경영통계

원광대학교 경영학부

담당교수: 정호일

제9장 통계적 추정:한 모집단

- 점추정과 구간추정
- 추정량의 결정기준
- 구간추정
- 모평균의 신뢰구간
- 모비율의 신뢰구간
- 표본크기의 결정
- 모분산의 신뢰구간

I. 점추정과 구간추정

추정량과 추정치

- 추정량(Estimator)이란 표본정보에 의존하는 확률변수로서 모수를 추정하는데 사용되는 표본통계량
- 추정치(Estimate)란 추정량을 평가하여 얻게 되는 특정한 수치

점추정치와 구간추정치

- 점추정치(point estimation)는 모르는 모수를 가장 잘 대표할 수 있는 표본을 추출하고 필요한 계산을 하여 얻는 하나의 수치
- 구간추정치(interval estimation)은 모수의 참값이 포함되리라고 기대하는 추정치를 일정한 범위로 나타내는 것
- 어느 정도 오차를 포함하는 구간추정 방법이 널리 사용.

II. 추정량의 결정기준

모수를 추정하는데 좋은 추정량이 되기 위해서는 다음 조건을 만족해야 한다.

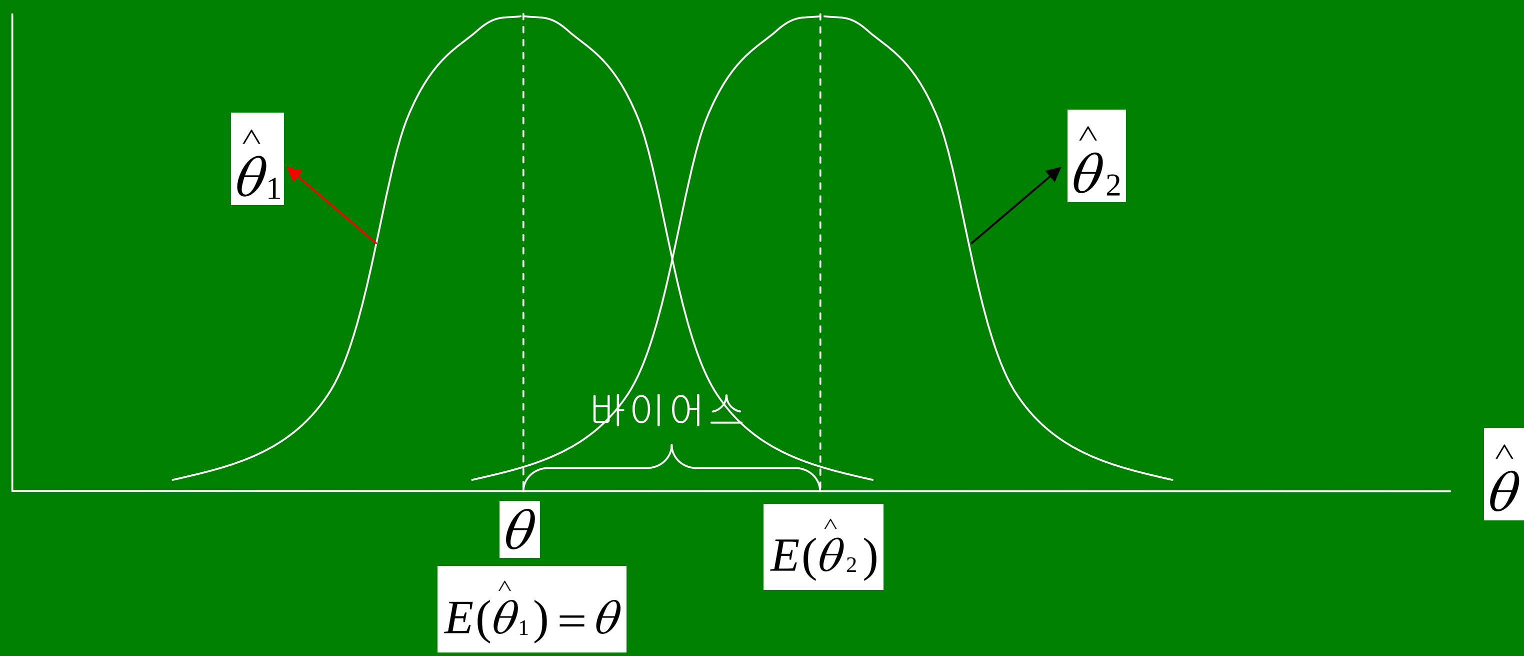
1. 불편성(unbiasedness)
2. 효율성(effectiveness)
3. 일치성(consistency)
4. 충족성(sufficiency)

