생생한 사례로 배우는 확률과 통계

[강의교안 이용 안내]

한빛아카데미 Hanbit Academy, Inc.

Chapter 09

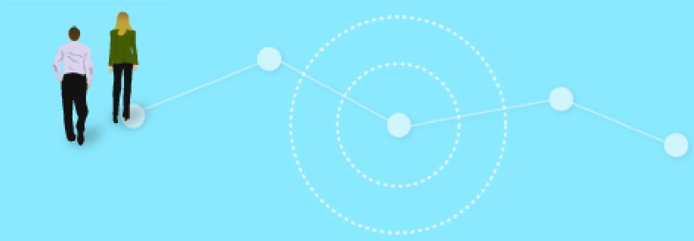
생생한 사례로 배우는
확률과 통계
PROBABILITY & STATISTICS

대표본 추정

Large Sample Estimation

목 차

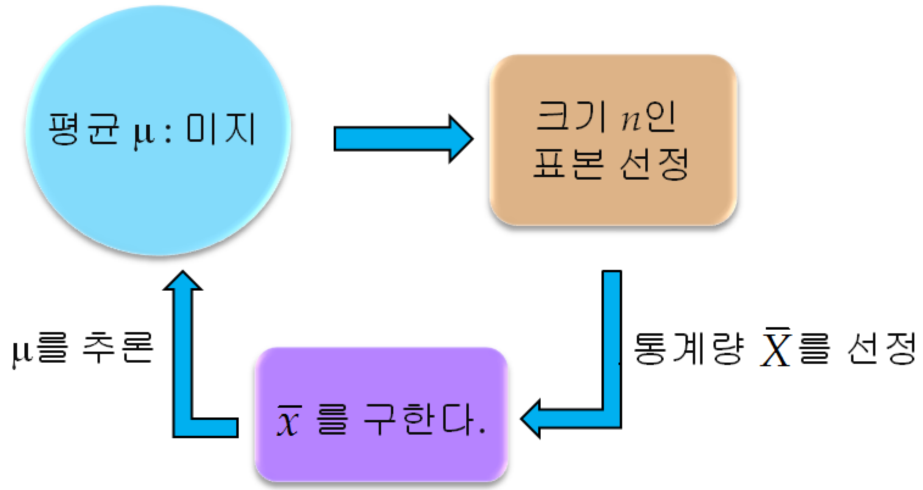
9.1 점추정과 구간추정



점추정과 구간추정

통계적 추론statistical inference : 표본을 이용하여 모집단의 특성을 나타내는 모수의 참값을 추론하거나 모수에 대한 주장에 대한 참 또는 거짓을 검정하는 과정

❖ 배터리 제조업체에서 생산된 모든 배터리의 하루 평균 사용시간을 알기 위하여, 크기 n인 표본을 임의로선정하여 얻은 표본평균 $\bar{x}$를 이용하여 모평균, 즉 생산되는 모든 배터리의 하루 평균 사용시간을 추론한다.



점추정량과 그 성질

예 어느 자동차 회사에서 새로 개발한 자동차의 평균 연비를 알기 위해 10대를 임의로 선정하여 다음을 얻었다고 하자.

17.4	17.2	18.1	17.5	17.7	17.6	17.5	17.1	17.8	17.6

$$\bar{x} = 17.55, \quad s = 0.288$$

결론 이 회사에서 새로 개발한 자동차의 평균 연비는 1리터당 17.55km이고 표준편차는 0.288km이라고 추론할 수 있다.

의문점 이러한 추론을 어느 정도 믿을 수 있는가?

모집단의 모수를 추론하는 방법을 살펴본다.

점추정량과 그 성질

추정량estimator : 모집단의 모수 θ를 추정하기 위하여 크기 n인 표본으로부터 선정한 통계량

❖ 예를 들어, 표본평균, 표본분산, 표본비율 등은 모평균, 모분산, 모비율에 대한 추정량이다.

NOTE

일반적으로 표본평균, 표본분산, 표본비율 등과 같은 추정량은 모집단으로부터 표본을 어떻게 선정하느냐에 따라 다른 값을 갖는다.

따라서 모수는θ에 대한 추정량은 다음과 같이 $i.i.d$ 확률변수 $X_1, X_2, \dots, X_n$의 함수 이다.

$$\hat{\Theta} = \hat{\Theta}(X_1, X_2, \cdots, X_n)$$

점추정량과 그 성질

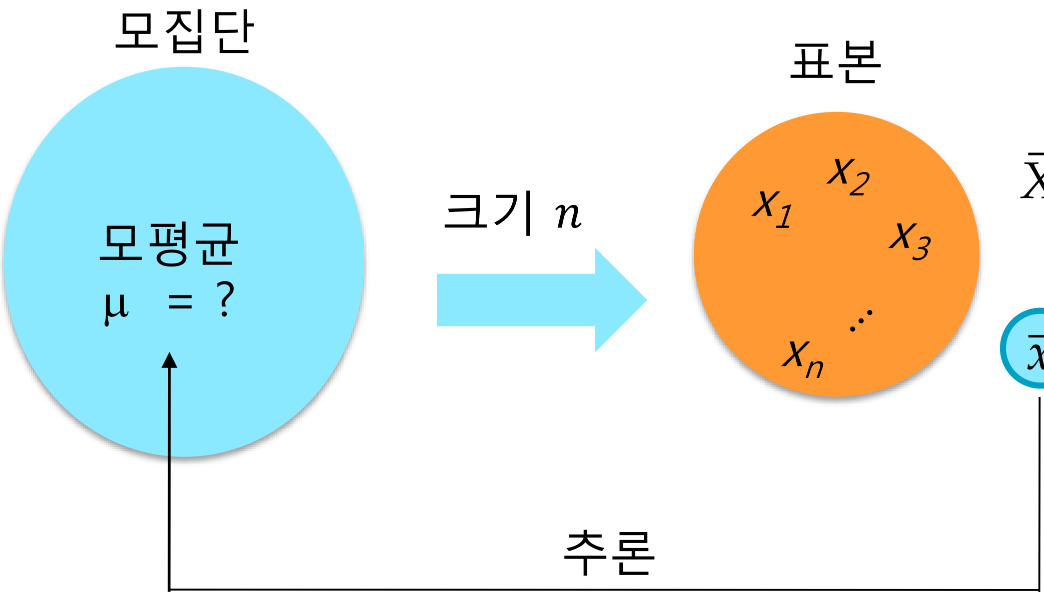
추정t estimate : 표본으로부터 얻은 추정량을 이용하여 모수 Q 를 추정하는 과정

점추정point estimate : 모수 θ 의 참값에 대해 최적의 추정값을 구하는 과정

점추정량point estimator : 모수 θ 의 참값인 수치를 추정하기 위해 표본에 기초하여 얻은 추정량 $\hat{\Theta} = \hat{\Theta}(X_1, X_2, \cdots, X_n)$

점추정값value of point estimator : 선정된 표본의 관찰값 $x_1, x_2, \ldots,$ x_n에 대한 정추정량값 $\hat{\theta} = \hat{\Theta}(x_1, x_2, \cdots, x_n)$

추정 절차



추정량

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

모평균 μ에 대한 점추정량

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

모평균 μ에 대한 점추정값

① 크기 n인 표본을 선정한다.

② 적당한 추정량을 선택한다.

③ 관찰값에 대한 추정값을 계산한다.

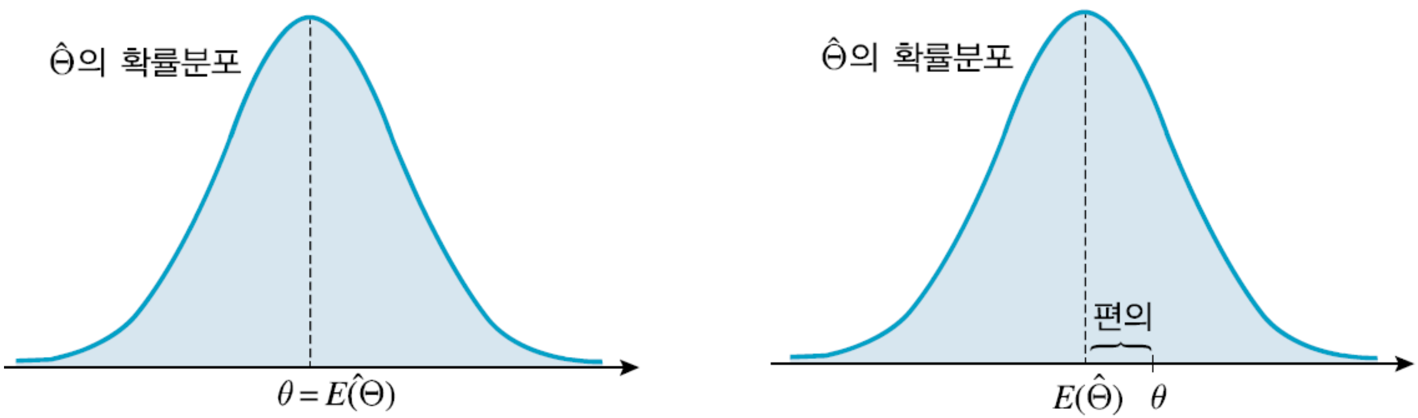
④ 추정값을 이용하여 모평균을 추정한다.

