

# 생생한 사례로 배우는 확률과 통계

## [강의교안 이용 안내]

- 본 강의교안의 저작권은 **이재원**과 **한빛아카데미(주)**에 있습니다.
- 이 자료를 무단으로 전제하거나 배포할 경우 저작권법 136조에 의거하여 벌금에 처할 수 있고 이를 병과(併科)할 수도 있습니다.



## Chapter 08

생생한 사례로 배우는

### 확률과 통계

PROBABILITY & STATISTICS

## 표본분포

Sampling Distribution

# 목 차

## 8.1 단일표본분포

## 8.2 이변량표본분포

## 8.1 단일표본분포



# 표본평균의 분포

**모집단 분포** population distribution :

어떤 통계적 실험 결과인 모집단의 자료가 가지는 확률분포를 의미한다.

❖ **모수** parameter : 모집단의 특성을 나타내는 수치

**모평균** population mean : 표본의 평균( $\mu$ )

**모분산** population variance : 표본의 분산( $\sigma^2$ )

**모표준편차** population standard deviation : 표본의 표준편차( $\sigma$ )

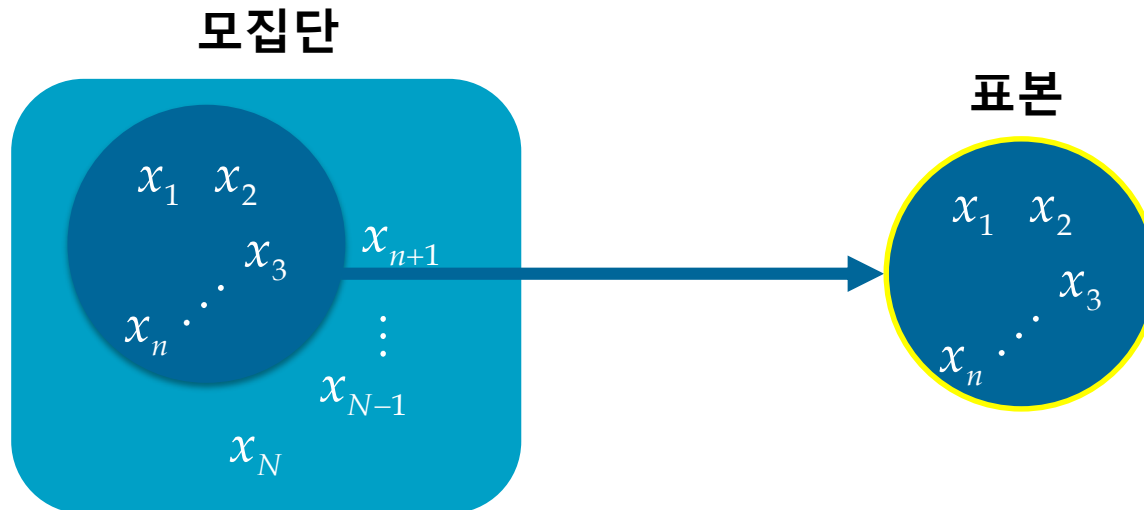
**모비율** population proportion : 표본의 비율( $p$ )

# 표본평균의 분포

- ❖ 대부분의 모집단 분포는 완전하게 알려진 것이 없으며, 따라서 모집단 분포의 정확한 중심의 위치나 산포도 등을 알 수 없다.
- ❖ 모집단의 확률분포를 비롯한 특성을 알기 위하여 전수조사를 한다는 것은 경제적, 공간적 또는 시간적인 제약에 의하여 거의 불가능



표본을 선정



# 표본평균의 분포

**표본분포** sampling distribution : 모집단에서 크기  $n$ 인 표본을 반복하여 선정할 때 얻어지는 통계량의 확률분포를 의미한다.

❖ **통계량** statistics : 표본의 특성을 나타내는 수치

**표본평균** sample mean : 표본의 평균( $\bar{x}$ )

**표본분산** sample variance : 표본의 분산( $s^2$ )

**표본표준편차** sample standard deviation : 표본의 표준편차( $s$ )

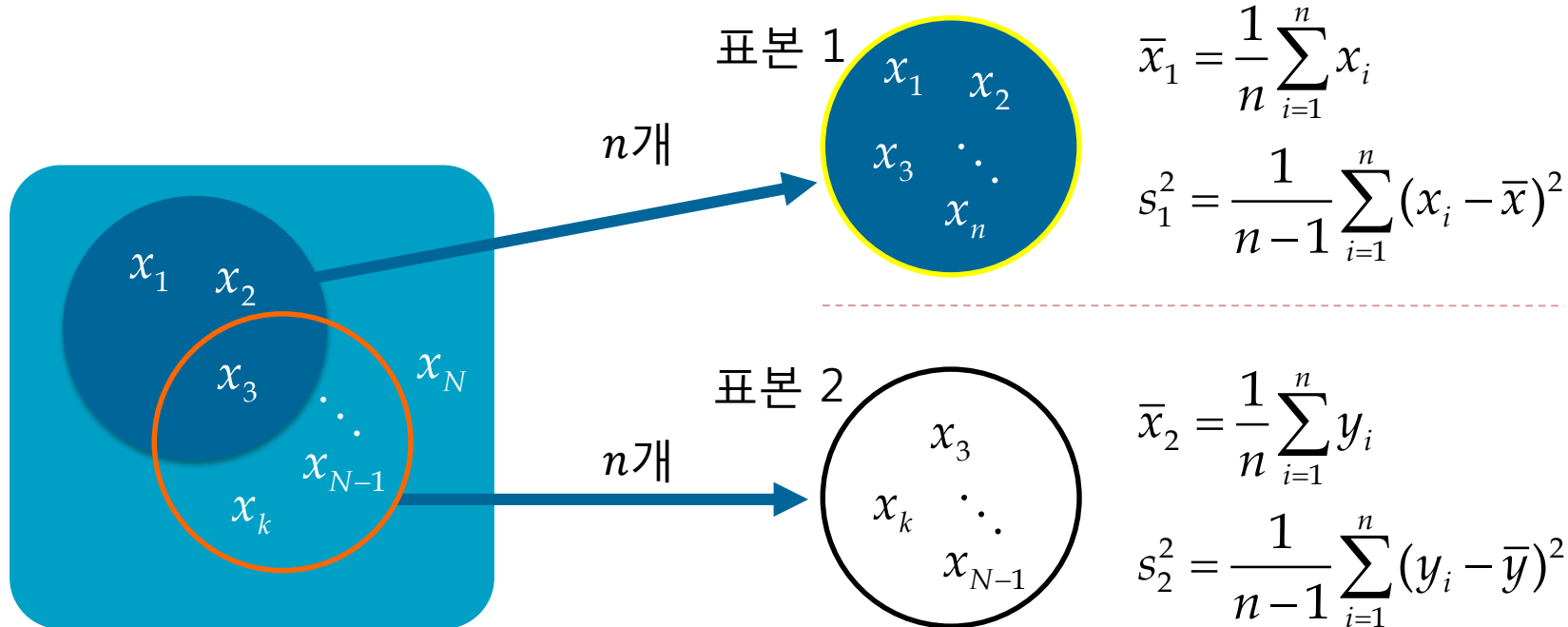
**표본비율** sample proportion : 표본의 비율( $\hat{p}$ )

# 표본평균의 분포

- ❖ 통계량은 표본을 어떻게 선정하느냐에 따라서 그 값이 다르게 나타난다. 즉, 동일한 모집단에서 동일한 크기의 표본을 선정하더라도 각 표본의 평균은 서로 다르게 나타남.

