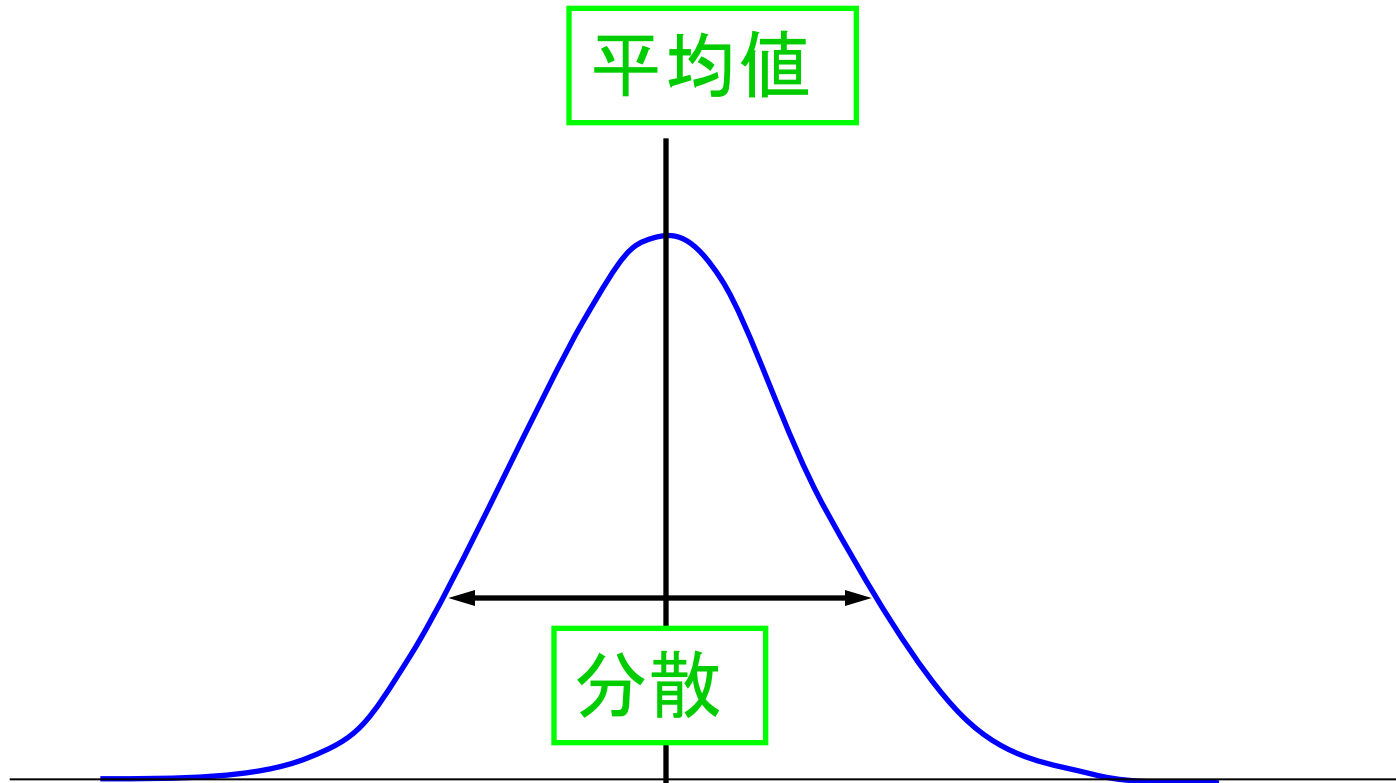


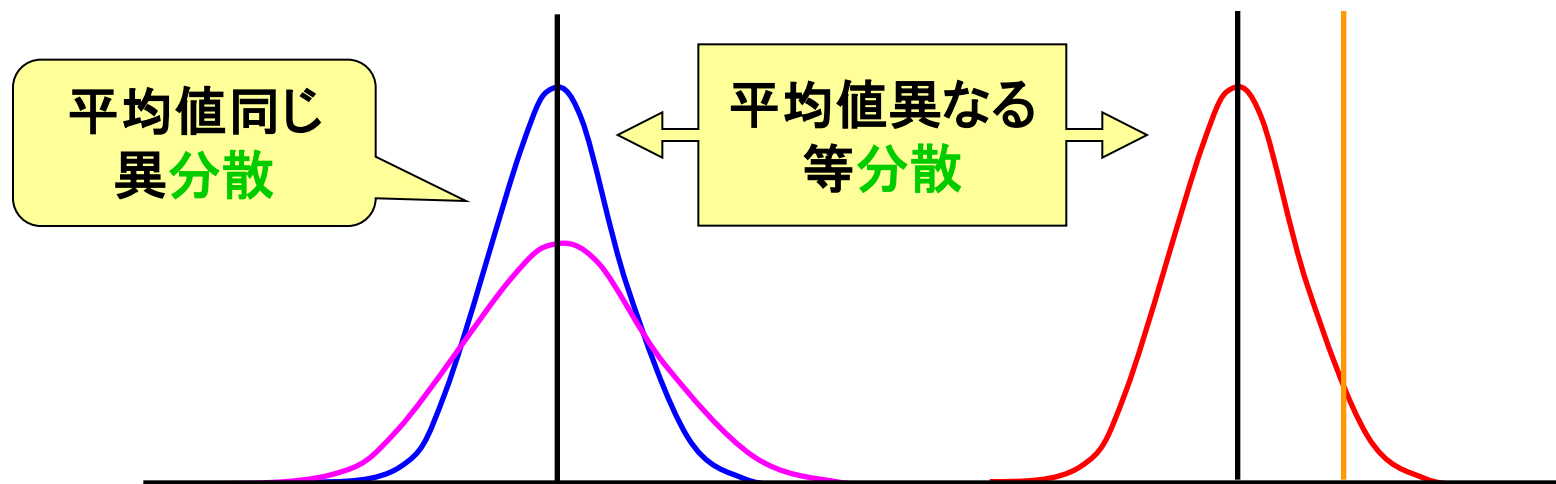
統計的検定 その1

分布は、平均値と分散(ばらつき)で決まる



検定とは、「比較によって母集団の様子を推測すること」

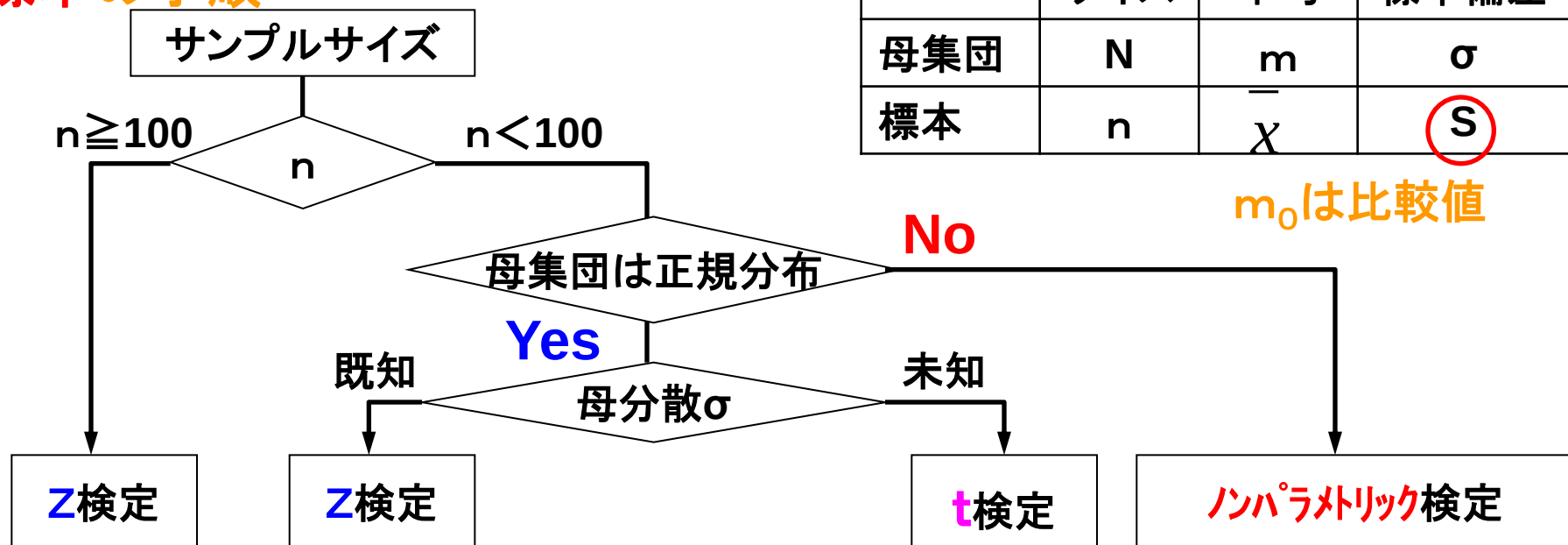
平均値、分散



- 検定
- 1 標本の検定・・・ ある一定値と比較し、
差があるか無いか
 - 2 標本の検定・・・ 2つの集団を比較し、
差があるか無いか

