問い

日本人の男子学生の身長の平均値は？

– 全員測れるか？
– サンプル測定して全員の平均値を推定できる？
– 何人サンプルすればいいのか？

１回だけ実験してデータを採る？

– 「たまたま」「偶然に」だろう？

↓　本当にその値か？

– 繰り返して実験/測定して平均を採る？

←　なぜ？　何回採ればいいのか

測定(標本) ⇒ 現象の推定

10人の男子学生の身長を測った結果

学生	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
身長 cm	175	170	169	165	164	179	172	168	175	170

であると (統計上) 何が言えるのか？

*母*平均 X＝標本平均 $(X_1+\ldots+X_{10})/10$ か？

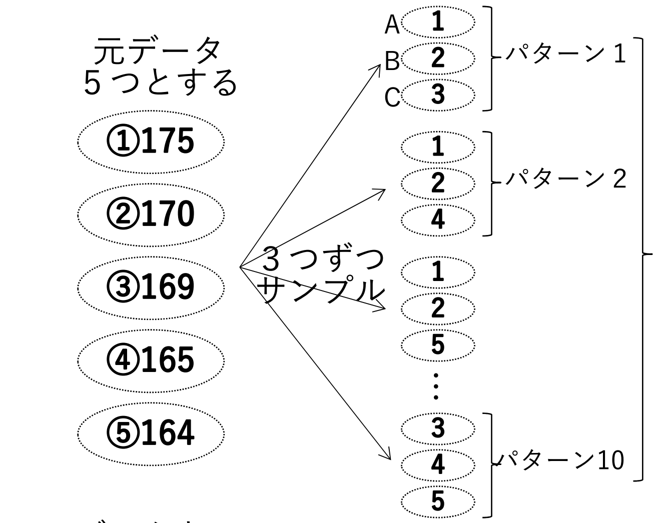
- *母*平均　*母*集団(全国の男子学生)の平均
- 標本平均　$(X_1+\ldots+X_{10})/10$

等しいか？

極端なケースは？？

標本(サンプル)

- 標本抽出(サンプリング)
  ⇒ サンプルが元の分布を表しているか？



10通りの組合せ

標本A, B, Cが無作為に抽出
されたものだとすれば
↓
どのパターンも
同じ確率で起こるはず
↓
統計量 t (例: 平均値)の分布は？

元データ上での
統計量 t (例: 平均値)

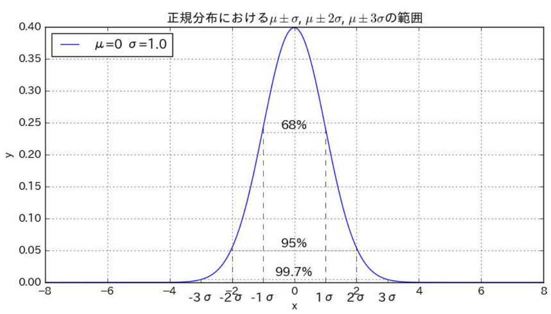
サンプル上では
統計量tが
どのような分布に
なるか？

サンプル上の平均＝母集団の平均

中心極限定理

前提条件：
　母集団の分布が何であっても
　サンプル X1, X2, ... Xn をランダム・独立に採るならば
　十分大きい n に対して

結論：
　サンプルの平均 (X1+X2+...+Xn)/n は
　$N(\mu, \sigma^2/n)$ に従う（近づく）
　　　$N(\mu, \sigma^2/n)$ は平均 μ、分散 σ^2/n の正規分布



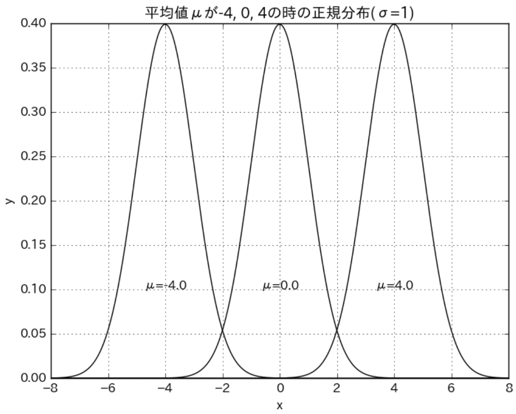
- 母集団の分布は何でもよい
- サンプルを繰り返し採って
  平均値を計算すれば、
  母集団の平均値 μ に近づく

