제 7장

표본분포

고려대학교 경영대학 박 광태

단순확률표본

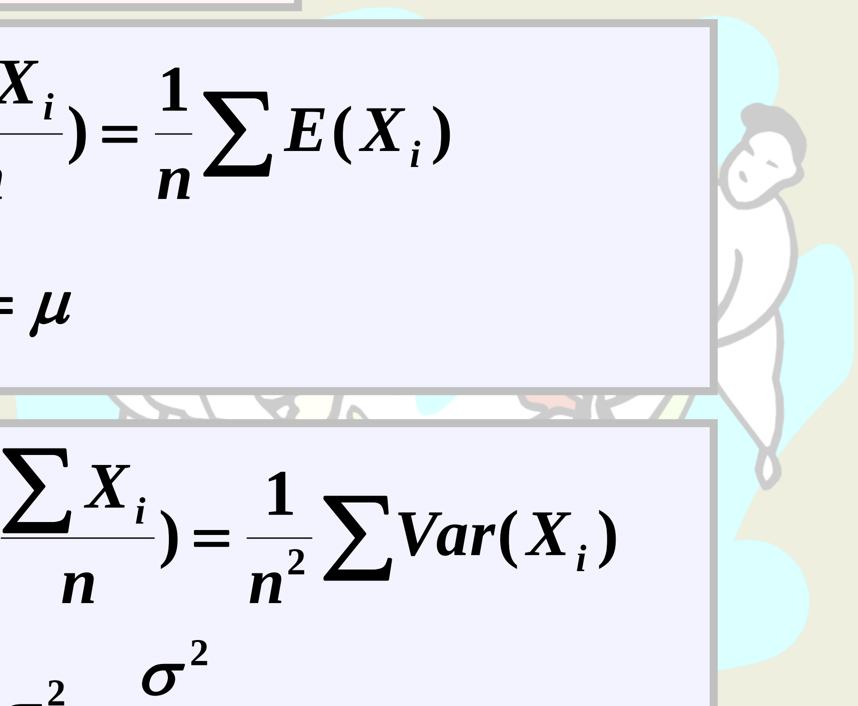
- 단순확률표본
  - 단순확률표본이란 매번 표본관찰치를 추출할 때마다 모집단의 각 원소들이 선택될 확률이 동일하도록 하여 얻어진 표본을 말한다.
  - 즉 난수표를 이용하여 표본을 무작위로 추출

표본평균의 분포

- 모집단의 분포가 정규분포 $X \sim N(\mu, \sigma^2)$ 일 때
- 표본평균의 분포는 $X \sim N(\mu, \frac{\sigma^2}{n})$ 임
- 참고

$$E(\overline{X}) = E(\frac{\sum X_i}{n}) = \frac{1}{n}\sum E(X_i)$$
$$= \frac{1}{n} n\mu = \mu$$

$$Var(\overline{X}) = Var(\frac{\sum X_i}{n}) = \frac{1}{n^2}\sum Var(X_i)$$
$$= \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$



중심극한정리 (central limit theorem)