母集団の分布 標本サイズ	正規分布		その他の分布
	σ既知	σ未知	
$n \geq 100$	Z検定	Z検定	Z検定
$n < 100$	Z検定	t検定	ノンパラメトリック検定

1標本の手順



	サイズ	平均	標準偏差
母集団	N	m	σ
標本	n	$\bar{x}$	S

m_0 は比較値

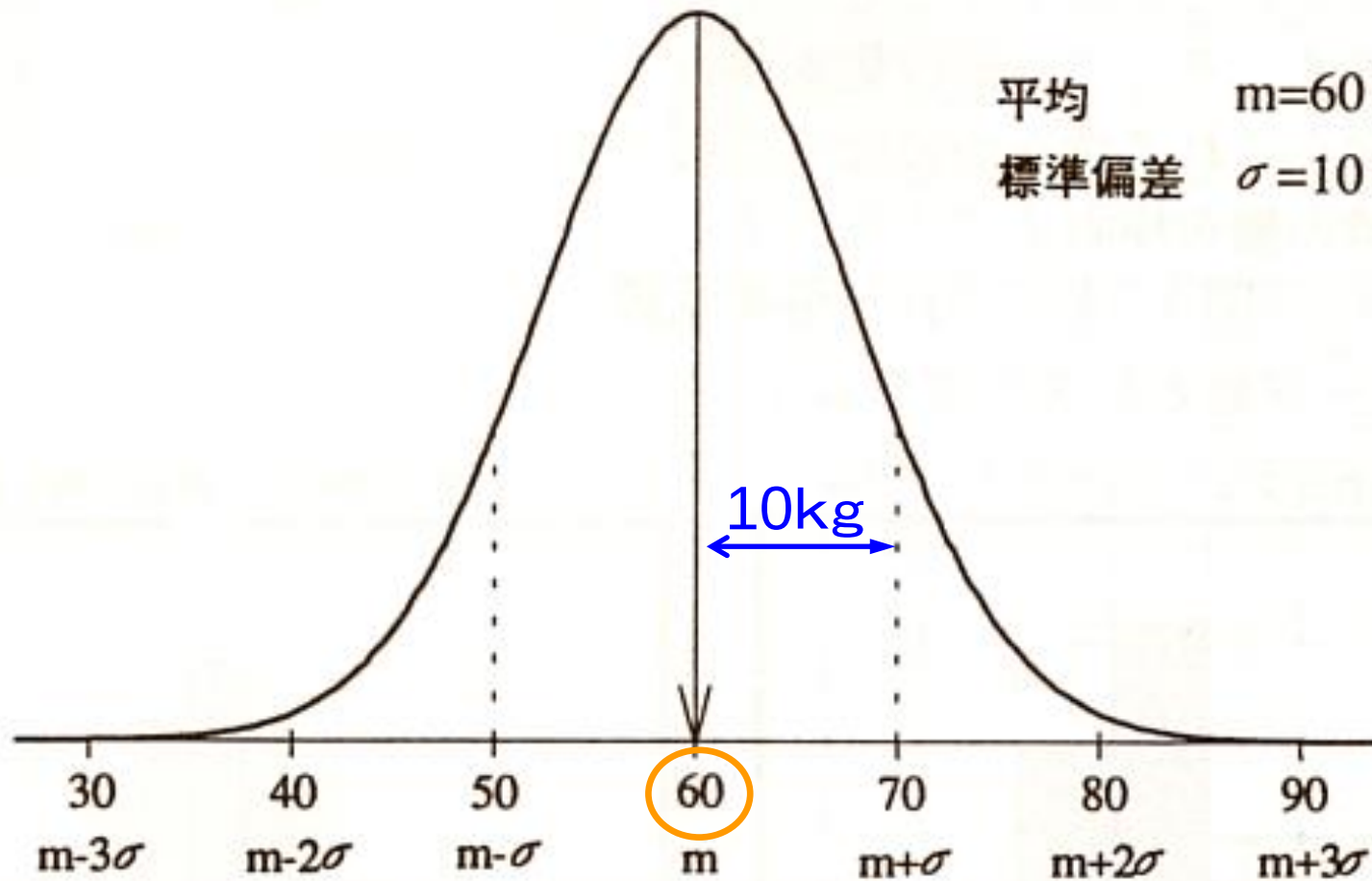
$$T = \frac{\bar{x} - m_0}{s/\sqrt{n}} \quad T = \frac{\bar{x} - m_0}{\sigma/\sqrt{n}}$$

