

제8장 단일 모집단의 추정

제1절 통계적 추론[Statistical Estimation]

표본의 특성을 분석하여 모집단의 특성을 추론 ... 통계량을 통해 모수를 추정

추정량(estimator)과 추정치(estimate)

추정량: 특정 모수를 추정하기 위해 사용하는 추정 방법 - 모수 μ 를 추정하기 위해 통계량 \bar{X} 를 사용

추정치: 모수를 구체적으로 추정한 값

점추정(Point Estimation)과 구간추정(Interval Estimation)

점추정 : 측정치를 단일 수치로 나타내는 것

구간추정 : 측정치를 구간으로 나타내는 것

제2절 점추정(Point Estimation)

1. 점추정의 의미

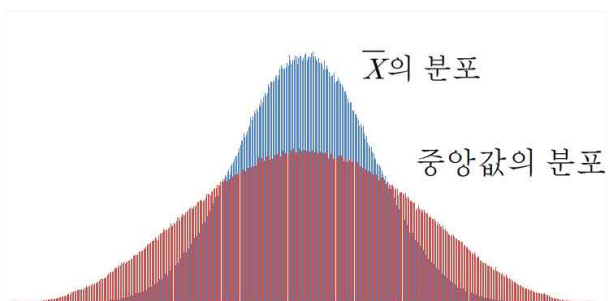
좋은 추정량의 조건

θ 는 모수, $\hat{\theta}$ 는 추정량을 의미한다.

[1] 불편성 (unbiasedness) $E(\hat{\theta}) = \theta$

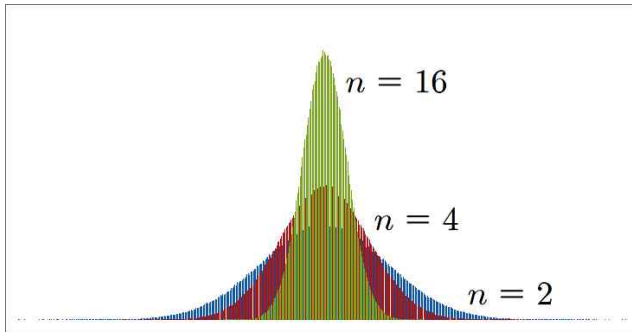
$$\mu_{\bar{X}} = E(\bar{X}) = \mu$$

[2] 효율성 (efficiency) 추정량 중에서 최소의 분산을 가진 추정량이 가장 효율적



관찰값: rand, 표본의 크기: 7, 표본의 수: 백만,

[3] 일치성(consistency) $\lim_{n \rightarrow \infty} E(\hat{\theta} - \theta)^2 = 0$



[4] 충족성(sufficiency) 추정량이 모수에 대하여 가장 많은 정보를 제공할 때 그 추정량을 충족추정량이라고 한다.

2. 모평균의 점추정

\bar{X} 는 μ 의 불편추정량

$$\begin{aligned} E(\bar{X}) &= E\left\{\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right\} \\ &= \frac{1}{n}\{E(X_1) + E(X_2) + \cdots + E(X_n)\} \\ &= \frac{1}{n} \times n\mu \\ &= \mu \end{aligned}$$

$n=5, X \sim U(0, 1)$

	X_1	X_2	X_3	X_4	X_5	\bar{X}_i
실험1	0.8938	0.0128	0.7009	0.8780	0.6236	0.6218
실험2	0.3123	0.9183	0.5234	0.9910	0.5035	0.6497
실험3	0.4896	0.8084	0.9932	0.8072	0.4337	0.7064
실험4	0.7939	0.3645	0.3456	0.1190	0.7795	0.4805
실험5	0.3611	0.4867	0.6715	0.7224	0.8170	0.6117
...
실험10,000	0.3736	0.1544	0.8509	0.5284	0.7115	0.5238
평균	0.5000	0.4974	0.5008	0.4968	0.5004	0.4991
	$E(X_1)=\mu$	$E(X_2)=\mu$	$E(X_3)=\mu$	$E(X_4)=\mu$	$E(X_5)=\mu$	
	$E(\bar{X}) = \frac{1}{n} \times n\mu = \mu$					

	X_1	X_2	...	X_5	\bar{X}_i
실험1	=RAND()	=RAND()	...	=RAND()	=AVERAGE(A1:E1)
실험2	=RAND()	=RAND()	...	=RAND()	=AVERAGE(A2:E2)
...
실험10,000	=RAND()	=RAND()	...	=RAND()	=AVERAGE(A10000:E10000)
평균	=AVERAGE(A1:A10000)	=AVERAGE(B1:B10000)	...	=AVERAGE(E1:E10000)	=AVERAGE(F1:F10000)

3. 모분산의 점추정

S^2 는 σ^2 의 불편추정량

$$\begin{aligned}
 V(X_i) &= E[(X_i - \mu)^2] \\
 &= E(X_i^2 - 2X_i \cdot \mu + \mu^2) \\
 &= E(X_i^2) - 2E(X_i) \cdot \mu + \mu^2 \\
 &= E(X_i^2) - \mu^2 \quad \text{since } E(X_i) = \mu
 \end{aligned}$$

$$\begin{aligned}
 V(\bar{X}) &= E[(\bar{X} - \mu_{\bar{X}})^2] \\
 &= E[(\bar{X} - \mu)^2] \quad \text{since } \mu_{\bar{X}} = \mu \\
 &= E(\bar{X}^2) - \mu^2
 \end{aligned}$$

$$\begin{aligned}
 E(S^2) &= E\left[\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}\right] \\
 &= \frac{1}{n-1} E\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] \\
 &= \frac{1}{n-1} E\left[\sum_{i=1}^n (X_i^2 - 2X_i \cdot \bar{X} + \bar{X}^2)\right] \\
 &= \frac{1}{n-1} \left\{ E\left[\sum_{i=1}^n X_i^2\right] - 2E\left[\sum_{i=1}^n X_i \cdot \bar{X}\right] + nE(\bar{X}^2) \right\} \\
 &= \frac{1}{n-1} \left\{ \sum_{i=1}^n E(X_i^2) - 2nE(\bar{X}^2) + nE(\bar{X}^2) \right\} \quad \text{since } \sum_{i=1}^n X_i = n\bar{X} \\
 &= \frac{1}{n-1} \left\{ \sum_{i=1}^n E(X_i^2) - nE(\bar{X}^2) \right\} \\
 &= \frac{1}{n-1} \left\{ \left[\sum_{i=1}^n V(X_i) + \mu^2 \right] - n[V(\bar{X}) + \mu^2] \right\} \quad \text{since } V(X_i) = E(X_i^2) - \mu^2, \\
 &\quad V(\bar{X}) = E(\bar{X}^2) - \mu^2
 \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{n-1} \left\{ n(\sigma^2 + \mu^2) - n \left(\frac{\sigma^2}{n} + \mu^2 \right) \right\} && \text{since } V(X_i) = \sigma^2, \quad V(\bar{X}) = \frac{\sigma^2}{n} \\
 &= \frac{1}{n-1} [(n-1) \cdot \sigma^2] = \sigma^2
 \end{aligned}$$

제3절 신뢰구간추정

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

문제 1. μ 의 추정치로 x 를 사용한다면, $P(x=\mu)$ 는?