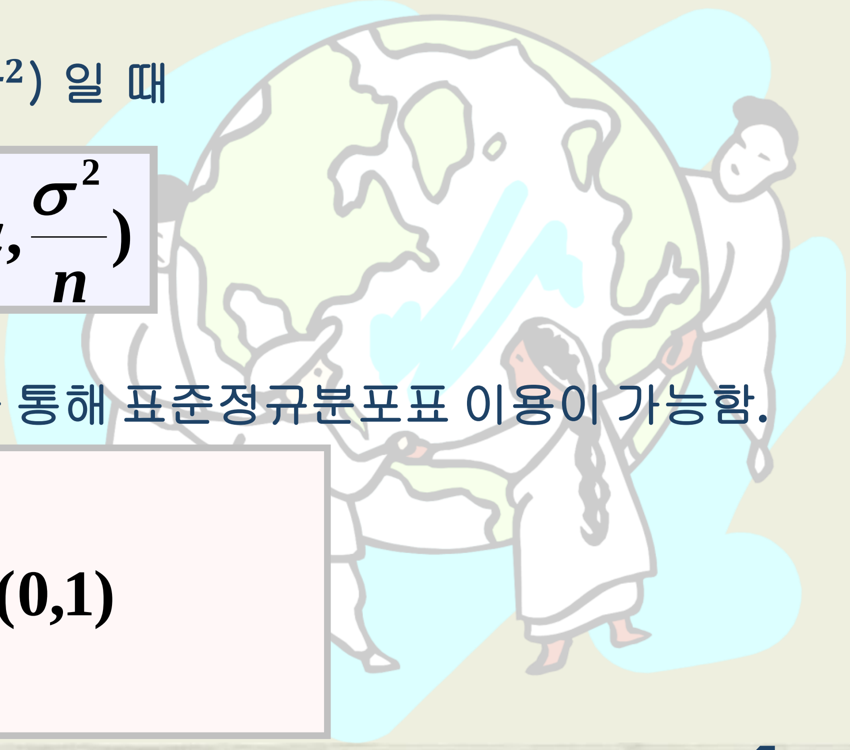
- 평균이 μ 이고, 분산이 σ^2 인 어떤 모집단으로부터 n개의 표본을 선택했을 때 표본평균 $\bar{X}$는 모집단의 분포에 상관없이 n이 충분히 클 때 평균이 μ 이고, 분산이 $\frac{\sigma^2}{n}$인 정규분포를 따른다.

- 모집단의 분포가 정규분포 $N(\mu\ ,\ \sigma^2)$ 일 때

  표본평균의 분포는 $$\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$$

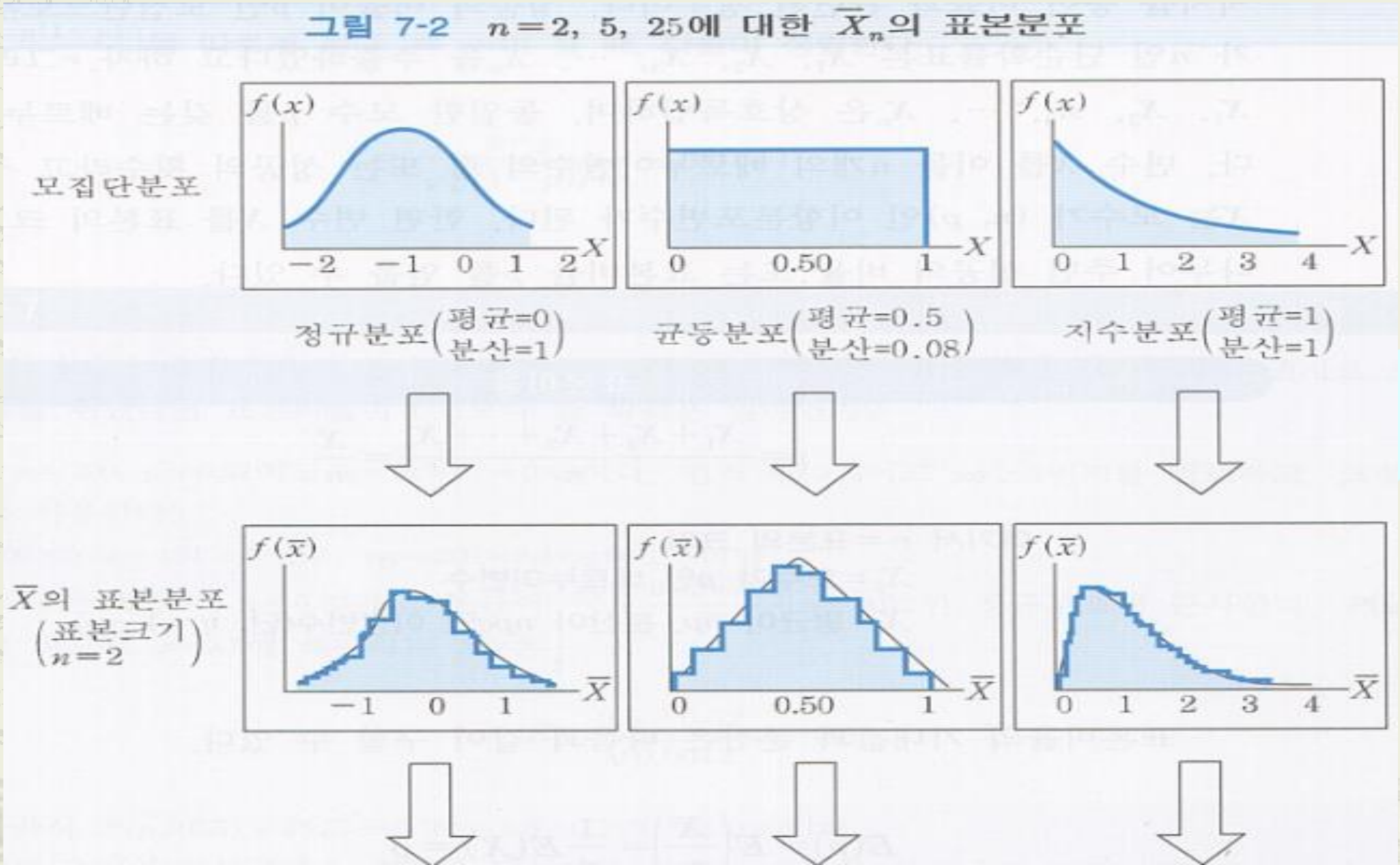
- 정규분포를 따른다면 표준화변환을 통해 표준정규분포표 이용이 가능함. 즉,