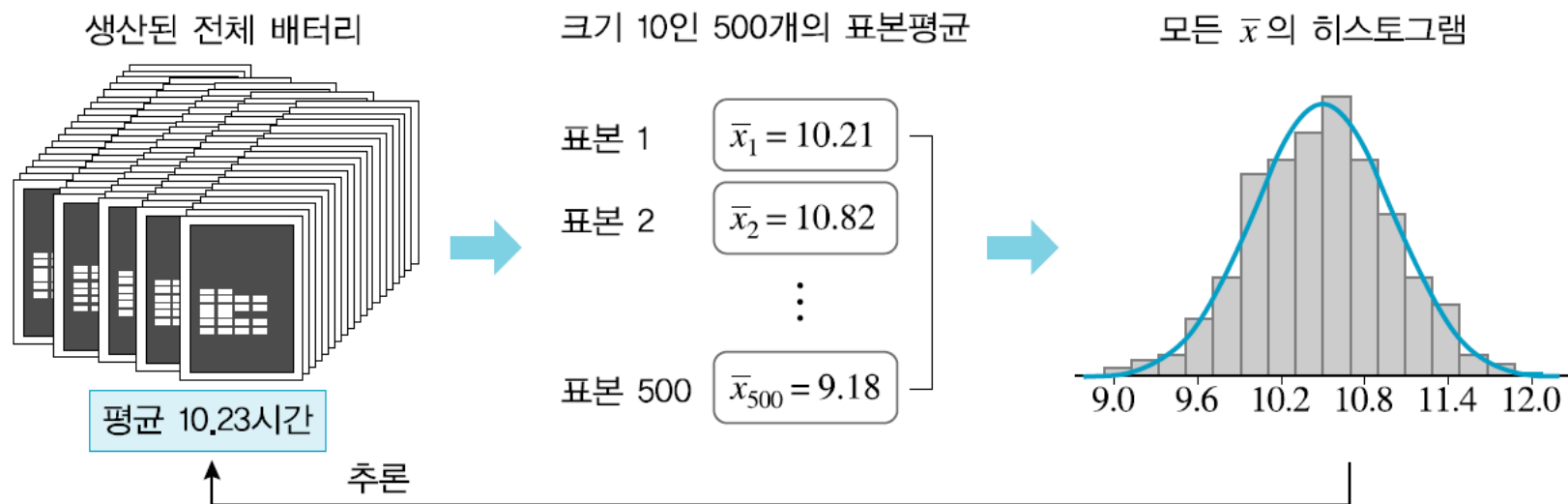


- 통계량은 확률변수
- 통계량의 확률분포를 표본분포라 한다.





# 표본평균의 분포



- ❖ 배터리 제조업체에서 생산된 모든 배터리의 하루 평균 사용시간을 알기 위하여, 10개로 구성된 표본 500개를 임의로 선정하여 각 표본의 평균을 구하면, 각 표본평균으로 구성된 500개의 자료가 갖는 특성, 특히 표본평균들의 평균을 얻을 수 있다. 그러면 표본평균들의 평균을 이용하여 이 회사에서 생산한 배터리의 하루 평균 사용시간(10.23시간)을 추론할 수 있다.

# 모분산이 알려진 정규모집단인 경우

모집단

$$X_i \sim N(\mu, \sigma^2)$$

$X_i$  : 독립

$i = 1, 2, \dots, N$

$\sigma^2$  : 기지



표본

$$X_i \sim N(\mu, \sigma^2)$$

$X_i$  : 독립

$i = 1, 2, \dots, n$

표본평균

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} E\left(\sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \frac{1}{n} (n\mu) = \mu$$

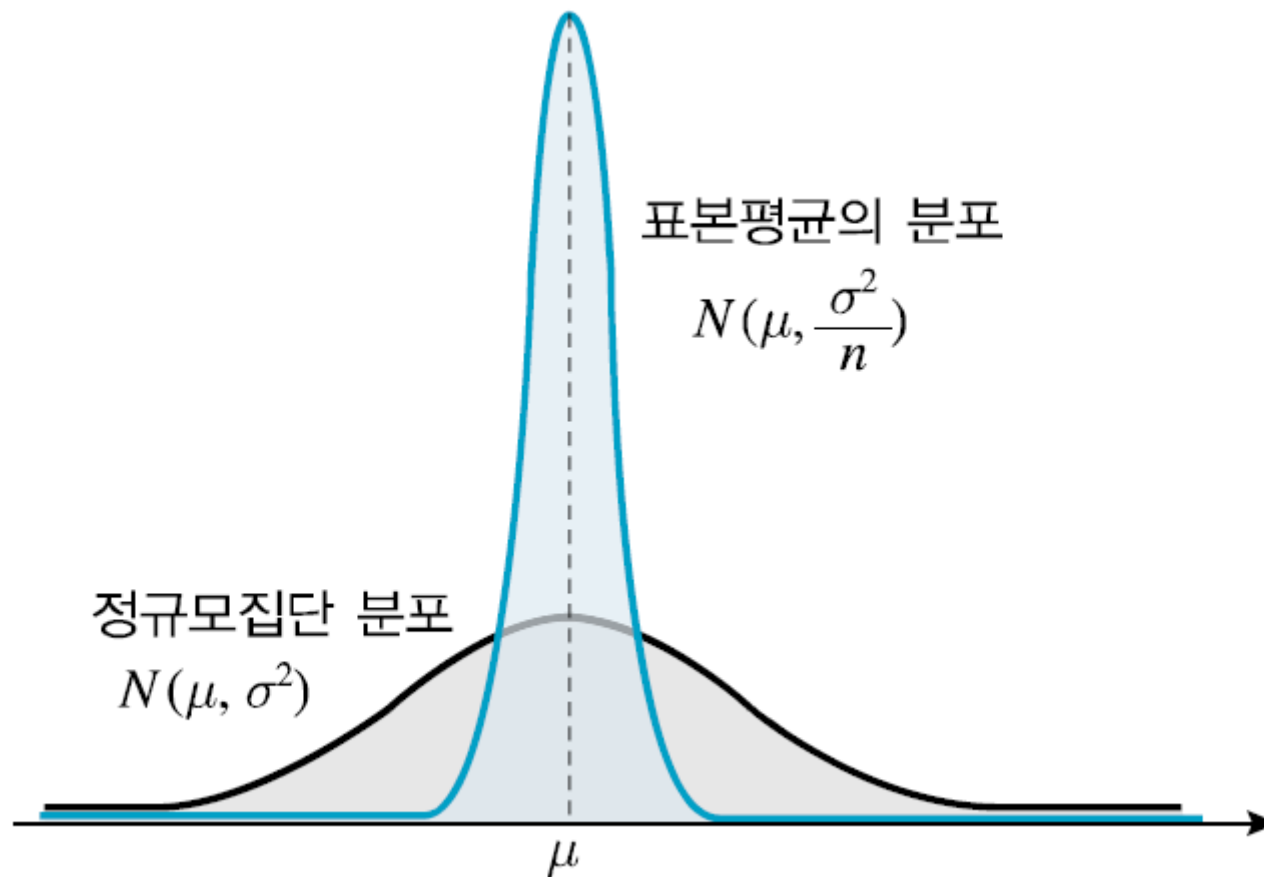
$$Var(\bar{X}) = Var\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n Var(X_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{1}{n^2} (n\sigma^2) = \frac{\sigma^2}{n}$$

# 모분산이 알려진 정규모집단인 경우

모평균  $\mu$ , 모분산  $\sigma^2$ 이 알려진 정규모집단에서 크기  $n$ 인 표본을 선정할 때, 표본평균에 관한 표본분포는 평균  $\mu$ , 분산  $\sigma^2/n$ 인 정규분포에 따른다. 즉, 다음과 같다.

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$$

# 모분산이 알려진 정규모집단인 경우



# 모분산이 알려진 정규모집단인 경우

## 예제 8-1

어떤 종류의 가솔린 차량 연료통의 용량은 평균 70리터, 표준편차 0.2리터인 정규분포를 따르는 것으로 알려져 있다. 이 종류의 차량 25대를 임의로 선정하였다.

- (a) 선정된 차량 연료통의 평균 용량에 대한 확률분포를 구하라.
- (b) 평균 용량이 69.95리터 이상 70.05리터 이하일 확률을 구하라.

