

6장. 표본과 통계량

모집단과 표본

큰 수의 법칙과 중심극한의 정리

정규분포로부터 생성되는 분포: 카이-제곱분포, t-분포, F-분포

표본평균과 표본분산의 분포

1 / 43

모집단(population)

- (1) 관심이나 사고의 대상이 되는 원소들의 모임으로, 그 집단구성물의 특징이 확률분포로 기술되는 모임을 **모집단**이라 한다.
- (2) 모집단의 모형은 모집단의 평균 및 분산과 같은 모집단의 특성값에 의하여 결정되는데 이와 같은 모집단의 특성값을 **모수**(parameter)라 한다.
- (3) 집단구성의 특성이 확률변수 X 나 확률질량함수 또는 확밀도함수 $f(x; \theta)$ 로 주어지면, 이 모집단을 간단히 '모집단 X ' 또는 '모집단 $f(x; \theta)$ '라 부른다.

2 / 43

표본, 확률표본, 통계량(sample, random sample, statistics)

- (1) 모집단 X 로부터 관찰된 n 개의 값 x_1, x_2, \dots, x_n 또는 이것을 실현값으로 하는 확률변수 X_1, X_2, \dots, X_n 을 모집단 X 로부터 추출된 크기 n 의 **표본**이라 한다. 여기서 X_i 는

“ $X_i = i$ -번째 추출로 얻는 모집단 X 의 값”

을 나타내는 확률변수이다.

- (2) X 가 확률변수일 때, 서로독립이고 X 와 같은 분포를 갖는 표본 X_1, X_2, \dots, X_n 을 **확률표본=임의표본**이라 한다.
- (3) X_1, X_2, \dots, X_n 을 X 로부터 추출된 표본이라 할 때, 모수를 포함하지 않는 표본의 함수를 **통계량**이라 한다.

3 / 43

표본평균, 표본분산(sample mean, sample variance)

X_1, X_2, \dots, X_n 이 모집단 X 로부터 추출된 확률표본일 때,

- (1) **표본평균**: $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$
- (2) **표본분산**: $S^2 = S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

Theorem

X_1, X_2, \dots, X_n 이 평균이 μ 이고 분산이 σ^2 인 모집단 X 로부터 추출된 확률표본이면,

$$(1) \quad E[\bar{X}] = \mu \qquad (2) \quad \text{Var}[\bar{X}] = \frac{\sigma^2}{n}$$

4 / 43

Example

어떤 항아리에 0, 2, 4, 6 이라고 쓰인 구슬 4개가 들어 있다.

- (1) 항아리에서 1개를 무작위로 뽑을 때 구슬에 적혀있는 수를 X 라 할 때, X 의 확률분포표, 평균, 분산
- (2) 항아리에서 2개를 무작위로 복원추출할 때 구슬에 적혀있는 수를 X_1, X_2 라 하고, 표본평균 $\bar{X} = \frac{1}{2}(X_1 + X_2)$ 의 확률분포표, 평균, 분산

Remark

일반적으로 표본분산을 $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ 보다 $\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ 을 더 많이 사용한다.

큰 수의 법칙=대수의 법칙(Law of Large Numbers=LLN)

X_1, X_2, \dots, X_n 이 평균이 μ 이고 분산이 σ^2 인 모집단 X 로부터 추출된 확률표본이면, 임의의 $\epsilon > 0$ 에 대하여

$$\lim_{n \rightarrow \infty} P[|\bar{X} - \mu| < \epsilon] = 1 \quad \left(\text{또는, } \bar{X} - \mu \xrightarrow{P} 0 \right)$$

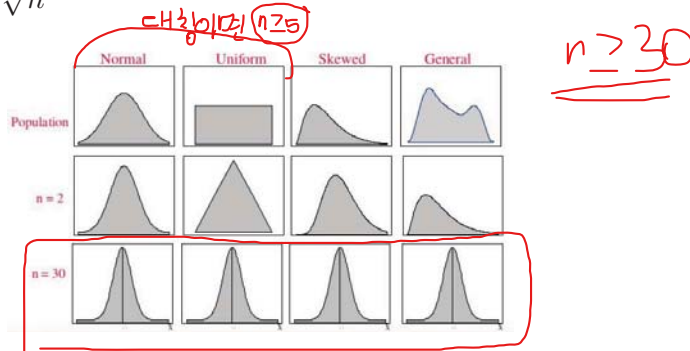
이다. 실제로, 임의의 $\epsilon > 0$ 에 대하여

$$\forall \delta \in (0, 1), \quad n \geq \frac{\sigma^2}{\epsilon^2 \delta} \rightarrow P[|\bar{X} - \mu| < \epsilon] \geq 1 - \delta.$$

중심극한의 정리(Central Limit Theorem)

X_1, X_2, \dots, X_n 이 평균이 μ 이고 분산이 σ^2 인 모집단 X 로부터 추출된 확률표본이면,

- (1) \bar{X} 는 근사적으로 정규분포 $N(\mu, \sigma^2/n)$ 를 따른다.
- (2) $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ 는 근사적으로 표준정규분포 $N(0, 1)$ 을 따른다.