脱線：平均μと分散σ2

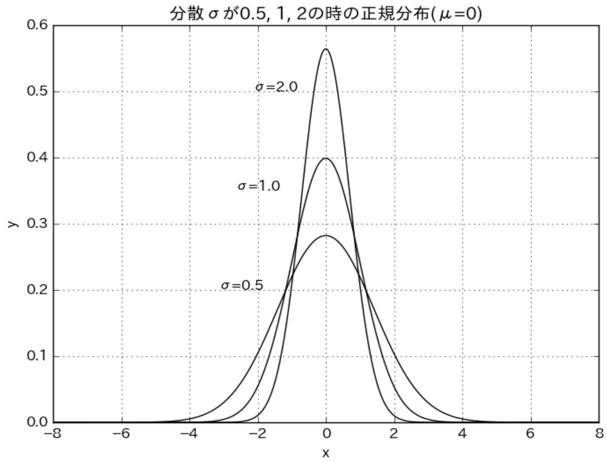
数式上の定義

$$\text{算術平均} = \frac{\sum_i x_i}{N}$$

$$\text{分散} = \frac{\sum_i (x_i - \overline{X})^2}{n}$$ （広がりの指標）



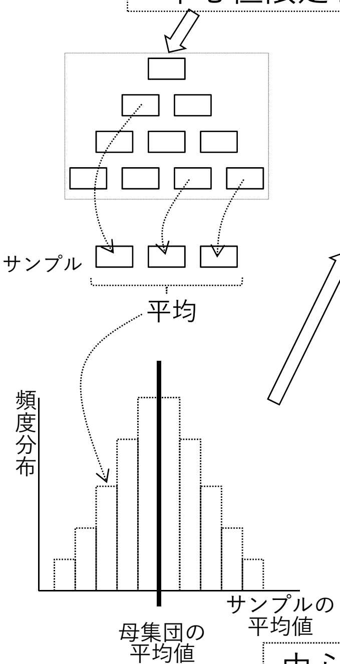
平均を変えると？



サンプルの平均の分布は？

母集団の平均の推定

中心極限定理：サンプルの平均の平均 → 母集団の平均



ランダムにサンプルを
採った時、その平均が
母集団（元のデータ）の
平均にどれほど近いか？

↑

サンプルの平均は
どのような散らばり方
をするのか
(標本平均の分布は何か？)

母集団の平均を
推定したい

↓

いくつサンプルを
採ればよいのか？

↑

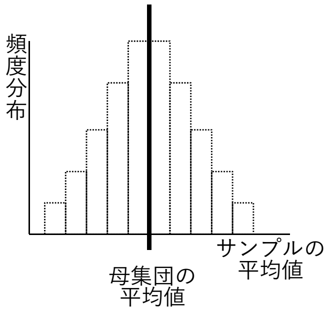
よいの定義？

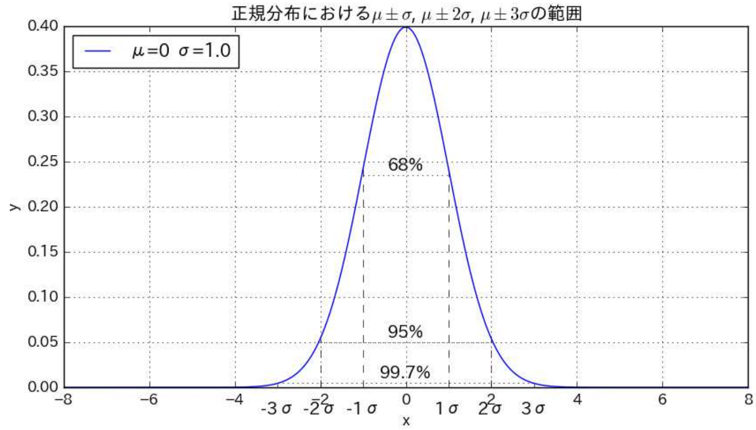
中心極限定理：サンプルの平均の分布は N(μ，σ2)

点推定 → 区間推定

母集団の平均の推定　　「まあ95%は当たってるね」？

サンプルの分布が正規分布であれば





ランダムにサンプルを採る

↓

上の分布で一様に発生

平均 μ、標準偏差 σ の正規分布では

$\mu\pm\sigma$ の範囲には、全体の約 68.27% のデータが含まれる。

$\mu\pm2\sigma$ の範囲には、全体の約 95.45% のデータが含まれる。

$\mu\pm3\sigma$ の範囲には、全体の約 99.73% のデータが含まれる。

中心極限定理：
　サンプルの平均の分布は $N(\mu, \sigma^2)$

95%当たるのは xが UからL の範囲内のとき

点推定 → 区間推定

母集団の平均の推定　　「まあ95%は当たってるね」？

95%当たるのは Xが UからL の範囲内のとき

$P(L \leqq X \leqq U) \geqq 1-\alpha$　　Xが $L \leqq X \leqq U$ に収まる確率が $1-\alpha$ 以上 → 95%