문제 2. μ 의 추정구간으로 $(x - \frac{b-a}{10}, x + \frac{b-a}{10})$ 을 설정했다면, 이 구간에 μ 가 포함될 확률,

$$P\left(x - \frac{b-a}{10} < \mu < x + \frac{b-a}{10}\right) \text{은?}$$

문제 3. μ 의 추정구간으로 $(x - \frac{b-a}{2}, x + \frac{b-a}{2})$ 을 설정했다면, 이 구간에 μ 가 포함될 확률,

$$P\left(x - \frac{b-a}{2} < \mu < x + \frac{b-a}{2}\right) \text{은?}$$

문제 4. μ 의 추정구간으로 $(-\infty, \infty)$ 을 설정했다면, 이 구간에 μ 가 포함될 확률, $P(-\infty < \mu < \infty)$ 은?

1. 신뢰구간의 의미

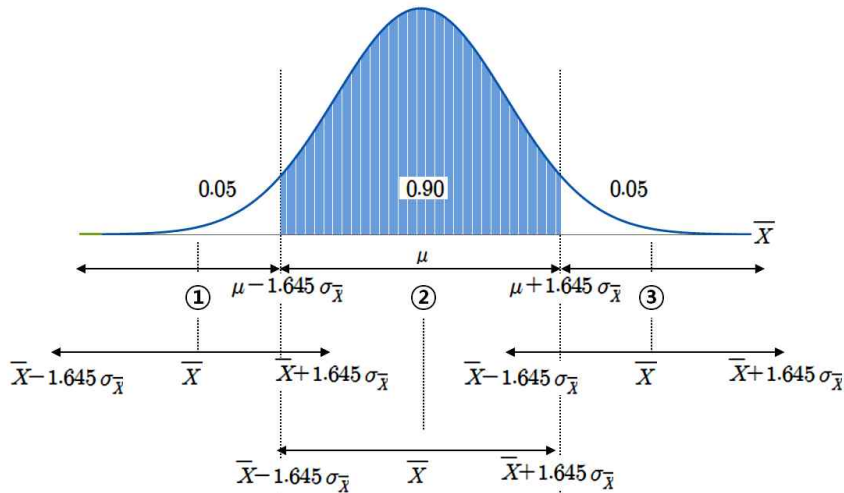
신뢰구간이란 구간 추정에 있어 일정한 확률범위 내에 모수가 포함될 가능성이 있는 구간이다.

2. 모평균의 신뢰구간

\bar{X} 의 표본분포와 90% 신뢰구간

임의로 표본을 구성하여 표본의 평균, \bar{X} 을 구하고 신뢰구간($\bar{X} \pm 1.645 \sigma_{\bar{X}}$)을 설정했을 때, 신뢰구간이 μ 를 포함하고 있을 확률은? ($\mu_{\bar{X}} = \mu_X = \mu$ 이다.)

이 확률을 수식으로 표현하면 $P(\bar{X} - 1.645 \sigma_{\bar{X}} < \mu < \bar{X} + 1.645 \sigma_{\bar{X}})$



구간	\bar{X} 가 해당 구간에서 관찰될 확률	$\bar{X} \pm 1.645 \sigma_{\bar{X}}$ 구간에 μ 포함 여부
$-\infty \sim \mu - 1.645 \sigma_{\bar{X}}$	$P(\bar{X} < \mu - 1.645 \sigma_{\bar{X}}) = 0.05$	포함되지 않는다.
$\mu - 1.645 \sigma_{\bar{X}} \sim \mu + 1.645 \sigma_{\bar{X}}$	$P(\mu - 1.645 \sigma_{\bar{X}} < \bar{X} < \mu + 1.645 \sigma_{\bar{X}}) = 0.90$	포함된다.
$\mu + 1.645 \sigma_{\bar{X}} \sim \infty$	$P(\bar{X} > \mu + 1.645 \sigma_{\bar{X}}) = 0.05$	포함되지 않는다.

$\bar{X} < \mu - 1.645 \sigma_{\bar{X}}$ 이면, $\bar{X} \pm 1.645 \sigma_{\bar{X}}$ 은 μ 를 포함하지 않는다.

$$\rightarrow P(\bar{X} < \mu - 1.645 \sigma_{\bar{X}}) = 0.05 \quad 1)$$

1) $P(\bar{X} < \mu - 1.645 \sigma_{\bar{X}})$ 값을 구하는 방법 (제 6장 연속확률분포 중 표준정규분포)

컴퓨터를 사용하면 이 값을 직접 구할 수 있지만 확률표를 사용하여 값을 구한다면,

① \bar{X} 를 z 로 변환한다.

$$P(\bar{X} < \mu - 1.645 \sigma_{\bar{X}})$$

$$= P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{\mu - 1.645 \sigma_{\bar{X}} - \mu}{\sigma/\sqrt{n}}\right) \quad \text{양변에서 } \mu \text{를 빼고 다시 } \sigma/\sqrt{n} \text{으로 나눈다.}$$

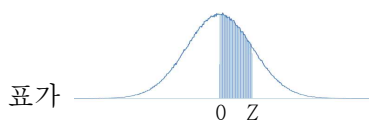
($\sigma/\sqrt{n} = \sigma_{\bar{X}}$ 임을 상기한다.)

$$= P(z < -1.645)$$

좌측 $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ 는 z 가 되고,

우측은 μ , σ , n 에 값을 대입하여 정리하면 -1.645가 된다.

