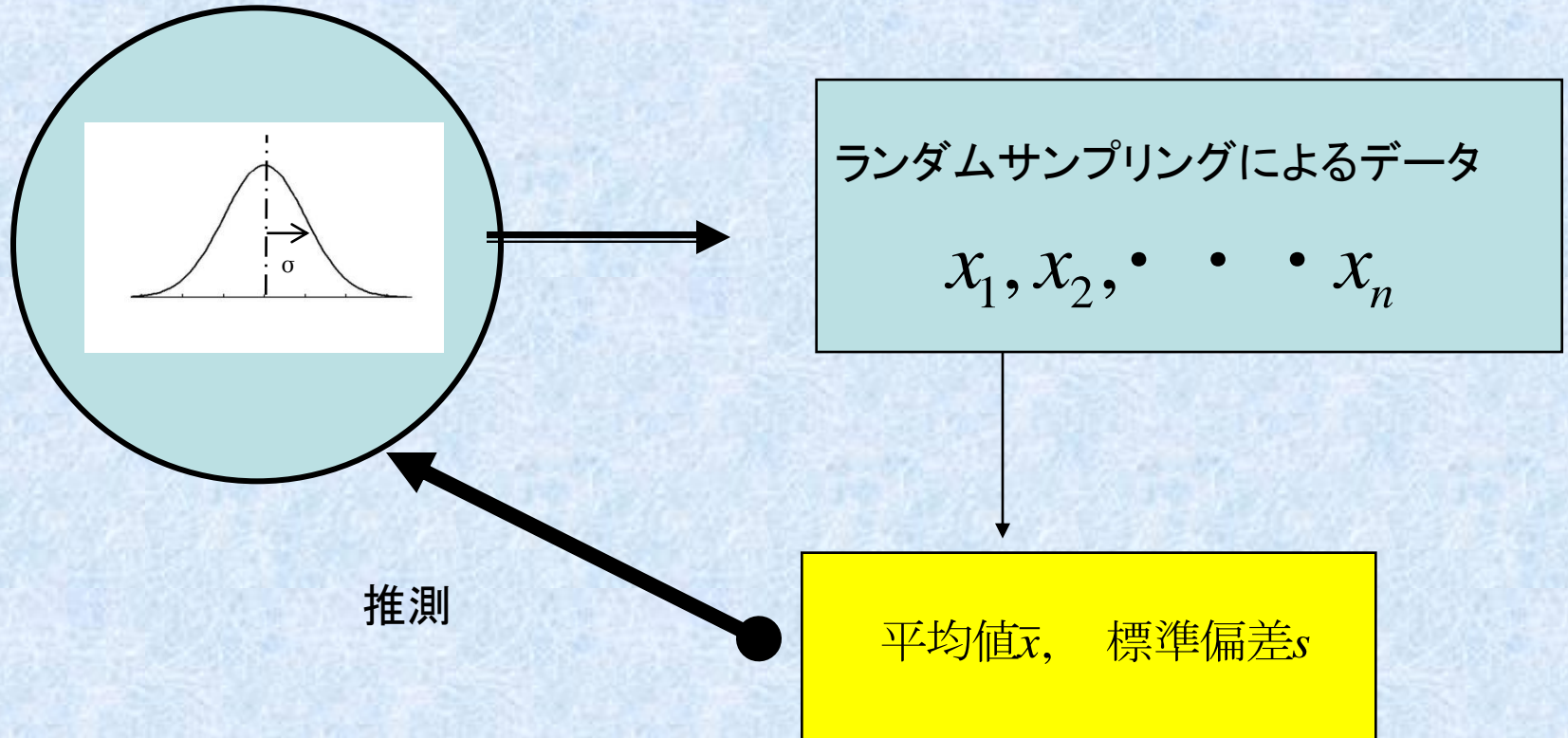


サンプリング

2009年4月28日

加藤洋一

一般的なサンプリング



平均値 \bar{x} , 標準偏差 s



①母平均の推定: 点推定

$$\hat{\mu} = \bar{x}$$



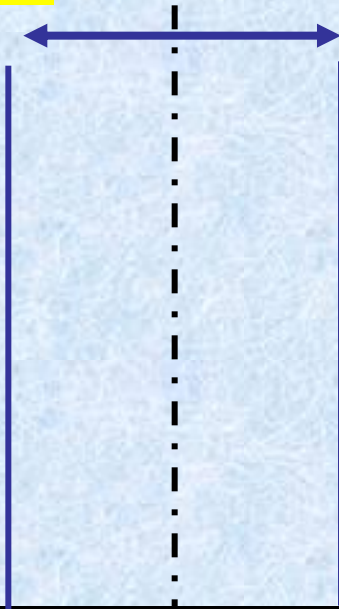