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



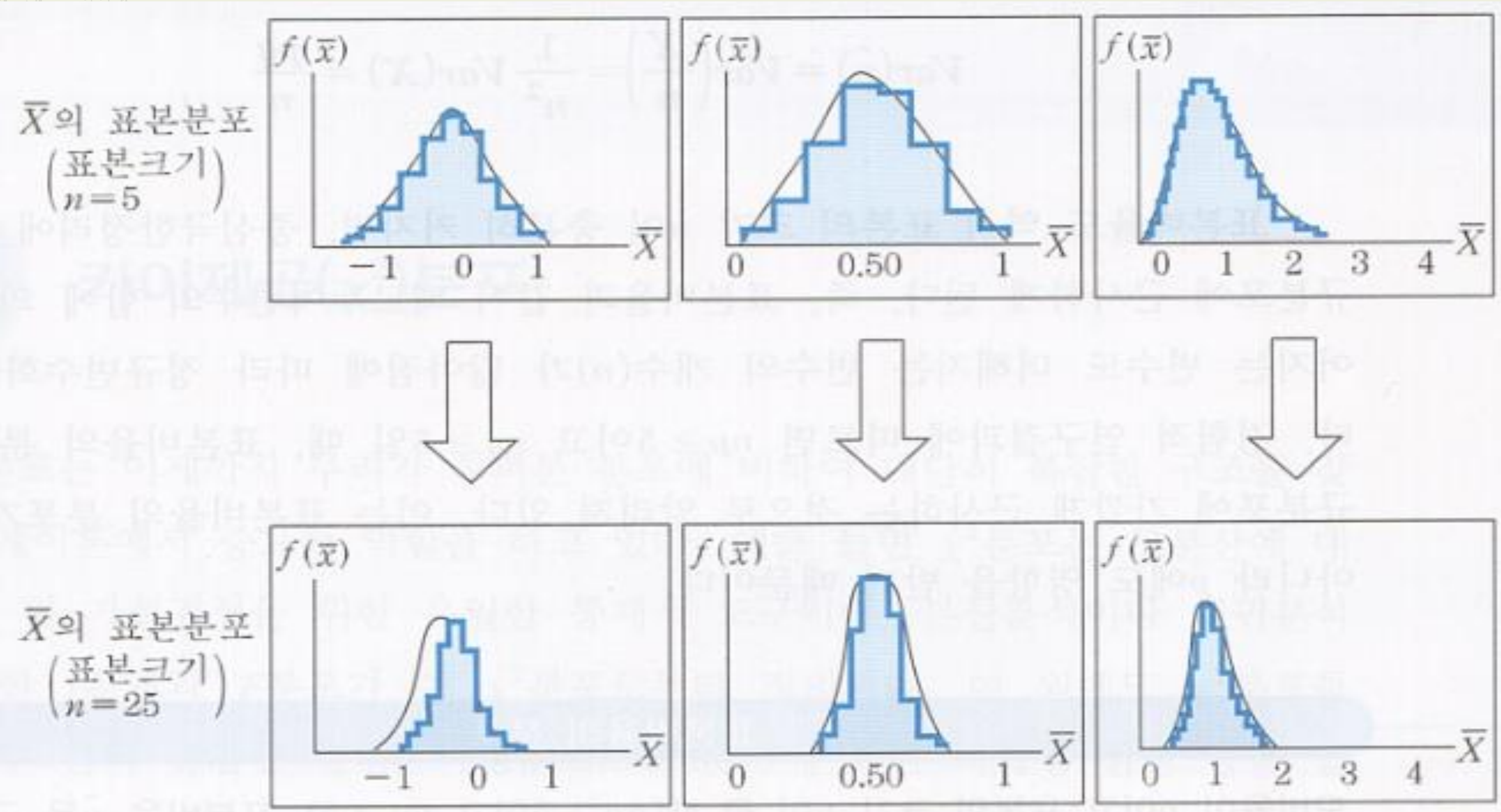
중심극한정리 (central limit theorem)