② 확률분포표를 사용하여 확률값을 구한다.



표가 z 방식으로 주어진 경우 $P(z < -1.645)$ 를 구하는 방법

z 의 분포는 0을 중심으로 좌우대칭이고 확률분포표는 $z > 0$ 인 구간에 대해 값을 주었으므로, $P(z < -1.645)$ 와 같이 z 가 음수 영역에 있는 경우는 그에 상응하는 양수 영역의 값을

$\mu - 1.645 \sigma_{\bar{X}} < \bar{X} < \mu + 1.645 \sigma_{\bar{X}}$ 이면, $\bar{X} \pm 1.645 \sigma_{\bar{X}}$ 은 μ 를 포함한다.

$$\rightarrow P(\mu - 1.645 \sigma_{\bar{X}} < \bar{X} < \mu + 1.645 \sigma_{\bar{X}}) = 0.90$$

$\bar{X} > \mu + 1.645 \sigma_{\bar{X}}$ 이면, $\bar{X} \pm 1.645 \sigma_{\bar{X}}$ 은 μ 를 포함하지 않는다.

$$\rightarrow P(\bar{X} > \mu + 1.645 \sigma_{\bar{X}}) = 0.05$$

$$\Rightarrow \bar{X} \pm 1.645 \sigma_{\bar{X}} \text{ 구간에 } \mu \text{ 포함될 확률} = 0.90$$

결론적으로 아래 등식이 성립한다.

$$P(\bar{X} - 1.645 \sigma_{\bar{X}} < \mu < \bar{X} + 1.645 \sigma_{\bar{X}}) = P(\mu - 1.645 \sigma_{\bar{X}} < \bar{X} < \mu + 1.645 \sigma_{\bar{X}})$$

구한다.

즉, 0을 중심으로 $z < -1.645$ 에 대칭되는 구간은 $z > 1.645$ 이며, $P(z < -1.645)$ 값은 $P(z > 1.645)$ 의 값과 동일하다.

	0.03	0.04	0.05	0.06
1.5	0.4370	0.4382	0.4394	0.4406
1.6	0.4484	0.4495	0.4505	0.4515
1.7	0.4582	0.4591	0.4599	0.4608

표에서 z 의 소숫점 첫째 자리의 값은 첫번째 행에서 찾고, z 의 두 번째 자리 값은 첫 번째 열에서 찾는다.

표에서 $P(0 < z < 1.64) = 0.4495$ 이고, $P(0 < z < 1.65) = 0.4505$ 이므로, $P(0 < z < 1.645) = 0.45$ 로 추정할 수 있다. 종합하면,

$$P(z < -1.645) = P(z > 1.645) = P(z > 0.5) - P(0 < z < 1.645) = 0.5 - 0.45 = 0.05$$

예: $P(\bar{X} - 1.28 \sigma_{\bar{X}} < \mu < \bar{X} + 1.28 \sigma_{\bar{X}}) = P(\mu - 1.28 \sigma_{\bar{X}} < \bar{X} < \mu + 1.28 \sigma_{\bar{X}})$ 실험

$$x \sim N(100, 10^2) \rightarrow \bar{x} \sim N(100, 5^2) \leftarrow \mu_{\bar{x}} = \mu_x, \sigma_{\bar{x}} = \sigma_x / \sqrt{n}$$

	\bar{x}	μ	$\bar{x} - 1.28 \sigma_{\bar{x}}$	$\bar{x} + 1.28 \sigma_{\bar{x}}$	μ 포함여부	$\mu - 1.28 \sigma_{\bar{x}}$	$\mu + 1.28 \sigma_{\bar{x}}$	\bar{x} 포함여부
1	97.26	100	90.86	103.66	1	93.60	106.40	1
2	99.99	100	93.59	106.39	1	93.60	106.40	1
3	91.40	100	85.00	97.80	0	93.60	106.40	0
4	93.01	100	86.61	99.41	0	93.60	106.40	0
5	101.62	100	95.22	108.02	1	93.60	106.40	1
6	94.26	100	87.86	100.66	1	93.60	106.40	1
7	107.51	100	101.11	113.91	0	93.60	106.40	0
8	102.76	100	96.36	109.16	1	93.60	106.40	1
9	100.54	100	94.14	106.94	1	93.60	106.40	1
10	91.42	100	85.02	97.82	0	93.60	106.40	0
...								
10,000	98.25	100	91.85	104.65	1	93.60	106.40	1
평균					0.7978			0.7978

예제. 확률변수 $X \sim N(\mu, \sigma^2)$ \bar{X} 에 대한 확률이면 μ 에 대한 확률로, μ 에 대한 확률이면 \bar{X} 에 대한 확률로 변환하시오.

문제 1. $P(\bar{X} < \mu - 1.28 \sigma_{\bar{X}})$

Ans. $P(\mu > \bar{X} + 1.28 \sigma_{\bar{X}})$

문제 2. $P(\bar{X} > \mu + 1.96 \sigma_{\bar{X}})$

Ans. $P(\mu < \bar{X} - 1.96 \sigma_{\bar{X}})$

문제 3. $P(\bar{X} - 1.645 \sigma_{\bar{X}} < \mu < \bar{X} + 1.645 \sigma_{\bar{X}})$

Ans. $P(\mu - 1.645 \sigma_{\bar{X}} < \bar{X} < \mu + 1.645 \sigma_{\bar{X}})$

연습문제 2. $X \sim N(100, 10^2)$, n (표본의 크기) = 25

문제 1. $\mu_{\bar{X}}$ 를 μ_x 로 표현하시오.

문제 2. $\mu_{\bar{X}}$ 를 계산하시오.

문제 3. $\sigma_{\bar{X}}$ 를 σ_x 와 n 으로 표현하시오.

문제 4. $\sigma_{\bar{X}}$ 를 계산하시오.

문제 5. \bar{X} 는 어떤 분포를 따르는가?

문제 6. $P(\bar{X} < 96.08)$

문제 7. $P(96.08 < \bar{X} < 103.92)$

문제 8. $P(103.92 < \bar{X})$

문제 9. $\bar{X} < 96.08$ 이 사실이면, $(\bar{X}-3.92, \bar{X}+3.92)$ 구간에 μ 가 포함되는가?

문제 10. $96.08 < \bar{X} < 103.92$ 이 사실이면, $(\bar{X}-3.92, \bar{X}+3.92)$ 구간에 μ 가 포함되는가?