오차가 적은 추정값을 얻기 위해 바람직한 추정량을 선정해야 한다.

불편추정량

불편추정량unbiased estimator: 추정량의 평균이 모수θ 의 참값과 일치하는 추정량, 즉 $\mu_{\hat{\Theta}} = E(\hat{\Theta}) = \theta$ 를 만족하는 추정량

편의추정량biased estimator : 불편추정량이 아닌 추정량
$b = E(\hat{\Theta}) - \theta$ 를 **편의**bias라 한다.



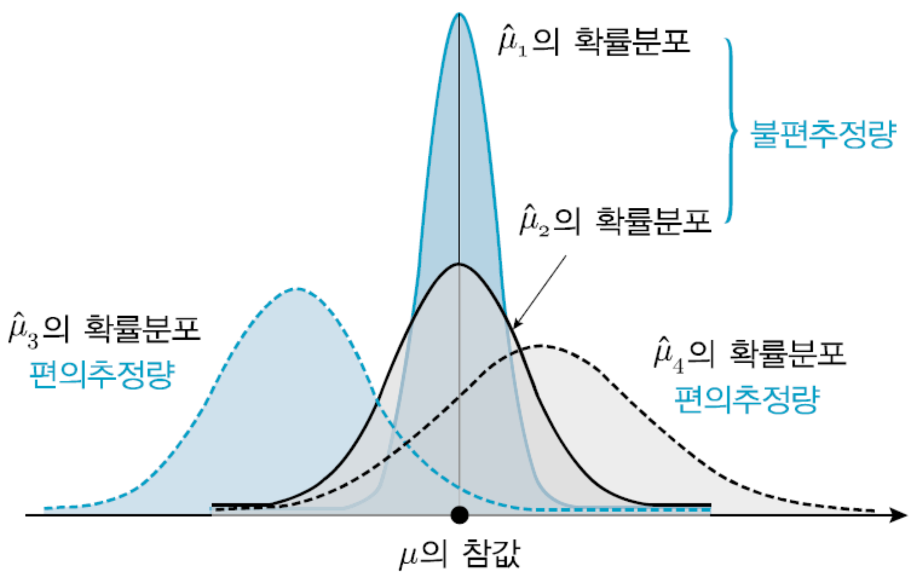
(a) 불편추정량　(b) 편의추정량

불편추정량

❖ 정규모집단에서 크기 n인 표본을 선정할 때, 표본평균 $\bar{X}$ 에 대하여 다음이 성립한다. 따라서 모평균μ에 대한 추정량 중에서 표본평균 $\hat{\mu} = \bar{X}$ 는 불편추정량이다.

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

❖ $E(\hat{\mu}_1) = E(\hat{\mu}_2) = \mu, \quad Var(\hat{\mu}_1) < Var(\hat{\mu}_2), \quad E(\hat{\mu}_3) < \mu < E(\hat{\mu}_4)$ 인 경우



불편추정량

예제 9-1

모집단으로부터 크기 2인 표본을 임의로 선정하여 $\{X_1, X_2\}$라 하자. 이때 양의 상수 a와 b에 대해 모평균 μ의 점추정량을 $\hat{\mu} = \dfrac{aX_1 + bX_2}{a+b}$라 하자. 이 점추정량이 μ에 대한 불편추정량인 것을 보여라.

풀이

$E(X_1) = E(X_2) = \mu$이므로 다음을 얻는다.

$$E(\hat{\mu}) = E\left(\frac{aX_1 + bX_2}{a+b}\right) = \frac{1}{a+b}\left[aE(X_1) + bE(X_2)\right] = \frac{a\mu + b\mu}{a+b} = \mu$$

따라서 점추정량 $\hat{\mu}$ 는 모평균 μ의 불편추정량이다.

❖ 가중표본평균과 표본평균 $\bar{X}$ 는 모평균μ에 대한 불편추정량이지만, 일반적으로 절사평균은 모평균에 대한 불편추정량이 아니다.

불편추정량

❖ 표본분산 S^2은 모분산σ^2의 불편추정량이다.

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left[(X_i-\mu)-(\bar{X}-\mu)\right]^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left[(X_i-\mu)^2 - \boxed{2(\bar{X}-\mu)(X_i-\mu)} + (\bar{X}-\mu)^2\right]$$

$\boxed{\sum_{i=1}^{n} X_i = n\mu \text{이므로}}$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(X_i-\mu)^2 - n(\bar{X}-\mu)^2\right]$$

$E(\bar{X}) = \mu \quad \Rightarrow \quad Var(\bar{X}) = E\left[(\bar{X}-\mu)^2\right] = \frac{\sigma^2}{n}$ 이므로

$$E(S^2) = \frac{1}{n-1}\left[\sum_{i=1}^{n}E\left[(X_i-\mu)^2\right] - nE\left[(\bar{X}-\mu)^2\right]\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}Var(X_i) - nVar(\bar{X})\right]$$

$$= \frac{1}{n-1}(n\sigma^2-\sigma^2) = \sigma^2$$

불편추정량

NOTE 표본분산을 다음과 같이 정의하면 S^2은 σ^2에 대한 편의추정량이다.

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \quad \Rightarrow \quad E(S^2) = \frac{n-1}{n}\sigma^2 (\neq \sigma^2)$$

$$bias = E(S^2) - \sigma^2 = -\frac{1}{n}\sigma^2$$

❖ 표본비율 $\hat{P}$ 은 모비율 p의 불편추정량이다.

$$X_i \sim B(1, p), \quad n = 1, 2, \cdots n, \quad \Rightarrow \quad X = \sum_{i=1}^{n} X_i \sim B(n, p)$$

이므로 $E(X) = np$이고 표본비율의 평균은 다음과 같다.

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

주요 모수에 대한 불편추정량

❖ 표본평균 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 는 모평균 μ에 대한 불편추정량이다.

❖ 표본분산 $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ 은 모분산σ^2에 대한 불편추정량이다.

❖ 표본비율 $\hat{p} = \frac{X}{n}$ 는 모비율 p에 대한 불편추정량이다.

주요 모수에 대한 불편추정량

예제 9-2

모평균이 μ인 모집단에서 크기 3인 확률표본 $\{X_1, X_2, X_3\}$를 추출하여 모평균에 대한 점추정량을 다음과 같이 정의하였다. 모평균 μ에 대한 불편추정량과 편의추정량을 구하라.

$$\hat{\mu}_1 = \frac{1}{3}(X_1 + X_2 + X_3), \quad \hat{\mu}_2 = \frac{1}{4}(X_1 + 2X_2 + X_3)$$

$$\hat{\mu}_3 = \frac{1}{5}(X_1 + X_2 + 2X_3), \quad \hat{\mu}_4 = \frac{1}{5}(2X_1 + 2X_2 + X_3)$$

풀이

X_1, X_2, X_3이 동일한 모집단 분포에 따르므로 $E(X_1) = E(X_2) = E(X_3) = \mu$이다.
따라서 각 추정량의 기댓값을 구하면 각각 다음과 같다.

$$E(\hat{\mu}_1) = \frac{1}{3}E(X_1 + X_2 + X_3) = \frac{1}{3}(\mu + \mu + \mu) = \mu$$

$$E(\hat{\mu}_2) = \frac{1}{4}E(X_1 + 2X_2 + X_3) = \frac{1}{4}(\mu + 2\mu + \mu) = \mu$$

$$E(\hat{\mu}_3) = \frac{1}{5}E(X_1 + X_2 + 2X_3) = \frac{1}{5}(\mu + \mu + 2\mu) = \frac{4}{5}\mu$$

$$E(\hat{\mu}_4) = \frac{1}{5}E(2X_1 + 2X_2 + X_3) = \frac{1}{5}(2\mu + 2\mu + \mu) = \mu$$

유효추정량

유효추정량efficient estimator : 추정량의 표본분포가 모수 θ의 참값에 가장 가까운 추정량, 즉 다음을 만족하는 추정량 $\hat{\Theta}$