II. 추정량의 결정기준

1. 불편추정량

점추정량 $\hat{\theta}$ 의 표본분포의 기대값이 모수 θ와 같을 때 점 추정량은 모수의 불편추정량(unbiased estimator)이라고 한다.

$$E(\hat{\theta}) = \theta$$

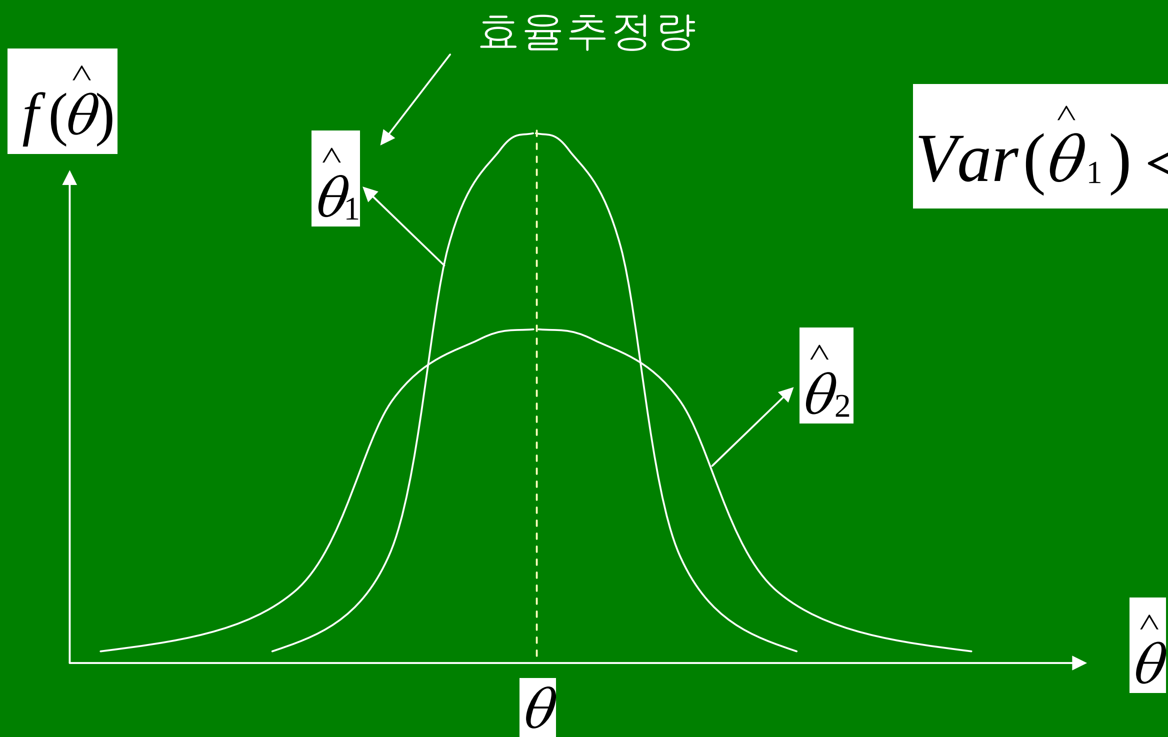


II. 추정량의 결정기준

2. 효율 추정량

불편추정량 중에서 분산이 작은 추정량을 효율추정량(efficient estimator)이라고 한다.

좋은 추정량이 되기 위해서는 불편성 요건 뿐만 아니라 추정량의 분산도 작아야 한다.