문제 11. $103.92 < \bar{X}$ 이 사실이면, $(\bar{X}-3.92, \bar{X}+3.92)$ 구간에 μ 가 포함되는가?

문제 12. $(\bar{X}-3.92, \bar{X}+3.92)$ 구간에 μ 가 포함될 확률은 얼마인가?

연습문제 3. $X \sim N(10, 4^2)$, n (표본의 크기) = 4

문제 1. \bar{X} 는 어떤 분포를 따르는가?

문제 2. $P(\bar{X} < 7.44)$

문제 3. $P(7.44 < \bar{X} < 12.56)$

문제 4. $P(12.56 < \bar{X})$

문제 5. $\bar{X} < 7.44$ 가 사실이면, $(\bar{X}-2.56, \bar{X}+2.56)$ 구간에 μ 가 포함되는가?

문제 6. $7.44 < \bar{X} < 12.56$ 이 사실이면, $(\bar{X}-2.56, \bar{X}+2.56)$ 구간에 μ 가 포함되는가?

문제 7. $12.56 < \bar{X}$ 가 사실이면, $(\bar{X}-2.56, \bar{X}+2.56)$ 구간에 μ 가 포함되는가?

문제 8. $(\bar{X}-2.56, \bar{X}+2.56)$ 구간에 μ 가 포함될 확률은 얼마인가?

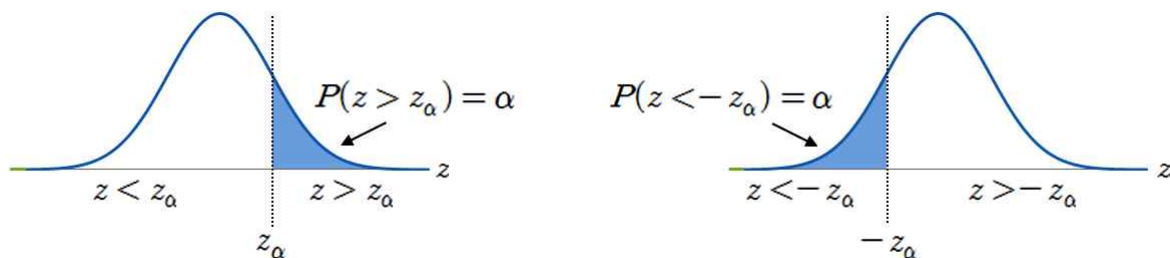
z_α 와 $-z_\alpha$ 개념 ($0 \leq \alpha \leq 1$)

$z \sim N(0, 1^2)$

z 는 확률변수이고, z_α 와 $-z_\alpha$ 는 다음을 만족시키는 특정한 값이다.

$P(z > z_\alpha) = \alpha \rightarrow z_\alpha$: 상위 α ($\alpha \cdot 100\%$)에 해당하는 z 의 특정한 값

$P(z < -z_\alpha) = \alpha \rightarrow -z_\alpha$: 하위 α ($\alpha \cdot 100\%$)에 해당하는 z 의 특정한 값



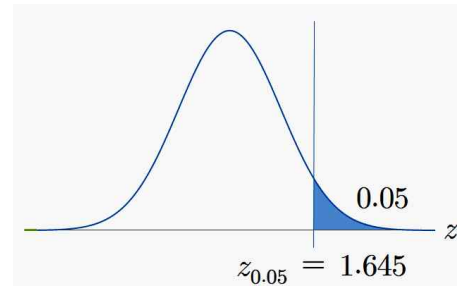
(1) z 는 정규분포(평균 0, 분산 1)를 따르는 확률변수이고, z_α 는 z 값이므로

$-\infty < z_\alpha < +\infty$ 이다. (z_α 는 -값도 가능하다.)

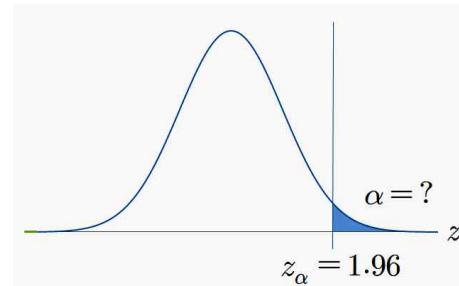
(2) α 는 확률이므로 $0 \leq \alpha \leq 1$ 이다. (α 는 -값이 불가능하다.)

z_α 의 값 구하기

(1) $z_{0.05} = 1.645$ (상위 5%에 해당하는 z 값)



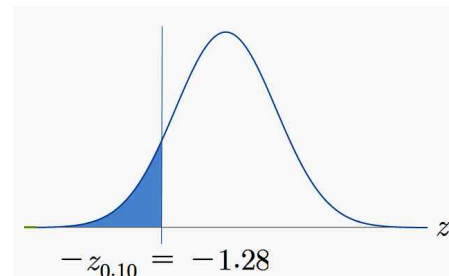
(2) $z_\alpha = 1.96$ 을 만족시키는 $\alpha = ?$



$P(z > 1.96) = 0.025$ 이므로, $\alpha = 0.025$

→ 1.96은 z 값 중 상위 몇 %에 속하는 가를 묻는 것

(3) $-z_{0.10} = ?$



z 값이 음수인 경우에는 $P(z \leftarrow z_\alpha) = \alpha$ 이다.

→ 하위 10%에 해당하는 z 값을 묻는 것

→ 하위 10%에 해당하는 z 값은 상위 10%에 해당하는 z 값에 -을 붙인 값과 같다.

→ 이러한 특성은 z 가 0을 중심으로 좌우대칭이기 때문이다.

→ 상위 10%에 해당하는 $z = 1.28$ 이므로, 하위 10%에 해당하는 $z = -1.28$

연습문제 4. 정규분포

문제 1. 상위 5%에 해당하는 z 값은?

문제 2. $x \sim N(\mu, \sigma^2)$. 상위 5%에 해당하는 x 값은?

문제 3. $x \sim N(\mu, \sigma^2)$. 상위 5%에 해당하는 x 값은 평균보다 몇 표준편차 큰 가?

문제 4. $x \sim N(100, 5^2)$. 상위 5%에 해당하는 x 값은?

문제 5. $x \sim N(100, 5^2)$. 하위 5%에 해당하는 x 값은?

문제 6. $x \sim N(100, 5^2)$. 110은 상위 %에 해당하는 가?