σ未知の時は標本のsを使う

$$T = \frac{\bar{x} - m_0}{s/\sqrt{n-1}}$$

平均体重が60kg、標準偏差が10kgの分布

図2 正規分布

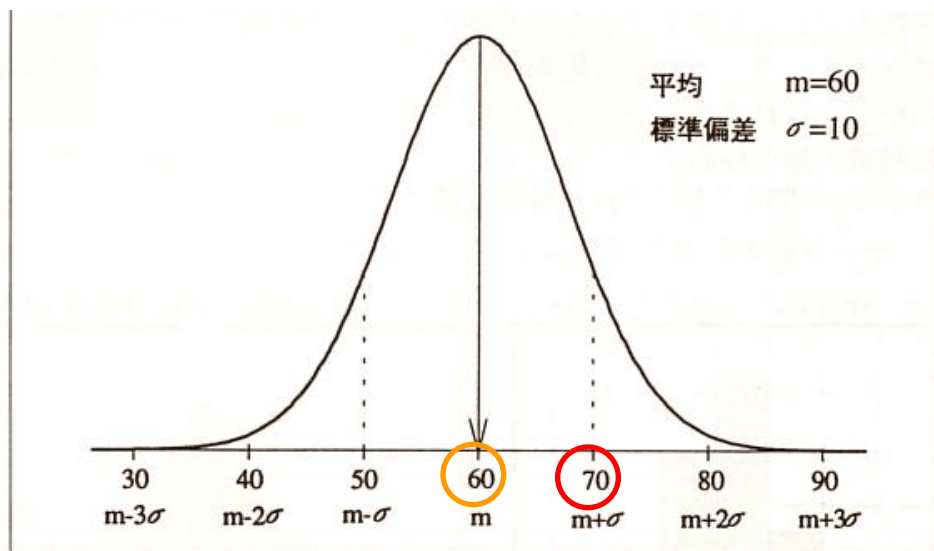


なぜ平方根にするか？



単位: kg

平均体重が60kg、標準偏差が10kgの分布



平均 $m=60$

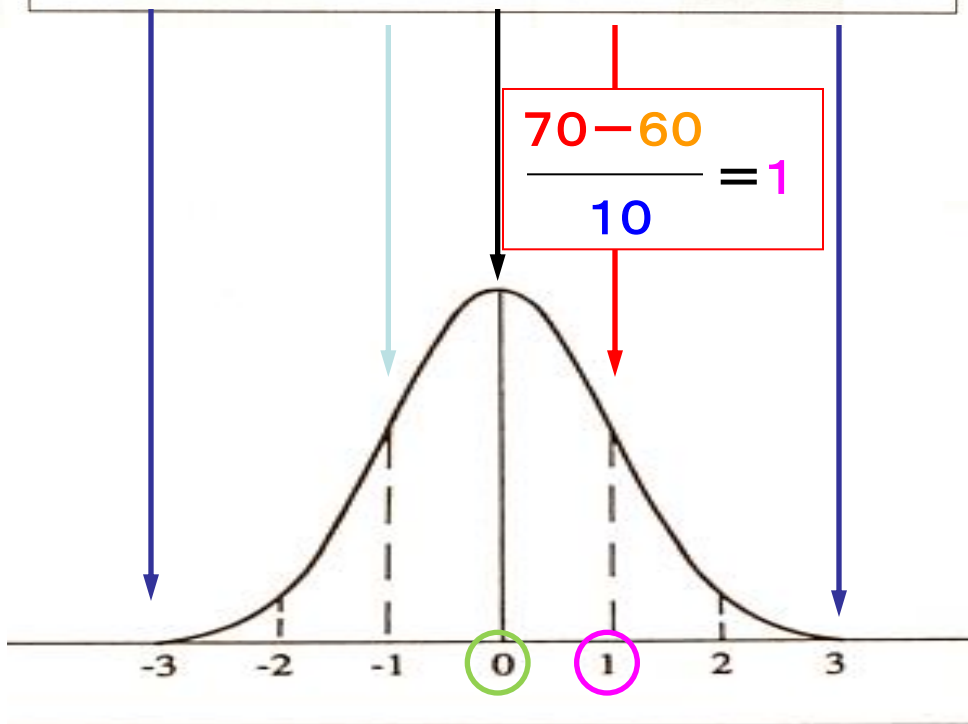