平均値 \bar{x} , 標準偏差 s



①母平均の推定: 点推定

②母平均の区間推定

$$\bar{x} \pm t(\phi, 0.05) \sqrt{\frac{V}{n}}$$

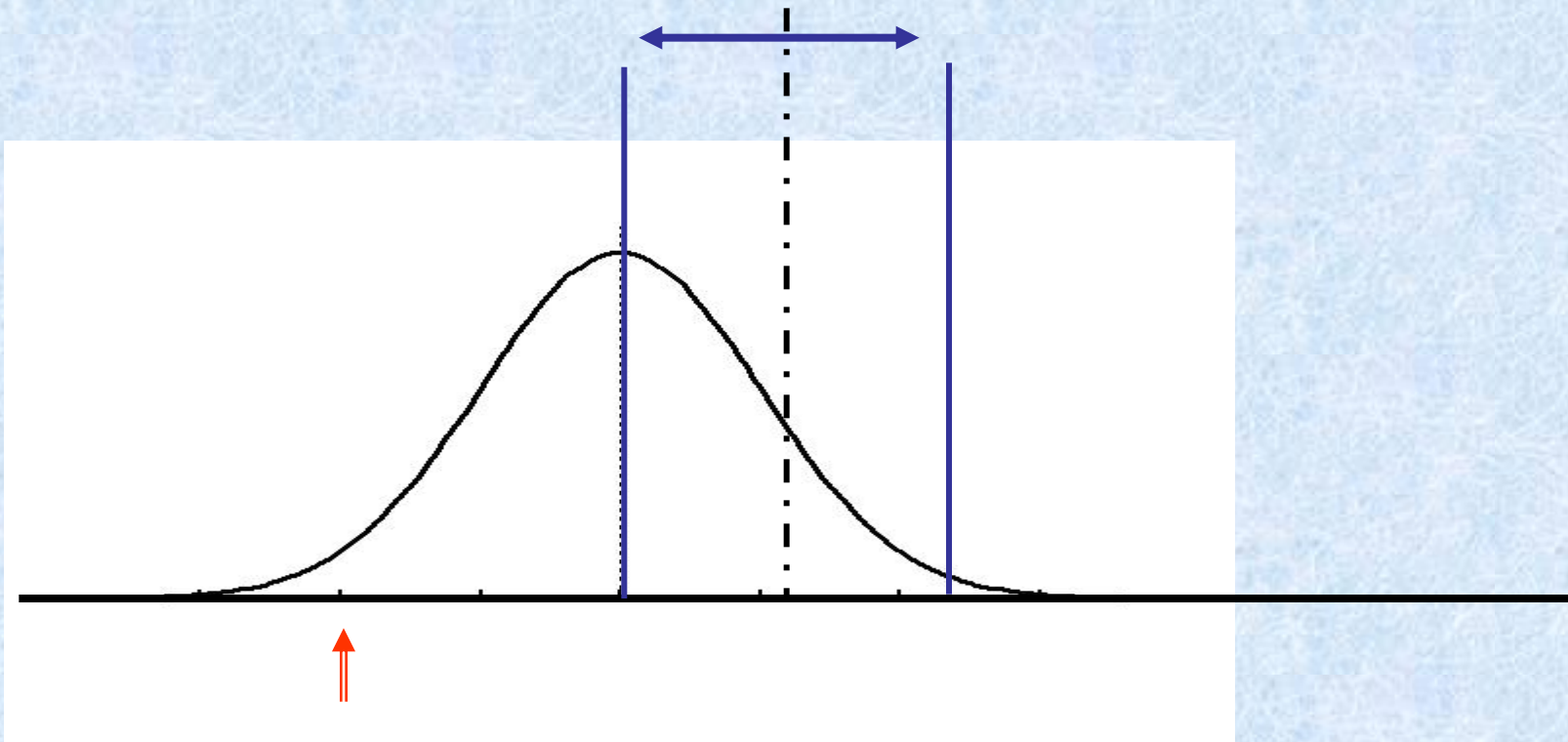


平均値 \bar{x} , 標準偏差 s