문제 7. 상위 2.228%에 해당하는 z 값은?

문제 8. 상위 2.228%에 해당하는 x 는 평균보다 몇 표준편차 큰 값인가?

문제 9. 하위 2.228%에 해당하는 x 는 평균보다 몇 표준편차 큰 값인가?

연습문제 5. 상위 5%에 해당하는 값을 구하시오.

문제 1. $x \sim U(0, 1)$

문제 2. $x \sim U(10, 20)$

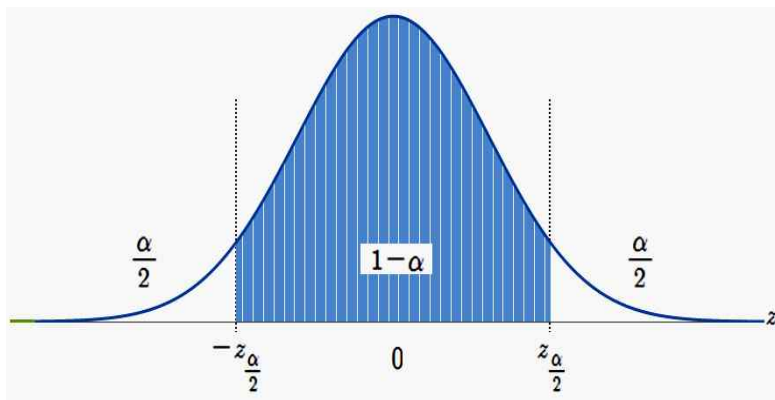
문제 3. $x \sim N(\mu, \sigma^2)$. 상위 5%에 해당하는 x 값은? Ans.

Note: 상위 몇 %에 해당하는 x 값을 구할 때 항상 z_α 값을 사용하는 것은 아니다.

문제 1과 2에서 보았듯이, 일양분포의 경우는 z_α 를 사용하지 않는다.

z_α 값은 x 가 정규분포를 따를 때만 사용한다. (정규분포가 아니면 사용할 수 없다.)

[1] 모집단 분산 σ^2 를 알고 있는 경우



① X 가 정규분포를 따르면, \bar{X} 도 정규분포를 따른다.

② σ 값이 알려져 있다면,

만일 σ_X 값이 알려져 있다면 $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ 이므로, 아래와 같이 신뢰구간을 구할 수 있다.

(표본에서 \bar{X} 값은 구할 수 있으나, $\sigma_{\bar{X}}$ 값은 구할 수 없다.)

100(1- α)% 신뢰수준에서 μ 에 대한 신뢰구간

$$\left[\bar{X} - z_{\frac{\alpha}{2}} \cdot \sigma_{\bar{X}}, \bar{X} + z_{\frac{\alpha}{2}} \cdot \sigma_{\bar{X}} \right] = \left[\bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right]$$

예제: $X \sim N(\mu, 10^2)$, $n = 30$, $\bar{X} = 150$; μ 에 대한 90% 신뢰구간

$$\bar{X} \pm z_{\frac{1-0.90}{2}} \cdot \sigma / \sqrt{n} = 150 \pm z_{0.05} \cdot 10 / \sqrt{30} = 150 \pm (1.65) \cdot (1.83)$$

연습문제 6.

- 문제 1. x 가 정규분포를 따르고 n 이 작으면, \bar{x} 는 정규분포를 항상 따른다.
 문제 2. x 가 정규분포를 따르고 n 이 크면, \bar{x} 는 정규분포를 항상 따른다.
 문제 3. x 가 정규분포를 따르지 않고 n 이 작으면, \bar{x} 는 항상 정규분포에 가깝다.
 문제 4. x 가 정규분포를 따르지 않고 n 이 크면, \bar{x} 는 항상 정규분포에 가깝다.
 문제 5. μ_x 와 σ_x^2 을 μ_x , σ_x^2 , n 으로 표현하시오.
 문제 6. $x \sim N(\mu_x, \sigma_x^2)$ 이다. $\bar{x} \sim N(\bar{x}$ 의 평균, \bar{x} 의 분산)을 μ_x , σ_x^2 , n 으로 표현하시오.

연습문제 7. $x \sim N(\mu, 10^2)$, 25개의 데이터를 조사한 결과, $\bar{x} = 35$ 이다.

- 문제 1. μ 에 대한 95% 신뢰구간
 문제 2. μ 에 대한 80% 신뢰구간

[2] 모집단 분산 σ^2 를 모르는 경우**[가] 대표본인 경우**

- ① n (표본의 크기)이 충분히 크다면, X 의 분포에 관계없이 \bar{X} 는 정규분포를 따른다.
 ② n (표본의 크기)이 충분히 크다면, σ 값이 알려져 있지 않아도 S 를 사용하여 추정한다.
 (표본에서 \bar{X} 값은 구할 수 있으나, $\sigma_{\bar{X}}$ 값은 구할 수 없다.)

만일 σ_X 값이 알려져 있지 않다면 $\frac{\sigma}{\sqrt{n}}$ 로 $\sigma_{\bar{X}}$ 값을 구하는 것은 불가능하다.

그렇지만 표본에서 구한 $S(=S_X)$ 값은 표본의 크기가 증가할수록 S 은 σ 에 접근하므로 σ 대신 S 를 사용하여 $\sigma_{\bar{X}}$ 를 추정한다. (좋은 추정량의 세 번째 조건은 일치성으로서 $\lim_{n \rightarrow \infty} E(\hat{\theta} - \theta)^2 = 0$ 이다. 이 조건을 σ 와 S 에 적용하면, $\lim_{n \rightarrow \infty} E(S - \sigma)^2 = 0$ 이다.)

$$\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} \leftarrow \frac{S_X}{\sqrt{n}} \quad (n \text{이 충분히 큰 경우 } S \text{로 } \sigma \text{를 추정})$$

100(1- α)% 신뢰수준에서 μ 에 대한 신뢰구간

$$\left[\bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}} \right]$$

예제. $n = 50$, $\bar{X} = 2,000$, $S = 100$; μ 에 대한 95% 신뢰구간

$$\bar{X} \pm z_{\frac{1-0.95}{2}} \cdot S / \sqrt{n} = 2,000 \pm z_{0.025} \cdot 100 / \sqrt{50} = 2,000 \pm (1.96)(1.4142)$$

연습문제 8.