표본비율(sample proportion)

X_1, X_2, \dots, X_n 을 $X \sim \text{Ber}(p)$ 로부터의 확률표본일때

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

를 표본비율이라 한다.

Theorem

X_1, X_2, \dots, X_n 이 $X \sim \text{Ber}(p)$ 로부터의 확률표본이면, 표본비율 \hat{p} 은 근사적으로 정규분포 $N\left(p, \frac{p(1-p)}{n}\right)$ 를 따른다.

$$\bar{X} \approx N(1000, \frac{2500}{100})$$

Example

1. 어떤 공장에서 생산되는 건전지의 평균수명은 1000시간, 분산이 2500시간이다. 만일 100개의 건전지를 임의표본으로 추출했을 때 표본평균이 995시간 이상 1010시간 미만일 확률을 구하시오.

$X \sim U(0,1)$
 $\mu = \frac{1}{2}$
 $\sigma^2 = \frac{1}{12}$
 $n=10$
 $\bar{X} \sim N(\frac{1}{2}, \frac{1}{120})$

2. X_1, X_2, \dots, X_{10} 은 구간 (0, 1) 상의 균등분포에서 뽑은 크기 10인 임의표본이다. $Y = X_1 + X_2 + \dots + X_{10}$ 이라 할 때 Y가 5.33 보다 작을 확률을 구하시오.

3. 도시 인구의 60%가 복지시설에 기부금을 내는 제안에 호의적이라고 한다. 임의로 뽑은 150명 중 이 제안에 호의적인 사람들의 비율이 0.52보다 작을 확률을 구하시오.

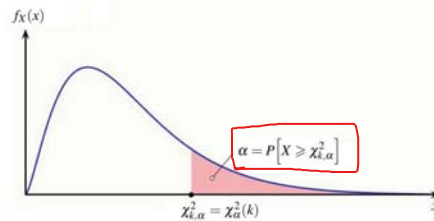
$X \sim \text{Ber}(0.6)$
 $\mu = p$
 $\sigma^2 = p(1-p)$
 $n=150$
 $\bar{X} \approx N(0.6, \frac{0.6 \times 0.4}{150})$

$P[\hat{p} < 0.52]$
 $= P[X' < 0.52] = P[\frac{X' - 0.6}{\sqrt{0.6 \times 0.4}} < \frac{0.52 - 0.6}{\sqrt{0.6 \times 0.4}}]$
 $= P[Z < -2]$
 $= P[Z > 2]$

카이제곱분포표 보는 방법

$X \sim \chi^2(k)$ 이라 하자.
 임의의 $\alpha \in (0, 1)$ 에 대하여,
 실수 $\chi^2_{k,\alpha} = \chi^2_\alpha(k)$ 는 다음을 만족하는 값이라 정의한다.

$$P[X \geq \chi^2_{k,\alpha}] = \alpha$$



Degrees of freedom	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188

카이-제곱분포(chi-square distribution)

$X \sim \text{Ga}(\frac{k}{2}, 2)$ 일 때, X를 자유도 k인 카이-제곱분포라 하고

$$X \sim \chi^2(k) \quad (k > 0)$$

로 나타낸다. 즉, X의 확률밀도함수가 다음과 같다.

$$f_X(x) = \begin{cases} \frac{x^{\frac{k}{2}-1} e^{-x/2}}{\Gamma(\frac{k}{2}) 2^{k/2}}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

Theorem

$X \sim \chi^2(k)$ 이면,

- (1) $E[X] = k$ (3) $m_X(t) = \left(\frac{1}{1-2t}\right)^{\frac{k}{2}}, \quad t < \frac{1}{2}$
 (2) $\text{Var}[X] = 2k$

Example

1. $X \sim \chi^2(3)$ 일 때, $P[X \geq x] = 0.05$ 인 x를 구하시오.