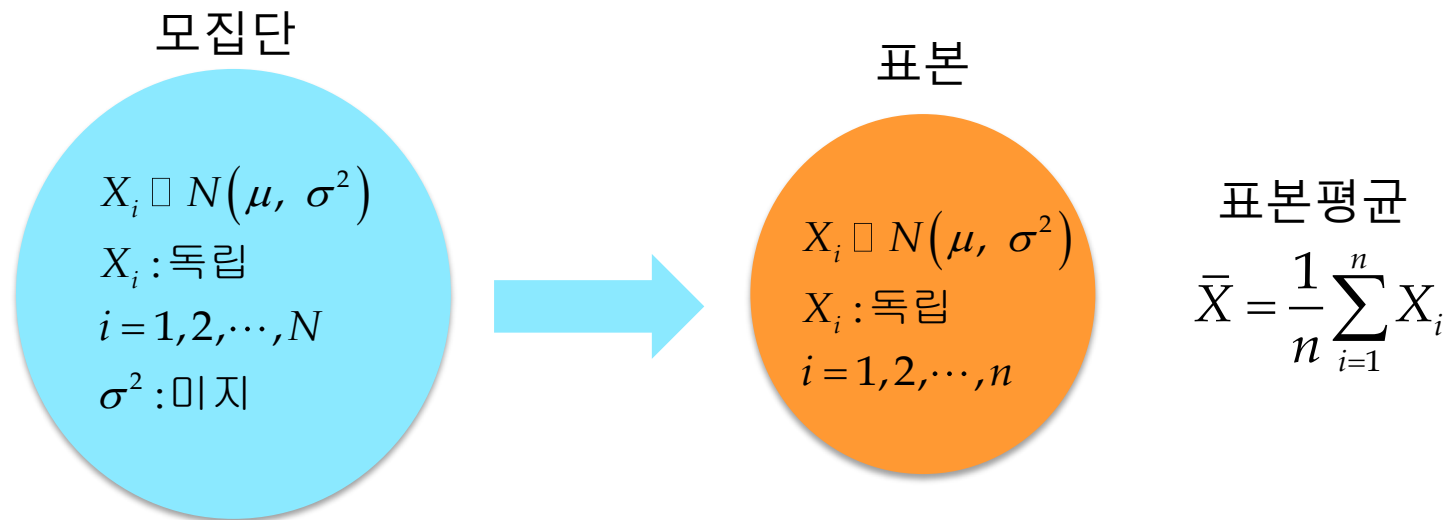
풀이

- (a)  $X \sim N(70, 0.22)$ 인 모집단으로부터 크기 25인 표본을 선정하였으므로 표본평균은 평균  $\mu_{\bar{X}} = 70$ , 분산  $\sigma_{\bar{X}}^2 = 0.04 / 25 = 0.04^2$ 인 정규분포에 따른다.  
즉, 표본평균은 정규분포  $\bar{X} \sim N(70, 0.04^2)$ 에 따른다.

- (b) 구하고자 하는 확률은 다음과 같다.

$$\begin{aligned} P(69.95 \leq \bar{X} \leq 70.05) &= P\left(\frac{69.95 - 70}{0.04} \leq Z \leq \frac{70.05 - 70}{0.04}\right) \\ &= P(-1.25 \leq Z \leq 1.25) = 2\Phi(1.25) - 1 \\ &= \Phi(1.78) - [1 - \Phi(0.57)] \\ &= 2 \times 0.8944 - 1 = 0.7888 \end{aligned}$$

# 모분산이 알려지지 않은 정규모집단인 경우



$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$$

$\sigma^2$ 이 미지이므로 사용할 수 없다.

# 모분산이 알려지지 않은 정규모집단인 경우

- ❖ 표본분산에 관련된 표본분포(※ 추후에 살펴봄)

$$V = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

- ❖ 표본평균에 관련된 표본분포

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$$

- ❖  $t$ -분포의 정의에 의해 다음을 얻는다.

$$T = \frac{Z}{\sqrt{V/(n-1)}} = \left( \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \right) / \sqrt{\frac{n-1}{\sigma^2} S^2 / (n-1)} = \frac{\bar{X} - \mu}{s / \sqrt{n}} \sim t(n-1)$$

- ❖ 대부분의 모집단은 모분산  $\sigma^2$ 이 알려지지 않으며, 따라서 표본평균이 정규분포를 이룬다고 할 수 없다.
- ❖ 이때 표본평균  $\bar{X}$ 의 표준화에서 모표준편차  $\sigma$ 를 표본표준편차  $s$ 로 대치하면,  $\bar{X}$ 의 표준화 확률변수는 자유도가  $n - 1$ 인  $t$ -분포가 된다.

# 모분산이 알려지지 않은 정규모집단인 경우

모평균이  $\mu$ 이고 모분산  $\sigma^2$ 이 알려지지 않은 정규모집단에서 크기  $n$ 인 표본을 선정할 때, 표본평균에 관한 표본분포는 다음과 같이 자유도  $n - 1$ 인  $t$ -분포이다. 즉, 다음과 같다.

$$T = \frac{\bar{X} - \mu}{s / \sqrt{n}} \sim t(n - 1)$$



# 모분산이 알려지지 않은 정규모집단인 경우

## 예제 8-2

제철회사에서 생산되는 콘크리트 보강용 강철봉의 강도는 평균 46psi인 정규분포를 따른다고 한다. 이 회사에서 생산된 강철봉 20개를 임의로 선정하여 강도를 측정한 결과가 다음과 같다.

45.6	43.6	40.3	44.2	42.5	41.0	49.1	44.3	49.3	44.1
45.1	48.7	43.9	48.7	47.4	45.4	42.6	42.1	44.6	44.9

- (a) 표본평균과 관련된 표본분포를 구하라.
- (b) 표본평균이 상위 5%인 강도를 구하라.

### 풀이

- (a) 표본으로 얻은 강철봉 20개의 평균과 분산은 각각 다음과 같다.

$$\bar{x} = \frac{1}{20} \sum x_i = 44.87, \quad s^2 = \frac{1}{19} \sum (x_i - 44.87)^2 = 6.97$$

# 모분산이 알려지지 않은 정규모집단인 경우

그러므로 표본표준편차는  $s = 2.64$ 이다. 그리고 정규모집단  $N(46, s^2)$ 으로부터 크기 20인 표본을 선정하여 표본표준편차가  $s = 2.64$ 이므로 다음을 얻는다.

$$T = \frac{\bar{X} - 46}{2.64 / \sqrt{20}} \sim t(19)$$

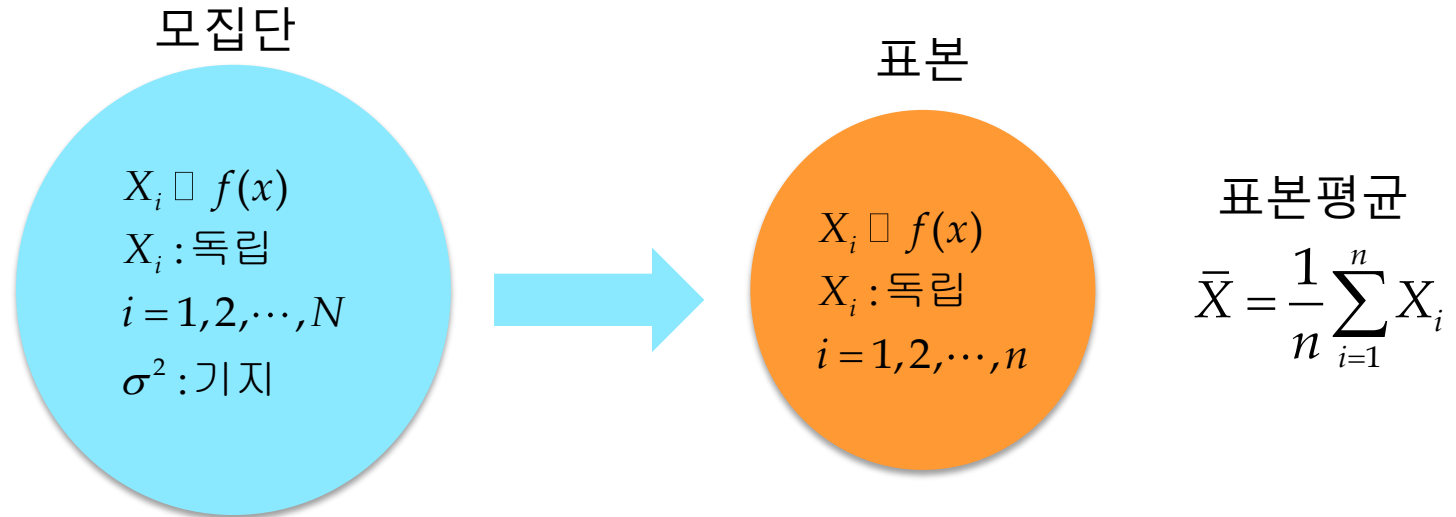
(b) 상위 5%인 강도를  $\bar{x}_0$ 라 하면  $P(\bar{X} \geq \bar{x}_0) = 0.05$ 이고  $T = \frac{\bar{X} - 46}{0.59} \sim t(19)$ 이므로 다음을 얻는다.