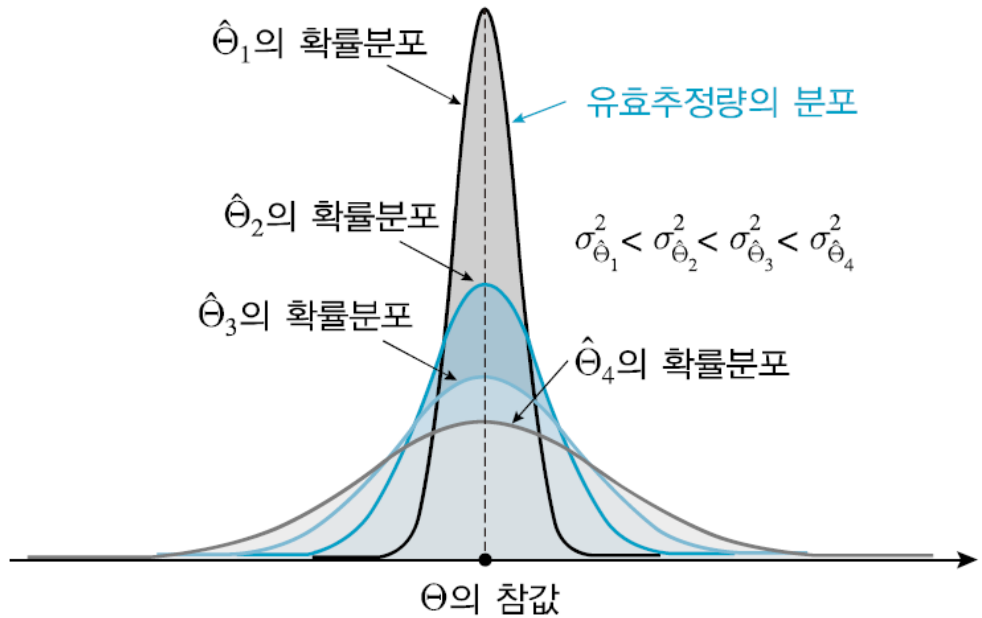
$$Var(\hat{\Theta}) = \min\{Var(\hat{\Theta}_1), Var(\hat{\Theta}_2), \cdots, Var(\hat{\Theta}_k)\}$$

유효추정량

표준오차standard error : 모수 θ를 추정하기 위해 사용되는 추정량의 표준편차

$$\mathrm{S.E}(\hat{\Theta}) = \sqrt{\mathrm{Var}(\hat{\Theta})}$$

유효추정량



❖ 표본평균 $\bar{X}$와 표본중위수 Me를 이용하여 모평균μ을 추정하는 경우 :

$$Var(\bar{X}) = \frac{1}{n}\sigma^2, \quad Var(Me) \approx \frac{1.57}{n}\sigma^2$$

❖ 표본평균 $\bar{X}$가 표본중위수 Me보다 더 효과적이다.

유효추정량

최소분산불편추정량minimum variance unbised estimator : 모수 θ의 불편성과 유효성을 갖는 추정량

❖ 표본의 크기가 n과 $2n$인 경우의 표본평균을 각각 $\bar{X}_n$, $\bar{X}_{2n}$ 이라 하면 두 표본평균의 분산은 각각 다음과 같다.

$$Var(\bar{X}_n) = \frac{1}{n^2} Var(X_1 + X_2 + \cdots + X_n) = \frac{1}{n}\sigma^2$$

$$Var(\bar{X}_{2n}) = \frac{1}{4n^2} Var(X_1 + X_2 + \cdots + X_{2n}) = \frac{1}{2n}\sigma^2$$

$$\Rightarrow \quad Var(\bar{X}_{2n}) < Var(\bar{X}_n)$$

표본의 크기가 클수록 유효성이 크다.

예제 9-3

[예제 9−2]에서 구한 불편추정량 중에서 유효추정량을 구하라.

풀이

$Var(X_1) = Var(X_2) = Var(X_3) = s^2$이고 모평균 m에 대한 불편추정량은 $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\mu}_4$ 이다. 그리고 이 추정량들의 분산을 구하면 각각 다음과 같다.

$$Var(\hat{\mu}_1) = \frac{1}{9} Var(X_1 + X_2 + X_3) = \frac{1}{3}\sigma^2$$

$$Var(\hat{\mu}_2) = \frac{1}{16} Var(X_1 + 2X_2 + X_3) = \frac{3}{8}\sigma^2$$

$$Var(\hat{\mu}_4) = \frac{1}{25} Var(2X_1 + 2X_2 + X_3) = \frac{9}{25}\sigma^2$$

$$\Rightarrow \quad Var(\hat{\mu}_1) < Var(\hat{\mu}_4) < Var(\hat{\mu}_2)$$

따라서 유효추정량은 $\hat{\mu}_1$ 이다.

유효추정량

구간추정interval estimation : 모수θ의 참값이 포함될 것으로 믿어지는 구간을 추정하는 것

❖ 표본에 의한 점추정량 $\hat{\Theta}$ 의 관찰값 $\hat{\theta}_l$, $\hat{\theta}_u$에 대하여 모수θ가 $\hat{\theta}_l < \theta < \hat{\theta}_u$를 만족하는 구간 $(\hat{\theta}_l, \hat{\theta}_u)$를 추정하는 방법

표본에 따라 점추정량 $\hat{\Theta}$의 관찰값 $\hat{\theta}_l$, $\hat{\theta}_u$의 값이 다르게 나타난다. 따라서 양 끝점은 $\hat{\Theta}$ 에 의해 결정되는 두 추정량 $\hat{\Theta}_l$, $\hat{\Theta}_u$ 의 관찰값으로 생각할 수 있다.

이때 모수θ의 참값이 포함될 확률이 다음과 같이 $1 - \alpha, 0 < \alpha < 1$이 되도록 두 추정량 $\hat{\Theta}_l$, $\hat{\Theta}_u$ 을 결정한다.

$$\boxed{P(\hat{\Theta}_l < \theta < \hat{\Theta}_u) = 1 - \alpha}$$

유효추정량

선정된 표본에 따라 $\hat{\Theta}_l$, $\hat{\Theta}_u$ 의 관찰값 $\hat{\theta}_l$, $\hat{\theta}_u$ 에 의한 구간 $\hat{\theta}_l < \theta < \hat{\theta}_u$ 를 얻는다.

신뢰도degree of confidence : 모수 θ 의 참값이 추정한 구간 안에 포함될 것으로 믿어지는 미리 정한 확신의 정도 $1 - \alpha$.
일반적으로 $100\,(1 - \alpha)\%$로 나타낸다.

유효추정량

신뢰구간confidence interval : 신뢰도 100 (1 – a)%에서 모수θ의 참값이 포함될 것으로 믿어지는 구간

$$\left(\hat{\theta}_l,\ \hat{\theta}_u\right)$$

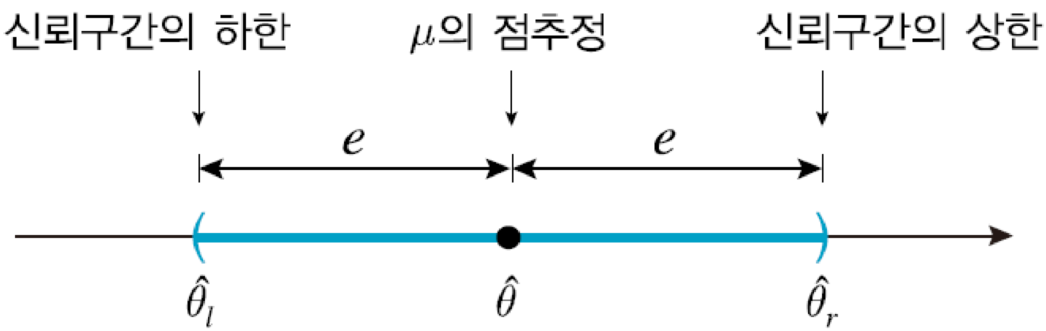
$\hat{\theta}_l$: 신뢰구간의 하한

$\hat{\theta}_u$: 신뢰구간의 상한

❖ 모수θ에 대한 신뢰구간은 점추정값 $\hat{\theta}$을 중심으로 동일한 길이 e만큼 떨어진 구간을 사용하며, 보편적으로 신뢰도는 *90%, 95%, 99%*를 자주 사용한다.

