$ 와 $V(\bar{X})$ 를 $E(X)$, $V(X)$ 와 n 으로 표현하시오.

모의실험 2. {0, 1, 2, ..., 9}로 구성된 집합에서 n 개 복원추출한 표본의 평균

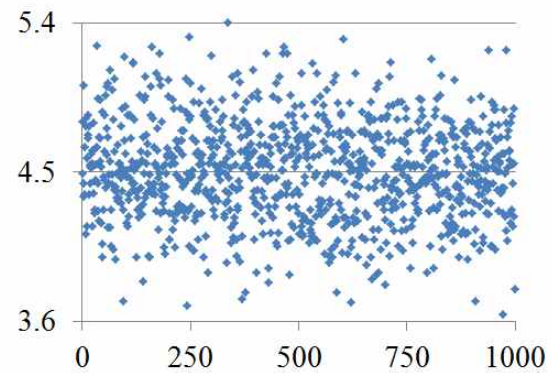
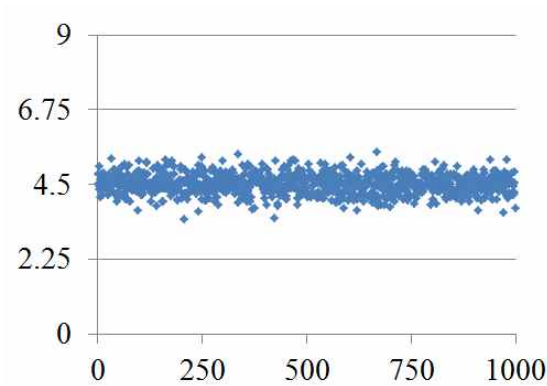
	n					
	1	10	100	1,000	10,000	100,000
1	5.0000	3.5000	4.4000	4.6070	4.5065	4.5071
2	4.0000	5.6000	4.5500	4.5290	4.4678	4.5139
3	5.0000	4.0000	3.8300	4.4780	4.5147	4.5059
4	5.0000	4.2000	4.7100	4.5500	4.5340	4.4939
5	9.0000	4.8000	4.1100	4.6040	4.5087	4.5113
6	7.0000	4.2000	4.2200	4.5660	4.5230	4.5175
7	-	3.5000	4.7200	4.4820	4.4743	4.4999
8	7.0000	4.5000	4.6600	4.5000	4.4600	4.5000
9	4.0000	3.1000	4.4800	4.5070	4.4708	4.4814
10	-	4.8000	4.1300	4.4920	4.5239	4.4883
...						
1,000	8.0000	2.3000	4.5700	4.5860	4.4858	4.4956
최대	9.0000	7.4000	5.3600	4.8130	4.6131	4.5241
최소	-	1.2000	3.5500	4.2190	4.4244	4.4766
실험 평균	4.5310	4.5208	4.5073	4.5009	4.4997	4.5000
실험 분산	8.5416	0.8094	0.0755	0.0083	0.0009	0.0001
실험 표준편차	2.9226	0.8996	0.2747	0.0913	0.0295	0.0096
이론 평균	4.5000	4.5000	4.5000	4.5000	4.5000	4.5000
이론 분산	8.25	0.825	0.0825	0.00825	0.000825	0.0000825
이론 표준편차	2.8723	0.9083	0.2872	0.0908	0.0287	0.0091



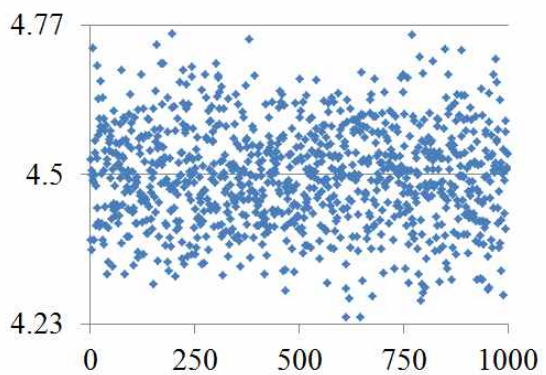
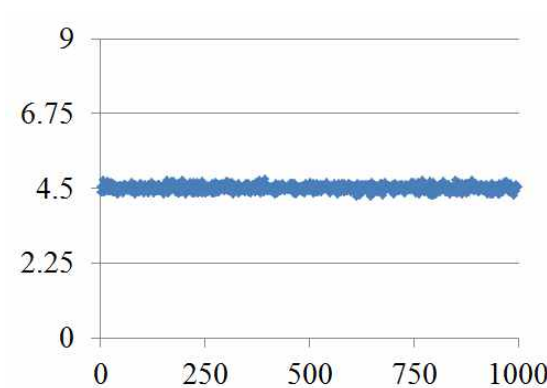
$n = 10$