$$P(\bar{X} \geq \bar{x}_0) = P\left(T \geq \frac{\bar{x}_0 - 46}{0.59}\right) = P(T \geq t_{0.05}(19)) = 0.05$$

한편 자유도 19인  $t$ -분포에서  $t_{0.05}(19) = 1.729$ 이므로 구하고자 하는  $\bar{x}_0$ 는 다음과 같다.

$$\frac{\bar{x}_0 - 46}{0.59} = 1.729; \quad \bar{x}_0 = 46 + 0.59 \times 1.729 = 47.02$$

# 임의의 모집단인 경우



중심극한정리에 의해  $n$ 이 충분히 크면 모평균  $\mu$ , 모분산  $\sigma^2$ 인 임의의 모집단에 대한 표본평균은 평균  $\mu$ , 분산  $\sigma^2/n$ 인 정규분포에 근사한다. 즉, 다음과 같다.

$$\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right), \quad Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \approx N(0, 1)$$

# 임의의 모집단인 경우

## 예제 8-3

해양의 파고  $X$ 는 다음과 같은 레일리 분포를 따른다. 임의로 50곳을 선정하여 파고를 측정하였다. 단, 단위는 미터이다.

$$f(x) = \frac{x}{100} e^{-x^2/200}, \quad x > 0$$

- (a) 표본평균과 관련된 근사표본분포를 구하라.
- (b) 평균 파고가 12미터 이상일 근사확률을 구하라.

풀이

- (a) 먼저 해양의 평균 파고를 구한다.

# 임의의 모집단인 경우

$$\begin{aligned}\mu &= \int_0^\infty \frac{x^2}{100} e^{-x^2/200} dx \left( \frac{x^2}{200} = t \Rightarrow x^2 = 200t, dx = 5\sqrt{2} t^{-1/2} dt \right) \\&= \frac{1}{100} \int_0^\infty 200te^{-t} 5\sqrt{2} t^{-1/2} dt \\&= 10\sqrt{2} \int_0^\infty t^{1/2} e^{-t} dt = 10\sqrt{2} \Gamma\left(\frac{3}{2}\right) \\&= 10\sqrt{2} \times \frac{1}{2} \Gamma\left(\frac{1}{2}\right) = 5\sqrt{2\pi}\end{aligned}$$

이제 분산을 구하기 위하여 2차 적률을 구한다.

$$\begin{aligned}E(X^2) &= \int_0^\infty \frac{x^3}{100} e^{-x^2/200} dx \left( \frac{x^2}{200} = t \Rightarrow x^2 = 200t, dx = 5\sqrt{2} t^{-1/2} dt \right) \\&= \frac{1}{100} \int_0^\infty 200te^{-t} 100 dt = 200 \int_0^\infty te^{-t} dt = 200\end{aligned}$$

# 임의의 모집단인 경우

따라서 분산은 다음과 같다.

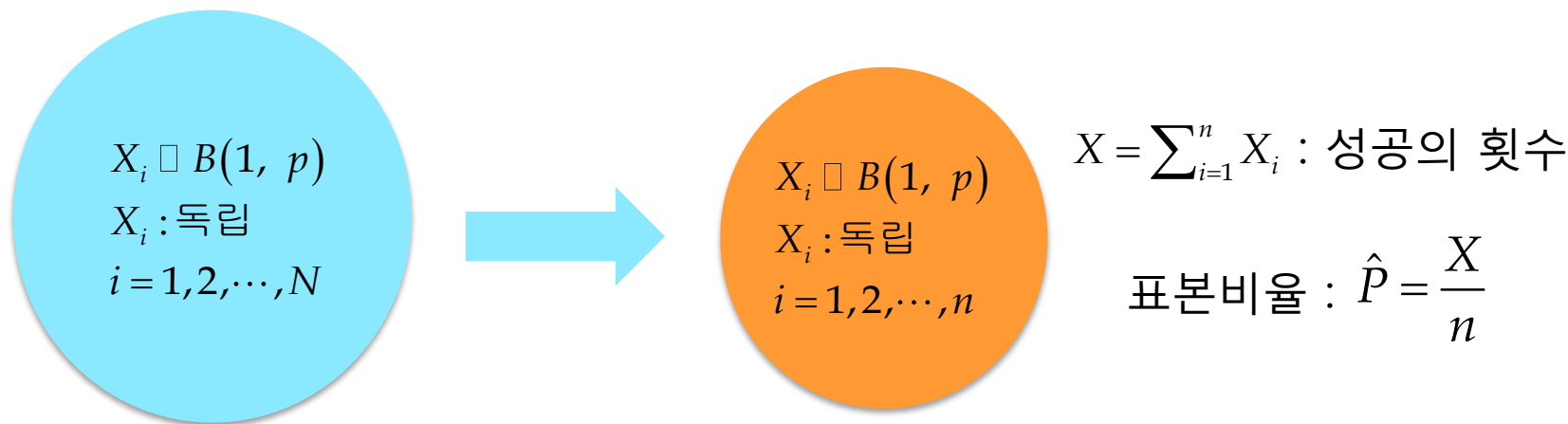
$$\sigma^2 = E(X^2) - E(X)^2 = 200 - (5\sqrt{2\pi})^2 = 50(4 - \pi)$$

그러므로 50곳의 평균 파고는  $\bar{X} \approx N(5\sqrt{2\pi}, 4 - \pi)$  이다.

(b) 구하고자 하는 근사확률은 다음과 같다.

$$\begin{aligned} P(\bar{X} \geq 15) &= P\left(Z \geq \frac{15 - 5\sqrt{2\pi}}{\sqrt{4 - \pi}}\right) = P(Z \geq 2.66) \\ &\approx 1 - \Phi(2.66) = 1 - 0.9961 = 0.0039 \end{aligned}$$

# 표본비율의 분포



$$\left. \begin{array}{l} X_i \sim B(1, p) \\ X_i : \text{독립}, i = 1, 2, \dots, n \end{array} \right\} \Rightarrow X = \sum_{i=1}^n X_i \sim B(n, p)$$

$$E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{1}{n}(np) = p$$

$$Var(\hat{p}) = Var\left(\frac{X}{n}\right) = \frac{1}{n^2}Var(X) = \frac{1}{n^2}(npq) = \frac{pq}{n}$$

$n$ 이 충분히 크면

$$\hat{p} \approx N\left(p, \frac{pq}{n}\right)$$

# 표본비율의 분포

## 예제 8-4

어느 제조회사에서 생산한 배터리의 5%가 불량품인 것으로 알려져 있다. 이 회사에서 생산한 배터리 100개를 임의로 선정하여 조사했을 때, 다음을 구하라.