2. 어느 의류판매장에 들르는 고객의 수는 시간당 30명인 푸아송분포를 따른다고 한다. 일곱 번째 고객이 들어오기까지 매장주인이 기다리는 시간이 6.57분 이상일 확률을 구하시오.

$X \sim \text{Po}(30)$
 $X \sim \text{Po}(\frac{1}{2})$
 $\text{co}, 1)$

$T \sim G(7, 2)$
 $T \sim \chi^2(14)$
 $P[T \geq 6.57]$

Theorem

- (1) Z_1, Z_2, \dots, Z_k 가 서로 독립이고 모든 i 에 대하여 $Z_i \sim N(0, 1)$ 이면,

$$U = \sum_{i=1}^k Z_i^2 = Z_1^2 + Z_2^2 + \dots + Z_k^2 \sim \chi^2(k)$$

- (2) X_1, X_2, \dots, X_k 가 서로 독립이고 모든 i 에 대하여 $X_i \sim N(\mu_i, \sigma_i^2)$ 이면,

$$U = \sum_{i=1}^k \left(\frac{X_i - \mu_i}{\sigma_i} \right)^2 \sim \chi^2(k)$$

Proof. (1) $k = 1$ 인 경우: $Z \sim N(0, 1)$ 이고 $U = Z^2$ 이라 하자.

Z 의 확률밀도함수는 $f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}z^2 \right\}$ 이므로,

$$\begin{aligned} m_U(t) &= E[e^{tU}] = E[e^{tZ^2}] \\ &= \int_{-\infty}^{\infty} \exp \{tz^2\} \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}z^2 \right\} dz \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{z^2}{2}(1-2t) \right\} dz \quad \left(\text{for } t < \frac{1}{2} \right) \\ &= \frac{1}{\sqrt{1-2t}} \int_{-\infty}^{\infty} \frac{\sqrt{1-2t}}{\sqrt{2\pi}} \exp \left\{ -\frac{z^2}{2}(1-2t) \right\} dz \quad \left\{ \begin{array}{l} y = \sqrt{1-2t} z, \\ dy = \sqrt{1-2t} dz \end{array} \right. \\ &= \frac{1}{\sqrt{1-2t}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{y^2}{2} \right\} dy \\ &= \frac{1}{\sqrt{1-2t}} = \left(\frac{1}{1-2t} \right)^{\frac{1}{2}} \quad \left(\text{for } t < \frac{1}{2} \right) \end{aligned}$$

따라서 $U = Z^2 \sim \chi^2(1)$.

Theorem

- (1) X_1, X_2, \dots, X_m 이 서로 독립이고 모든 i 에 대하여 $X_i \sim \chi^2(k_i)$ 이면,

$$V = \sum_{i=1}^m X_i \sim \chi^2(k_1 + k_2 + \dots + k_m)$$

- (2) $X_1 \sim \chi^2(k_1)$, $X_2 \sim \chi^2(k_2)$ 이고 $k_2 > k_1$ 이라 하자.
 $Y = X_2 - X_1$ 와 X_1 이 서로 독립이면

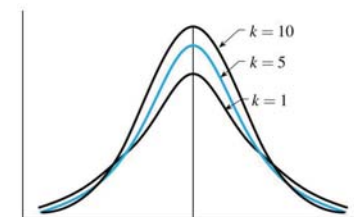
$$Y \sim \chi^2(k_2 - k_1)$$

 t -분포(t -distribution)

확률변수 X 의 확률밀도함수가

$$f_X(x) = \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} \cdot \frac{1}{\sqrt{k\pi}} \cdot \frac{1}{\left(1 + \frac{x^2}{k}\right)^{\frac{k+1}{2}}}, \quad -\infty < x < \infty$$

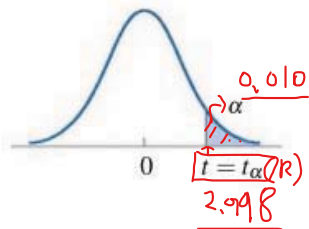
이면, X 는 자유도 k 인 t -분포를 따른다고 하고 $X \sim t(k)$ 로 나타낸다.



t-분포표 보는 방법

$X \sim t(k)$ 이라 하자. 임의의 $\alpha \in (0, 1)$ 에 대하여, 실수 $t_\alpha(k) = t_\alpha$ 는 다음을 만족하는 값이라 정의한다.

$$P[X \geq t_\alpha(k)] = P[X \geq t_\alpha] = \alpha$$



d.f.	t _{.100}	t _{.050}	t _{.025}	t _{.010}	t _{.005}
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947