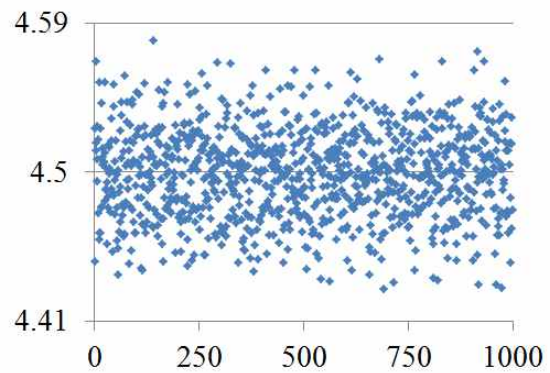
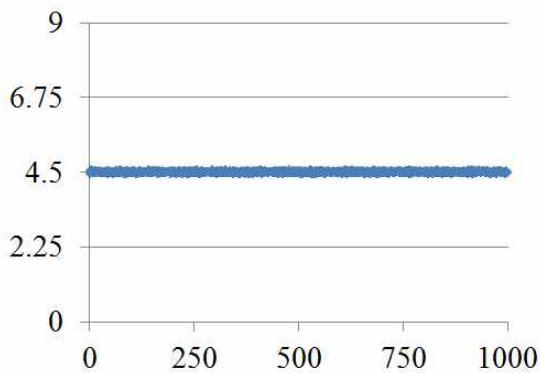
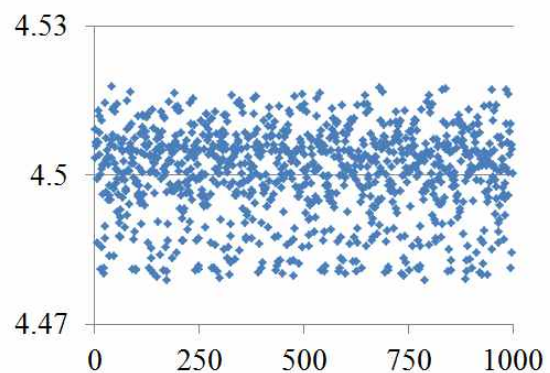
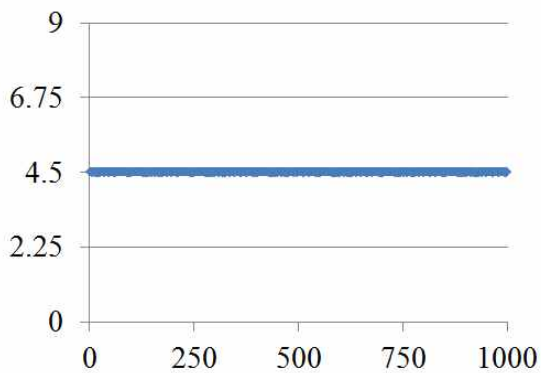


$n = 100$



$n = 1,000$



$n = 10,000$  $n = 100,000$ 

- ① n 과 무관하게 $E(\bar{X}) = E(X)$ 이다.
- ② n 이 증가함에 따라 $V(\bar{X})$ 는 $1/n$ 꼴로 감소한다. 즉 $V(\bar{X}) = \frac{V(X)}{n}$ 이다.
- ③ n 이 증가함에 따라 $\sigma_{\bar{X}}$ 는 $1/\sqrt{n}$ 꼴로 감소한다. 즉 $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ 이다.

연습문제 2. $x \sim U[0, 1]$ (확률변수 x 는 최소값이 0, 최대값이 1인 일양분포를 따른다.

문제 1. $E(x)$ 와 $V(x)$ 는?

문제 2. $n=100$ 이라 하자. $E(\bar{x})$ 와 $V(\bar{x})$ 는?

3. 모집단분포와 평균의 표본분포

[1] 모집단이 정규분포일 때

모집단이 정규분포일 때 평균의 표본분포는 표본의 크기 n 에 관계없이 언제나 정규분포를 이루며, 표본분포의 평균 $\mu_{\bar{X}}$ 와 표준편차 $\sigma_{\bar{X}}$ 는 다음과 같다.

$$\mu_{\bar{X}} = \mu_X$$

$$\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}}$$

즉, $X \sim N(\mu_X, \sigma_X^2)$ 이면, $\bar{X} \sim N\left(\mu_X, \frac{\sigma_X^2}{n}\right)$ 이 성립한다.