- (a) 선정된 배터리 중에 불량품이 없을 확률
- (b) 선정된 배터리 중에 10% 이상이 불량품일 확률

### 풀이

- (a) 모비율이  $p = 0.05$ 이므로 100개의 배터리 중에서 불량품의 수를  $X$ 라 하면,  $np = 5, npq = 4.75$ 이므로  $X \approx N(5, 4.75)$ 라서 구하고자 하는 확률은 다음과 같다.

$$\begin{aligned} P(X = 0) &= P(-0.5 \leq X \leq 0.5) = P\left(\frac{-0.5 - 5}{\sqrt{4.75}} \leq Z \leq \frac{0.5 - 5}{\sqrt{4.75}}\right) \\ &= P(-2.52 \leq Z \leq -2.06) \approx \Phi(-2.06) - \Phi(-2.52) \\ &= \Phi(2.52) - \Phi(2.06) = 0.9941 - 0.9803 = 0.0138 \end{aligned}$$

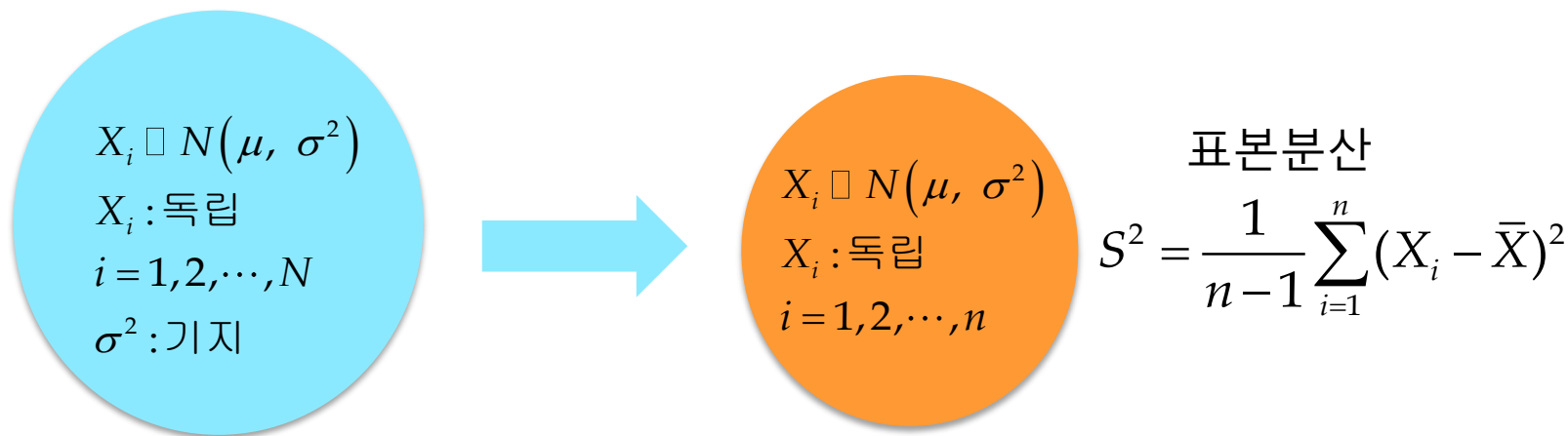


# 표본비율의 분포

- (b)  $p = 0.05, n = 100$ 이므로 표본비율의 확률분포는  $\hat{P} \approx N(0.05, 0.0218^2)$ 이다. 그러므로 구하고자 하는 확률은 다음과 같다.

$$\begin{aligned} P(\hat{P} \geq 0.1) &= P\left(Z \geq \frac{0.1 - 0.05}{0.0218}\right) = P(Z \geq 2.29) \\ &\approx 1 - \Phi(2.29) = 1 - 0.9890 = 0.011 \end{aligned}$$

# 표본분산의 분포



- ❖ 모평균  $\mu$ , 모분산  $\sigma^2$ 이 알려진 정규모집단에서 크기  $n$ 인 표본을 선정할 때, 표본평균에 대해 다음을 얻는다.

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1) \Rightarrow Z^2 = \left( \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \right)^2 \sim \chi^2(1)$$

# 표본분산의 분포

❖ 독립인 확률변수  $X_i \sim N(m, \sigma^2), i = 1, 2, \dots, n$ 의 표준화 확률변수의 제곱에 대하여  $\sum_{i=1}^n Z_i^2 \sim \chi^2(n)$  이 성립한다.

$$\begin{aligned}\sum_{i=1}^n Z_i^2 &= \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 = \frac{1}{\sigma^2} \sum_{i=1}^n [(X_i - \bar{X}) + (\bar{X} - \mu)]^2 \\&= \frac{1}{\sigma^2} \left[ \sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^n (\bar{X} - \mu)^2 + 2(\bar{X} - \mu) \sum_{i=1}^n (X_i - \bar{X}) \right] \\&= \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 + \frac{n}{\sigma^2} (\bar{X} - \mu)^2 \\&= \frac{n-1}{\sigma^2} \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 + \left( \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \right)^2 \\&= \frac{n-1}{\sigma^2} S^2 + \left( \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \right)^2\end{aligned}$$

$$\sum_{i=1}^n Z_i^2 \sim \chi^2(n), \quad Z^2 = \left( \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \right)^2 \sim \chi^2(1) \quad \text{이므로} \quad \frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1) \quad \text{이어야 한다.}$$

## 표본분산의 분포

모평균  $\mu$ , 모분산  $\sigma^2$ 인 정규모집단에서 크기  $n$ 인 표본을 선정할 때, 표본분산에 관한 표본분포는 자유도  $n - 1$ 인 카이제곱분포에 따른다. 즉, 다음과 같다.

$$V = \frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$$

# 표본분산의 분포

## 예제 8-5

감기약의 무게는 분산이  $0.00066\text{ g}$ 인 정규분포를 따른다고 한다. 시중에서 판매되는 감기약 20개를 임의로 수거하여 무게를 측정한 결과가 다음과 같다.

4.33	4.23	4.24	4.29	4.21	4.25	4.23	4.18	4.30	4.27
4.20	4.26	4.23	4.24	4.27	4.19	4.26	4.23	4.25	4.24

- (a) 표본분산과 관련된 통계량  $V$ 의 분포를 구하라.
- (b) 관찰된 표본분산의 값  $s_0^2$ 을 구하라.
- (c) 이 표본을 이용하여 통계량의 관찰값  $v_0 = \frac{(n-1)s_0^2}{\sigma^2}$ 을 구하라.
- (d) 표본분산  $S^2$ 이 (b)에서 구한  $s_0^2$ 보다 클 근사확률을 구하라.

풀이

# 표본분산의 분포

- (a)  $\sigma^2 = 0.00066$ 이고 확률표본의 크기가 20이므로 표본분산에 관련된 표본분포는 자유도 19인 카이제곱분포이다. 즉, 다음과 같다.

$$V = \frac{19S^2}{0.00066} \sim \chi^2(19)$$

- (b) 표본평균과 표본분산은 각각 다음과 같다.

$$\bar{x} = \frac{1}{20} \sum x_i = 4.245, \quad s_0^2 = \frac{1}{19} \sum (x_i - 4.245)^2 = \frac{0.0255}{19} = 0.00134$$

- (c)  $n = 20$ ,  $s^2 = 0.00066$ 이므로 통계량의 관찰값은 다음과 같다.

$$v_0 = \frac{(n-1)s_0^2}{\sigma^2} = \frac{19 \times 0.00134}{0.00066} = 38.576$$

- (d) 카이제곱분포표로부터 구하고자 하는 확률은 다음과 같다.

$$\begin{aligned} P(S^2 > s_0^2) &= P(S^2 > 0.00134) = P\left(\frac{19S^2}{\sigma^2} > \frac{19 \times 0.00134}{0.00066}\right) \\ &\approx P(V > 38.58) = 0.005 \end{aligned}$$

## 8.2 이변량표본분포



# 두 모분산이 알려진 정규모집단인 경우

❖ 6장 학습 내용:  $X, Y$ : 독립,  $X \sim N(\mu_1, \sigma_1^2)$ ,  $Y \sim N(\mu_2, \sigma_2^2)$



$$X - Y \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$$

$$\bar{X}, \bar{Y} : \text{독립}, \bar{X} \sim N\left(\mu_1, \frac{\sigma_1^2}{n}\right), \bar{Y} \sim N\left(\mu_2, \frac{\sigma_2^2}{m}\right)$$



$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)$$



## 두 표본평균 차의 분포

모평균  $\mu_1$ , 모분산  $\sigma_1^2$  인 정규모집단과 모평균  $\mu_2$ , 모분산  $\sigma_2^2$  인 정규모집단에서 각각 크기  $n, m$ 인 표본을 선정할 때, 두 표본평균의 차  $\bar{X} - \bar{Y}$ 에 관한 표본분포는 평균  $\mu_1 - \mu_2$ , 분산  $\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}$ 인 정규분포에 따른다. 즉, 다음과 같다.