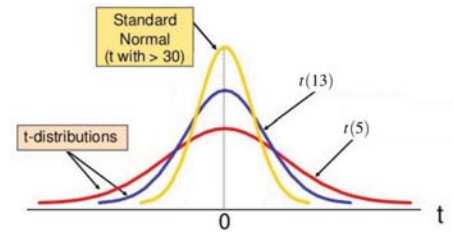
Theorem

$Z \sim N(0, 1)$ 이고 $U \sim \chi^2(n)$ 이고 Z 와 U 가 서로 독립이면,

$$T = \frac{Z}{\sqrt{U/n}} \sim t(n)$$

Remark (t-분포와 표준정규분포의 확률밀도함수 비교)

$t(n) \rightarrow Z(0, 1)$



F-분포(F-distribution)

확률변수 X 의 확률밀도함수가

$$f_X(x) = \begin{cases} \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \cdot \left(\frac{m}{n}\right)^{\frac{m}{2}} \cdot \frac{x^{\frac{m}{2}-1}}{(1 + \frac{m}{n}x)^{\frac{m+n}{2}}}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

이면, X 는 자유도 m, n 인 F -분포를 따른다고 하고 $X \sim F(m, n)$ 로 나타낸다.

Theorem

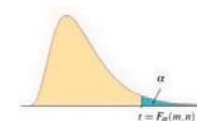
$U \sim \chi^2(m)$, $V \sim \chi^2(n)$ 이고 U 와 V 가 서로 독립이면,

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

F-분포표 보는 방법

$X \sim F(m, n)$ 이라 하자. 임의의 $\alpha \in (0, 1)$ 에 대하여, 실수 $F_\alpha(m, n)$ 는 다음을 만족하는 값이라 정의한다.

$$P[X \geq F_\alpha(m, n)] = \alpha$$



		Degrees of freedom in the numerator (m)								
p		1	2	3	4	5	6	7	8	9
.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	
.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	
.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.99	8.90	
.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	
.001	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00	48.47	
.100	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	
.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	
.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	
.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	
.001	47.18	37.12	33.20	31.09	29.75	28.83	28.16	27.65	27.24	
.100	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	
.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	
.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	
.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	
.001	35.51	27.00	23.70	21.92	20.80	20.03	19.46	19.03	18.69	
.100	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	
.050	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	
.025	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	
.010	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	
.001	29.25	21.69	18.77	17.20	16.21	15.52	15.02	14.63	14.33	
.100	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	
.050	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	
.025	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	
.010	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	
.001	25.41	18.49	15.83	14.39	13.48	12.86	12.40	12.05	11.77	

Theorem

$X \sim F(m, n)$ 에 대해 $F_\alpha(m, n)$ 을 $P[X \geq t] = \alpha$ 를 만족하는 t 라 할 때, 즉, $P[F \geq F_\alpha(m, n)] = \alpha$ 라 할 때,

$$F_\alpha(m, n) = \frac{1}{F_{1-\alpha}(n, m)}$$

Example

- (1) $X \sim F(3, 5)$ 일 때, $P[X \geq f] = 0.05$ 를 만족시키는 f 의 값을 구하시오.
- (2) $X \sim F(3, 5)$ 일 때, $F_{0.95}(5, 3)$ 를 구하시오.

Theorem(F -분포와 t -분포와의 관계)

$Z \sim N(0, 1)$, $U \sim \chi^2(n)$ 이고 Z 와 U 가 서로 독립이면,

- (1) $T = \frac{Z}{\sqrt{U/n}} \sim t(n)$
- (2) $T^2 = \frac{Z^2/1}{U/n} \sim F(1, n)$

표본평균의 분포

X_1, X_2, \dots, X_n 이 정규모집단 $N(\mu, \sigma^2)$ 에서 추출된 확률표본이면,

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

Proof.

$$\begin{aligned} m_{\bar{X}}(t) &= E[e^{t\bar{X}}] = E\left[\exp\left\{\frac{t}{n}\sum_{i=1}^n X_i\right\}\right] = E\left[\prod_{i=1}^n \exp\left\{\frac{t}{n}X_i\right\}\right] \\ &= \prod_{i=1}^n E\left[\exp\left\{\frac{t}{n}X_i\right\}\right] = \prod_{i=1}^n m_{X_i}\left(\frac{t}{n}\right) = \prod_{i=1}^n \exp\left\{\mu\frac{t}{n} + \frac{1}{2}\sigma^2\left(\frac{t}{n}\right)^2\right\} \\ &= \exp\left\{\sum_{i=1}^n \left(\mu\frac{t}{n} + \frac{1}{2}\sigma^2\left(\frac{t}{n}\right)^2\right)\right\} = \exp\left\{\mu t + \frac{1}{2}\left(\frac{\sigma}{\sqrt{n}}\right)^2 t^2\right\} \end{aligned}$$