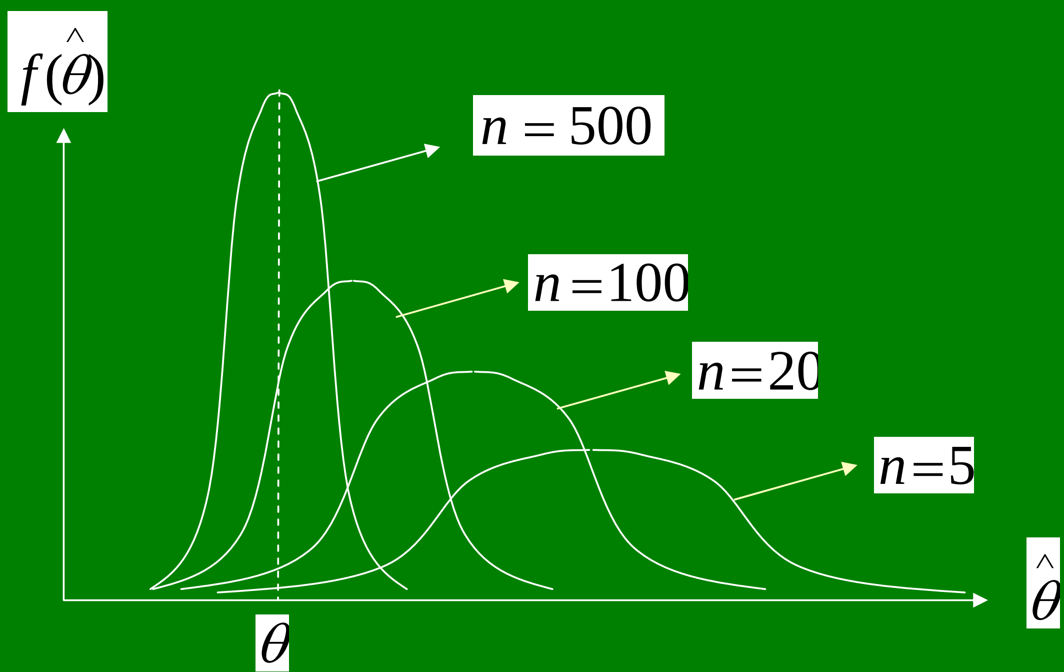


$$Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$$

Ⅱ. 추정량의 결정기준

3. 일치추정량

표본크기가 증가할수록 추정량 $\hat{\theta}$이 모수 θ에 더욱 근접하는 추정량을 일치추정량(consistent estimator)이라고 한다.



4. 충족추정량

모수를 추정하기 위하여 추출하는 동일한 크기의 표본으로부터 가장 많은 정보를 제공하는 추정량을 말한다.

Ⅲ. 점추정량

모수와 점추정량

모 수	점추정량
모평균(μ)	표본평균($\overline{X}$)
모분산(σ^2)	표본분산(S^2)
모표준편차(σ)	표본표준편차(S)
모비율(p)	표본비율($\hat{p}$)

Ⅳ. 구간추정

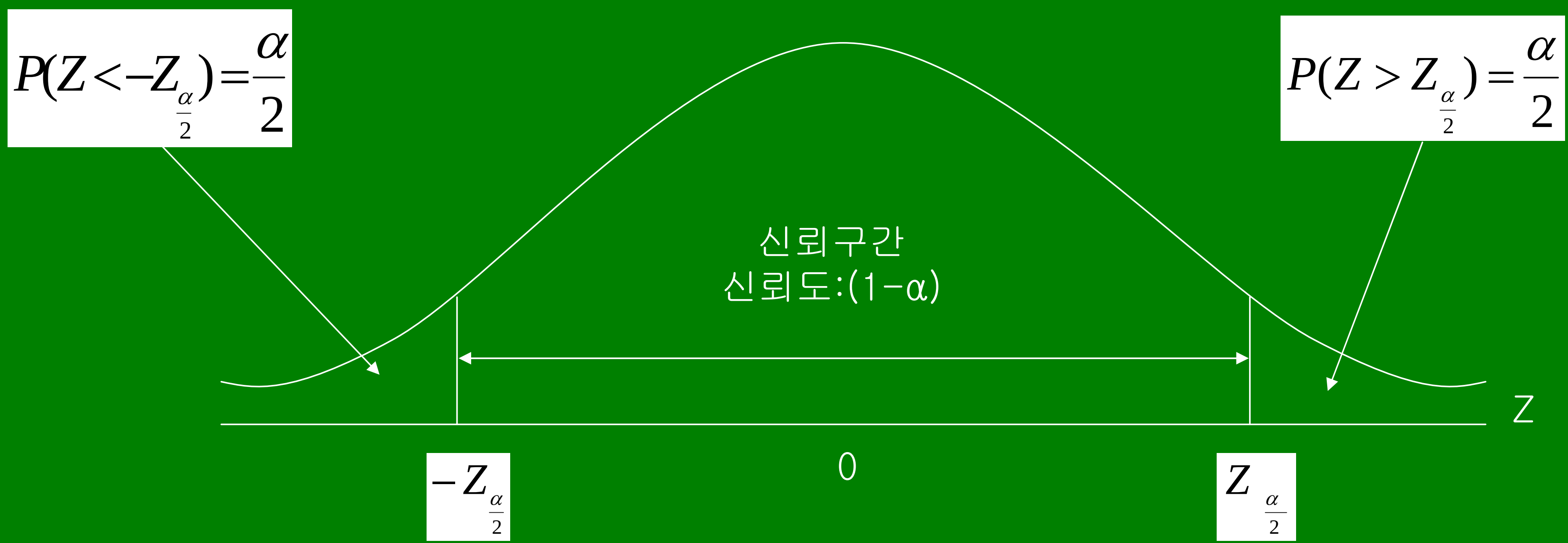
구간추정은 모수가 포함되리라고 보는 범위(구간)를 확률을 가지고 제시함으로써 추정치에 대한 불확실성을 표현한다.