유효추정량

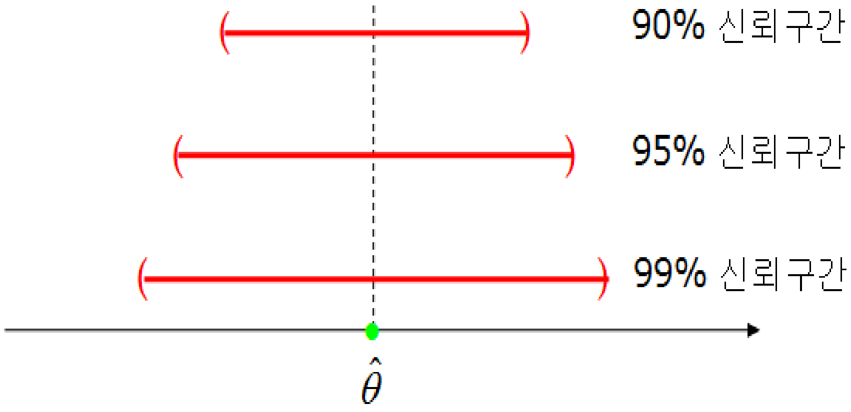
오차한계margin of error : 모수θ의 점추정값을 중심으로 동일한 거리만큼 떨어진 길이 e



[그림 9-4] 신뢰도 $100(1-\alpha)\%$의 신뢰구간

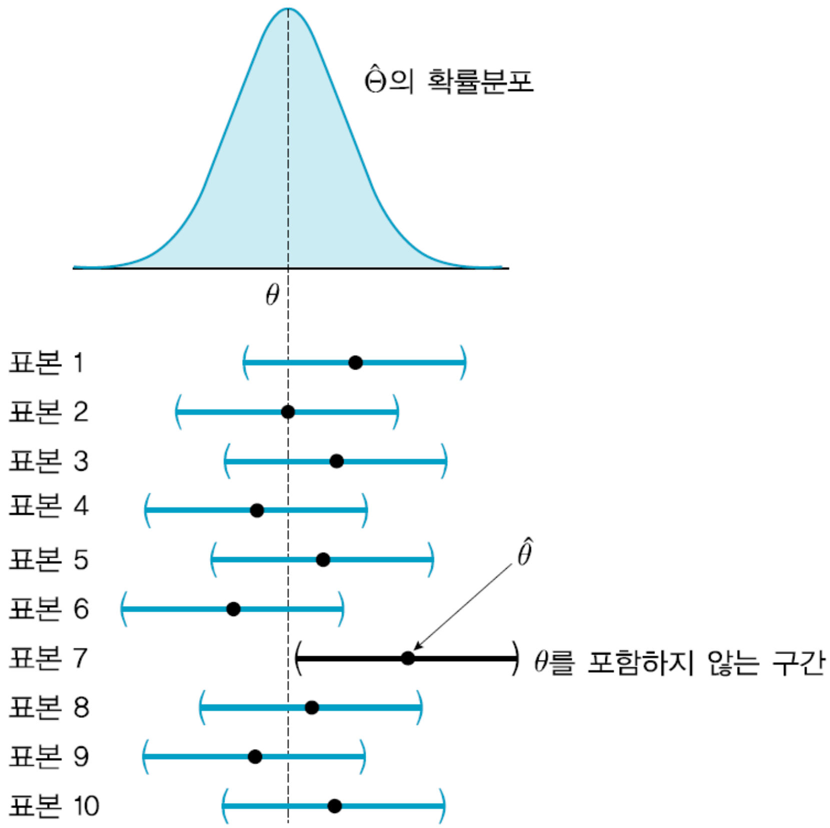
유효추정량

❖ 신뢰도가 클수록 모수θ에 대한 신뢰구간은 커진다.

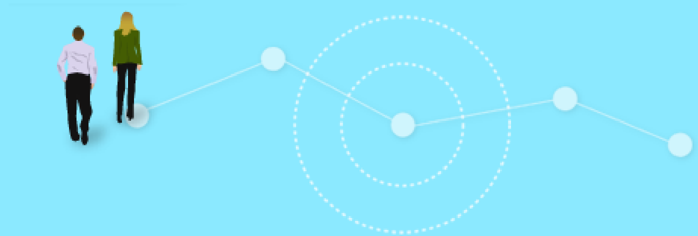


95% 신뢰도의 의미

❖ 동일한 모집단으로부터 동일한 크기의 표본 10개를 임의로 추출하였을 때, 이 표본들로부터 얻은 신뢰구간들 중에서 90%에 해당하는 9개의 구간이 모평균의 참값을 포함하고 최대 10%에 해당하는 1개의 구간은 모수의 참값을 포함하지 않을 수 있음을 의미한다.

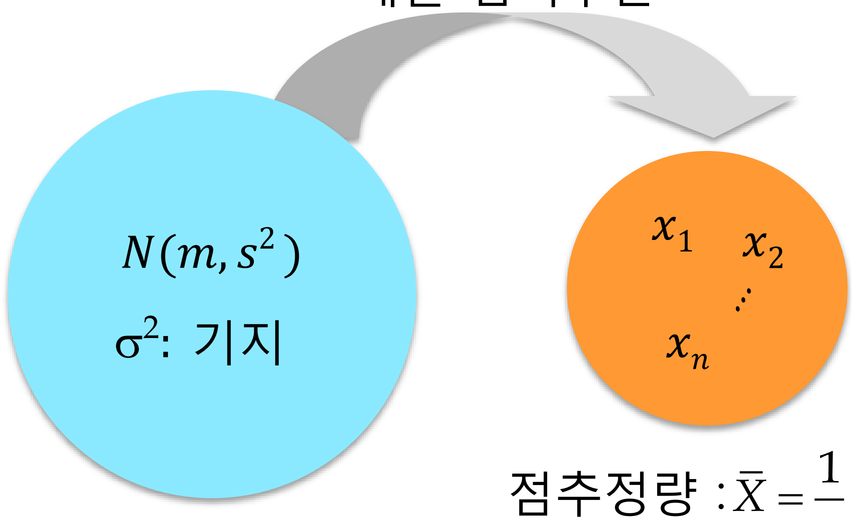


9.2 모평균의 추정



모분산이 알려진 단일 정규모집단인 경우

n개를 임의추출



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

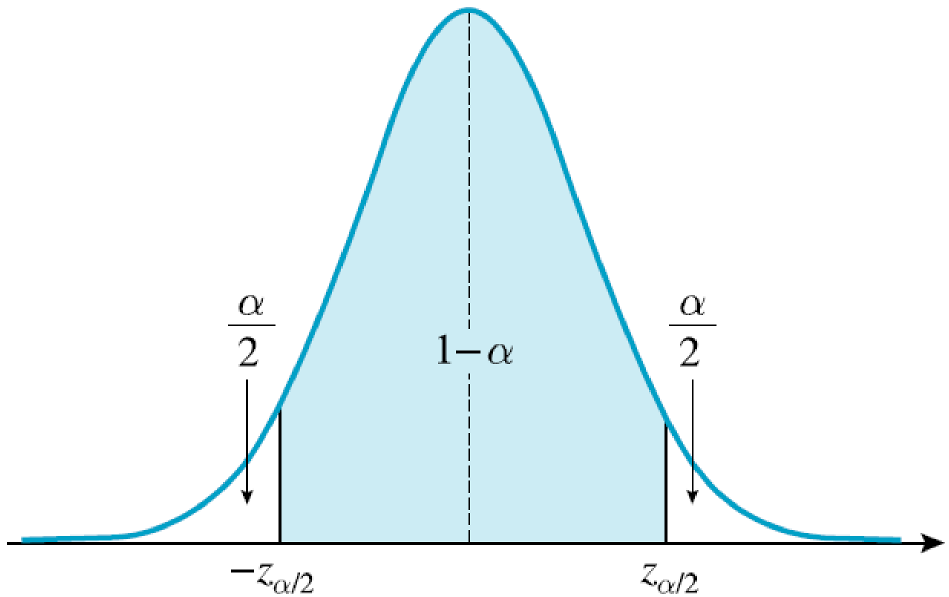
표본평균의 확률분포 : $\bar{X} \sim N\left(\mu,\ \frac{\sigma^2}{n}\right),\quad Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,\ 1)$

모평균의 신뢰구간

❖ **모평균 μ에 대한 100(1 – α)% 신뢰구간 :**

모평균에 대한 추정값 $\bar{x}$를 중심으로 신뢰구간의 하한 $\hat{\mu}_l$과 상한 $\hat{\mu}_u$에 대해 다음을 만족하는 구간 $(\hat{\mu}_l, \hat{\mu}_u)$를 구한다.

표준정규분포에서 양쪽 꼬리확률이 각각 $a/2$인 임계점 : $-z_{a/2}, z_{a/2}$



모평균의 신뢰구간

$$P\left(\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| < z_{\alpha/2}\right) = P\left(\left|\bar{X}-\mu\right| < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