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)$$

표준화



$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0, 1)$$

# 두 표본평균 차의 분포

## 예제 8-6

모분산이  $\sigma_1^2 = 4$ 인 정규모집단에서 크기 25인 표본의 표본평균을  $\bar{X}$ 라 하고, 모분산이  $\sigma_2^2 = 9$ 인 정규모집단에서 크기 36인 표본의 표본평균을  $\bar{Y}$ 라 할 때,  $|\bar{X} - \bar{Y}|$ 가 1.67보다 클 확률을 구하라. 단, 두 모평균은 동일하다.

**풀이** 두 모평균이 동일하므로  $\bar{X} - \bar{Y}$ 의 평균은  $\mu_{\bar{X}-\bar{Y}} = 0$ 이고, 두 표본평균의 분산은 각각 다음과 같다.

$$\sigma_{\bar{X}}^2 = \frac{4}{25}, \quad \sigma_{\bar{Y}}^2 = \frac{9}{36}$$

따라서  $\bar{X} - \bar{Y}$ 의 분산은 다음과 같다.

$$\sigma_{\bar{X}-\bar{Y}}^2 = \frac{4}{25} + \frac{9}{36} = \frac{41}{100} = 0.64^2$$

그러므로 두 표본평균 차는  $\bar{X} - \bar{Y} \sim N(0, 0.64^2)$ 이고, 구하고자 하는 확률은 다음과 같다.

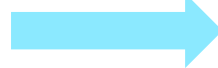
$$\begin{aligned} P(|\bar{X} - \bar{Y}| \geq 1.67) &= P\left(|Z| \geq \frac{1.67}{0.64}\right) = P(|Z| \geq 2.61) \\ &= 2[1 - \Phi(2.61)] = 2(1 - 0.9955) = 0.009 \end{aligned}$$

## 두 표본평균 차의 분포

모평균  $\mu_1$ , 모분산  $\sigma^2$ 인 정규모집단과 모평균  $\mu_2$ , 모분산  $\sigma^2$ 인 정규모집단에서 각각 크기  $n, m$ 인 표본을 선정할 때, 두 표본평균의 차에 관한 표본분포는 평균  $\mu_1 - \mu_2$ , 분산  $\left(\frac{1}{n} + \frac{1}{m}\right)\sigma^2$ 인 정규분포에 따른다. 즉, 다음과 같다.

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \left(\frac{1}{n} + \frac{1}{m}\right)\sigma^2\right)$$

표준화



$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim N(0, 1)$$

# 두 표본평균 차의 분포

## 예제 8-7

모평균이  $\mu_1 = 8$ ,  $\mu_2 = 7$ 이고 모분산이  $\sigma_1^2 = \sigma_2^2 = 9$ 인 정규모집단에서 동일한 크기 32인 표본을 선정했을 때, 다음을 구하라.

- (a)  $\bar{X} - \bar{Y}$ 의 확률분포
- (b)  $P(\bar{X} - \bar{Y} \geq a) = 0.05$ 를 만족하는  $a$

### 풀이

- (a) 모평균이  $\mu_1 = 8$ ,  $\mu_2 = 7$ 이고 모분산이  $\sigma_1^2 = \sigma_2^2 = 9$  그리고 크기가 동일한 32이므로  $\bar{X} - \bar{Y}$ 의 평균과 분산은 각각 다음과 같다.

$$\mu_{\bar{X}-\bar{Y}} = \mu_{\bar{X}} - \mu_{\bar{Y}} = 8 - 7 = 1$$

$$\sigma_{\bar{X}-\bar{Y}}^2 = \frac{9}{32} + \frac{9}{32} = \frac{9}{16} = 0.75^2$$

그러므로 두 표본평균 차는  $\bar{X} - \bar{Y} \sim N(1, 0.75^2)$  이다.

## 두 표본평균 차의 분포

(b)  $z_{0.05} = 1.645$ 이고 다음이 성립한다.

$$P(\bar{X} - \bar{Y} \geq a) = P\left(Z \geq \frac{a-1}{0.75}\right) = P(Z \geq z_{0.05}) = 0.05$$

따라서 구하고자 하는  $a$ 는 다음과 같다.

$$\frac{a-1}{0.75} = z_{0.05} = 1.645; \quad a = 1 + 0.75 \times 1.645 = 2.23$$

# 두 모분산이 같지만 알려지지 않은 정규모집단인 경우

❖  $\sigma_1^2 = \sigma_2^2 = \sigma^2$  이면

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \left(\frac{1}{n} + \frac{1}{m}\right)\sigma^2\right) \xrightarrow{\text{표준화}} \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim N(0, 1)$$

❖  $\sigma_1^2 = \sigma_2^2 = \sigma^2$  인 두 정규모집단에서 표본의 크기  $n, m$ 인 표본을 추출할 때, **합동표본분산** **pooled sample variance**을 다음과 같이 정의하고, 양의 제곱근을 합동표본표준편차라 한다.

$$S_p^2 = \frac{1}{n+m-2} \left[ \sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{j=1}^m (Y_j - \bar{Y})^2 \right]$$
$$= \frac{1}{n+m-2} \left[ (n-1)S_1^2 + (m-1)S_2^2 \right]$$

$$\text{단, } S_1^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad S_2^2 = \frac{1}{m-1} \sum_{j=1}^m (Y_j - \bar{Y})^2$$

# 두 모분산이 같지만 알려지지 않은 정규모집단인 경우

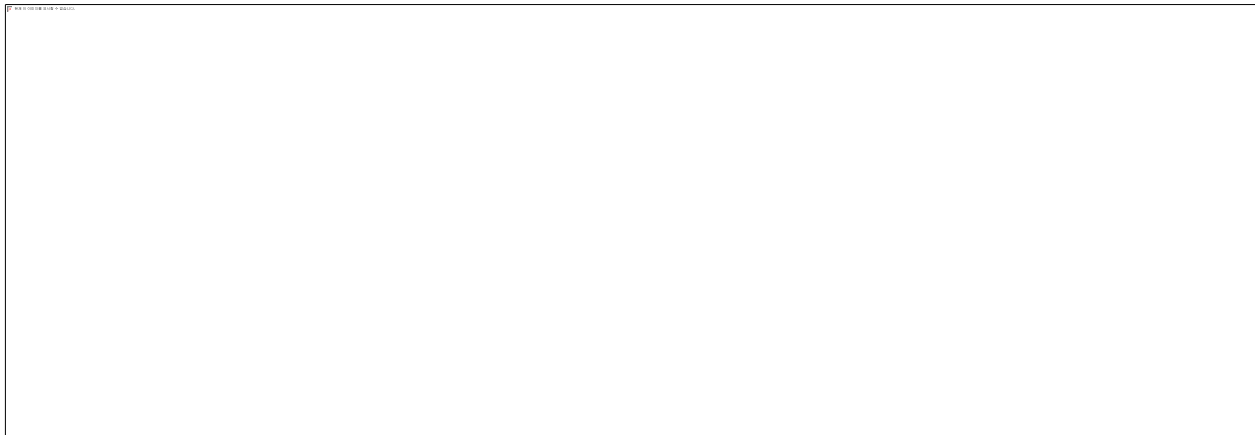
❖ 합동표본분산의 분포 : (7.2.3절에서 살펴봄)

$$V = \frac{n+m-2}{\sigma^2} S_p^2 \sim \chi^2(n+m-2)$$

❖  $\sigma_1^2 = \sigma_2^2 = \sigma^2$  인 두 표본평균 차의 분포 :

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim N(0, 1)$$

❖  $t$  - 분포의 정의 :



## 두 모분산이 같지만 알려지지 않은 정규모집단인 경우

모평균  $\mu_1, \mu_2$  이고, 모분산  $\sigma_1^2 = \sigma_2^2 = \sigma^2$  인 두 정규모집단에서 각각 크기  $n, m$ 인 표본을 선정할 때, 두 표본평균의 차  $\bar{X} - \bar{Y}$ 에 관한 표본분포는 자유도  $n + m - 2$ 인  $t$  - 분포에 따른다. 즉, 다음과 같다.

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t(n + m - 2)$$



# 두 표본평균 차의 분포

## 예제 8-8

모평균이 각각  $\mu_1 = 19$ ,  $\mu_2 = 16$ 이고 모분산이 동일하다고 알려진 두 정규모집단에서 각각 표본을 선정하여 다음을 얻었다. 두 표본평균의 차이가 4.5765 이상일 확률을 구하라.