문제 1. n 이 충분히 클 때, σ 의 추정치로 s 를 사용하는 것은 추정량의 조건 중 _____과 관련있다.

문제 2. σ_x 는 단일한 값이지만, s_x 는 표본마다 그 값이 달라질 수 있다. T, F

문제 3. σ 값이 알려져 있어도 n 이 충분히 크면 $\sigma_{\bar{x}}$ 를 구할 때 $\frac{s}{\sqrt{n}}$ 으로 한다. T, F

문제 4. μ 에 대한 $100(1-\alpha)\%$ 신뢰구간을 아래와 같은 공식으로 구할 조건을 모두 나열하시오.

$$\left[\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \right]$$

연습문제 9. $x \sim N(\mu, \sigma^2)$, 64개의 데이터를 조사한 결과, $\bar{x} = 35$ 이고, $s_x = 16$ 이다.

문제 1. μ 에 대한 95% 신뢰구간

문제 2. μ 에 대한 90% 신뢰구간

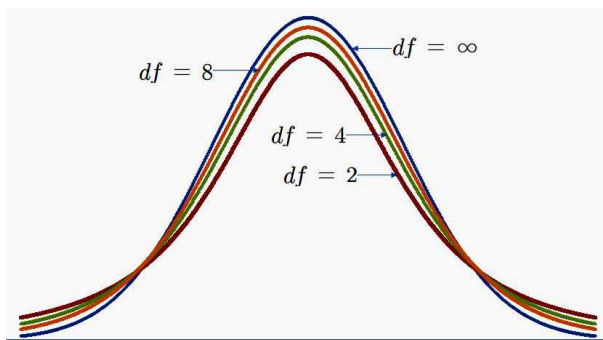
문제 3. μ 에 대한 80% 신뢰구간

[나] 소표본인 경우

① X 가 정규분포를 따른다. (\bar{X} 도 정규분포를 따르게 된다.)

② σ 가 알려져 있지 않고, n 이 충분히 크지 않다.

$$t_{n-1} = \frac{\bar{X} - \mu_{\bar{X}}}{S_{\bar{X}}} = \frac{\bar{X} - \mu_{\bar{X}}}{S / \sqrt{n}}$$

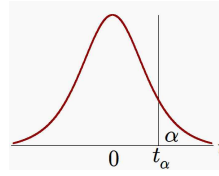


① 정규분포와 유사한 모양을 가진다. 평균 중심으로 좌우대칭이다.

② 자유도에 따라 분포모양이 결정된다.

③ 자유도가 커질수록 정규분포에 접근한다. (자유도가 $+\infty$ 이면, 정규분포가 된다.)

t-분포표 읽기



df	α			
	0.250	0.100	0.050	0.025
1	1.0000	3.0777	6.3138	12.7062
2	0.8165	1.8856	2.9200	4.3027
3	0.7649	1.6377	2.3534	3.1824
4	0.7407	1.5332	2.1318	2.7764
5	0.7267	1.4759	2.0150	2.5706
6	0.7176	1.4398	1.9432	2.4469
7	0.7111	1.4149	1.8946	2.3646
9	0.7064	1.3968	1.8595	2.3060
30	0.6828	1.3104	1.6973	2.0423
100,000	0.6745	1.2816	1.6449	1.9600

$$\leftarrow t_{df=2, \alpha=0.025} = 4.3027$$

 σ^2 을 모를 때 표본의 신뢰구간

$$\left[\bar{X} - t_{\frac{\alpha}{2}, n-1} \cdot \frac{S}{\sqrt{n}}, \bar{X} + t_{\frac{\alpha}{2}, n-1} \cdot \frac{S}{\sqrt{n}} \right]$$

예제. $n = 7$ 인 샘플 $\{400, 415, 418, 409, 382, 391, 406\}$ 모집단은 정규분포를 따를 때 μ 의 95% 신뢰구간은?

$$\bar{X} = 403, S = 12.9615$$

$$\bar{X} \pm t_{\frac{\alpha}{2}, n-1} \cdot \frac{S}{\sqrt{n}} = 403 \pm t_{0.025, 6} \cdot 12.9615 / \sqrt{7} = 403 \pm 2.4469 \cdot 4.8990$$

연습문제 10. $x \sim N(\mu, \sigma^2)$

문제 1. $\{5, 5, 10, 9, 7, 13, 12, 11\}$ μ 에 대한 95% 신뢰구간은?

문제 2. $\{14, 5, 7, 10, 14, 13\}$ μ 에 대한 90% 신뢰구간은?

문제 3. μ 에 대한 $100(1-\alpha)\%$ 신뢰구간을 아래와 같은 공식으로 구할 조건을 모두 나열하시오.

$$\left[\bar{x} - t_{\frac{\alpha}{2}, n-1} \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_{\frac{\alpha}{2}, n-1} \cdot \frac{s}{\sqrt{n}} \right]$$

연습문제 11.

문제 1. $P(\bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}})$ 와 $P(\bar{X} - t_{\frac{\alpha}{2}, n-1} \cdot \frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{\frac{\alpha}{2}, n-1} \cdot \frac{s}{\sqrt{n}})$ 의 비교하시오.

신뢰구간 비교

신뢰구간	필요한 가정	비고
$\bar{X} \pm z_{\frac{\alpha}{2}} \cdot \sigma_{\bar{X}}$	\bar{X} 는 정규분포를 따른다.	X 가 정규분포를 따르거나 n 이 충분히 크다. (중심극한정리)
	σ 가 알려져 있다.	σ 는 표본으로부터 구할 수 없다.
$\bar{X} \pm t_{\frac{\alpha}{2}, n-1} \cdot s_{\bar{X}}$	\bar{X} 는 정규분포를 따른다.	X 가 정규분포를 따르거나 n 이 충분히 크다. (중심극한정리)
	σ 를 모른다.	σ 를 표본에서 구하는 s 로 추정한다.

$\bar{X} \pm t_{\frac{\alpha}{2}, n-1} \cdot s_{\bar{X}}$ 의 경우, n 이 충분히 크면 $t_{\frac{\alpha}{2}, n-1}$ 이 $z_{\frac{\alpha}{2}}$ 에 근접하므로 $\bar{X} \pm z_{\frac{\alpha}{2}} \cdot \sigma_{\bar{X}}$ 로 대체하기도 한다.