- ◆ 구간추정치: 점추정량에 허용오차를 감안한 것(=신뢰구간)
  점추정량±오차한계
- ◆ 오차한계: 신뢰구간의 중심인 점추정치로부터 상한(하한)사이의 거리 폭
- ◆ 신뢰구간(Confidence interval):모수가 포함 될 것으로 기대하는 범위(점추정치를 중심으로 하한과 상한을 범위로 나타낸 것)
- ◆ 신뢰상한(하한): 점추정치로부터 모수가 포함 될 상한(하한)까지의 구간
- ◆ 신뢰수준(=신뢰도): 신뢰구간에 모수의 참값이 포함될 확률[=1−오차율(α)]
- ◆ 오차확률(=오차율): 구간추정의 부정확도 즉, 신뢰구간이 모수를 포함하지 않을 확률(=α). 오차율이 작아지면 신뢰구간이 커져 모수가 포함될 가능성이 증가하지만 정보가치가 감소하게 된다.
  (보통 1%, 5%, 10% 수준을 활용한다.)

IV. 구간추정

신뢰구간을 설정하기 위해서는 모수가 신뢰구간의 상한과 하한을 벗어나는 실수를 저지를 확률 즉, 오차확률 (=α) 을 주관적으로 설정해야 한다.
오차확률 (=α)은 신뢰구간의 양끝에 위치하는 확률로서 오차확률의 절반(**α/2**) 에 해당하는 분포의 양끝의 Z값을 구하여야 한다.

[분포의 양 끝에 있는 Z값과 그 때의 오류확률의 크기와 위치]



V. 모평균의 신뢰구간

※ 모집단 평균이 μ이고 분산이 σ^2(**알려짐**)인 **정규분포일때**, μ의 신뢰구간

(i) 우선, μ에 대한 추정량은 $\overline{X}=\sum_{i=1}^{n} X_i/n$이다.

(ii) $\overline{X}\sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1^2)$. : 반드시 추정량의 분포를 이용해야 한다.

(iii) $\Pr\left(-z_{\alpha/2} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right)=1-\alpha$.

(iv) $\Pr\left(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \overline{X}-\mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)=1-\alpha$

$$\Rightarrow \Pr\left(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu-\overline{X} < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)=1-\alpha$$

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

(v) 그러므로, $1-\alpha$ 신뢰구간 $\left(\overline{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\ ,\ \overline{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$

이다. 또는 $\overline{X}\pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ 으로도 표현한다.

V. 모평균의 신뢰구간

◆ 모집단이 정규분포이고 모표준편차(σ)가 알려진 경우

모평균 μ에 대한 100(1−α)% 신뢰구간

$$P\left(\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

예1)다음 자료는 엑셀대학교 통계학 1반에서 실시한 퀴즈시험에서 무작위로 추출한 9명의 학생이 받은 20점 만점의 성적이다, 성적의 표준편차 2점으로 정규분포를 따른다고 한다. (6, 8, 10, 14, 14, 16, 18, 18, 20)

1) 모평균μ에 대한 90%, 95%, 99%의 신뢰구간을 구하라.