표본 1	$n_1 = 15, s_1^2 = 2.82$
표본 2	$n_2 = 12, s_2^2 = 2.51$

풀이

표본1과 표본2의 표본평균을 각각  $\bar{X}$ ,  $\bar{Y}$  라 하면 다음 합동표본분산을 얻는다.

$$S_p^2 = \frac{1}{15+12-2} (14 \times 2.82 + 11 \times 2.51) = 2.6836, \quad s_p = \sqrt{2.6836} = 1.638$$

## 두 표본평균 차의 분포

$n_1 = 15, n_2 = 12$ 이므로 다음을 얻는다.

$$\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \sqrt{\frac{1}{15} + \frac{1}{12}} = \sqrt{0.15} = 0.3873, \quad s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 1.638 \times 0.3873 = 0.6344$$

한편  $\mu_1 - \mu_2 = 3$ 이므로 구하고자 하는 확률은 다음과 같다.

$$P(\bar{X} - \bar{Y} \geq 4.5765) = P\left(\frac{\bar{X} - \bar{Y} - 3}{0.6344} \geq \frac{4.5765 - 3}{0.6344}\right) = P(T \geq 2.485) = 0.01$$

# 두 모분산이 다르고 알려지지 않은 정규모집단인 경우

❖ 통계량  $U$ 를 다음과 같이 정의한다.

$$U = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}}$$

❖ 확률변수  $U$ 는 다음과 같이 정의되는 자유도  $r$ 을 갖는  $t$ -분포에 따른다. 이때 자유도  $r$ 은 아래 식에서 소수점 이하를 잘라낸 정수이다.

$$r = \frac{\left[ \frac{(s_1^2 / n) + (s_2^2 / m)}{\frac{(s_1^2 / n)^2}{n-1} + \frac{(s_2^2 / m)^2}{m-1}} \right]^2}{1}$$

$$\text{단, } s_1^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad s_2^2 = \frac{1}{m-1} \sum_{j=1}^m (Y_j - \bar{Y})^2$$

# 두 모분산이 다르고 알려지지 않은 정규모집단인 경우

## 예제 8-9

모평균이 각각  $\mu_1 = 19$ ,  $\mu_2 = 16$ 이고 모분산이 알려지지 않은 두 정규모집단에서 각각 표본을 선정하여 다음을 얻었다. 두 표본평균의 차이가 4.5765 이상일 근사확률을 구하라.

표본 1	$n = 15, s_1^2 = 2.82$
표본 2	$m = 12, s_2^2 = 2.51$

**풀이** 표본1과 표본2의 표본평균을 각각  $\bar{X}$ ,  $\bar{Y}$  라 하면 다음을 얻는다.

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{2.82}{15} + \frac{2.51}{12}} = \sqrt{0.3972} = 0.6302$$

$$r = \frac{[(2.82/15) + (2.51/12)]^2}{\frac{(2.82/15)^2}{14} + \frac{(2.51/12)^2}{11}} = 24.26 \approx 24$$

따라서  $U = \frac{(\bar{X} - \bar{Y}) - 3}{0.6032} \sim t(24)$  이고 다음을 얻는다.

## 두 모분산이 다르고 알려지지 않은 정규모집단인 경우

$$P(\bar{X} - \bar{Y} \geq 4.5765) = P\left(\frac{\bar{X} - \bar{Y} - 3}{0.6302} \geq \frac{4.5765 - 3}{0.6302}\right) = P(T \geq 2.501)$$

한편 자유도 24인  $t$ -분포에서  $P(T \geq 2.797) = 0.005, P(T \geq 2.492) = 0.01$ 이므로 두 점  $(2.797, 0.005), (2.492, 0.01)$ 을 지나는 일차방정식은 다음과 같다.

$$y = 0.01 + \frac{0.005 - 0.01}{2.797 - 2.492}(x - 2.492); \quad y \approx 0.0509 - 0.0164x$$

따라서  $x = 2.501$ 을 대입하면 구하고자 하는 근사확률은 다음과 같다.

$$P(\bar{X} - \bar{Y} \geq 4.5765) = P(T \geq 2.501) \approx 0.0099$$

# 두 모분산이 알려진 임의의 정규모집단인 경우

❖ 표본의 크기  $n, m$ 이 충분히 크면 중심극한정리에 의하여 다음과 같다.

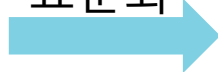
$$\bar{X} \approx N\left(\mu_1, \frac{\sigma_1^2}{n}\right), \quad \bar{Y} \approx N\left(\mu_2, \frac{\sigma_2^2}{m}\right)$$



$X$ 와  $Y$ 가 독립이므로

$$\bar{X} - \bar{Y} \approx N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)$$

표준화



$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \approx N(0, 1)$$

❖  $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 인 경우 :

$$\bar{X} - \bar{Y} \approx N\left(\mu_1 - \mu_2, \left(\frac{1}{n} + \frac{1}{m}\right)\sigma^2\right)$$

표준화



$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \approx N(0, 1)$$

# 두 모분산이 알려진 임의의 정규모집단인 경우

## 예제 8-10

독립인 두 모집단의 모평균과 모분산이 각각 다음과 같다. 모집단 1과 2에서 각각 크기 64인 표본을 임의로 선정할 때, 두 표본평균의 차이가 1.56 이상일 근사확률을 구하라.

모집단 1	$\mu_1 = 6, \sigma_1^2 = 3.2$
모집단 2	$\mu_2 = 5, \sigma_2^2 = 2.6$

### 풀이

표본1과 표본2의 표본평균을 각각  $\bar{X}$ ,  $\bar{Y}$  라 하면  $\mu_1 - \mu_2 = 1$ 이고 다음을 얻는다.

$$\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} = \sqrt{\frac{3.2}{64} + \frac{2.6}{64}} = \sqrt{0.090625} = 0.301$$

따라서  $\bar{X} - \bar{Y} \sim N(1, 0.301^2)$ 이고 구하고자 하는 확률은 다음과 같다.

$$P(\bar{X} - \bar{Y} \geq 1.56) = P\left(\frac{\bar{X} - \bar{Y} - 1}{0.301} \geq \frac{1.56 - 1}{0.301}\right) \approx P(Z \geq 1.86) = 1 - 0.9686 = 0.0314$$

## 두 모분산이 알려진 정규모집단인 경우

- ❖ 독립이고 모비율이 각각  $p_1, p_2$ 인 두 모집단에서 표본의 크기  $n, m$ 인 표본을 추출할 때, 각각의 표본비율은 다음과 같은 정규분포에 근사한다.

$$\hat{p}_1 \approx N\left(p_1, \frac{p_1 q_1}{n}\right), \quad \hat{p}_2 \approx N\left(p_2, \frac{p_2 q_2}{m}\right), \quad q_1 = 1 - p_1, q_2 = 1 - p_2$$

두 표본비율의 차  $\hat{p}_1 - \hat{p}_2$  는 다음 정규분포에 근사한다.

$$\hat{p}_1 - \hat{p}_2 \approx N\left(p_1 - p_2, \frac{p_1 q_1}{n} + \frac{p_2 q_2}{m}\right) \xrightarrow{\text{표준화}} \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n} + \frac{p_2 q_2}{m}}} \approx N(0, 1)$$



## 두 모분산이 알려진 정규모집단인 경우

모비율  $p_1, p_2$ 인 두 모집단에서 각각 크기  $n, m$ 인 표본을 선정할 때,  
두 표본비율의 차  $\hat{p}_1 - \hat{p}_2$ 에 관한 표본분포는 다음 정규분포에 근사한다.

$$\frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n} + \frac{p_2 q_2}{m}}} \approx N(0, 1)$$

# 두 모분산이 알려진 정규모집단인 경우

## 예제 8-11

불량률이  $p_1 = 0.005$ 인 생산라인 1에서 크기 100인 표본 1을 임의로 선정한다. 그리고 불량률이  $p_2 = 0.004$ 인 생산라인 2에서 크기 144인 표본 2를 임의로 선정한다. 두 표본의 불량률의 차이가  $-0.0159$ 와  $0.0186$  사이일 근사확률을 구하라.