- 그림 7-2



중심극한정리 (central limit theorem)

- 그림 7-2(계속)



예제 7-6

천마연구소는 연구원의 평균월급(단위: 천 원)을 추정하고자 한다. 모평균을 추정하기 위하여 연구소는 50명의 연구원을 확률표본으로 취하였다. 모평균이 2,000이고 모분산은 300^2이라고 가정할 때, 표본평균과 모평균의 차가 80 이내일 확률은 얼마인가?

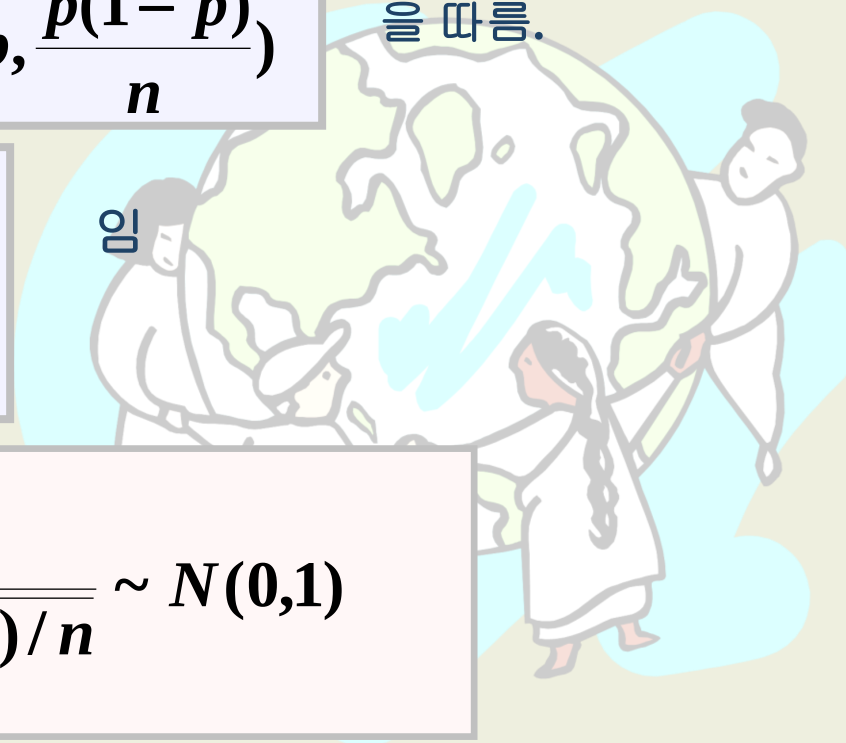
표본비율의 분포

- 모비율이 p 이고, 표본의 크기 n이 클 때

$np \geq 5$ 이고, $nq \geq 5$ $(q = 1 - p)$ 일 때

표본비율 $\hat{P}$ 는 근사적으로 $N(p, \frac{p(1-p)}{n})$ 을 따름.

