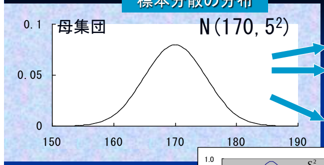
標本分散

- $x_1, x_2, \cdots, x_n \sim N(\mu, \sigma^2)$
- $$S^2 = \frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2}{n-1}$$

  はσ^2の不偏な推定値
- S^2：標本(不偏)分散
- σ^2：母分散
- S^2の期待値は，n，μ，σ^2の値にかかわらずσ^2
- S^2はσ^2の不偏推定値

標本分散の分布



サイズ10の標本

165, …, 175

180, …, 158

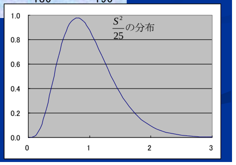
…

x_1, …, x_{10}

分散

22, 28,

…, S^2



推測統計学への扉を開いたt分布

$x_1, \cdots\cdots, x_n \sim N(\mu, \sigma^2)$

$$\bar{x} = \frac{x_1+x_2+\cdots+x_n}{n} \sim N(\mu, \frac{\sigma^2}{n})$$ 標本平均の定理

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$ 標準化定理

$$S^2 = \frac{(x_1-\bar{x})^2+\cdots\cdots+(x_n-\bar{x})^2}{n-1}$$ 分母に注意!

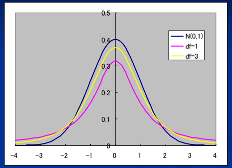
$$t = \frac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}}$$ ～t分布;自由度 $n-1$

σ に依存しない！

シミュレーションで解説

p62

t 分布とN(0,1)の比較



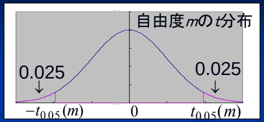
t分布は裾が厚くてとがり気味

$$t(m) \longrightarrow N(0,1)$$

t分布は自由度が大きくなるとN(0,1)

p61

μの信頼区間：σは未知



x_1, …, $x_n \sim N(\mu, \sigma^2)$

μ の95%信頼区間は

$$\left(\bar{x}-t_{0.05}(n-1)\frac{S}{\sqrt{n}}, \bar{x}+t_{0.05}(n-1)\frac{S}{\sqrt{n}}\right)$$

σ既知　なら　$\left(\bar{x}-1.96\frac{\sigma}{\sqrt{n}}, \bar{x}+1.96\frac{\sigma}{\sqrt{n}}\right)$

p61

信頼区間の使用例

- ある山において野ネズミ16匹を採取し，ヒゲの長さを測定したところ，

  $\bar{x} = 3.5, S = 0.4$　であった。
- μ =野ネズミ全体での平均のヒゲの長さの95%信頼区間を求めよ

  $$\left(3.5 - t_{0.05}(15)\frac{0.4}{\sqrt{16}}, 3.5 + t_{0.05}(15)\frac{0.4}{\sqrt{16}}\right)$$

  $t_{0.05}(15)$=2.13 を代入して

  (3.29, 3.71)

母平均の区間推定，検定

- 標本が十分に大，母分散が既知
  - 正規分布が利用可能
- 標本があまり大きくない
  - 標本分散が母分散に等しいと仮定できない
  - t分布を利用
  - 標本平均が母平均の推定値
  - どれだけの誤差を含んでいるか評価できない
  - したがって，信頼区間をつけて区間推定

W.S.ゴセット

- ギネスビール社の醸造技術者
- 「小標本」の問題について言及し，t分布を発見
- ビールの品質改善の問題を考える場合，一番確実な方法論は「全部のビールを調べる」
  - 商売が成り立たない
- 少数のサンプルを抜き出し，出荷された全体のビールに関して言及する
  - t分布の発見

t-分布の名前の由来

- 発見者W.S.ゴセットのペンネームに由来？
  - ギネスビール社に本名での論文発表を禁止
  - ペンネームStudentで論文を発表
  - ゴセットは確率密度関数をs分布と呼んでいた
  - sは標準偏差を指す事にも使われ，紛らわしい
  - Studentの第2文字目からt分布とした？？？
    - 推測統計学の父R.A.フィッシャー