# 統計学 01

早稲田大学政治経済学部

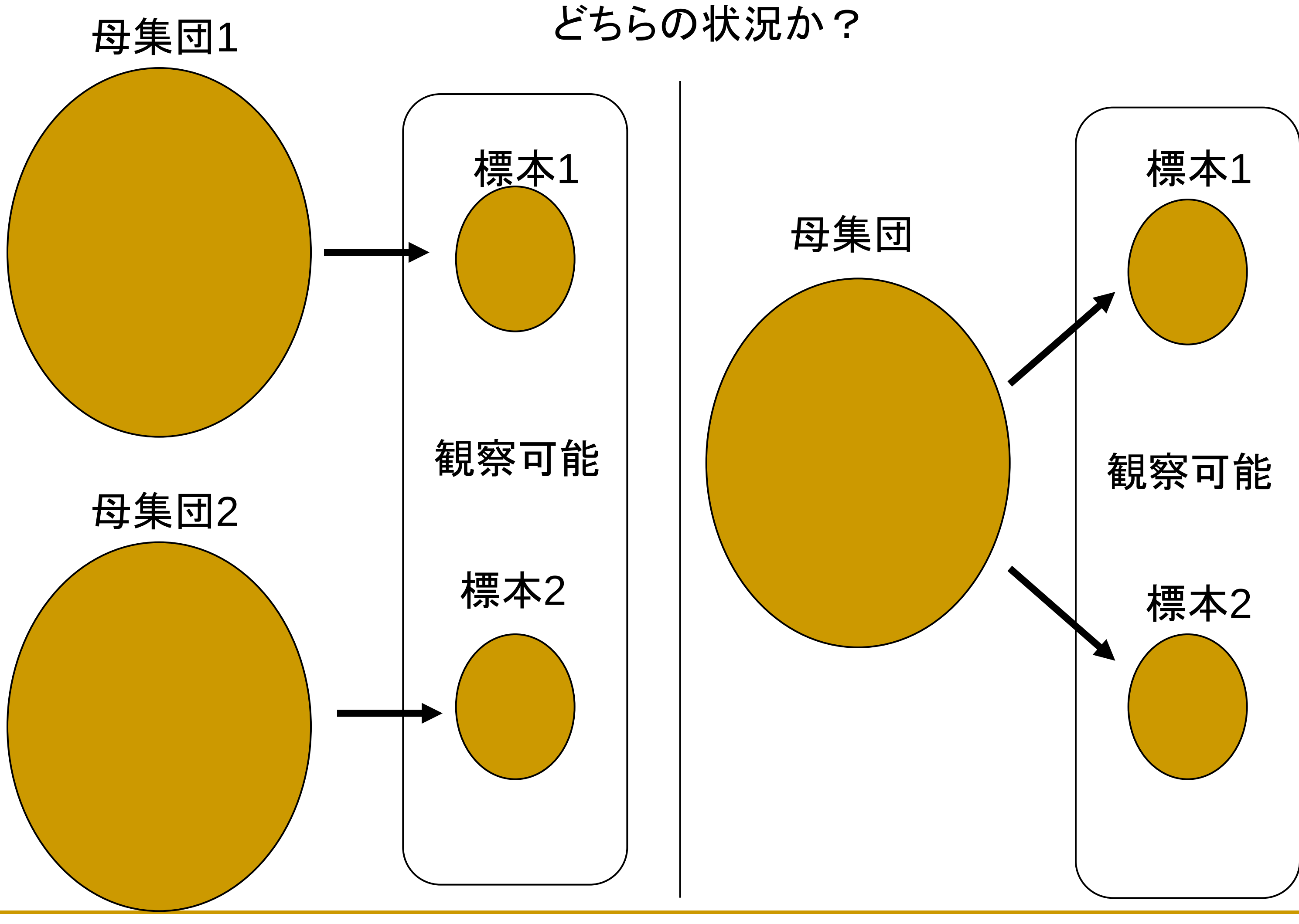
第9回

西郷 浩

# 本日の目標

- 10.5 二標本問題
  - 二標本問題とは？
  - 2つの標本平均の差の標本分布
  - 2つの標本分散の比の差の標本分布

# 二標本問題とは



# ２つの標本平均の差の標本分布(1)

$$X_i \sim N\left(\mu_1, \sigma_1^2\right) (i = 1, 2, \ldots, m)$$かつ相互に独立

$$Y_j \sim N\left(\mu_2, \sigma_2^2\right) (j = 1, 2, \ldots, n)$$かつ相互に独立

標本平均 $\bar{X} = \frac{1}{m}\sum_{i=1}^{m} X_i, \bar{Y} = \frac{1}{n}\sum_{j=1}^{n} Y_j$ について、