여기서 $\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$ 임



- 이 경우

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

예제 7-7

대통령 선거에서 유권자의 52%가 후보 K를 선호하는 것으로 나타났다. 200명의 유권자로 이루어진 확률표본을 취한다면 표본비율이 0.5보다 클 확률은 얼마인가?

χ^2 분포

- χ^2변수의 정의: **k**개의 상호독립하는 정규변수

$$x_i \sim N(\mu_i, \sigma_i^2), i - 1, 2, 3, \cdots, k$$

가 있을때

$$\sum_{i=1}^{k} Z_i^2 = \left(\frac{X_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{X_2 - \mu_2}{\sigma_2}\right)^2 + \left(\frac{X_3 - \mu_3}{\sigma_1}\right)^2 + \cdots + \left(\frac{X_k - \mu_k}{\sigma_k}\right)^2$$

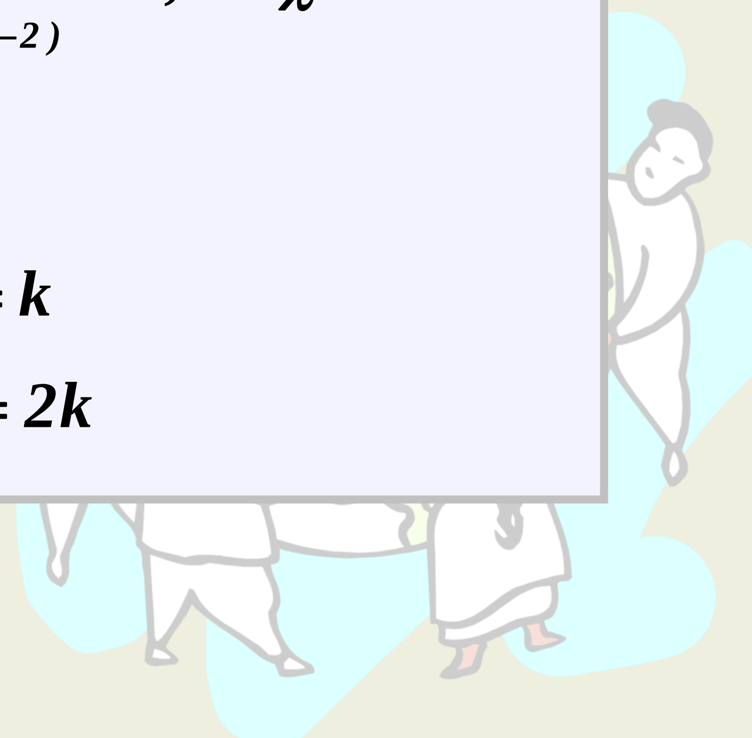
은 자유도가 **k** 인 χ^2변수이다.

χ^2 분포

$$f_k(\chi^2) = \frac{(\chi^2)^{\frac{(k-2)}{2}} \exp\left(\frac{-\chi^2}{2}\right)}{\left(\frac{k-2}{2}\right)! 2^{(k-2)}}, 0 < \chi^2 < \infty$$