標準偏差 $\sigma=10$

基準化(標準化) ★

$$\text{統計量 } T = \frac{\text{データ} - \text{平均値}}{\text{標準偏差}}$$

平均 $m=0$

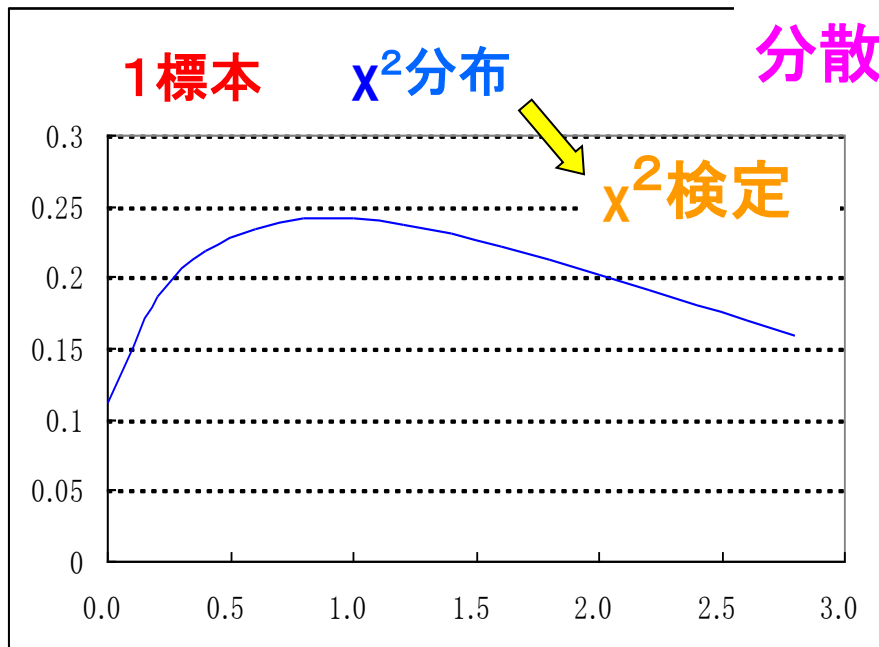
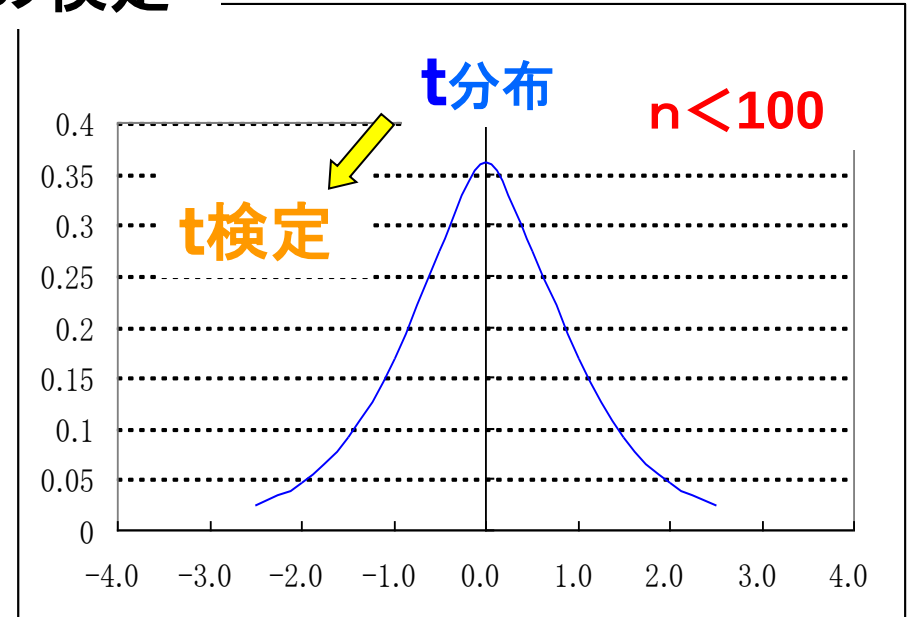
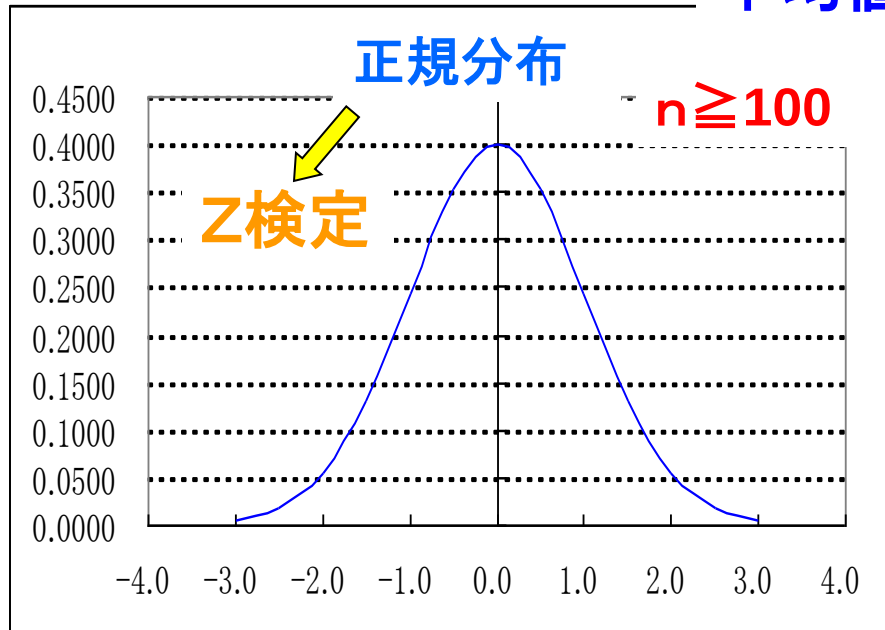
標準偏差 $\sigma=1$