(풀이) 90% 신뢰구간

n=9, σ=2

$\sigma_{\overline{X}} = 2/\sqrt{9} = 0.667$

$\overline{X} = 13.778$

$$\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$$

$$13.778 - Z_{0.05}(0.667) \le \mu \le 13.778 + Z_{0.05}(0.667)$$

$$13.778 - (1.645)(0.667) \le \mu \le 13.778 + (1.645)(0.667)$$

$$= 12.681 \le \mu \le 14.875$$

V. 모평균의 신뢰구간

예1)다음 자료는 엑셀대학교 통계학 1반에서 실시한 퀴즈시험에서 무작위로 추출한 9명의 학생이 받은 20점 만점의 성적이다, 성적의 표준편차 2점으로 정규분포를 따른다고 한다. (6, 8, 10, 14, 14, 16, 18, 18, 20)

1) 모평균μ에 대한 90%, 95%, 99%의 신뢰구간을 구하라.

(풀이) 95% 신뢰구간

n=9, σ=2

$$\sigma_{\overline{X}} = 2/\sqrt{9} = 0.667$$

$$\overline{X} = 13.778$$

$$\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$$

$$13.778 - Z_{0.025}(0.667) \le \mu \le 13.778 + Z_{0.025}(0.667)$$

$$13.778 - (1.96)(0.667) \le \mu \le 13.778 + (1.96)(0.667)$$

$$= 12.471 \le \mu \le 15.085$$

99%신뢰구간

$$\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$$

$$13.778 - Z_{0.005}(0.667) \le \mu \le 13.778 + Z_{0.005}(0.667)$$

$$13.778 - (2.575)(0.667) \le \mu \le 13.778 + (2.575)(0.667)$$

$$= 12.060 \le \mu \le 15.496$$

V. 모평균의 신뢰구간

◆ 모집단이 정규분포이고 모표준편차(σ)를 모르는 경우:소표본

모집단의 평균과 표준편차를 모르고 표본크기가 n<30 인 경우는 표준정규분포를 따르지 않고 자유도 (n-1)의 t-분포를 따른다.

평균 μ 인 정규모집단으로부터 크기 n의 표본을 무작위로 추출했을 때 그 평균이 $\overline{X}$ 이고 표본표준편차가 s 일 때 표본통계량 t는

$$t = \frac{\overline{X} - \mu_{\overline{X}}}{S / \sqrt{n}}$$

으로 자유도 (n-1)인 t-분포를 따른다.

V. 모평균의 신뢰구간

◆ 모집단이 정규분포이고 모표준편차(σ)를 모르는 경우

모평균 μ에 대한 100(1−α)% 신뢰구간(소표본)

$$P(\overline{X} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}} \le \mu \le \overline{X} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}) = 1-\alpha$$

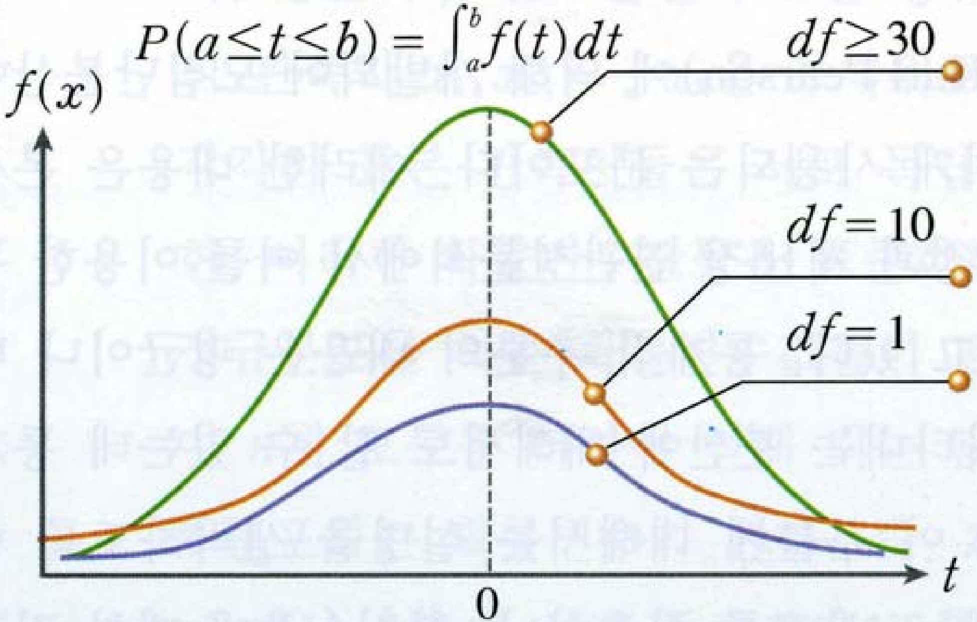
모평균 μ에 대한 100(1−α)% 신뢰구간(대표본)