신뢰구간의 하한

$$P\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

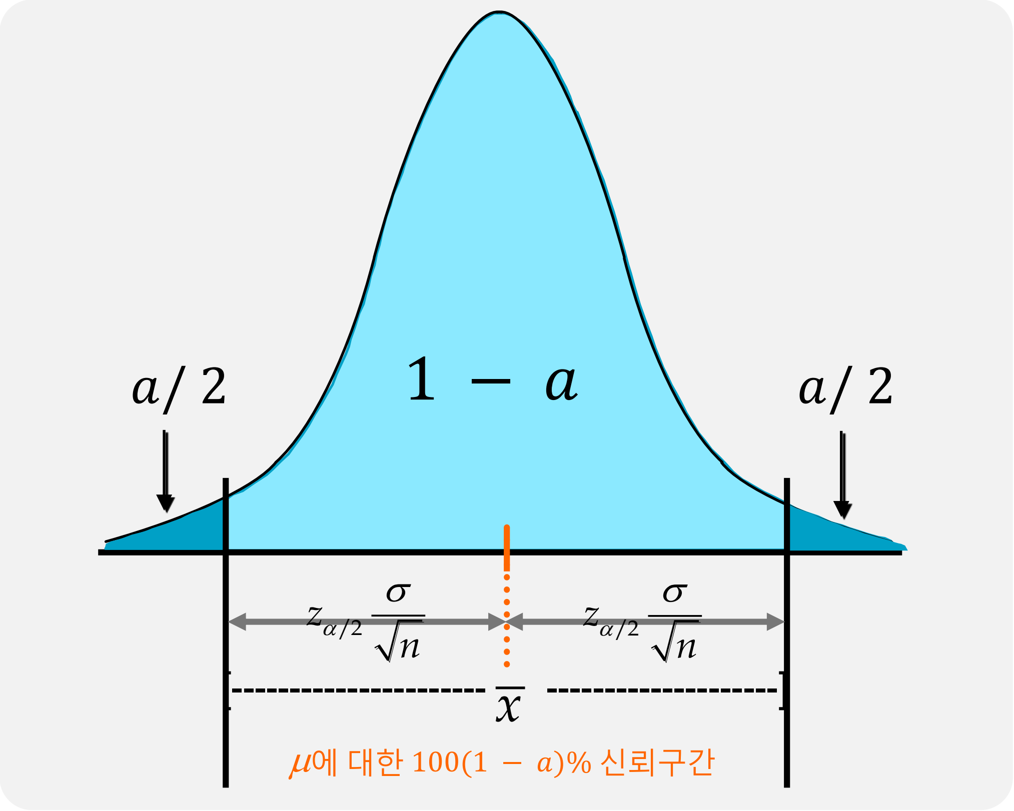
신뢰구간의 상한

모평균의 신뢰구간

$$P\left(\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| < z_{\alpha/2}\right) = P\left(|\bar{X}-\mu| < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

표준오차 : $\mathrm{S.E}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$

100 (1 – a)% 신뢰도에서 모수 μ 에 대한 오차한계



모평균의 신뢰구간

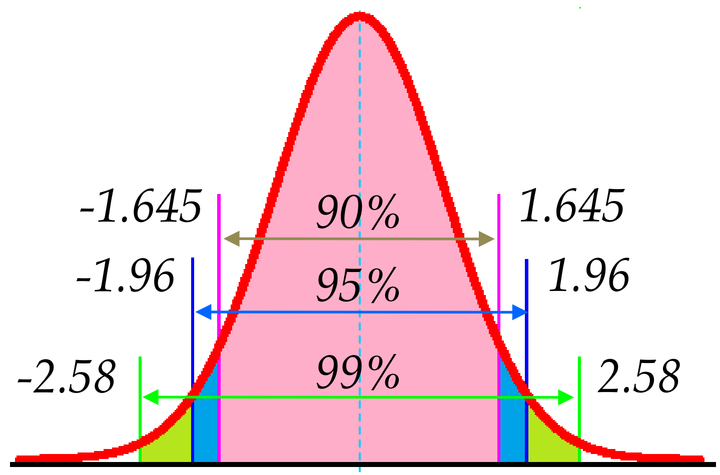
❖ 모분산 σ^2이 알려진 정규모집단의 모평균 μ에 대한 100(1 – α)% 신뢰구간

$$\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\quad \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

$$P\left(\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| < 1.645\right) = P\left(\left|\bar{X}-\mu\right| < 1.645\frac{\sigma}{\sqrt{n}}\right) = 0.90$$

$$P\left(\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| < 1.96\right) = P\left(\left|\bar{X}-\mu\right| < 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

$$P\left(\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| < 2.58\right) = P\left(\left|\bar{X}-\mu\right| < 2.58\frac{\sigma}{\sqrt{n}}\right) = 0.99$$



모평균의 신뢰구간

❖ **신뢰도에 따른 오차한계**

$|\bar{X}-\mu|$ 에 대한 90% 오차한계 : $e_{90\%} = 1.645\dfrac{\sigma}{\sqrt{n}}$

$|\bar{X}-\mu|$ 에 대한 95% 오차한계 : $e_{95\%} = 1.96\dfrac{\sigma}{\sqrt{n}}$

$|\bar{X}-\mu|$ 에 대한 99% 오차한계 : $e_{99\%} = 2.58\dfrac{\sigma}{\sqrt{n}}$

모평균의 신뢰구간

❖ **모분산 σ^2이 알려진 정규모집단의 모평균 μ에 대한 90%, 95%, 99% 신뢰구간**

모수 μ에 대한 90% 신뢰구간 : $\left(\bar{x} - 1.645\dfrac{\sigma}{\sqrt{n}},\quad \bar{x} + 1.645\dfrac{\sigma}{\sqrt{n}}\right)$

모수 μ에 대한 95% 신뢰구간 : $\left(\bar{x} - 1.96\dfrac{\sigma}{\sqrt{n}},\quad \bar{x} + 1.96\dfrac{\sigma}{\sqrt{n}}\right)$

모수 μ에 대한 99% 신뢰구간 : $\left(\bar{x} - 2.58\dfrac{\sigma}{\sqrt{n}},\quad \bar{x} + 2.58\dfrac{\sigma}{\sqrt{n}}\right)$

모평균의 신뢰구간

예제 9-4

$\sigma^2 = 16$ 인 정규모집단의 모평균을 추정하기 위해 크기 50 인 표본을 추출하였다. 표본평균이 25 일 때, 모평균에 대한 90% 신뢰구간을 구하라.

풀이

$\bar{x} = 25, \sigma^2 = 16, n = 50$이므로 표본평균 $\bar{X}$ 의 표준오차는 다음과 같다.

$$S.E(\bar{X}) = \frac{4}{\sqrt{50}} = 0.5657$$

모평균 μ에 대한 90% 신뢰구간의 오차한계 :

$$e = 1.645 \times S.E(\bar{X}) = 1.645 \times 0.5657 = 0.93$$

모평균 μ에 대한 90% 신뢰구간 :

$$(25 - 0.93, \quad 25 + 0.93) = (24.07, \quad 25.93)$$

모분산이 알려진 단일 임의의 모집단인 경우

❖ 모분산σ^2이 알려진 임의의 모집단에 대해 다음을 얻는다.(중심극한정리)

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

❖ 모분산 σ^2 이 알려진 임의의 모집단의 모평균μ에 대한 90%, 95%, 99% 근사신뢰구간

모수 μ에 대한 90% 근사신뢰구간 : $\left(\bar{x} - 1.645 \frac{\sigma}{\sqrt{n}}, \quad \bar{x} + 1.645 \frac{\sigma}{\sqrt{n}} \right)$

모수 μ에 대한 95% 근사신뢰구간 : $\left(\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}}, \quad \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}} \right)$

모수 μ에 대한 99% 근사신뢰구간 : $\left(\bar{x} - 2.58 \frac{\sigma}{\sqrt{n}}, \quad \bar{x} + 2.58 \frac{\sigma}{\sqrt{n}} \right)$

모분산을 모르는 단일 임의의 모집단인 경우

❖ **표본의 크기가 충분히 크면 $s^2 \approx \sigma^2$이므로 다음을 얻는다.**

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

❖ **모분산σ^2이 알려진 임의의 모집단의 모평균 μ에 대한 90%, 95%, 99% 근사신뢰구간**

모수 μ에 대한 90% 근사신뢰구간 : $\left(\bar{x} - 1.645 \frac{s}{\sqrt{n}}, \quad \bar{x} + 1.645 \frac{s}{\sqrt{n}} \right)$

모수 μ에 대한 95% 근사신뢰구간 : $\left(\bar{x} - 1.96 \frac{s}{\sqrt{n}}, \quad \bar{x} + 1.96 \frac{s}{\sqrt{n}} \right)$

모수 μ에 대한 99% 근사신뢰구간 : $\left(\bar{x} - 2.58 \frac{s}{\sqrt{n}}, \quad \bar{x} + 2.58 \frac{s}{\sqrt{n}} \right)$

모평균의 신뢰구간

예제 9-5

$\sigma^2 = 9$인 모집단에서 크기 64인 표본을 추출하여 $\overline{x} = 15$를 얻었다. 이때 모평균에 대한 90% 근사신뢰구간을 구하라.