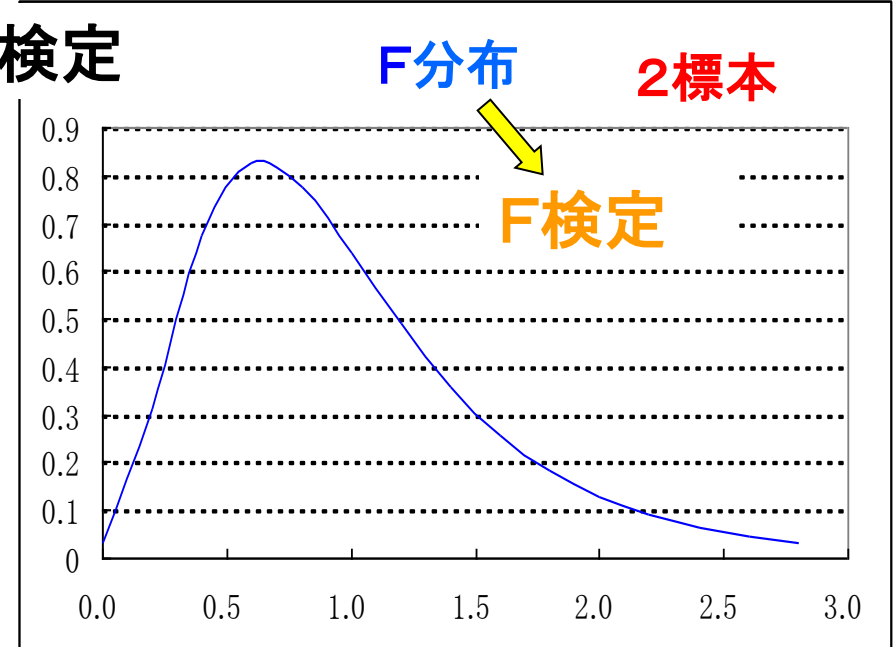
$$\frac{70 - 60}{10} = 1$$

検定に使用する分布は？

平均値の検定

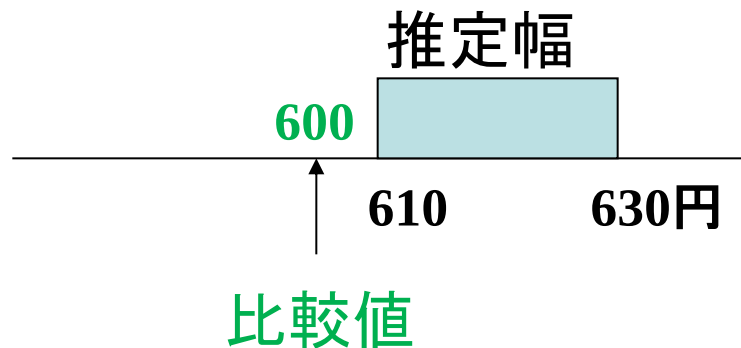


分散の検定

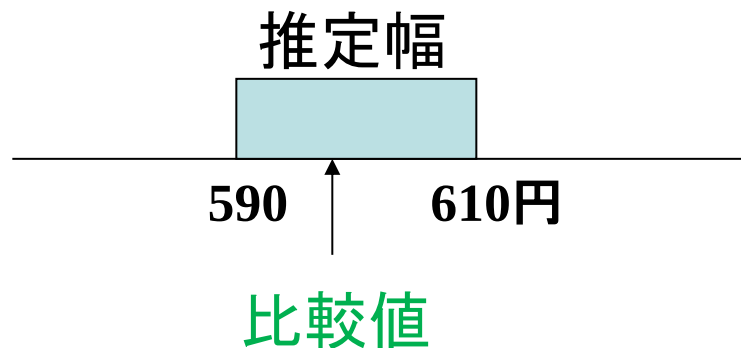


母平均の検定、1標本の検定

差があるといえる



差があるといえない



「帰無仮説」

主張したい内容の逆の仮説
説をなきものにしたい逆説
そこで無に帰したい仮説
差がない
効果がない

「対立仮説」

証明したい仮説
主張したい仮説

検定の手順

- ①帰無仮説、対立仮説を立てる
- ②有意水準(危険率) α を定める
- ③統計量を求める
- ④判定する

帰無仮説	Aの母平均 μ は比較値 m_0 と差がない	∴1標本
対立仮説	$\mu \neq m_0$ 、 $\mu > m_0$ 、 $\mu < m_0$	

例 1標本の場合

帰無仮説 : Aのラーメンの平均価格は、比較値600円に等しい

対立仮説a: Aのラーメンの平均価格は、比較値600円に等しくない → 両側検定

対立仮説b: Aのラーメンの平均価格は、比較値600円よりも高い → 右側検定

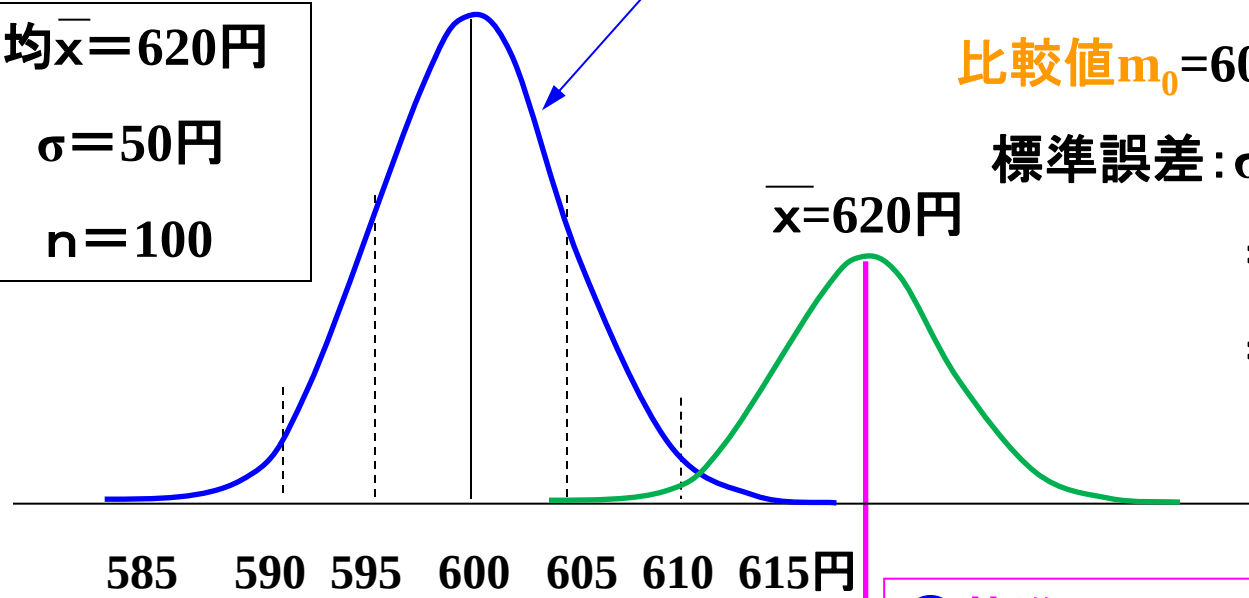
対立仮説c: Aのラーメンの平均価格は、比較値600円よりも低い → 左側検定

東京(A)のラーメン価格の平均値が μ
全国平均(比較値)が m_0

1 標本の場合、 $n \geq 100$ 母集団の標準偏差($\sigma = 50$ 円)既知 → Z検定

①帰無仮説: 東京のラーメンの平均価格は600円に等しい(差がない)