$$\bar{x} \pm t(\phi, 0.05) \sqrt{\left(1 + \frac{1}{n}\right) V}$$

③データの区間推定

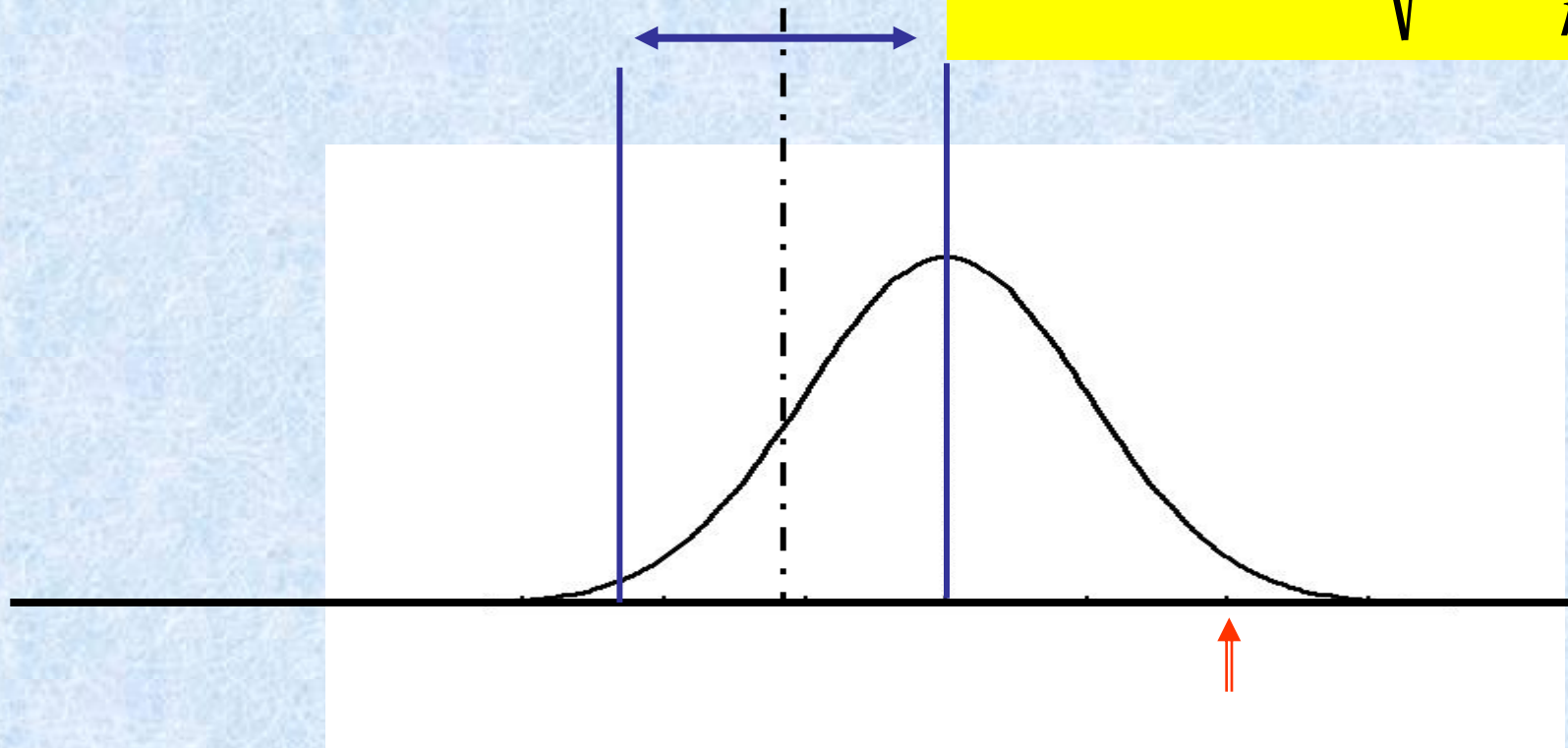


平均値 \bar{x} , 標準偏差 s



③データの区間推定

$$\bar{x} \pm t(\phi, 0.05) \sqrt{\left(1 + \frac{1}{n}\right) V}$$