풀이

$\overline{x} = 15, \sigma^2 = 9, n = 64$이므로 표본평균 $\overline{X}$의 표준오차는 다음과 같다.

$$S.E(\overline{X}) \approx \frac{3}{\sqrt{64}} = 0.375$$

모평균 μ에 대한 90% 근사신뢰구간의 오차한계 :

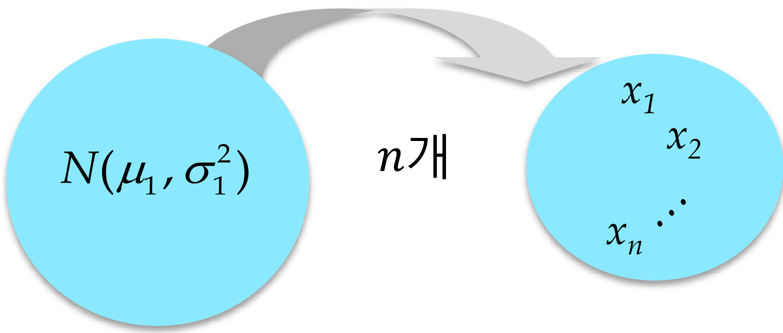
$$e = 1.645 \times S.E(\overline{X}) = 1.645 \times 0.375 = 0.617$$

모평균 μ에 대한 90% 근사신뢰구간 :

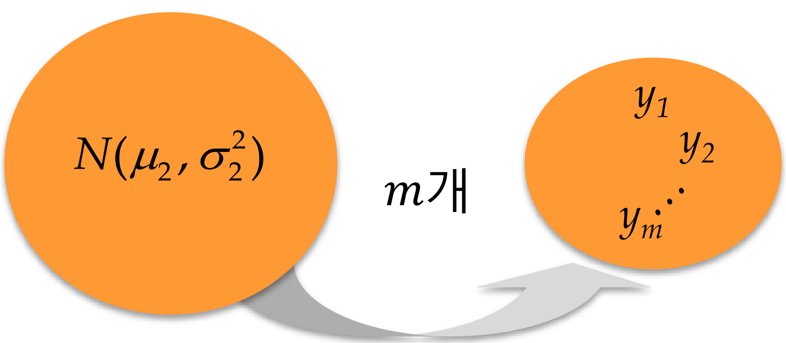
$$(15 - 0.617, \quad 15 + 0.617) = (14.383, \quad 15.617)$$

두 모분산이 알려진 정규모집단인 경우

두 모분산 σ_1^2, σ_2^2을 알고 있는 독립인 정규모집단 $N(\mu_1, \sigma_1^2)$, $N(\mu_2, \sigma_2^2)$에서 각각 크기 n과 m인 표본을 선정하여 표본평균을 $\bar{X}, \bar{Y}$라 하자.



표본평균 : $\bar{X} \sim N\left(\mu_1, \dfrac{\sigma_1^2}{n}\right)$



표본평균 : $\bar{Y} \sim N\left(\mu_2, \dfrac{\sigma_2^2}{m}\right)$

두 모분산이 알려진 정규모집단인 경우

두 표본평균의 차 $\bar{X}-\bar{Y}$ **의 확률분포는 다음 정규분포에 따른다.**

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

$$P\left(\left|\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{(\sigma_1^2/n)+(\sigma_2^2/m)}}\right|<z_{\alpha/2}\right)=P\left(\left|(\bar{X}-\bar{Y})-(\mu_1-\mu_2)\right|<z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\right)=1-\alpha$$

100 (1 – α)% 신뢰도에서 모수 μ_1 - μ_2에 대한 오차한계

표준오차 : S.E$(\bar{X}-\bar{Y})$

$$P\left((\bar{X}-\bar{Y})-z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}<\mu_1-\mu_2<(\bar{X}-\bar{Y})+z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\right)=1-\alpha$$

모평균 차의 신뢰구간

❖ **모수 μ_1 - μ_2에 대한 *100 (1 – α)%* 신뢰구간**

$$\left((\bar{X}-\bar{Y}) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}, \qquad (\bar{X}-\bar{Y}) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}} \right)$$

❖ **신뢰도에 따른 오차한계**

$\left|(\bar{X}-\bar{Y})-(\mu_1-\mu_2)\right|$ 에 대한 90% 오차한계 : $e_{90\%} = 1.645\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}$

$\left|(\bar{X}-\bar{Y})-(\mu_1-\mu_2)\right|$ 에 대한 95% 오차한계 : $e_{95\%} = 1.96\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}$

$\left|(\bar{X}-\bar{Y})-(\mu_1-\mu_2)\right|$ 에 대한 99% 오차한계 : $e_{99\%} = 2.58\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}$

모평균 차의 신뢰구간

❖ **모분산이 알려진 두 정규모집단의 모평균의 차에 대한 90%, 95%, 99% 신뢰구간**

$\mu_1 - \mu_2$ 에 대한 *90%* 신뢰구간 : $\left((\bar{x}-\bar{y})-e_{90\%},\ \ (\bar{x}-\bar{y})+e_{90\%}\right)$

$\mu_1 - \mu_2$ 에 대한 *95%* 신뢰구간 : $\left((\bar{x}-\bar{y})-e_{95\%},\ \ (\bar{x}-\bar{y})+e_{95\%}\right)$

$\mu_1 - \mu_2$ 에 대한 *99%* 신뢰구간 : $\left((\bar{x}-\bar{y})-e_{99\%},\ \ (\bar{x}-\bar{y})+e_{99\%}\right)$

* $\sigma_1^2 = \sigma_2^2 = \sigma^2$일 때, 100(1 – a)% 신뢰구간

$$\left((\bar{x}-\bar{y})-z_{\alpha/2}\sigma\sqrt{\frac{1}{n}+\frac{1}{m}},\ \ (\bar{x}-\bar{y})+z_{\alpha/2}\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}\right)$$

모평균 차의 신뢰구간

예제 9-6

동급인 자동차 A와 B의 연비에 차이가 있는지 알아보기 위해 자동차 A와 B를 각각 50대씩 임의로 선정하여 조사하였다. 선정된 자동차들의 1리터당 평균 거리가 각각 21.5km와 20.6km일 때, 두 자동차의 1리터당 평균 거리의 차에 대한 95% 신뢰구간을 구하라. 단, 자동차 A와 B의 1리터당 운행거리는 모표준편차가 각각 1.2km와 1.5km인 정규분포를 따른다고 한다.

풀이

선정된 자동차 A와 B의 표본평균을 각각 $\bar{X}, \bar{Y}$ 라 하면 $\bar{x}=21.5, \bar{y}=20.6$ 이므로 $\bar{x}-\bar{y}=0.9$ 이다. $\sigma_1^2=1.44,\ \sigma_2^2=2.25,\ n=m=50$ 이므로 표준오차는 다음과 같다.

$$S.E(\bar{X}-\bar{Y})=\sqrt{\frac{1.44}{50}+\frac{2.25}{50}}=\sqrt{0.0738}=0.2717$$

두 모평균 차에 대한 95% 신뢰구간의 오차한계 :

$$e=1.96\times S.E(\bar{X}-\bar{Y})=1.96\times 0.2717=0.5325$$

두 모평균 $\mu_1-\mu_2$ 에 대한 95% 신뢰구간 :

$$(0.9-0.5325,\quad 0.9+0.5325)=(0.3657,\quad 1.4325)$$

모분산이 알려진 임의의 두 모집단인 경우

모분산 σ_1^2, σ_2^2 이 알려진 임의의 두 모집단에 대해 다음을 얻는다.(중심극한정리)

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

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

$$P\left((\bar{X}-\bar{Y}) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} < \mu_1 - \mu_2 < (\bar{X}-\bar{Y}) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}\right) \approx 1 - \alpha$$

❖ 모분산 σ_1^2, σ_2^2이 알려진 임의의 모집단의 모평균 차 $\mu_1 - \mu_2$에 대한 근사신뢰구간

$$\left((\bar{X}-\bar{Y}) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}, \quad (\bar{X}-\bar{Y}) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}\right)$$

모분산이 알려진 임의의 두 모집단인 경우

모분산 σ_1^2, σ_2^2 이 알려진 임의의 모집단의 모평균 차 $\mu_1 - \mu_2$에 대한 *90%, 95%, 99%* 근사신뢰구간

모평균 차 $\mu_1 - \mu$에 대한 *90%* 근사신뢰구간 :

$$\left((\bar{x} - \bar{y}) - 1.645\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}},\quad (\bar{x} - \bar{y}) + 1.645\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \right)$$

모평균 차 $\mu_1 - \mu$에 대한 *90%* 근사신뢰구간 :

$$\left((\bar{x} - \bar{y}) - 1.96\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}},\quad (\bar{x} - \bar{y}) + 1.96\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \right)$$

모평균 차 $\mu_1 - \mu$에 대한 *90%* 근사신뢰구간 :

$$\left((\bar{x} - \bar{y}) - 2.58\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}},\quad (\bar{x} - \bar{y}) + 2.58\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \right)$$

모분산이 알려진 임의의 두 모집단인 경우

예제 9-7

어느 기업의 연구소에서 두 종류의 휴대전화용 배터리 A와 B를 개발하였다. 두 종류의 배터리를 시험한 결과가 다음과 같을 때, 이때 두 종류의 배터리의 평균 사용시간 차에 대한 90% 신뢰구간을 구하라.

$n=64$	$\overline{x}=26.5$시간, $\sigma_1=1.3$시간
$m=100$	$\overline{y}=25.6$시간, $\sigma_2=0.9$시간

풀이

$s_1^2=1.69,\ s_2^2=0.81,\ n=64,\ m=100$ 이므로 표준오차는 다음과 같다.

$$\mathrm{S.E}(\bar{X}-\bar{Y})=\sqrt{\frac{1.3^2}{64}+\frac{0.9^2}{100}}\approx\sqrt{0.0345}=0.186$$

모분산이 알려진 임의의 두 모집단인 경우

두 배터리의 표본평균 차는 $\bar{x}-\bar{y}=0.9$ 이다.