$1-\alpha$ ⇒ 信頼係数

区間 $[L, U]$ ⇒ 信頼区間　　と呼ぶ

母平均の区間推定

サンプル平均の分布は $N(\mu, \sigma^2/n)$ に従う

変換変換 $Z = (X-\mu)/(\sigma/\sqrt{n})$ をすると

（途中省略）

μ（母平均）の（信頼係数 $1-\alpha$ の）信頼区間

$[X - Z_{\alpha/2} \cdot \sigma/\sqrt{n},\quad X + Z_{\alpha/2} \cdot \sigma/\sqrt{n}]$

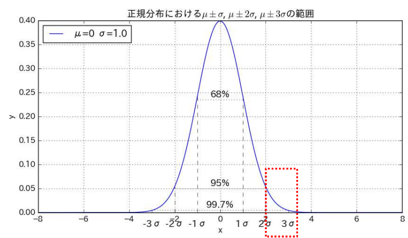
σ（母集団の分散）が分かっていない時は

サンプル平均は自由度n-1の t 分布になり

μ の信頼区間は

$[X - t_{\alpha/2\,(n-1)} \cdot s/\sqrt{n},\quad X + t_{\alpha/2\,(n-1)} \cdot s/\sqrt{n}]$

$Z_{\alpha/2}$ ～ 正規分布N(0,1)で
$P(U \leqq X)$ が $\alpha/2$となる点



踏み込んで「仮説検定」

ちょっと面倒だし サボってもいいけれど

仮説「母集団の平均値はxxxである」は成立つか？

仮説検定 (有意性の検定)：

例：母集団は正規分布とする
7回の測定で
24.5, 25.3, 26.2, 25.7, 24.4
25.1, 25.6
(サンプル平均 X=25.21)
であったときに、
「母集団平均が25.0 」と
いう仮説は支持できるか

仮説検定 (母分散が既知の場合)

$Z = (X - \mu)/(\sigma/\sqrt{n})$ とすると
仮説が正しければZはN(0,1)に従う
↓ $Z_{\alpha/2}$ を表で調べて
$|Z| > Z_{\alpha/2}$では仮説は棄却され
$|Z| \leqq Z_{\alpha/2}$では仮説は棄却されない

仮説検定 (母分散が未知の場合)

$t = (X - \mu)/(s/\sqrt{n})$ とすると
仮説が正しければ t は t(n-1)に従う
↓ $t(7)_{\alpha/2}$ を表で調べて
$|t| > t(7)_{\alpha/2}$では仮説は棄却され
$|t| \leqq t(7)_{\alpha/2}$では仮説は棄却されない

例： 母集団平均 $\mu = 25.0$
サンプル平均 X=25.21

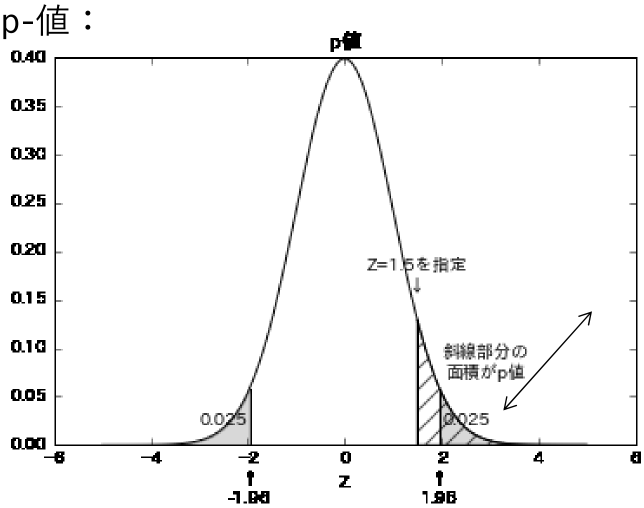
母集団分散 $\sigma = 0.71$	母集団分散未知
↓	↓
$Z = (X - \mu)/(\sigma/\sqrt{n})$	$t = (X - \mu)/(s/\sqrt{n})$
$= 0.21/(0.71/\sqrt{7})$	$= 0.21/(0.715/\sqrt{7})$
$= 0.78$	$= 0.77$
$Z_{0.025} = 1.96$で 棄却されない	$t(7)_{0.025} = 2.45$で 棄却されない

踏み込んで「仮説検定」

ちょっと面倒だし サボってもいいけれど

仮説「母集団の平均値はxxxである」は成立つか？

p-値：



母分散既知
Z=0.78から p=0.44 > α=0.05
で棄却されない

母分散未知
t(7) =0.77から p=0.43 > α=0.05
で棄却されない

$|Z| > Z_{\alpha/2}$では仮説は棄却され
$|Z| \leqq Z_{\alpha/2}$では仮説は棄却されない

αからZ$_{\alpha/2}$を求めて Z と比較する

の代わりに

Z からα相当値 (p値と呼ぶ) を求め
p値と、設定したいαとを比較する