平均値 \bar{x} , 標準偏差 s

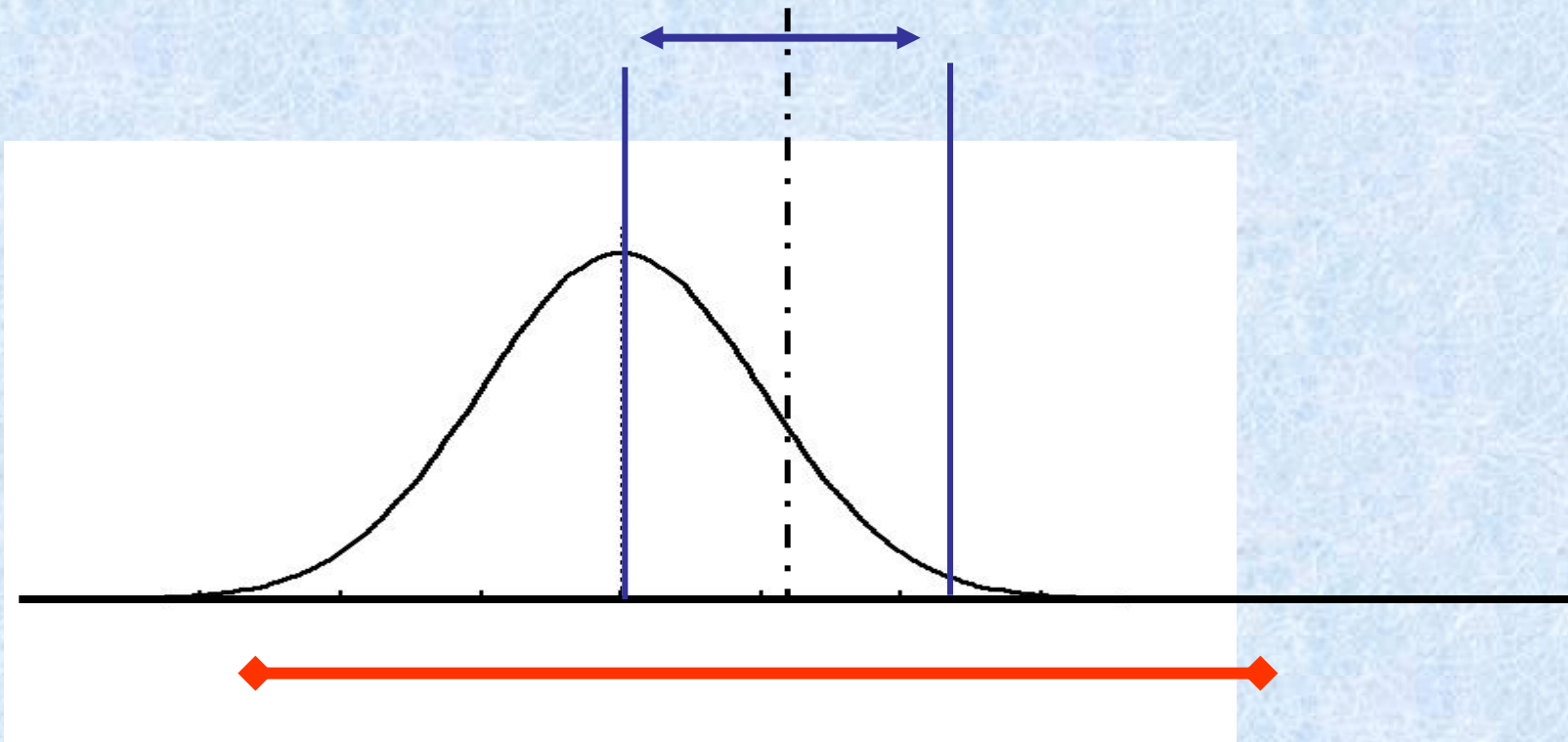


①母平均の推定: 点推定

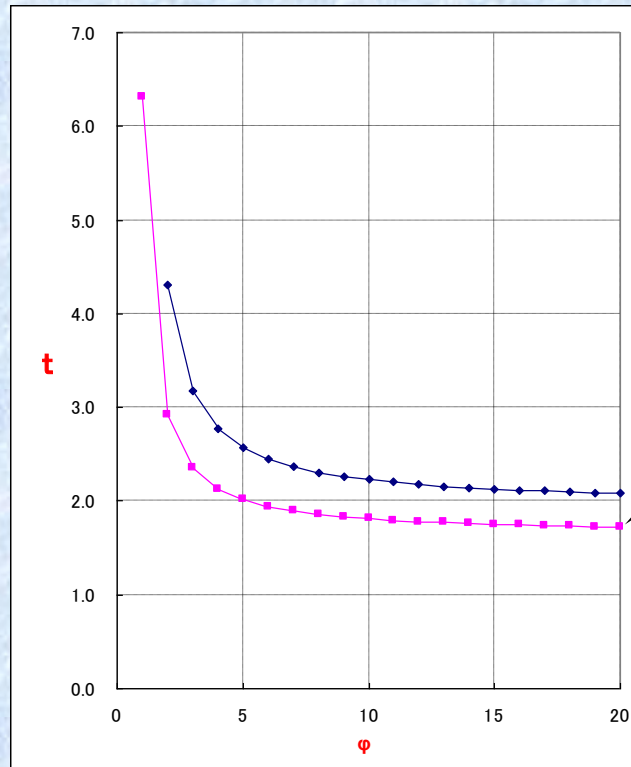
②母平均の区間推定

③データの区間推定