標本の平均 $\bar{x} = 620$ 円
標準偏差 $\sigma = 50$ 円
サンプル数 $n = 100$

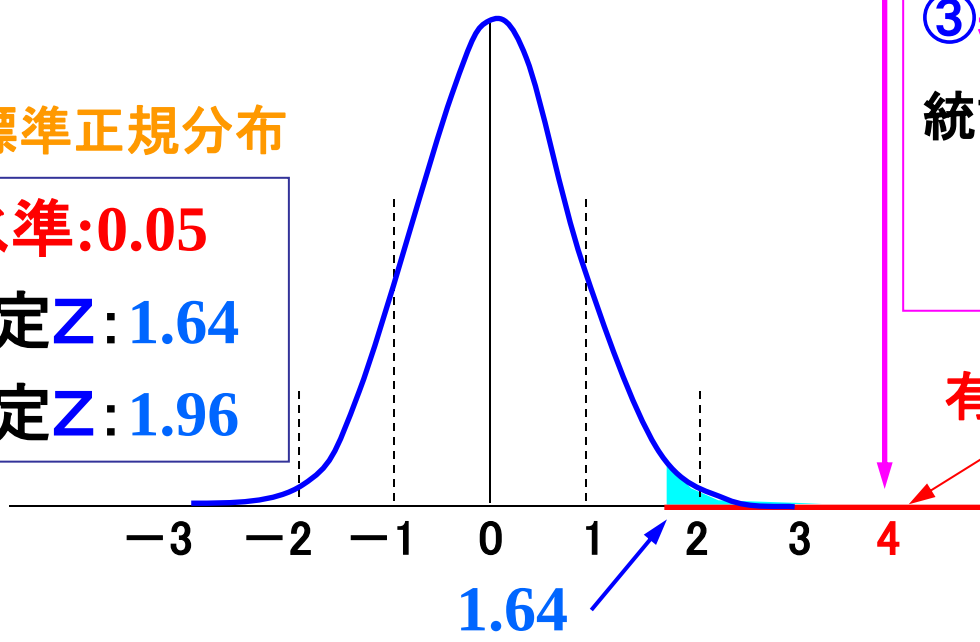


比較値 $m_0 = 600$ 円

標準誤差: $\sigma / \sqrt{n}$
 $= 50 / \sqrt{100}$
 $= 5$ 円

標準正規分布

②有意水準: 0.05
片側検定 Z: 1.64
両側検定 Z: 1.96



③基準化
統計量 $T = (\bar{x} - m_0) / (\sigma / \sqrt{n})$
 $= (620 - 600) / 5$
 $= 4$

有意水準0.05の棄却域

④ $T = 4 > 1.64$ (棄却)
P値 < 0.05 (棄却)

平均価格は600円より高い

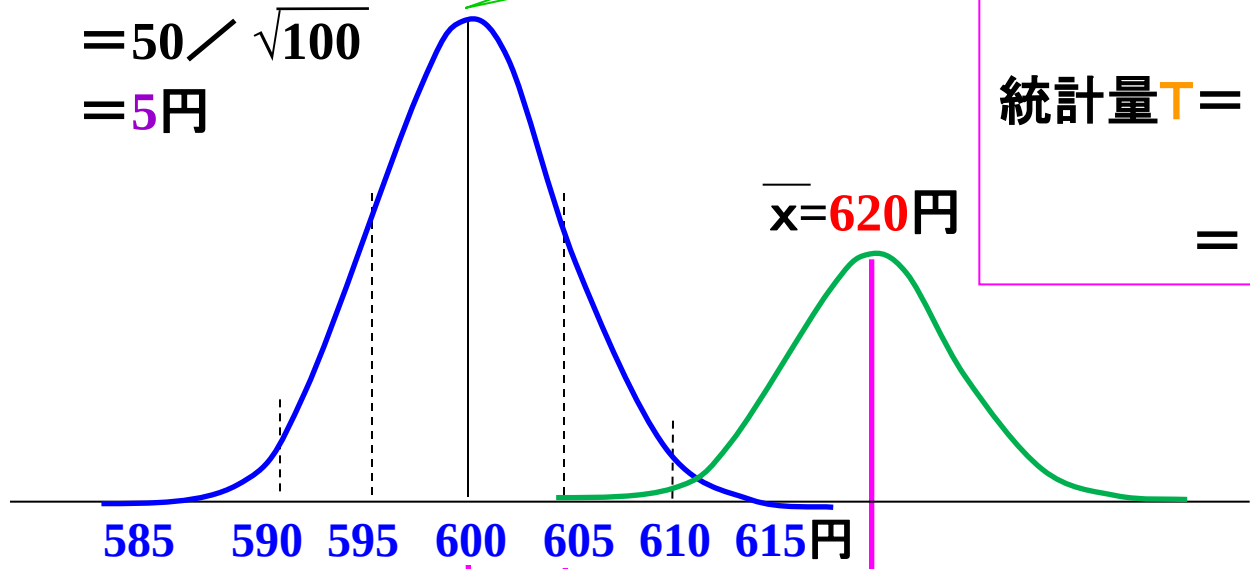
標本平均の標準偏差(標準誤差)