\bar{X} 에 해당하는 Z값

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}}$$

Z값의 정의: $\frac{\text{변수} - \text{변수의 평균}}{\text{변수의 표준편차}}$

현재 변수는 \bar{X} 이고, \bar{X} 의 평균은 $\mu_{\bar{X}}$, 표준편차는 $\sigma_{\bar{X}}$ 이다.

$$= \frac{\bar{X} - \mu_X}{\sigma_X / \sqrt{n}}$$

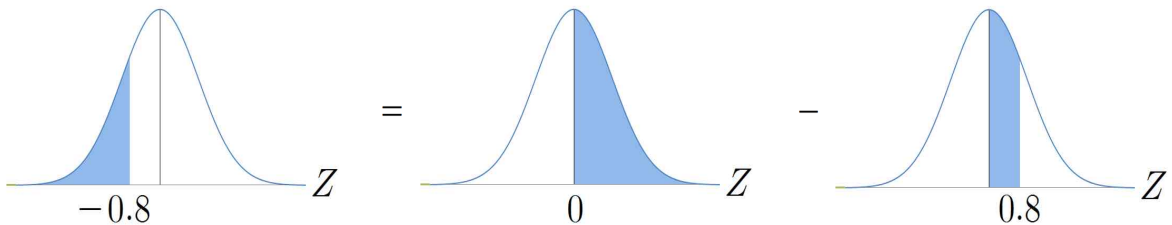
μ_X 와 σ_X 값은 알려져 있고 $\mu_{\bar{X}}$ 값과 $\sigma_{\bar{X}}$ 값이 알려져 있지 않은

경우, μ_X 값과 σ_X / \sqrt{n} 값을 사용하여 $\mu_{\bar{X}}$ 과 $\sigma_{\bar{X}}$ 값을 구한다.

Example. 3학년 전체 2,000명의 체중(X) 조사. $X \sim N(42, 5^2)$

① 체중이 38kg 이하인 학생은 몇 명인가?

$$\begin{aligned} P(X \leq 38) &= P\left(\frac{X - \mu}{\sigma} \leq \frac{38 - \mu}{\sigma}\right) = P\left(Z \leq \frac{38 - 42}{5}\right) = P(Z \leq -0.8) \\ &= 0.5 - P(0 \leq Z \leq 0.8) = 0.5 - 0.2881 = 0.2119 \end{aligned}$$



$$2,000\text{명} \times 0.2119 = 423.8\text{명}$$

② 두 명씩 임의로 추출시 \bar{X} 분포는?

$$\bar{X} \sim N(\mu_{\bar{X}}, \sigma_{\bar{X}}^2) \text{인데 } \mu_{\bar{X}} = \mu_X = 42 \text{이고 } \sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n} = \frac{25}{2} = 12.5 \text{이므로}$$

$\bar{X} \sim N(42, 12.5)$ 이다.

③ 세 명씩 임의로 추출시 \bar{X} 분포는?

$$\bar{X} \sim N(42, 6.25)$$

④ 한 학급은 40명이다. 학급 평균 체중이 40kg 미만일 확률은?

$$\bar{X} \sim N\left(42, \frac{25}{40}\right) = N(42, 0.7906^2)$$

$$P(\bar{X} < 40) = P\left(z < \frac{40 - 42}{0.7906}\right) = P(z < -2.5298) = 0.0057$$

[2] 모집단이 정규분포가 아닐 때

모집단이 정규분포가 아니라면 표본분포가 반드시 정규분포를 따르는 것은 아니다. 그렇지만 표본의 크기 n 이 충분히 커지면 표본분포가 정규분포에 근접해 간다.

중심극한정리(The Central Limit Theorem)

표본 크기 n 이 증가함에 따라 평균의 표본분포는 모집단의 분포 모양에 관계없이 정규분포에 접근한다.

연습문제 3. $x \sim U[0, 1]$, $n = 100$

문제 1. \bar{x} 는 어떤 분포를 따르는가?

문제 2. 0~1 사이의 난수를 100개 뽑아 평균을 구했을 때, 그 값이 0.6 이상일 확률은?

문제 3. 0~1 사이의 난수를 100개 뽑아 평균을 구했을 때, 상위 2.5%에 해당하는 \bar{x} 는?

문제 4. 0~1 사이의 난수를 100개 뽑아 평균을 구했을 때, 하위 5%에 해당하는 \bar{x} 는?

연습문제 4.