$$\bar{x} \pm t(\phi, 0.05) \sqrt{\left(1 + \frac{1}{n}\right) V}$$

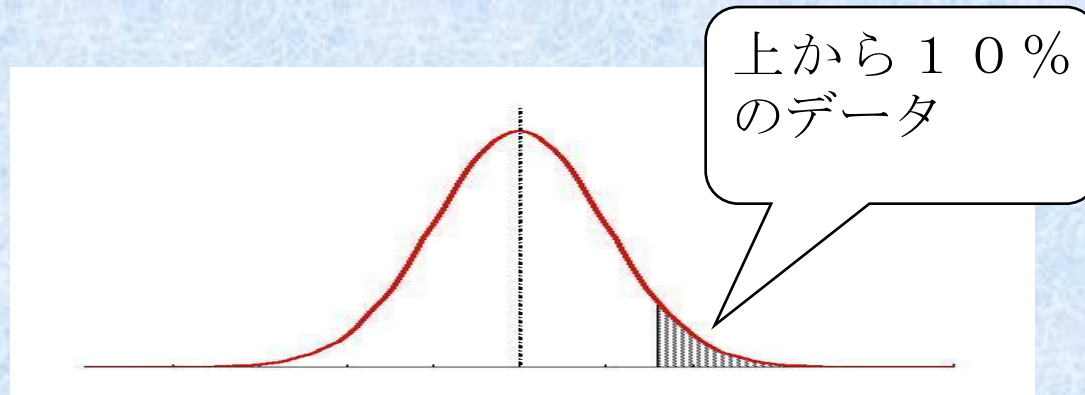


t分布から考えたサンプルサイズ



両側5%
正規分布 1.960に
該当

上から10%の大きいデータが、
確率90%でサンプルに含まれる



のに必要なサンプルサイズ

NIOSHの値の例:

$N=32$ から、 $n=16$ のサンプルをとる
〈計算は、超幾何分布で計算されている〉

計量値抜取検査の考え方

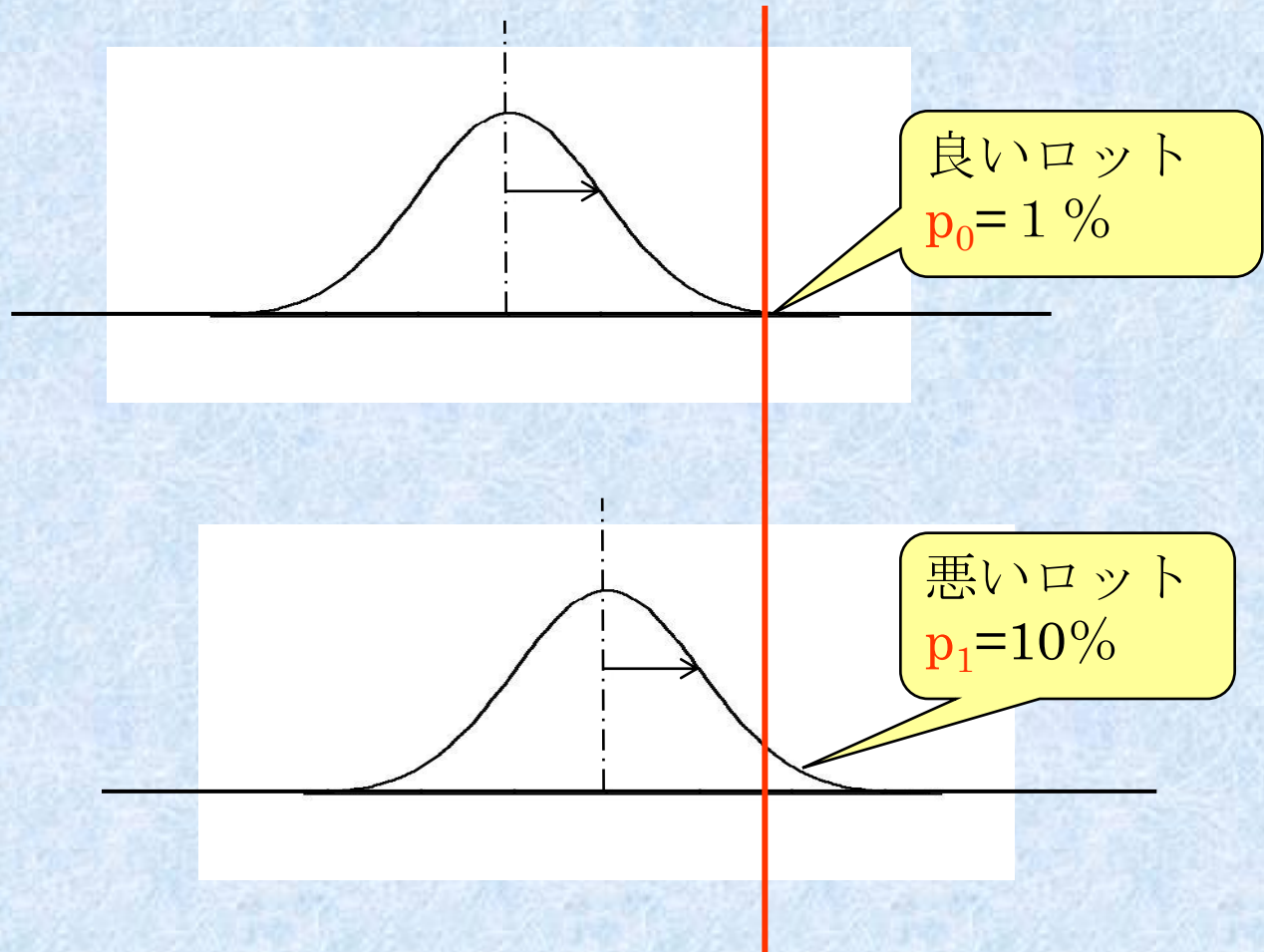
- 規格から外れる「不適合品」が，設定した率よりも大きいとき，
⇒不合格 と判定