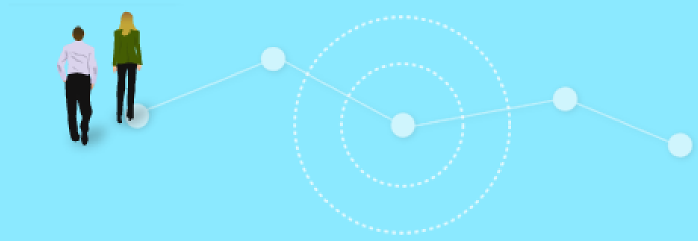
두 모평균 차에 대한 90% 신뢰구간의 오차한계 :

$$e=1.645\times S.E(\bar{X}-\bar{Y})=1.645\times 0.186=0.306$$

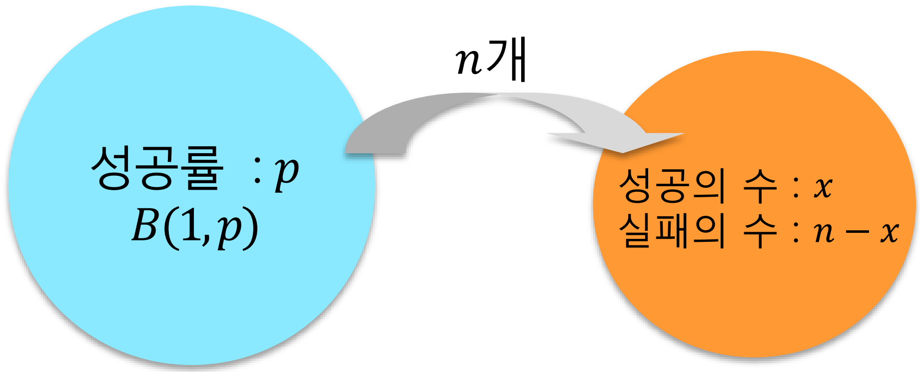
두 모평균 차 $\mu_1 - \mu_2$ 에 대한 90% 신뢰구간 :

$$(0.9-0.306,\quad 0.9+0.306)=(0.594,\quad 1.206)$$

9.3 모비율의 추정



모비율의 신뢰구간



표본비율 : $\hat{P} = \dfrac{x}{n}$

$$E(\hat{p}) = E(X/p) = p$$
$$Var(\hat{p}) = Var(X/p) = pq/n$$

표본의 크기 n이 충분히 크다면,
$np > 5, n(1-p) > 5$이면
중심극한정리에 의하여

표본비율의 확률분포 : $\hat{P} \approx N\left(p,\ \dfrac{pq}{n}\right),\quad Z = \dfrac{\hat{P}-p}{\sqrt{pq/n}} \approx N(0,\ 1)$

모비율의 신뢰구간

❖ **표본의 크기 n이 클수록 표본비율은 모비율에 근사한다. 즉,** $\hat{p} \approx p$ **이므로 다음이 성립한다.**

$$Z = \frac{\hat{P} - p}{\sqrt{\hat{p}\hat{q}/n}} \approx N(0,\ 1)$$

$$P\left(\left|\frac{\hat{P} - p}{\sqrt{\hat{p}\hat{q}/n}}\right| < z_{\alpha/2}\right) = P\left(\left|\hat{P} - p\right| < z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right) \approx 1 - \alpha$$

100 (1 – α)% 신뢰도에서 모수 p에 대한 오차한계

$$P\left(\hat{P} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{P} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right) \approx 1 - \alpha$$

모비율의 신뢰구간

❖ 모비율에 대한 100 (1 – α)% 신뢰구간

$$\left(\hat{P} - z_{\alpha/2}\sqrt{\frac{\hat{p}\,\hat{q}}{n}},\quad \hat{P} + z_{\alpha/2}\sqrt{\frac{\hat{p}\,\hat{q}}{n}}\right)$$

❖ 신뢰도에 따른 오차한계

$|\hat{p}-p|$에 대한 90% 오차한계 : $e_{90\%} = 1.645\sqrt{\frac{\hat{p}\,\hat{q}}{n}}$

$|\hat{p}-p|$에 대한 95% 오차한계 : $e_{95\%} = 1.96\sqrt{\frac{\hat{p}\,\hat{q}}{n}}$

$|\hat{p}-p|$에 대한 99% 오차한계 : $e_{99\%} = 2.58\sqrt{\frac{\hat{p}\,\hat{q}}{n}}$

모비율의 신뢰구간

❖ 모비율에 대한 90%, 95%, 99% 신뢰구간

모수 p에 대한 90% 신뢰구간 : $\left(\hat{p}-1.645\sqrt{\frac{\hat{p}\hat{q}}{n}},\quad \hat{p}+1.645\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$

모수 p에 대한 95% 신뢰구간 : $\left(\hat{p}-1.96\sqrt{\frac{\hat{p}\hat{q}}{n}},\quad \hat{p}+1.96\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$

모수 p에 대한 99% 신뢰구간 : $\left(\hat{p}-2.58\sqrt{\frac{\hat{p}\hat{q}}{n}},\quad \hat{p}+2.58\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$

모비율의 신뢰구간

예제 9-7

어느 기업의 연구소에서 두 종류의 휴대전화용 배터리 A와 B를 개발하였다. 두 종류의 배터리를 시험한 결과가 다음과 같을 때, 이때 두 종류의 배터리의 평균 사용시간 차에 대한 90% 신뢰구간을 구하라.

$n=64$	$\overline{x}=26.5$시간, $\sigma_1=1.3$시간
$m=100$	$\overline{y}=25.6$시간, $\sigma_2=0.9$시간

풀이

500개의 배터리 중에서 3개가 불량이므로 $\hat{p}=\dfrac{3}{500}=0.006, \hat{q}=0.994$ 이고,
표준오차는 다음과 같다.

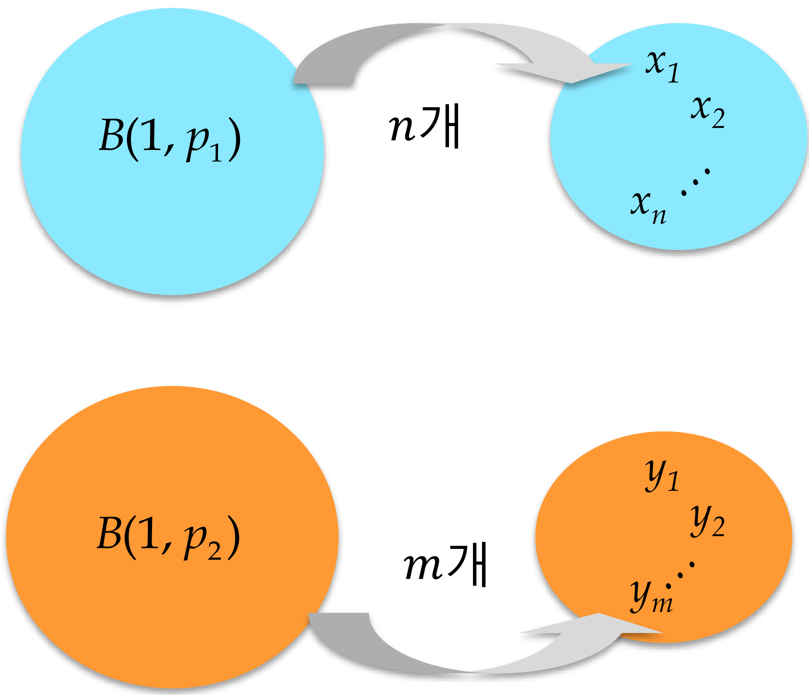
$$\mathrm{S.E}(\hat{P})=\sqrt{\frac{0.006\times 0.994}{500}}=0.003$$

95% 신뢰구간의 오차한계 : $e=1.96\times S.E(\hat{P})=1.96\times 0.003=0.00588$

모비율 p에 대한 95% 신뢰구간 : $(0.006-0.00588,\quad 0.006+0.00588)=(0.00012,\quad 0.01188)$

모비율 차의 신뢰구간

❖ 두 모비율이 p_1, p_2인 모집단에서 각각 크기 n과 m인 표본을 선정하여 표본비율을 각각 $\hat{P}_1, \hat{P}_2$ 라 하자.