따라서, $\bar{X} \sim N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$

표본분산의 분포

X_1, X_2, \dots, X_n 이 정규모집단 $N(\mu, \sigma^2)$ 에서 추출된 확률표본이면,

$$\frac{(n-1)}{\sigma^2} S^2 \sim \chi^2(n-1) \quad \text{단, } S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Lemma

X_1, X_2, \dots, X_n 이 정규모집단 $N(\mu, \sigma^2)$ 에서 추출된 확률표본이고, 표본평균을 \bar{X} , 표본분산을 S^2 이라 하자.

- (1) \bar{X} 와 $X_1 - \bar{X}, X_2 - \bar{X}, \dots, X_n - \bar{X}$ 은 독립이다.
- (2) \bar{X} 와 S^2 은 독립이다.

Proof.
$$\begin{aligned}\sum_{i=1}^n (X_i - \mu)^2 &= \sum_{i=1}^n (X_i - \bar{X} + \bar{X} - \mu)^2 \\ &= \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2 + 2(\bar{X} - \mu) \left(\sum_{i=1}^n X_i - n\bar{X} \right) \\ &= \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2\end{aligned}$$

이므로

$$\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 + \frac{1}{\sigma^2} n(\bar{X} - \mu)^2 = \frac{(n-1)}{\sigma^2} S^2 + \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2$$

이다. Lemma에 의해 S^2 과 \overline{X} 가 독립이고, Theorem에 의해

$$\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(n), \quad \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2 \sim \chi^2(1)$$

이므로 Theorem에 의해

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

Theorem

X_1, X_2, \dots, X_n 이 정규모집단 $N(\mu, \sigma^2)$ 에서 추출된
 확률표본이면,

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

Proof. Theorem에 의해

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

이때 Z, U 가 서로 독립이므로, Theorem에 의해

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)}{\sigma^2} S^2 / (n-1)}} = \frac{Z}{\sqrt{U/(n-1)}} \sim t(n-1)$$

Example

$$\frac{d}{dx} (600 - 49x^2)$$

1. 어떤 공장에서 공급되는 형광등은 평균수명이 600시간, 표준편차가 49시간인 정규분포를 따른다고 한다. 49개의 형광등을 임의로 뽑아 평균수명을 잴 때 그 값이 607시간 이상일 확률을 구하시오.

일 확률을 구하시오. $\bar{X} \sim N(600, \left(\frac{49}{n}\right))$
 $P[\bar{X} \geq 607] = P[Z \geq 1]$

2. 어떤 공장에서 생산하는 스마트폰 건전지의 두께는 분산 $\sigma^2 = 0.5$ 인 정규분포를 따른다고 한다. 21개의 확률표본을 추출할 때, 건전지 두께의 표본분산이 0.2713 보다 클 확률을 구하시오.

구하시오.
 $X \sim N(0, 5) \rightarrow V = \frac{20}{0.6} S^2 \sim \chi^2(20)$

$$P[S^2 > 0.2713] = P[\underbrace{40}_{\substack{11 \\ \checkmark}} S^2 > \underbrace{40 \times 0.2713}_{\substack{11 \\ 10.852}}]$$

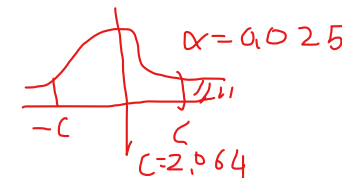
Example

$n=25$

1. 정규분포 $N(\mu, \sigma^2)$ 을 따르는 모집단에서 추출된 크기 25인 임의표본에서 표본평균 \bar{X} 와 표본분산 S^2 에 대하여 다음을 만족하는 상수 c 를 구하시오.

$$P\left[-c < \frac{5(\bar{X}-\mu)}{S} < c\right] = 0.95$$

$$T = \frac{\bar{x} - \mu}{S/\sqrt{25}} \sim t_{(24)}$$



2. 어떤 공장에서 엔진의 회발유 소비량은 평균 16인 정규분포를 따른다고 하자. 16번 엔진을 시험가동했을 때 회발유 소비량의 표본표준편차가 4였다. 이때 표본평균 \bar{X} 가 13.053 보다 크지 않을 확률을 구하시오. $n=16$