$$P(\overline{X} - Z_{\alpha/2}\frac{S}{\sqrt{n}} \le \mu \le \overline{X} + Z_{\alpha/2}\frac{S}{\sqrt{n}}) = 1-\alpha$$

t분포

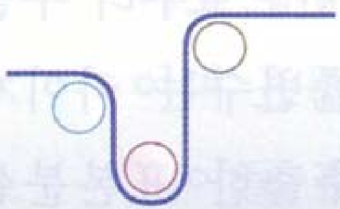
- 평균이 0인 구릉이나 종모양의 대칭분포
- 자유도(df)에 따라 그 모양이 변하고,
  자유도가 30개 이상($df \geq 30$)이면 표준정규분포(Z분포)와 거의 일치함

자유도에 따른 t분포의 모양



t분포의 특징

- 표준정규분포(Z분포)와 같이 평균이 0임
- 자유도(df)에 따라 분포의 모양이 변함
- 자유도가 30 미만($df < 30$)인 경우, 표준정규분포에 비해 양쪽 끝이 평평하고 두터운 꼬리모양을 가짐
- 자유도(df)가 증가함에 따라 분산은 1에 접근함
- 표본의 크기(n)가 커질수록 자유도가 증가하여 표본크기가 30개 이상일 경우($df \geq 30$)에는 표준정규분포(Z분포)와 거의 동일한 분포를 함



확률변수 $t_{(n-1)} = \dfrac{\bar{x} - \mu}{\dfrac{S}{\sqrt{n}}}$

t분포표

자유도	오른쪽 꼬리면적 α							
	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	3.078	6.314	12.706	31.821	63.656	127.321	318.289	636.578
2	1.886	2.920	4.303	6.965	9.925	14.089	22.328	31.600
3	1.638	2.353	3.182	4.541	5.841	7.453	10.214	12.924
4	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	4.773	5.894	6.869
6	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.689
28	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.660
30	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373

예2) t값과 그들의 자유도 및 단측 또는 양측이 다음과 같을 때 t-분포표를 이용하여 그들의 확률 α를 구하라.

t=1.476, df=5, 단측 (풀이) P(t(5)>1.476)= α=0.1

1) t=1.699, df=29, 양측 (풀이) P(t(29)>1.699)= α=0.05 x 2=0.1

예3) 다음과 같이 오차율과 지유도가 주어졌을 때 t-분포표를 이용하여 t값을 구하라.

1) α=0.05, df=15 (풀이) t=1.735

2) α=0.01, df=1 (풀이) t=31.821

예4) 정규분포를 따르는 모집단에서 표본을 12개 추출하여 어떤 물건의 무게를 측정한 결과 다음과 같은 자료를 얻었다. 이 물건의 모평균 무게 μ에 대한 95% 신뢰구간을 구하라.(13, 20, 19, 40, 25, 10, 21, 29, 17, 11, 7, 18)

(풀이) $\overline{X}$ =(13+20+……..+18)/12=19.167

S^2 =(∑X²−$n\overline{X}^2$)=82.879 S=9.104 $t_{11,\,0.025} = 2.201$

∴ 신뢰하한: 19.167−2.201 x (9.104/ $\sqrt{12}$)=13.382

신뢰상한: 19.167+2.201 x (9.104/ $\sqrt{12}$)=24,592

예5) 전화회사는 주말 장거리 전화의 평균 통화시간을 추정하려고 한다. 40통화를 무작위로 추출하여 조사한 결과 다음과 같은 자료를 얻었다. (S=4.4503 $\overline{X}$ =8.8)

Z분포와 t분포를 이용하여 주말 장거리 전화의 평균 통화시간 μ에 대한 95% 신뢰구간을 구하라.

(풀이)

∴

$$\overline{X} \pm Z_{\alpha/2} \frac{S}{\sqrt{n}}$$

$$8.8 \pm 1.96\left(\frac{4.4503}{\sqrt{40}}\right) \quad \therefore \text{신뢰하한} : 7.421 \quad \text{신뢰상한} : 10.179$$

$$\overline{X} \pm t_{n-1,\alpha/2} \frac{S}{\sqrt{n}}$$

$$8.8 \pm 2.021\left(\frac{4.4503}{\sqrt{40}}\right) \quad \therefore \text{신뢰하한} : 7.378 \quad \text{신뢰상한} : 10.222$$