ただし，これは平均値，標準偏差で判断するので，サンプルに「不適合品」が1つも含まれない場合でも，

不合格 と判断することがある。

(データに不適合が入ることは条件ではない。)

抜取検査



左下の数字は n , 右上の数字は k

		$p_i(\%)$	代表値		
			0.80	1.00	1.2
		$p_c(\%)$	範囲		
			0.71	0.91	1.1
代表値	範囲				
0.100	0.090~ 0.112		2.71 87	2.67 68	2.2 54
0.125	0.113~ 0.140			2.64 80	2.2 62
0.160	0.141~ 0.180			2.60 98	2.2 74

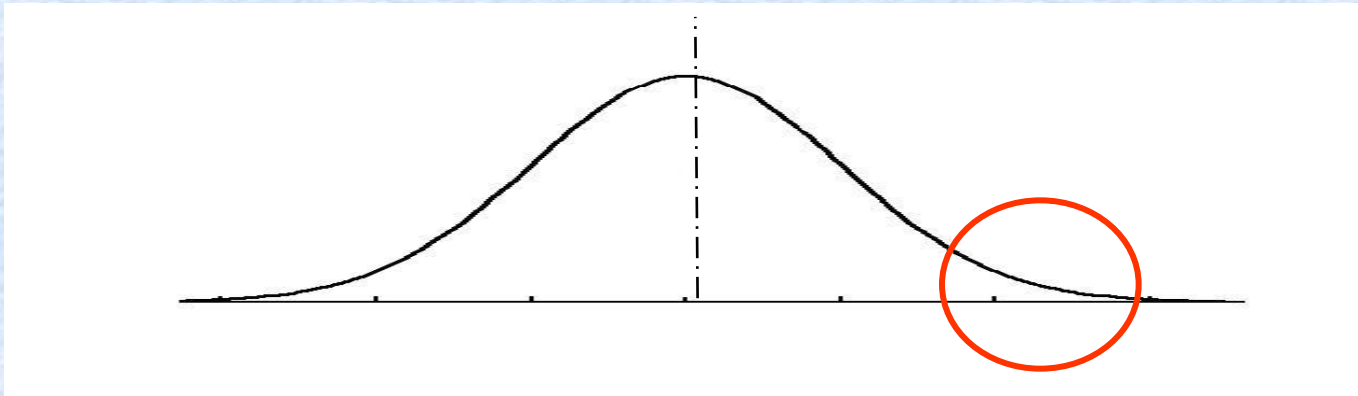
8.00	10.00	12.50	16.00
7.11	9.01	11.30	14.11
9.00	11.20	14.00	18.00
2.19 11	2.11 9	2.07 8	1.1 6
2.16 12	2.10 10	2.02 8	1.1 7
2.10 12	2.04 10	2.00 9	1.1 7

0.800	0.711~ 0.900			
1.000	0.901~ 1.120			
1.250	1.130~ 1.400			
1.600	1.410~ 1.800			
2.000	1.810~ 2.240			

1.86 24	1.79 18	1.72 14	1.1 11
1.83 28	1.76 21	1.69 16	1.1 13
1.78 32	1.72 24	1.65 18	1.1 15
1.74 40	1.67 28	1.60 20	1.1 15
1.69	1.63	1.56	1.1 11

4. サンプルを小さくする工夫

(1)母集団から, なんらかの情報で, 値の大きいものをピックアップする.