$x \sim N(16, 6^2) \rightarrow S=4 \rightarrow \sim t(15)$
 $P[\bar{x} \leq 13.053] = \Phi\left[\frac{\bar{x}-16}{4/\sqrt{16}} \leq 13.053-16\right]$

Theorem

X_1, X_2, \dots, X_m 이 정규모집단 $N(\mu_1, \sigma_1^2)$ 에서 추출된 확률표본이고
 Y_1, Y_2, \dots, Y_n 이 정규모집단 $N(\mu_2, \sigma_2^2)$ 에서 추출된 확률표본이며,

$$\bar{X} = \frac{1}{m} \sum_{i=1}^m X_i, \quad \bar{Y} = \frac{1}{n} \sum_{j=1}^n Y_j,$$

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

이라 하자. $X_1, X_2, \dots, X_m, Y_1, Y_2, \dots, Y_n$ 이 서로 독립이면,

$$F = \frac{S_X^2 / \sigma_1^2}{S_Y^2 / \sigma_2^2} \sim F(m-1, n-1)$$

특히, $\sigma_1^2 = \sigma_2^2$ 이면, $F = \frac{S_X^2}{S_Y^2} \sim F(m-1, n-1)$

Proof. Theorem에 의해

$$\frac{(m-1)}{\sigma_1^2} S_X^2 \sim \chi^2(m-1), \quad \frac{(n-1)}{\sigma_2^2} S_Y^2 \sim \chi^2(n-1)$$

이고 서로 독립이므로, Theorem에 의해

$$\frac{S_X^2 / \sigma_1^2}{S_Y^2 / \sigma_2^2} = \frac{\frac{(m-1)}{\sigma_1^2} S_X^2 / (m-1)}{\frac{(n-1)}{\sigma_2^2} S_Y^2 / (n-1)} \sim F(m-1, n-1)$$

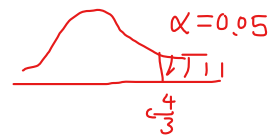
Example

분산이 각각 $\sigma_1^2 = 8, \sigma_2^2 = 6$ 인 두 정규모집단 X, Y 에서 크기
 $m = 5, n = 7$ 의 각 확률표본의 표본분산을 각각 S_X^2, S_Y^2 이라

하자. 이 때, $P\left[\frac{S_X^2}{S_Y^2} > c\right] = 0.05$ 를 만족하는 상수 c 를 구하시오.

$$F = \frac{S_X^2 / 8}{S_Y^2 / 6} \sim F(4, 6)$$

$$P\left[F > c \times \frac{8}{6}\right] = 0.05$$



$$c \frac{4}{3} = 4.53$$

$$c = 4.53 \times \frac{3}{4}$$

Theorem

X_1, X_2, \dots, X_m 이 모집단 $N(\mu_1, \sigma_1^2)$ 에서 추출된 확률표본이고,
 Y_1, Y_2, \dots, Y_n 이 모집단 $N(\mu_2, \sigma_2^2)$ 에서 추출된 확률표본이라 하자.
 $X_1, X_2, \dots, X_m, Y_1, Y_2, \dots, Y_n$ 이 서로 독립이면, 표본평균의 차
 $\bar{X} - \bar{Y}$ 는 정규분포를 따른다. 즉,

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

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

또한, 모집단이 정규분포를 따르지 않아도 m 과 n 이 충분히 크면
 표본평균의 차 $\bar{X} - \bar{Y}$ 는 정규분포에 근사한다.

Example

평균이 같고 분산이 각각 $\sigma_1^2 = 81$, $\sigma_2^2 = 64$ 인 두 정규모집단 X, Y 에서 크기 $m = 25$, $n = 16$ 의 각 확률표본의 표본평균을 각각 \bar{X} , \bar{Y} 라 할 때, 다음을 구하시오.

$$(1) P[\bar{X} - \bar{Y} \geq 10] = P\left[\frac{(\bar{X} - \bar{Y}) - (0)}{\sqrt{\frac{81}{25} + \frac{64}{16}}} \geq \frac{10}{\sqrt{181/25}}\right]$$

$$(2) P[|\bar{X} - \bar{Y}| \leq 5]$$

$$= P\left[-5 \leq \frac{\bar{X} - \bar{Y}}{\sqrt{181/25}} \leq \frac{5}{\sqrt{181/25}}\right]$$