$$E(\chi_k^2) = k$$

$$Var(\chi_k^2) = 2k$$

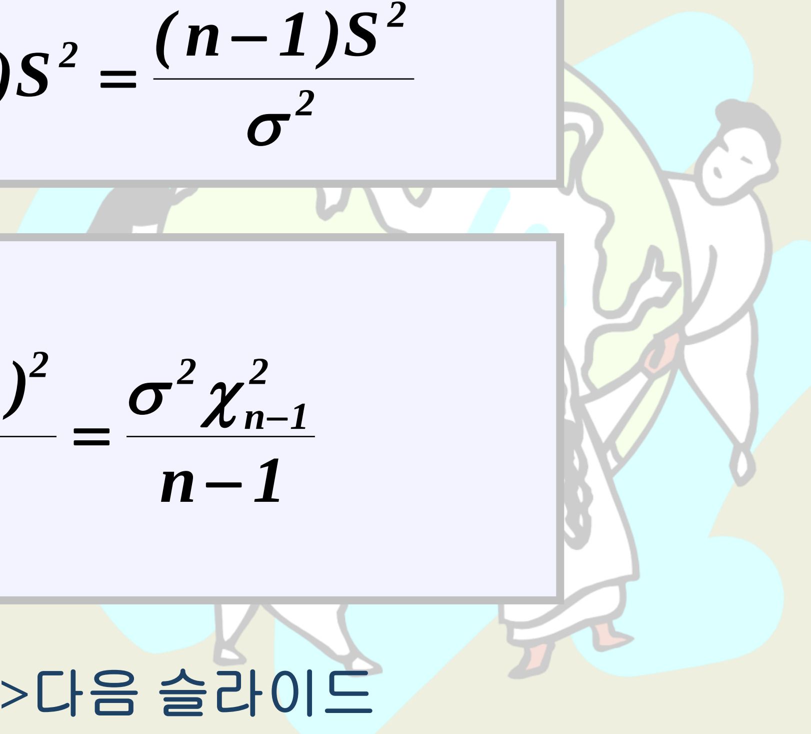


χ^2 변수와 표본분산의 관계

$$\chi^2_{n-1} = \frac{1}{\sigma^2}\sum (X_i - \overline{X}_n)^2$$

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

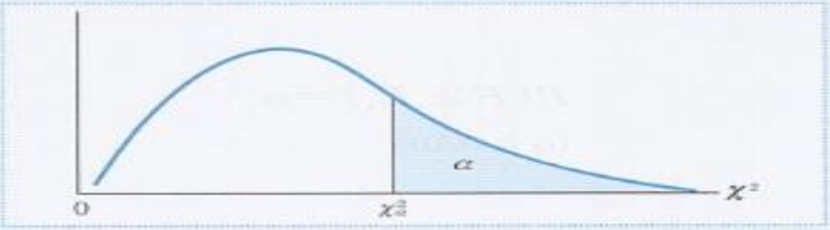
$$S^2 = \frac{\sum (x_i - \overline{x})^2}{n-1} = \frac{\sigma^2 \chi^2_{n-1}}{n-1}$$



※ 참조　χ^2분포표 보는 방법 ->다음 슬라이드

χ^2 분포표

$$P(\chi^2 \geq \chi^2_\alpha) = \alpha$$



$d.f.$	$\chi^2_{0.995}$	$\chi^2_{0.990}$	$\chi^2_{0.975}$	$\chi^2_{0.950}$	$\chi^2_{0.900}$	$\chi^2_{0.100}$	$\chi^2_{0.050}$	$\chi^2_{0.025}$	$\chi^2_{0.010}$	$\chi^2_{0.005}$
1	.0000390	.000157	.000982	.00393	.0158	2.705	3.841	5.024	6.635	7.879
2	.0100	.0201	.0506	.103	.210	4.605	5.991	7.378	9.210	10.597
3	.0717	.115	.216	.352	.584	6.251	7.815	9.348	11.345	12.838
4	.207	.297	.484	.711	1.063	7.779	9.488	11.143	13.277	14.860
5	.412	.554	.831	1.145	0.610	9.236	11.070	12.832	15.086	16.750
6	.676	.872	1.237	1.635	2.204	10.644	13.592	14.449	16.812	18.548
7	.989	1.239	1.690	2.167	2.283	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.489	13.361	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.683	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
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.041	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.789	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.650	27.204	30.114	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.558
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.088	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.706	22.164	24.433	26.509	29.050	51.805	55.758	59.342	63.691	66.765
50	27.990	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.535	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	91.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

예제 7-10

표본평균으로 대체하여 도출한 χ^2변수는 자유도가 하나 감소한다는 사실을 $n=2$인 경우에 대하여 증명하시오.