$\sigma / \sqrt{n} \doteq s / \sqrt{n}$
 $= 50 / \sqrt{100}$
 $= 5 \text{ 円}$

比較値

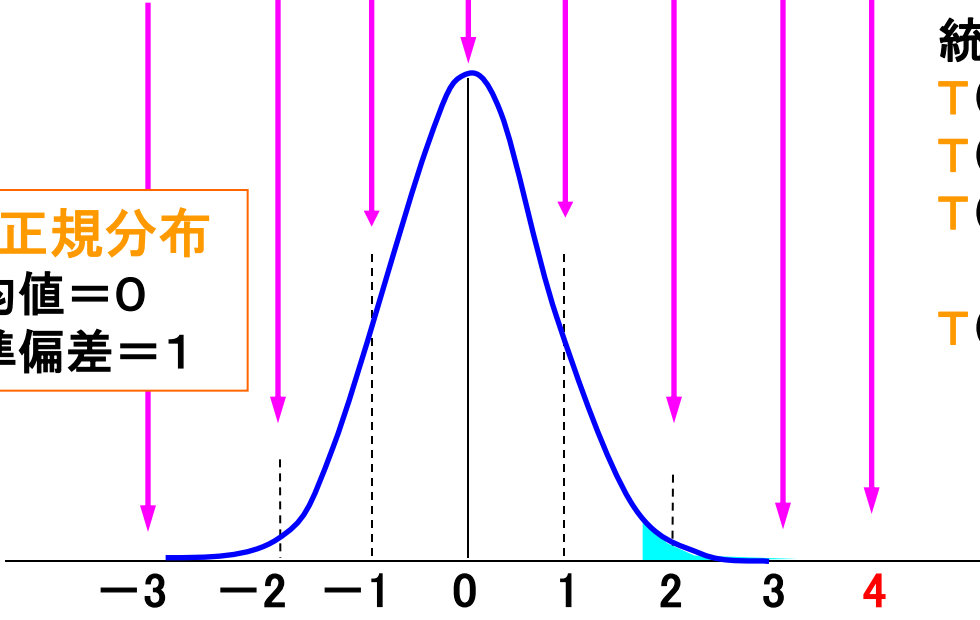
基準化

統計量 $T = \frac{\text{データ} - \text{比較値}}{\text{標準誤差}}$
 $= (\bar{x} - m_0) / (s / \sqrt{n})$



標準正規分布
平均値 = 0
標準偏差 = 1

統計量	データ	比較値
$T(585)$	$(585 - 600) / 5 = -3$	
$T(600)$	$(600 - 600) / 5 = 0$	
$T(615)$	$(615 - 600) / 5 = 3$	
$T(620)$	$(620 - 600) / 5 = 4$	



棄却域限界のZ値

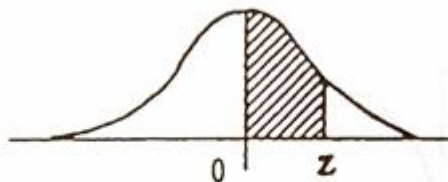
付表(1) 標準正規分布表

有意水準0.05(5%) = 信頼度95%

片側検定 $50 - 5 = 45\%$ → 1.64

両側検定 $50 - 5 / 2 = 47.5\%$ → 1.96

値Zに対する確率



1σの確率

$$0.34134 \times 2$$

$$= 0.68268$$

2σの確率

$$0.47725 \times 2$$

$$= 0.9545$$

3σの確率

$$0.49865 \times 2$$

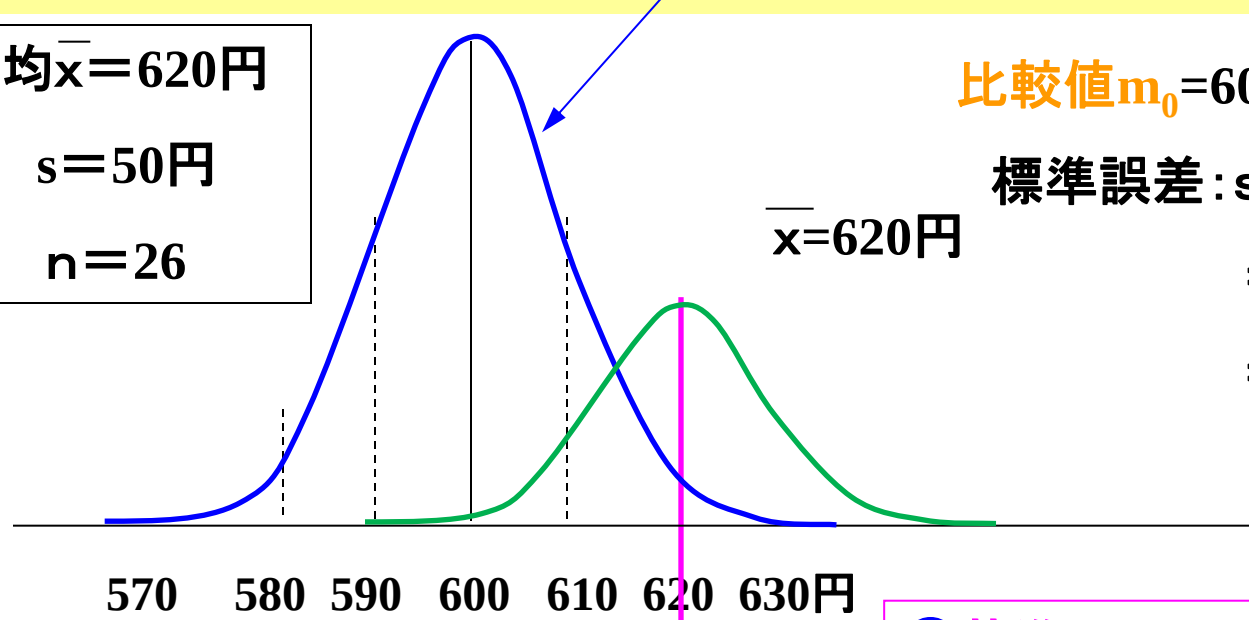
$$= 0.9973$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.00000	0.00399	0.00798	0.01197	0.01595	0.01994	0.02392	0.02790	0.03188	0.03586
0.1	.03983	.04380	.04776	.05172	.05567	.05962	.06356	.06749	.07142	.07535
0.2	.07926	.08317	.08706	.09095	.09483	.09871	.10257	.10642	.11026	.11409
0.3	.11791	.12172	.12552	.12930	.13307	.13683	.14058	.14431	.14803	.15173
0.4	.15542	.15910	.16276	.16640	.17003	.17364	.17724	.18082	.18439	.18793
0.5	.19146	.19497	.19847	.20194	.20540	.20884	.21226	.21566	.21904	.22240
0.6	.22575	.22907	.23237	.23565	.23891	.24215	.24537	.24857	.25175	.25490
0.7	.25804	.26115	.26424	.26730	.27035	.27337	.27637	.27935	.28230	.28524
0.8	.28814	.29103	.29389	.29673	.29955	.30234	.30511	.30785	.31057	.31327
0.9	.31594	.31859	.32121	.32381	.32639	.32894	.33147	.33398	.33646	.33891
1.0	.34134	.34375	.34614	.34850	.35083	.35314	.35543	.35769	.35993	.36214
1.1	.36433	.36650	.36864	.37076	.37286	.37493	.37698	.37900	.38100	.38298
1.2	.38493	.38686	.38877	.39065	.39251	.39435	.39617	.39796	.39973	.40147
1.3	.40320	.40490	.40658	.40824	.40988	.41149	.41309	.41466	.41621	.41774
1.4	.41924	.42073	.42220	.42364	.42507	.42647	.42786	.42922	.43056	.43189
1.5	.43319	.43448	.43574	.43699	.43822	.43948	.44062	.44179	.44295	.44408
1.6	.44520	.44630	.44738	.44845	.44950	.45058	.45154	.45254	.45352	.45449
1.7	.45543	.45637	.45728	.45818	.45907	.45994	.46080	.46164	.46246	.46327
1.8	.46407	.46485	.46562	.46638	.46712	.46784	.46856	.46926	.46995	.47062
1.9	.47128	.47193	.47257	.47320	.47381	.47441	.47500	.47558	.47615	.47670
2.0	.47725	.47778	.47831	.47882	.47932	.47982	.48030	.48077	.48124	.48169
2.1	.48214	.48257	.48300	.48341	.48382	.48422	.48461	.48500	.48537	.48574
2.2	.48610	.48645	.48679	.48713	.48745	.48778	.48809	.48840	.48870	.48899
2.3	.48928	.48956	.48983	.49010	.49036	.49061	.49086	.49111	.49134	.49158
2.4	.49180	.49202	.49224	.49245	.49266	.49286	.49305	.49324	.49343	.49361
2.5	.49379	.49396	.49413	.49430	.49446	.49461	.49477	.49492	.49506	.49520
2.6	.49534	.49547	.49560	.49573	.49585	.49598	.49609	.49621	.49632	.49643
2.7	.49653	.49664	.49674	.49683	.49693	.49702	.49711	.49720	.49728	.49736
2.8	.49744	.49752	.49760	.49767	.49774	.49781	.49788	.49795	.49801	.49807
2.9	.49813	.49819	.49825	.49831	.49836	.49841	.49846	.49851	.49856	.49861
3.0	.49865	.49869	.49874	.49878	.49882	.49886	.49889	.49893	.49897	.49900
3.1	.49903	.49906	.49910	.49913	.49916	.49918	.49921	.49924	.49926	.49929