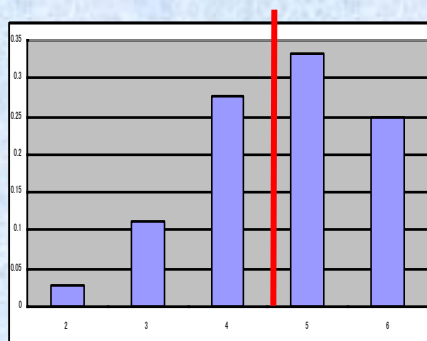
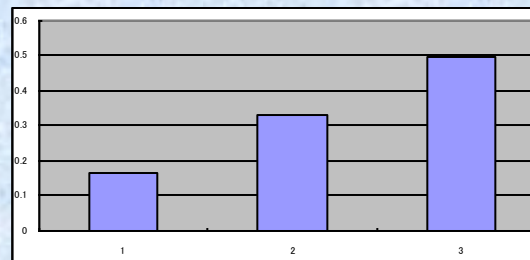
- ・平均やばらつきの情報は得られないが, 重点的にデータをとることができる.

ポイント:

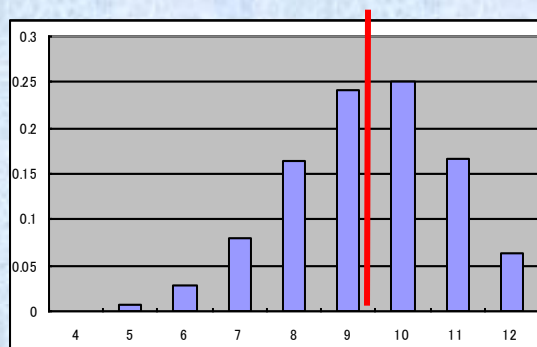
本当に大きな値のものを, 予測し選択できるか.

正規分布ではないが

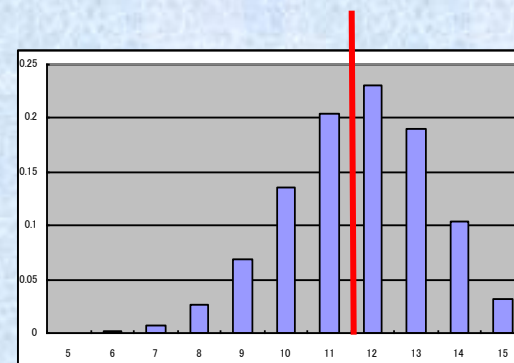
- データを多く採れば,
- それなりに処理可能



$n=2$



$n=4$



$n=5$

5. 現実の場面で想定されること

- 全体からのサンプリング？
- 対象から、いくつかの組織を選び、その中から更にサンプリング ⇒ 2段サンプリング
- 少数の組織を選び、そこは全部をサンプル ⇒ 集落サンプリング
- 何らかの基準で、層別 ⇒ 層別サンプリング
- 大きそうな組織の、大きそうなものを選ぶ ⇒ 有為サンプリング

基準の明確な設定

- 基準を超えるものが1つでもあるかどうかを、判定したい。

⇒検査で、全品良品に該当

全数検査が必要。

(不適合の事前選別が可能なら、

全数は 不要)

基準の明確な設定

- 異常値, 承知した違反状態も, 対象とするか

対象: 考えられる最悪の状態を調べればよい.

 サンプリングというより, 実験による検証

 (ここまで, ありうる, というデータになる)

対象としない: サンプリングを検討

基準の明確な設定

- 基準を超える比率を設定

⇒ 抜取検査の条件と似てくる.

① 基準の設定: p_0 , p_1 の設定

② 判定の誤りの設定

 サンプリングであるので、必ず誤りは含まれる.

抜取検査で考える判定の誤り

- α (第一種の誤り)

検査では, よいロットを, 不合格としてしまう誤りの確率(たまたまサンプルに悪いデータが入りすぎたという確率) $\alpha=0.05$ が一般的

- β (第二種の誤り)

悪いロットを, 合格としてしまう誤りの確率(たまたまサンプルに良いデータが入りすぎたという確率) $\beta=0.10$ が一般的

<目的によっては, こちらが重要になる>

β をもっと小さく \Rightarrow 設定値で, サンプルサイズを設計