\parallel
 $\frac{5}{\sqrt{181/25}}$

33 / 43

Theorem

X_1, X_2, \dots, X_m 이 모집단 $N(\mu_1, \sigma_1^2)$ 에서 추출된 확률표본이고, Y_1, Y_2, \dots, Y_n 이 모집단 $N(\mu_2, \sigma_2^2)$ 에서 추출된 확률표본이라 하자. $X_1, X_2, \dots, X_m, Y_1, Y_2, \dots, Y_n$ 이 서로 독립이고 $\sigma^2 = \sigma_1^2 = \sigma_2^2$ 이면,

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

여기서,

$$s_p^2 = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}$$

34 / 43

Proof. 모든 i, j 에 대하여, $X_i \sim N(\mu_1, \sigma^2)$ 이고 $Y_j \sim N(\mu_2, \sigma^2)$ 이므로, Theorem에 의해

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

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

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

이다. 또한 Theorem에 의해

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

$$V = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{\sigma^2} \sim \chi^2(m+n-2)$$

이다.

35 / 43

그러므로 Theorem에 의해

$$T = \frac{Z}{\sqrt{V/(m+n-2)}} \sim t(m+n-2)$$

이고

$$\begin{aligned} & \frac{Z}{\sqrt{V/(m+n-2)}} \\ &= \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2}{m} + \frac{\sigma^2}{n}}} \times \frac{1}{\sqrt{\frac{(m-1)S_X^2 + (n-1)S_Y^2}{\sigma^2} / (m+n-2)}} \\ &= \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{m} + \frac{1}{n}} \cdot \sqrt{\frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}}} = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_p \cdot \sqrt{\frac{1}{m} + \frac{1}{n}}} \end{aligned}$$

36 / 43

$$X \sim N(19, 0.2^2) \quad Y \sim N(16, 0.6)$$

Example

모평균이 각각 $\mu_1 = 19$, $\mu_2 = 16$ 이고 모분산이 같다고 알려진 두 정규모집단 A, B에서 표본을 선정하여 다음을 얻었다.

모집단 A : 표본의 크기 $m = 15$, 표본분산 $s_X^2 = 2.82$

모집단 B : 표본의 크기 $n = 12$, 표본분산 $s_Y^2 = 2.51$

두 표본평균의 차이가 4.5675 이상일 확률을 구하시오.

$$P[|\bar{X} - \bar{Y}| \geq 4.5675] = P[\bar{X} - \bar{Y} \leq -4.5675] + P[\bar{X} - \bar{Y} \geq 4.5675]$$

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

Theorem

X_1, X_2, \dots, X_m 이 모집단 $N(\mu_1, \sigma_1^2)$ 에서 추출된 확률표본이고, Y_1, Y_2, \dots, Y_n 이 모집단 $N(\mu_2, \sigma_2^2)$ 에서 추출된 확률표본이라 하자. $X_1, X_2, \dots, X_m, Y_1, Y_2, \dots, Y_n$ 이 서로 독립이고 $\sigma_1^2 \neq \sigma_2^2$ 이면,

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_X^2}{m} + \frac{s_Y^2}{n}}} \sim t(\nu)$$

여기서,

$$\nu = \frac{\left(\frac{s_X^2}{m} + \frac{s_Y^2}{n}\right)^2}{\frac{\left(\frac{s_X^2}{m}\right)^2}{m-1} + \frac{\left(\frac{s_Y^2}{n}\right)^2}{n-1}} \quad (\text{반올림})$$

순서 통계량(order statistic)

X_1, X_2, \dots, X_n 이 모집단 X 로부터 추출된 표본이고 $X_{(k)}$ 를 표본 안에서 k -번째로 작은 관측값일 때, $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ 을 **순서통계량**이라 한다.

특히,

- ▶ $X_{(1)}$ 은 표본의 최소값
- ▶ $X_{(n)}$ 은 표본의 최댓값
- ▶ $R = X_{(n)} - X_{(1)}$ 을 표본의 범위

라 한다.