### 풀이

표본1과 표본2의 표본비율을 각각  $\hat{p}_1, \hat{p}_2$ 라 하면  $p_1 - p_2 = 0.001$ 이고 다음을 얻는다.

$$\sqrt{\frac{p_1 q_1}{n} + \frac{p_2 q_2}{m}} = \sqrt{\frac{0.005 \times 0.995}{100} + \frac{0.004 \times 0.996}{144}} = \sqrt{0.0000774} = 0.0088$$

따라서  $\frac{(\hat{p}_1 - \hat{p}_2) - 0.001}{0.0088} \approx N(0, 1)$  이고 구하고자 하는 근사확률은 다음과 같다.

$$\begin{aligned} P(-0.0159 \leq \hat{p}_1 - \hat{p}_2 \leq 0.0186) &= P\left(\frac{-0.0159 - 0.001}{0.0088} \leq \frac{\hat{p}_1 - \hat{p}_2 - 0.001}{0.0088} \leq \frac{0.0186 - 0.001}{0.0088}\right) \\ &\approx P(-1.92 < Z < 2.00) = \Phi(2.00) - [1 - \Phi(1.92)] \\ &= 0.9772 - (1 - 0.9726) = 0.9498 \end{aligned}$$

# 합동표본분산의 분포

- ❖ 8.1절 학습 내용: 모분산  $\sigma^2$ 인 정규모집단에서 크기  $n$ 인 표본을 선정할 때 다음을 얻는다.

$$\frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$$

- ❖  $\sigma_1^2 = \sigma_2^2 = \sigma^2$  인 두 정규모집단에서 표본의 크기  $n, m$ 인 표본을 추출할 때, 두 표본분산을 각각  $S_1^2, S_2^2$  이라 하면 다음을 얻는다.

$$\frac{n-1}{\sigma^2} S_1^2 \sim \chi^2(n-1), \quad \frac{m-1}{\sigma^2} S_2^2 \sim \chi^2(m-1)$$

- ❖ 6장 학습 내용: 독립인  $V \sim c^2(n), U \sim c^2(m)$ 에 대하여  $V + U \sim c^2(n + m)$ 이다.

- ❖  $\sigma_1^2 = \sigma_2^2 = \sigma^2$  인 두 정규모집단에서 표본의 크기  $n, m$ 인 표본을 추출할 때, 두 표본분산을 각각  $S_1^2, S_2^2$  이라 하면 합동표본분산을 다음과 같이 정의한다.

$$S_p^2 = \frac{1}{n+m-2} [(n-1)S_1^2 + (m-1)S_2^2]$$

# 합동표본분산의 분포

❖ 합동표본분산  $S_p^2$ 은 다음과 같이 표현할 수 있다.

$$\begin{aligned}\frac{n+m-2}{\sigma^2} S_p^2 &= \frac{n+m-2}{\sigma^2} \cdot \frac{1}{n+m-2} [(n-1)S_1^2 + (m-1)S_2^2] \\ &= \frac{n-1}{\sigma^2} S_1^2 + \frac{m-1}{\sigma^2} S_2^2 \sim \chi^2(n+m-2)\end{aligned}$$

$\sigma_1^2 = \sigma_2^2 = \sigma^2$ 인 두 정규모집단에서 표본의 크기  $n, m$ 인 표본에 대한 합동표본분산은 자유도  $n + m - 2$ 인 카이제곱분포를 이룬다.

$$\frac{n+m-2}{\sigma^2} S_p^2 \sim \chi^2(n+m-2)$$

# 합동표본분산의 분포

## 예제 8-12

독립인 두 정규모집단  $N(\mu_1, 36)$ 과  $N(\mu_2, 36)$ 에서 각각 크기 6과 8인 확률표본을 추출한다.

(a)  $P(S_p^2 > s_0) = 0.05$ 를 만족하는  $s_0$ 를 구하라.

(b)  $P(S_p^2 > 84.9)$ 를 구하라.

풀이

(a)  $n = 6, m = 8$ 이고  $\sigma_1^2 = \sigma_2^2 = 36$ 이므로  $\frac{n+m-2}{\sigma^2} = \frac{1}{3}$  이고  $\frac{1}{3}S_p^2 \sim \chi^2(12)$  이다.

자유도 12인 카이제곱분포에서  $\chi_{0.05}^2(12) = 21.03$  이므로 다음을 얻는다.

$$P(S_p^2 > s_0) = P\left(\frac{1}{3}S_p^2 > \frac{1}{3}s_0\right) = P\left(\frac{1}{3}S_p^2 > \chi_{0.05}^2(12)\right) = 0.05$$

# 합동표본분산의 분포

따라서 구하고자 하는  $s_0$ 은 다음과 같다.

$$\frac{1}{3}s_0 = \chi_{0.05}^2(12) = 21.03; \quad s_0 = 3 \times 21.03 = 63.09$$

(b) 카이제곱분포표로부터 구하고자 하는 확률은 다음과 같다.

$$P(S_p^2 > 84.9) = P\left(\frac{1}{3}S_p^2 > \frac{84.9}{3}\right) = P\left(\frac{1}{3}S_p^2 > 28.3\right) = 0.005$$

# 표본분산 비의 분포

- ❖ 7장 학습 내용: 독립인  $U \sim c^2(n), V \sim c^2(m)$ 에 대하여 다음  $F$  - 분포를 얻는다.

$$\frac{U/n}{V/m} \sim F(n, m)$$

- ❖  $\sigma_1^2, \sigma_2^2$  인 두 정규모집단에서 표본의 크기  $n, m$ 인 표본을 추출할 때, 두 표본분산을 각각  $S_1^2, S_2^2$ 이라 하면 다음을 얻는다.

$$U = \frac{n-1}{\sigma_1^2} S_1^2 \sim \chi^2(n-1), \quad V = \frac{m-1}{\sigma_2^2} S_2^2 \sim \chi^2(m-1)$$

따라서 다음을 얻는다.

$$\frac{U/(n-1)}{V/(m-1)} = \frac{\frac{n-1}{\sigma_1^2} S_1^2 / (n-1)}{\frac{m-1}{\sigma_2^2} S_2^2 / (m-1)} = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \sim F(n-1, m-1)$$

# 표본분산 비의 분포

$\sigma_1^2, \sigma_2^2$  인 두 정규모집단에서 표본의 크기  $n, m$ 인 표본을 추출할 때, 두 표본분산을 각각  $S_1^2, S_2^2$  이라 하면 표본분산의 비는 분자, 분모의 자유도가 각각  $n - 1, m - 1$ 인  $F$  - 분포를 이룬다.

$$\frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \sim F(n-1, m-1)$$



# 표본분산 비의 분포

## 예제 8-13

독립인 두 정규모집단  $N(15, 9.61)$ 과  $N(15, 10.89)$ 에서 각각 크기 4와 6인 확률표본을 추출할 때,  $P\left(\frac{S_1^2}{S_2^2} > s_0\right) = 0.05$ 를 만족하는  $s_0$ 를 구하라.

풀이

$\sigma_1^2 = 9.61$ ,  $\sigma_2^2 = 10.89$  이고  $n = 4, m = 6$ 이므로  $\frac{S_1^2 / 9.61}{S_2^2 / 10.89} = \frac{10.89}{9.61} \frac{S_1^2}{S_2^2} \sim F(3, 5)$  이다.

분자와 분모의 자유도가 각각 3과 5인  $F$ - 분포에서  $f_{0.05}(3, 5) = 5.41$ 이므로 다음을 얻는다.

$$P\left(\frac{S_1^2}{S_2^2} > s_0\right) = P\left(\frac{10.89}{9.61} \frac{S_1^2}{S_2^2} > \frac{10.89}{9.61} s_0\right) = P\left(\frac{10.89}{9.61} \frac{S_1^2}{S_2^2} > f_{0.05}(3, 5)\right) = 0.05$$

따라서 구하고자 하는  $s_0$ 은 다음과 같다.

$$\frac{10.89}{9.61} s_0 = f_{0.05}(3, 5) = 5.41; \quad s_0 = \frac{9.61}{10.89} \times 5.41 = 4.774$$

# Q&A