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \sigma_1^2/m + \sigma_2^2/n\right)$$

$$\Leftrightarrow Z = \frac{\left(\bar{X} - \bar{Y}\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}} \sim N(0, 1)$$

# ２つの標本平均の差の標本分布(2)

もし、$\sigma_1^2$と$\sigma_2^2$とが未知で、これらを標本推定値

$$s_1^2 = \frac{1}{m-1}\sum_{i=1}^{m}\left(X_i - \overline{X}\right)^2,\ s_2^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(Y_j - \overline{Y}\right)^2$$

でおきかえたとき、

$$T = \frac{\left(\overline{X} - \overline{Y}\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{s_1^2/m + s_2^2/n}}$$ ～とても複雑な分布

歴史的な難題（Behrens - Fisher問題）

Welchの近似が有名

# 2つの標本平均の差の標本分布(3)

ただし、（未知ながら）$\sigma_1^2 = \sigma_2^2 (= \sigma^2)$であれば、

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2 (1/m + 1/n)}} \sim N(0,1)$$

において、$\sigma^2$を標本推定値

$$s^2 = \frac{\sum_{i=1}^{m} (X_i - \bar{X})^2 + \sum_{j=1}^{n} (Y_j - \bar{Y})^2}{m + n - 2} = \frac{(m-1)s_1^2 + (n-1)s_2^2}{m + n - 2}$$

でおきかえたとき、

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{s^2 (1/m + 1/n)}} \sim t(m + n - 2)$$

# ２つの標本平均の差の標本分布(3)

- 例：
  - ２つの大学 W, K からそれぞれ10人ずつの学生を無作為抽出して所持金を調べた。

$$\overline{X} = 5000,\ s_1^2 = 1000000, \quad \overline{Y} = 6000,\ s_2^2 = 1440000$$

  - ２つの大学で平均的な所持金に差があると考えるべきか？
- 解答：
  - 「２つの大学で学生の所持金の平均値・バラツキともに等しい」と想定する。所持金の分布は正規分布で近似できると仮定する。

# 2つの標本平均の差の標本分布(4)

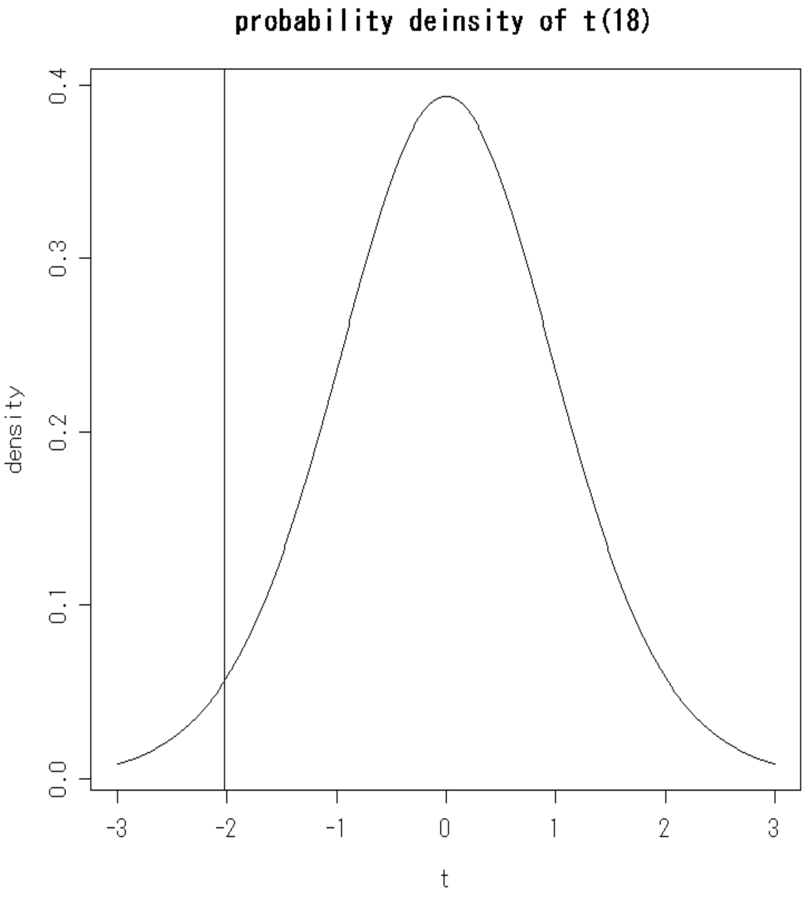
$$T=\frac{\left(\bar{X}-\bar{Y}\right)}{\sqrt{s^2\left(1/m+1/n\right)}}=\frac{\left(\bar{X}-\bar{Y}\right)}{\sqrt{\dfrac{(m-1)s_1^2+(n-1)s_2^2}{m+n-2}\left(\dfrac{1}{m}+\dfrac{1}{n}\right)}}$$

$$\sim t(m+n-2)\quad (\text{前提条件が正しい場合})$$

$$T_{obs}=\frac{(5000-6000)}{\sqrt{\dfrac{9\times1000000+9\times1440000}{18}\left(\dfrac{1}{10}+\dfrac{1}{10}\right)}}=-2.02$$