값들의 결정	상수, 변수, 계수	비고
알고 싶은 값	μ	μ 는 상수로서 모집단 전체를 알 때만 구할 수 있다. 모집단이 아닌 표본만 가지고 있다면, μ 를 확정적으로 알 수는 없고 추정만 가능하다.
의사결정자가 결정하는 값	α	α 는 의사결정자가 선택하는 계수이다.
자동 결정되는 값	$z_{\frac{\alpha}{2}}, t_{\frac{\alpha}{2}, n-1}$	$z_{\frac{\alpha}{2}}$ 는 α 에 의해, $t_{\frac{\alpha}{2}, n-1}$ 는 α 와 n 에 의해 자동 결정된다.
표본으로부터 구하는 값	\bar{X}, s	\bar{X} 와 s 는 표본에서 구하는 값으로, 변수이다.
알려져 있는 값 (표본에서 구할 수 없는 값)	σ	σ 는 모집단으로부터 구할 수 있는 상수이다. (표본에서는 구할 수 없다.)

연습문제 정답

1. (1) 0

Note: x 는 연속확률변수이다.

μ 에 일치하는 x 가 발생할 확률은 0이다.

→ μ 의 추정치로 x 를 사용한다면, 그 추정치가 μ 와 일치 확률은 0이다.

예: $x \sim U[0, 1] \rightarrow \mu = \frac{a+b}{2} = 0.5$

x	μ	μ 추정치	일치 여부
0.00806	0.50000	0.00806	불일치
0.62351	0.50000	0.62351	불일치
0.52600	0.50000	0.52600	불일치
0.39788	0.50000	0.39788	불일치
0.94201	0.50000	0.94201	불일치
0.01856	0.50000	0.01856	불일치
0.45828	0.50000	0.45828	불일치

(2) 0.2

예: 이 확률은 $P\left(\mu - \frac{b-a}{10} < x < \mu + \frac{b-a}{10}\right)$ 와 같다.

① x 가 $\left(\mu - \frac{b-a}{10}, \mu + \frac{b-a}{10}\right)$ 구간에 속해 있으면, μ 는 $\left(x - \frac{b-a}{10}, x + \frac{b-a}{10}\right)$ 구간에 속한다.

$$\begin{aligned} \mu - \frac{b-a}{10} < x < \mu + \frac{b-a}{10} &\rightarrow \mu - \frac{b-a}{10} - x - \mu < x - x - \mu < \mu + \frac{b-a}{10} - x - \mu \\ &\rightarrow x - \frac{b-a}{10} < \mu < x + \frac{b-a}{10} \end{aligned}$$

② x 가 $\mu - \frac{b-a}{10}$ 보다 작으면, $x < \mu - \frac{b-a}{10} \rightarrow \mu > x + \frac{b-a}{10}$

x 가 $\mu + \frac{b-a}{10}$ 보다 크면, $x > \mu + \frac{b-a}{10} \rightarrow \mu < x - \frac{b-a}{10}$

둘을 결합하면, x 가 $\left(\mu - \frac{b-a}{10}, \mu + \frac{b-a}{10}\right)$ 구간에 속하지 않으면, μ 는 $\left(x - \frac{b-a}{10}, x + \frac{b-a}{10}\right)$ 구간에 속하지 않는다.

①과 ②로부터 $P\left(x - \frac{b-a}{10} < \mu < x + \frac{b-a}{10}\right) = P\left(\mu - \frac{b-a}{10} < x < \mu + \frac{b-a}{10}\right)$ 이 성립

예: $P\left(x - \frac{b-a}{10} < \mu < x + \frac{b-a}{10}\right) = P\left(\mu - \frac{b-a}{10} < x < \mu + \frac{b-a}{10}\right)$ 에 관한 실험

$x \sim U[0, 20]$

	x	μ	$x-2$	$x+2$	μ 포함여부	$\mu-2$	$\mu+2$	x 포함여부
1	4.02	10	2.02	6.02	-	8	12	-
2	8.91	10	6.91	10.91	1	8	12	1
3	0.50	10	-1.49	2.50	-	8	12	-
4	11.35	10	9.35	13.35	1	8	12	1
5	11.73	10	9.73	13.73	1	8	12	1
6	2.24	10	0.24	4.24	-	8	12	-
7	10.33	10	8.33	12.33	1	8	12	1
8	13.24	10	11.24	15.24	-	8	12	-
9	0.74	10	-1.25	2.74	-	8	12	-
10	3.27	10	1.27	5.27	-	8	12	-
...								
1,000	12.83	10	10.83	14.83	-	8	12	-
평균					0.21	8	12	0.21