표본비율 : $\hat{P}_1 \approx N\left(p_1, \dfrac{p_1 q_1}{n}\right)$

표본비율 : $\hat{P}_2 \approx N\left(p_2, \dfrac{p_2 q_2}{m}\right)$

모비율 차의 신뢰구간

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

❖ 표본의 크기 n이 클수록 표본비율은 모비율에 근사한다. 즉, $\hat{p}_1 \approx p_1, \hat{p}_2 \approx p_2$ 이므로 다음이 성립한다.

$$\frac{(\hat{P}_1 - \hat{P}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n} + \frac{\hat{p}_2 \hat{q}_2}{m}}} \approx N(0,\ 1) \Rightarrow P\left(\left|\frac{(\hat{P}_1 - \hat{P}_2) - (p_1 - p_2)}{\sqrt{(\hat{p}_1 \hat{q}_1 / n) + (\hat{p}_2 \hat{q}_2 / m)}}\right| < z_{\alpha/2}\right) = 1 - \alpha$$

$$P\left((\hat{P}_1 - \hat{P}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n} + \frac{\hat{p}_2 \hat{q}_2}{m}} < p_1 - p_2 < (\hat{P}_1 - \hat{P}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n} + \frac{\hat{p}_2 \hat{q}_2}{m}}\right) = 1 - \alpha$$

모비율 차의 신뢰구간

❖ 두 모비율의 차 $p_1 - p_2$에 대한 100 (1 – α)% 신뢰구간

$$\left((\hat{P}_1 - \hat{P}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n} + \frac{\hat{p}_2\hat{q}_2}{m}},\quad (\hat{P}_1 - \hat{P}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n} + \frac{\hat{p}_2\hat{q}_2}{m}} \right)$$

❖ 신뢰도에 따른 오차한계

$\left|(\hat{P}_1 - \hat{P}_2) - (p_1 - p_2)\right|$ 에 대한 90% 오차한계 : $e_{90\%} = 1.645\sqrt{\frac{\hat{p}_1\hat{q}_1}{n} + \frac{\hat{p}_2\hat{q}_2}{m}}$

$\left|(\hat{P}_1 - \hat{P}_2) - (p_1 - p_2)\right|$ 에 대한 95% 오차한계 : $e_{95\%} = 1.96\sqrt{\frac{\hat{p}_1\hat{q}_1}{n} + \frac{\hat{p}_2\hat{q}_2}{m}}$

$\left|(\hat{P}_1 - \hat{P}_2) - (p_1 - p_2)\right|$ 에 대한 99% 오차한계 : $e_{99\%} = 2.58\sqrt{\frac{\hat{p}_1\hat{q}_1}{n} + \frac{\hat{p}_2\hat{q}_2}{m}}$

모비율 차의 신뢰구간

❖ 모비율 차에 대한 90%, 95%, 99% 신뢰구간

모수 p에 대한 90% 신뢰구간 : $\left((\hat{p}_1 - \hat{p}_2) - e_{90\%},\ \ (\hat{p}_1 - \hat{p}_2) + e_{90\%}\right)$

모수 p에 대한 95% 신뢰구간 : $\left((\hat{p}_1 - \hat{p}_2) - e_{95\%},\ \ (\hat{p}_1 - \hat{p}_2) + e_{95\%}\right)$

모수 p에 대한 99% 신뢰구간 : $\left((\hat{p}_1 - \hat{p}_2) - e_{99\%},\ \ (\hat{p}_1 - \hat{p}_2) + e_{99\%}\right)$

모비율 차의 신뢰구간

예제 9-9

자동차를 생산하는 조립라인 A와 B의 불량률의 차이를 알아보기 위해 조사한 결과가 다음과 같았다. 이때 두 조립라인의 불량률의 차에 대한 99% 신뢰구간을 구하라.

조립라인	조사 대상 자동차 수	불량인 자동차 수
A	250대	16대
B	150대	6대

풀이

조립라인 A와 B에 대한 표본조사 결과는 다음과 같다.

조립라인 A	$n=250,\ \hat{p}_1=\dfrac{16}{250}=0.064,\ \hat{q}_1=0.936$
조립라인 B	$m=150,\ \hat{p}_2=\dfrac{6}{150}=0.04,\ \hat{q}_2=0.96$

따라서 두 조립라인에 대한 불량률의 차에 대한 점추정값은 $\hat{p}_1-\hat{p}_2=0.024$이고 표준오차는 다음과 같다.

모비율 차의 신뢰구간

$$S.E(\hat{P}_1 - \hat{P}_2) = \sqrt{\frac{0.064 \times 0.936}{250} + \frac{0.04 \times 0.96}{150}} = \sqrt{0.000496} = 0.0223$$

99% 신뢰구간의 오차한계 :

$$e = 2.58 \times S.E(\hat{P}_1 - \hat{P}_2) = 2.58 \times 0.0223 = 0.0575$$

모비율 차에 대한 99% 신뢰구간 :

$$(0.024 - 0.0575, \quad 0.024 + 0.0575) = (-0.0335, \quad 0.0815)$$

9.4 표본의 크기 결정

단일 모집단인 경우

❖ 모분산σ^2이 알려진 경우에 100(1 – α)% 신뢰도에 모평균 μ에 오차한계 :

$$e = z_{\alpha/2}\frac{\sigma^2}{\sqrt{n}}$$

모평균에 대한 오차한계를 d이하로 추정하기 위한 표본의 크기 n :

$$z_{\alpha/2}\frac{\sigma^2}{\sqrt{n}} \le d \quad \Rightarrow \quad n \ge \left(z_{\alpha/2}\frac{\sigma}{d} \right)^2$$

모평균 추정을 위한 표본의 크기

예제 9-10

$\sigma = 0.25$ 인 정규모집단의 모평균에 대한 95% 신뢰구간을 얻기 위한 표본의 크기를 구하라. 이때 최대 오차한계는 $d = 0.05$ 이다.

풀이

$\sigma = 0.25$, $d = 0.05$이므로 모평균에 대한 95% 신뢰구간을 얻기 위한 표본의 크기는 다음과 같다.

$$n \geq \left(1.96 \times \frac{0.25}{0.05}\right)^2 = 96.4; \quad n = 97$$

표본의 크기가 동일한 모집단인 경우

❖ 두 모분산 σ_1^2, σ_2^2 이 알려진 경우에 100(1 – α)% 신뢰도에 모평균 차에 오차한계 :

$$e = z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$$

모평균 차에 대한 오차한계를 d이하로 추정하기 위한 표본의 크기 $n = m$:

$$z_{\alpha/2}\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{n}} \le d \quad \Rightarrow \quad n = m \ge \left(\frac{z_{\alpha/2}}{d}\right)^2 \left(\sigma_1^2 + \sigma_2^2\right)$$

표본의 크기가 동일한 모집단인 경우

예제 9-11

동급인의 자동차 A와 B의 1리터당 운행거리는 표준편차가 각각 1.2km와 1.5km인 정규분포를 따른다고 한다. 두 자동차 A와 B의 1리터당 평균 운행거리의 차에 대한 95% 신뢰구간을 구하기 위해 동일한 대수의 자동차 A와 B를 임의로 선정하여 조사하고자 한다. 최대 오차한계를 0.5km 이하라 할 때, 표본조사해야 하는 자동차 수를 구하라.

풀이

$\sigma_1 = 1.2, \sigma_2 = 1.5, d = 0.5$ 이므로 평균 운행거리의 차에 대한 95% 신뢰구간을 얻기 위한 표본의 크기는 다음과 같다.

$$n = m \geq \left(\frac{1.96}{0.5}\right)^2 \left(1.2^2 + 1.5^2\right) = 56.702; \quad n = 57$$

모비율 추정을 위한 표본의 크기

모비율 p에 대한 $100(1-\alpha)\%$ 신뢰구간의 오차한계 :

$$e = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

모비율에 대한 오차한계를 d이하로 추정하기 위한 표본의 크기 n :

$$z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} \le d \quad \Rightarrow \quad n \ge \left(\frac{z_{\alpha/2}}{d}\right)^2 \hat{p}\hat{q}$$

NOTE 표본비율 $\hat{p}$ 은 표본을 선정해야만 알 수 있는 수치이므로 다음과 같은 방법에 의해 근사적으로 표본의 크기를 결정한다.

① 과거 조사결과 p^*를 알고 있는 경우에 $\hat{p} = p^*$ 를 사용한다.
② 과거 조사 결과가 없는 경우에 크기 $n \ge 30$인 표본을 선정하여 예비로 얻은 $\hat{p}$ 를 사용한다.
③ 과거 조사 결과가 없는 경우에 $n \ge \left(\frac{z_{\alpha/2}}{2d}\right)^2$ 을 택한다.

모비율 추정을 위한 표본의 크기

예제 9-12

대통령 후보의 지지도에 대한 최대 오차범위 $\pm 2\%$ 에서 신뢰도 95% 인 신뢰구간을 구하기 위해 조사해야 하는 유권자 수를 구하라.

풀이

$d = 0.02$이고 $z_{0.05} = 1.96$이므로 조사해야 할 유권자의 수는 다음과 같다.

$$n \geq \left(\frac{1.96}{2 \times 0.02} \right)^2 = 49^2 = 2401; \quad n = 2401$$

Q&A