# 2つの標本平均の差の標本分布(5)

- 観察された*T*の値が、
  - 「滅多にないような珍しいこと」なのか
  - 「それほど珍しいことではない」ということなのか。
- 本格的な解答は統計的仮説検定によってえられる。
  - 第12章で登場



# 2つの標本分散の比の標本分布(1)

$X_i \sim N\left(\mu_1, \sigma_1^2\right)\left(i = 1, 2, \ldots, m\right)$かつ相互に独立

$Y_j \sim N\left(\mu_2, \sigma_2^2\right)\left(j = 1, 2, \ldots, n\right)$かつ相互に独立

このとき、

$$\frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2} \sim F\left(m-1, n-1\right)$$

とくに、$\sigma_1^2 = \sigma_2^2$が成り立つなら

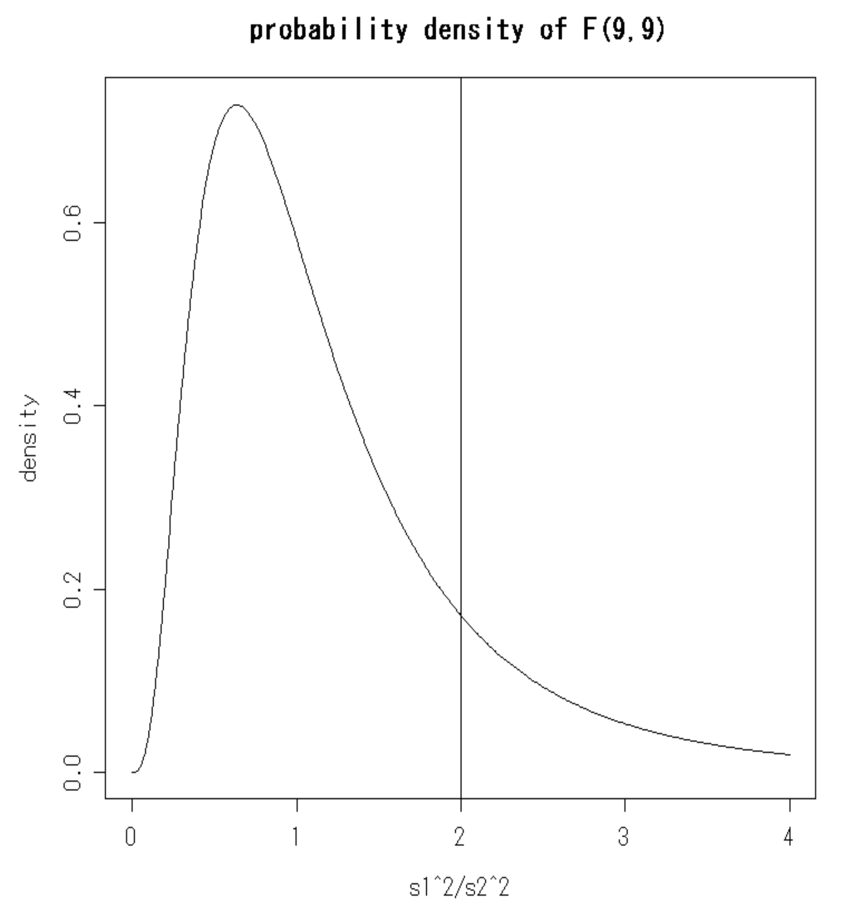
$$\frac{s_1^2}{s_2^2} \sim F\left(m-1, n-1\right)$$

# 2つの標本分散の比の標本分布(2)

- 例:
  - ある標本($m$=10)から計算した標本分散 $s_1^2=200$
  - 別の標本($n$=10)から計算した標本分散 $s_2^2=100$
  - これらは別々の母集団から発生したと考えるべきか?
- 解答
  - 「同じ母集団から発生した」と考えれば、
    - $s_1^2/s_2^2 \sim F(9, 9)$

# 2つの標本分散の比の標本分布(3)

- $(s_1^{\,2}/s_2^{\,2})_{obs} = 2$ はそれほど珍しいこととはいえない。
  - 標本分散の比だけでは、異なる母集団からの標本であるという強い証拠はえられない。
    - 本格的な解答は統計的仮説検定からえられる。



# 練習問題

- 10.1-10.7
  - 第10章の復習と統計表の読み方の練習
  - 時間があれば講義中に答え合わせをします。