Theorem

확률변수 X 의 분포함수는 $F(x)$ 이고 $F'(x) = f(x)$ 이라 하자. $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ 이 X 에서 추출된 확률표본의 순서통계량이면,

$$F_{X_{(k)}}(x) = \sum_{j=k}^n \binom{n}{j} [F(x)]^j [1 - F(x)]^{n-j}$$

Proof.

$$\begin{aligned} F_{X_{(k)}}(x) &= P[X_{(k)} \leq x] \\ &= P[\text{적어도 } k \text{개의 확률변수 } X_j \leq x] \\ &= P[k \text{개의 확률변수 } X_j \leq x] \\ &\quad + P[(k+1) \text{개의 확률변수 } X_j \leq x] + \dots \\ &\quad + P[n \text{개의 확률변수 } X_j \leq x] \\ &= \sum_{j=k}^n \binom{n}{j} [F(x)]^j [1 - F(x)]^{n-j} \end{aligned}$$

Theorem

연속확률변수 X 의 분포함수는 $F(x)$ 이고 $F'(x) = f(x)$ 이라 하자.
 $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ 이 X 에서 추출된 확률표본의 순서통계량이면,

$$F_{X_{(1)}}(x) = \sum_{j=1}^n \binom{n}{j} [F(x)]^j [1 - F(x)]^{n-j} = 1 - [1 - F(x)]^n$$

$$F_{X_{(n)}}(x) = \sum_{j=n}^n \binom{n}{j} [F(x)]^j [1 - F(x)]^{n-j} = [F(x)]^n$$

$$f_{X_{(1)}}(x) = n [1 - F(x)]^{n-1} f(x)$$

$$f_{X_{(n)}}(x) = n [F(x)]^{n-1} f(x)$$


41 / 43

Example

확률변수 X 의 확률밀도함수가 다음과 같다. (단, $\theta > 0$)

$$f_X(x) = \begin{cases} \frac{2x}{\theta^2}, & 0 < x \leq \theta \\ 0, & \text{otherwise} \end{cases}$$

X_1, X_2, \dots, X_n 을 모집단 X 로부터 추출된 확률표본일 때,
순서통계량 $X_{(1)}$ 과 $X_{(n)}$ 의 분포함수와 확률밀도함수를 각각
구하시오.

$$F_X(x) = P[X \leq x] = \int_{-\infty}^x f_X(t) dt = \begin{cases} \int_0^x \frac{2t}{\theta^2} dt = \frac{x^2}{\theta^2}, & 0 < x \leq \theta \\ 1, & x > \theta \end{cases}$$


42 / 43

Example



n 개의 동일한 부품이 직렬로 연결되어 있는 전자제품은 어느 한
부품이라도 고장이 나면 제 기능을 발휘하지 못한다. 각 부품의
수명은 지수분포 $\exp(\beta)$ 를 따르고 서로 독립이다. 이 제품의
수명분포를 구하시오.

$$X \sim \exp(\beta) \quad f_X(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta}, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-x/\beta}, & x \geq 0 \end{cases}$$

$$X_{(n)} = \min\{X_1, X_2, \dots, X_n\}$$

$$\begin{aligned} f_{X_{(n)}}(x) &= n [1 - F_X(x)]^{n-1} f_X(x) \\ &= \begin{cases} n [1 - (1 - e^{-x/\beta})]^{n-1} \cdot \frac{1}{\beta} e^{-x/\beta}, & x \geq 0 \\ 0, & x < 0 \end{cases} \\ &= \begin{cases} \frac{n}{\beta} e^{-nx/\beta}, & x \geq 0 \\ 0, & x < 0 \end{cases} \end{aligned}$$

$$\therefore U \sim \exp(\beta/n)$$

43 / 43

$$F_{X_{(1)}}(x) = 1 - (1 - F_X(x))^n = \begin{cases} 0, & x < 0 \\ 1 - (1 - \frac{x^2}{\theta^2})^n, & 0 < x \leq \theta \\ 1, & x > \theta \end{cases}$$

$$f_{X_{(1)}}(x) = n [1 - F_X(x)]^{n-1} f_X(x) = \begin{cases} \frac{2nx}{\theta^2} (1 - \frac{x^2}{\theta^2})^{n-1}, & 0 < x \leq \theta \\ 0, & x > \theta \end{cases}$$

$$F_{X_{(n)}}(x) = [F_X(x)]^n = \begin{cases} 0, & x < 0 \\ (\frac{x^2}{\theta^2})^n, & 0 < x \leq \theta \\ 1, & x > \theta \end{cases}$$

$$f_{X_{(n)}}(x) = n [F_X(x)]^{n-1} f_X(x) = \begin{cases} n (\frac{x^2}{\theta^2})^{n-1} \cdot \frac{2x}{\theta^2}, & 0 < x \leq \theta \\ 0, & x > \theta \end{cases}$$