- (3) 1.0 (4) 1.0
2. (1) μ_X (2) 100 (3) $\frac{\sigma_X}{\sqrt{n}}$ (4) $\frac{10}{\sqrt{25}} = 2$ (5) $\bar{X} \sim N(100, 5^2)$
- (6) $P(\bar{X} < 96.08) = P\left(\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} < \frac{96.08 - 100}{2}\right) = P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{96.08 - 100}{10/\sqrt{25}}\right)$
 $= P(Z < -1.96) = 0.025$
- (7) $P(96.08 < \bar{X} < 103.92) = P\left(\frac{96.08 - 100}{10/\sqrt{25}} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{103.92 - 100}{10/\sqrt{25}}\right)$
 $= P(-1.96 < Z < 1.96) = 0.95$
- (8) $P(103.92 < \bar{X}) = P\left(\frac{103.92 - 100}{10/\sqrt{25}} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right) = P(Z > 1.96) = 0.025$
- (9) 포함되지 않는다. 예로 $\bar{X} = 90$ 이라면, $(\bar{X} - 3.92, \bar{X} + 3.92)$ 은 $(90 - 3.92, 90 + 3.92) = (86.08, 93.92)$ 가 되어 $\mu = 100$ 가 포함되지 않는다.
 $\bar{X} = 96.08$ 이면 $(\bar{X} - 3.92, \bar{X} + 3.92) = (92.16, 100)$ 이 되어 μ 가 포함되지만, 현재 \bar{X} 는 96.08보다 작은 값이므로, $(\bar{X} - 3.92, \bar{X} + 3.92)$ 구간에 μ 가 포함될 수 없다.
- (10) 항상 포함된다. 예로 $\bar{X} = 98$ 이라면, $(\bar{X} - 3.92, \bar{X} + 3.92)$ 은 $(98 - 3.92, 98 + 3.92) = (94.08, 101.92)$ 가 되어 $\mu = 100$ 가 포함된다.
- (11) 포함되지 않는다. 예로 $\bar{X} = 110$ 이라면, $(\bar{X} - 3.92, \bar{X} + 3.92)$ 은 $(110 - 3.92, 110 + 3.92) = (106.08, 113.92)$ 가 되어 $\mu = 100$ 를 포함하지 않는다.
 $\bar{X} = 103.92$ 이면 $(\bar{X} - 3.92, \bar{X} + 3.92) = (100, 103.92)$ 가 되어 μ 가 포함되지만, 현재 \bar{X} 는 103.92보다 큰 값이므로, $(\bar{X} - 3.92, \bar{X} + 3.92)$ 구간에 μ 가 포함될 수 없다.
- (12) 문제 10의 구간, 즉 $96.08 < \bar{X} < 103.92$ 에서만 μ 가 포함되므로, 구간에 μ 가 포함될 확률은 $P(96.08 < \bar{X} < 103.92)$ 와 동일하다. 답은 0.95.
3. (1) $\bar{X} \sim N(10, 2^2)$ (2) 0.1 (3) 0.8 (4) 0.1 (5) 포함되지 않는다. (6) 포함된다.
 (7) 포함되지 않는다. (8) 0.8
4. (1) 1.645 (2) $\mu + 1.645\sigma$ (3) $\frac{(\mu + 1.645\sigma) - \mu}{\sigma} = 1.645$ (4) $100 + 1.645(5) = 108.225$
 (5) $100 - 1.645(5) = 91.775$ (6) $\frac{x - \mu}{\sigma} = \frac{110 - 100}{10} = 2 \rightarrow P(z > 2) = 0.0228$ (7)
 $P(z > z_{0.0228}) = 0.0228 \rightarrow z_{0.0228} = 2$ (8) 2 (9) -2
5. (1) 0~1 사이의 일양분포에서 상위 5%에 해당하는 값은 0.95 ($a + (b - a) \cdot 0.95$)
 (2) 10~20 사이의 일양분포에서 상위 5%에 해당하는 값은 19.5 ($a + (b - a) \cdot 0.95$)
 (3) $\mu + 1.645\sigma$
6. (1) T x 가 정규분포를 따르면, \bar{x} 는 (n 에 무관하게) 항상 정규분포를 따른다.
 (2) T x 가 정규분포를 따르면, \bar{x} 는 (n 에 무관하게) 항상 정규분포를 따른다.
 (3) F 정규분포에 가깝지 않을 수 있다.
 (4) T 중심극한의 정리(The Central Limit Theorem): 표본 크기 n 이 증가함에 따라 평균의 표본분포는 모집단의 분포 모양에 관계없이 정규분포에 접근한다.

$$(5) \mu_{\bar{x}} = \mu_x, \sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n} \quad (6) \bar{x} \sim N(\bar{x} \text{의 평균}, \bar{x} \text{의 분산}) = \bar{x} \sim N(\mu_x, \frac{\sigma_x^2}{n})$$

$$7. (1) \bar{x} \pm z_{0.025} \cdot \frac{\sigma}{\sqrt{n}} \rightarrow 35 \pm 1.96 \cdot 10/5 \rightarrow 35 \pm 3.920$$

$$(2) \bar{x} \pm z_{0.1} \cdot \frac{\sigma}{\sqrt{n}} \rightarrow 35 \pm 1.28 \cdot 10/5 \rightarrow 35 \pm 2.56$$

8. (1) 일치성

(2) True

$x \sim N(100, 10^2)$ 이고 $n = 64$ 이면 $\sigma_{\bar{x}} = \frac{10}{\sqrt{64}}$ 이다. \rightarrow 단일한 값, 즉 상수이다.

그렇지만 표본으로부터 s 를 구하는 경우, 아래 모의실험에서 보듯이 s 값은 실험마다 다르다. 다만, n 이 증가할수록 s 값은 $\sigma=10$ 에 근접한다. (n 이 무한대이면 s 값은 σ 에 일치한다.)

모의실험: $x \sim N[100, 10^2]$ 일때 s 값

n	s				
	실험 1	실험 2	실험 3	실험 4	실험 5
5	3.258	11.183	11.443	7.375	9.906
10	5.891	7.835	13.304	11.871	8.972
30	10.477	9.499	11.245	9.653	7.959
100	10.872	9.446	10.963	10.261	7.774
1,000	10.319	10.007	9.658	10.369	9.646
10,000	9.980	9.853	10.032	10.079	9.959

(3) False 위의 모의실험에서 보듯 s 는 σ 의 추정치이지 같은 값이 아니다.

(4) n 이 충분히 크다. (n 이 충분히 크면, x 가 정규분포를 따르거나 그렇지 않아도 \bar{x} 는 정규분포를 따르고, n 이 충분히 크면 σ 의 추정치로 s 를 사용할 수 있다.)

$$9. (1) 35 \pm (1.96) \frac{16}{8} = 35 \pm 3.92 \quad (2) 35 \pm (1.645) \frac{16}{8} = 35 \pm 3.29 \quad (3) 35 \pm (1.645) \frac{16}{8} = 35 \pm 2.56$$

$$10. (1) \bar{x} = 9, s = 3.07, t_{0.025, 8-1} = 2.3646 \rightarrow 9 \pm 2.3646 \frac{3.07}{\sqrt{8}} = 9 \pm 2.57$$

$$(2) \bar{x} = 10.5, s = 3.83, t_{0.050, 6-1} = 2.0150 \rightarrow 10.5 \pm 2.0150 \frac{3.83}{\sqrt{6}} = 9 \pm 3.15$$

(3) x 가 정규분포를 따르고 σ 가 알려져 있지 않다.

11. (1) 두 확률 모두 $1-\alpha$ 로 동일하다.

$E(s) = \sigma$ 이고 $t_{\frac{\alpha}{2}, n-1} > z_{\frac{\alpha}{2}}$ 이므로, $t_{\frac{\alpha}{2}, n-1} \cdot \frac{s}{\sqrt{n}}$ 의 기대값은 $z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ 보다 크다. 확률이 동일하다면 신뢰구간이 작은 것을 선호하므로 σ 가 알려져 있는 경우에는 신뢰구간으로 $\bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ 를 사용한다.