예제 7-11

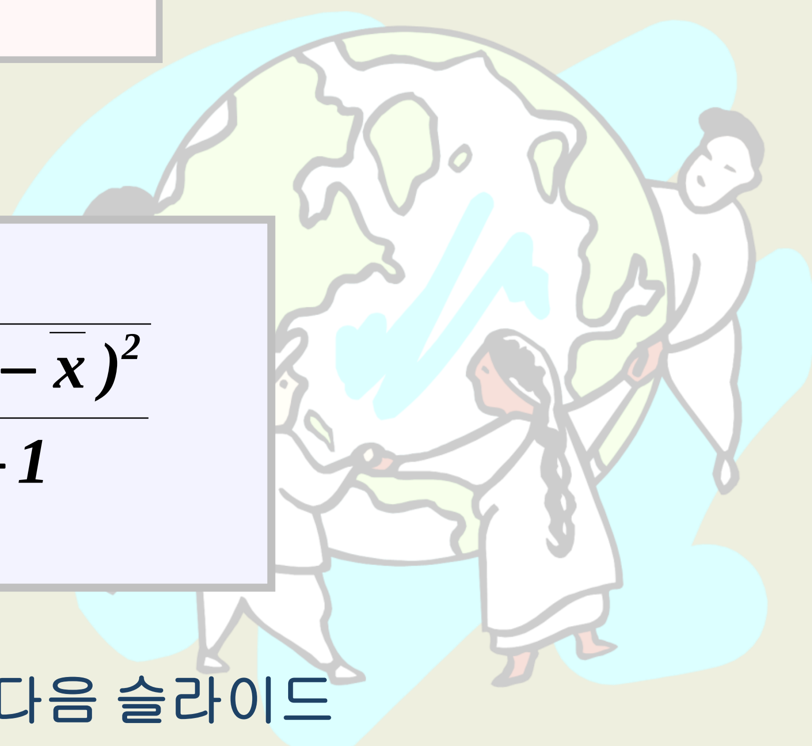
경영대학 구내의 커피자판기는 한 컵당 μg씩의 커피가 나오도록 조정되어 있으며 표준편차가 1인 정규분포를 가지는 것으로 알려져 있다. 어느 날 커피자판기에서 나오는 커피의 양에 대해서 10개를 무작위로 추출하여 각각에 대한 무게를 측정하였다. 이때 계산되는 표본분산이 1.88보다 클 확률은 얼마인가?

t 분포

$$t = \frac{\overline{X} - \mu}{S / \sqrt{n}} \sim t_{n-1}$$

여기서

$$S = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n - 1}}$$



※ t 분포표 보는 방법 -> 다음 슬라이드

t분포표

$P(t \geq t_\alpha) = \alpha$

$d.f.$	$t_{0.100}$	$t_{0.050}$	$t_{0.025}$	$t_{0.010}$	$t_{0.005}$	$d.f.$
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	4.303	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	1.533	2.132	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	2.365	2.998	3.499	7
8	1.397	1.860	2.306	2.896	3.355	8
9	1.383	1.833	2.262	2.821	3.250	9
10	1.372	1.812	2.228	2.764	3.169	10
11	1.363	1.796	2.201	2.718	3.106	11
12	1.356	1.782	2.179	2.681	3.055	12
13	1.350	1.771	2.160	2.650	3.012	13
14	1.345	1.761	2.145	2.624	2.977	14
15	1.341	1.753	2.131	2.602	2.947	15
16	1.337	1.746	2.120	2.583	2.921	16
17	1.333	1.740	2.110	2.567	2.898	17
18	1.330	1.734	2.101	2.552	2.878	18
19	1.328	1.729	2.093	2.539	2.861	19
20	1.325	1.725	2.086	2.528	2.845	20
21	1.323	1.721	2.080	2.518	2.831	21
22	1.322	1.717	2.074	2.508	2.819	22
23	1.319	1.714	2.069	2.500	2.807	23
24	1.318	1.711	2.064	2.492	2.797	24
25	1.316	1.708	2.060	2.485	2.787	25
26	1.315	1.706	2.056	2.479	2.779	26
27	1.314	1.703	2.052	2.473	2.771	27
28	1.313	1.701	2.048	2.467	2.763	28
29	1.311	1.699	2.045	2.462	2.756	29
inf.	1.282	1.645	1.960	2.326	2.576	inf.

예제 7-12

천호기업에서 생산하는 철사의 강도는 정규분포를 따른다. 생산공정으로부터 일정한 길이의 철사 6가닥을 임의로 추출하여 강도를 측정하였다. 표본평균을 $\overline{X}_6$, 표본분산을 S^2이라고 할 때, $\overline{X}_6$과 미지의 모수 μ와의 차이가 $\frac{2S}{\sqrt{n}}$ 이내일 확률은 얼마인가?