번호	X	$P(X)$
1	2.0	1/3
2	3.0	1/3
3	4.0	1/3

$$E(X) = 3.0, V(X) = 2/3$$

문제 1. $n=30$ 일 때, \bar{x} 는 어떤 분포를 따르는가?

문제 2. $n=30$ 일 때, $P(\bar{x} > 3.3)$

문제 3. $n=30$ 일 때, 상위 10%에 해당하는 \bar{x} 는?

문제 4. $n=30$ 일 때, 하위 20%에 해당하는 \bar{x} 는?

[3] 유한모집단 수정계수(finite population correction)

유한모집단의 \bar{X} 의 표준편차

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}, \text{ 여기서 } N \text{은 모집단의 크기이고 } n \text{은 표본의 크기이다.}$$

일반적으로 표본의 크기가 모집단 크기의 5% 이상인 경우 유한모집단 수정계수를 적용한다.

제4절 비율의 표본분포

1. 비율의 표본분포의 의미

모집단 비율 $\mu_p = \frac{X}{N} = \pi$

표본 비율 $\bar{X}_p = \frac{x}{n} = p$

비율의 표본분포의 기댓값과 분산 $E(P) = \pi$, $V(P) = \frac{\pi(1-\pi)}{n}$

비율의 표본분포에서 표준화변수 $Z = \frac{p-\pi}{\sigma_p}$,
 $n \geq 100$ 이고 $n\pi \geq 10$, $n(1-\pi) \geq 10$ 인 경우

연습문제 정답

- (1) $E(X) = 2(1/3) + 3(1/3) + 4(1/3) = 3.0$
 $V(X) = (2-3)^2(1/3) + (3-3)^2(1/3) + (4-3)^2(1/3) = 2/3$

(2) $E(\bar{X}) = 3.0$, $V(\bar{X}) = 1/3$ (3) $E(\bar{X}) = E(X)$, $V(\bar{X}) = \frac{V(X)}{n}$
- (1) $E(x) = \frac{0+1}{2} = 0.5$, $V(x) = \frac{(1-0)^2}{12} = 0.0833$

(2) $E(\bar{x}) = E(x) = 0.5$, $V(\bar{x}) = \frac{V(x)}{100} = 0.00083$
- (1) $n \geq 30$ 이므로, $\bar{x} \sim N\left(\mu_x, \frac{\sigma_x^2}{n}\right) = N(0.5, 0.00083) = N(0.5, 0.0289^2)$

(2) $P(\bar{x} > 0.6) = P\left(z > \frac{0.6-0.5}{0.0289}\right) = P(z > 3.4642) = 0.0003$

(3) $P(z > Z) = 0.025$ 를 충족시키는 $Z = 1.96$ 이다.
 상위 2.5%에 해당하는 $\bar{x} = \mu_{\bar{x}} + 1.96 \cdot \sigma_{\bar{x}} = 0.5 + 1.96(0.0289) = 0.5566$

(4) $P(z < Z) = 0.050$ 를 충족시키는 $Z = -1.645$ 이다.
 하위 5%에 해당하는 $\bar{x} = \mu_{\bar{x}} - 1.645 \cdot \sigma_{\bar{x}} = 0.5 - 1.645(0.0289) = 0.4525$

3) $E(P) = E\left(\frac{x}{n}\right) = \frac{1}{n} E(x) = \frac{1}{n} \cdot n\pi = \pi$

4) $V(P) = V\left(\frac{x}{n}\right) = \frac{1}{n^2} V(x) = \frac{1}{n^2} \cdot n\pi(1-\pi) = \frac{\pi(1-\pi)}{n}$

4. (1) $n \geq 30$ 이므로, $\bar{x} \sim N\left(\mu_x, \frac{\sigma_x^2}{n}\right) = N(3.0, 2/90) = N(3.0, 0.1491^2)$

(2) $P(\bar{x} > 3.3) = P\left(z > \frac{3.3 - 3.0}{0.1491}\right) = P(z > 2.0125) = 0.0221$

(3) $P(z > Z) = 0.100$ 를 충족시키는 $Z = 1.28$ 이다.

상위 10%에 해당하는 $\bar{x} = \mu_x + 1.28 \cdot \sigma_x = 3.0 + 1.28(0.1491) = 3.1908$

(4) $P(z < Z) = 0.20$ 를 충족시키는 $Z = -0.8416$

하위 20%에 해당하는 $\bar{x} = \mu_x - 0.8416 \cdot \sigma_x = 3.0 - 0.8416(0.1491) = 2.8745$