1 標本の場合、 $n < 100$ 母集団の標準偏差未知 → t検定

①帰無仮説: 母集団のラーメンの平均価格は600円に等しい(差がない)

標本の平均 $\bar{x} = 620$ 円
標準偏差 $s = 50$ 円
サンプル数 $n = 26$

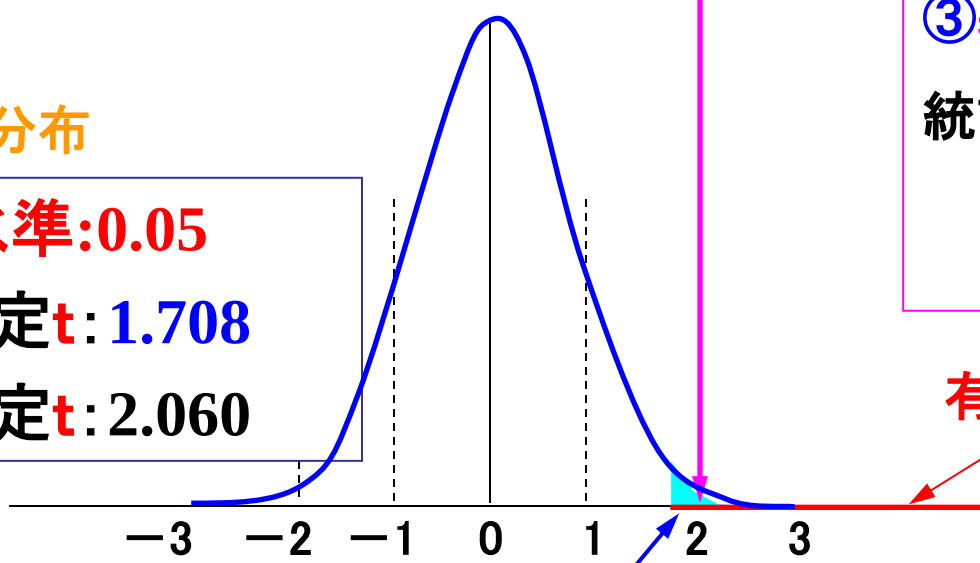


比較値 $m_0 = 600$ 円

標準誤差: $s / \sqrt{n-1}$
 $= 50 / \sqrt{26-1}$
 $= 10$ 円

t分布

②有意水準: 0.05
片側検定 $t: 1.708$
両側検定 $t: 2.060$



③基準化
統計量 $T = (\bar{x} - m_0) / (\sigma / \sqrt{n})$
 $= (620 - 600) / 10$
 $= 2$

有意水準0.05の棄却域

④ $T = 2 > 1.708$ (棄却)
 P 値 < 0.05 (棄却)

自由度: $26 - 1 = 25$

1.708

平均価格は600円より高い

棄却域限界のt値

Excelの関数

TINV(確率、自由度)

自由度	両側検定 $\alpha / 2 = 0.05$ 片側検定 α 0.1	両側検定 $\alpha / 2 = 0.025$ 片側検定 α 0.05
1	6.314	12.706
2	2.920	4.303
3	2.353	3.182
4	2.132	2.776
5	2.015	2.571
6	1.943	2.447
7	1.895	2.365
8	1.860	2.306
9	1.833	2.262
10	1.812	2.228
11	1.796	2.201
12	1.782	2.179
13	1.771	2.160
14	1.761	2.145
15	1.753	2.131
16	1.746	2.120
17	1.740	2.110
18	1.734	2.101
19	1.729	2.093
20	1.725	2.086
21	1.721	2.080
22	1.717	2.074
23	1.714	2.069
24	1.711	2.064
25	1.708	2.060
26	1.706	2.056
27	1.703	2.052
28	1.701	2.048
29	1.699	2.045
30	1.697	2.042

添付資料

母集団

母集団サイズ	N
母平均	μ
母分散	$V = \sigma^2$
母標準偏差	σ

標本(サンプル)

標本(サンプル)サイズ	n
標本平均	$\bar{x}$
標本分散	$V = \sigma^2$
標本標準偏差	σ

統計的推定

標本平均の平均(標本平均の期待値)

$$E(\bar{x}) = \mu$$

標本平均の標準偏差=標準誤差

$$D(\bar{x}) = \sigma / \sqrt{n}$$

標本分散の平均(標本分散の期待値)

$$E(u^2) = \sigma^2$$

