統計学入門 補助資料
～分散・相関・比率に関する検定～

2022年度1学期:　　月曜2限
担当教員:　　　　石垣　司

分散・相関・比率に関する検定

① 1つの正規母集団の等分散性の検定
- 統計量 $U=(n-1)\frac{S^2}{\sigma^2}$ は自由度 $n-1$ の χ^2分布に従う

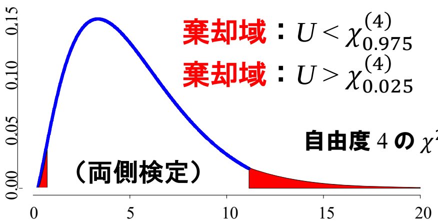
② 2つの正規母集団の等分散性の検定
- 2つの母分散が等しいかどうかの検定

③ 比率の検定

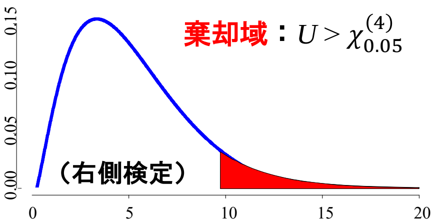
④ 相関係数の検定

① 1つの正規母集団の分散の検定

- **検定統計量**　$U=(n-1)\frac{S^2}{\sigma_0^2}$　(自由度 $n-1$ の χ^2 分布)
  - 帰無仮説　$\sigma^2=\sigma_0^2$ (σ_0^2 は帰無仮説の下での分散)
  - 例 $n=5$ での有意水準5%の両側検定 (対立仮説 $\sigma^2\neq\sigma_0^2$)
    - 境界値 $R_+=\chi^{(4)}_{0.025}=11.1$ (χ^2 分布表より)
    - 境界値 $R_-=\chi^{(4)}_{0.975}=0.48$ (χ^2 分布表より)
  - 例 $n=5$ での有意水準5%の右側検定 (対立仮説 $\sigma^2>\sigma_0^2$)
    - 境界値 $R=\chi^{(4)}_{0.05}=9.49$ (χ^2 分布表より)



自由度 4 の χ^2分布



演習問題

- **問題**
  - ある工場では重さ10.0gの薬品を大量生産している。その重さは $N(10,0.1^2)$ となるように品質管理している。
    あるとき、サンプル9個を無作為に抜き打ち検査したところ、
    $\bar{X}=9.9, S^2=0.15^2$ であった。
  - 母分散は $\sigma^2=0.1^2$ に管理されているか5%有意水準で両側検定しなさい。

$\chi^{(8)}_{0.025}=17.53, \chi^{(9)}_{0.025}=19.02, \chi^{(8)}_{0.05}=15.51, \chi^{(9)}_{0.05}=16.92,$
$\chi^{(8)}_{0.975}=2.18, \chi^{(9)}_{0.975}=2.70, \chi^{(8)}_{0.95}=2.73, \chi^{(9)}_{0.95}=3.33$

② 等分散性の検定とF分布

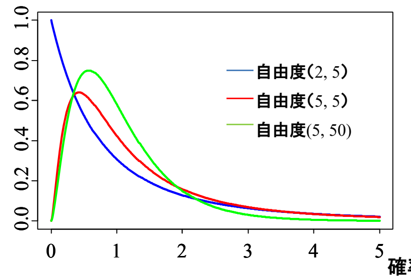
- 帰無仮説 H_0 と対立仮説 H_1
  - 両側検定　$H_0 : \sigma_x^2 = \sigma_y^2,\quad H_1 : \sigma_x^2 \neq \sigma_y^2$
  - 右側検定　$H_0 : \sigma_x^2 = \sigma_y^2 \text{ or } \sigma_x^2 \leq \sigma_y^2,\quad H_1 : \sigma_x^2 > \sigma_y^2$
  - 左側検定　$H_0 : \sigma_x^2 = \sigma_y^2 \text{ or } \sigma_x^2 \geq \sigma_y^2,\quad H_1 : \sigma_x^2 < \sigma_y^2$
- 検定統計量に標本分散の比 $F = S_x^2/S_y^2$ を採用
  - $F \approx 1$ ならば帰無仮説 $\sigma_x^2 = \sigma_y^2$ が正しい
    - 補足　平均の検定とは異なり $D = S_x^2 - S_y^2$ が従う分布は不明
- F 分布
  - χ^2分布に従う確率変数の比の分布
    - 帰無仮説 $\sigma_x^2 = \sigma_y^2$ の下で F が従う分布

F 分布 #1

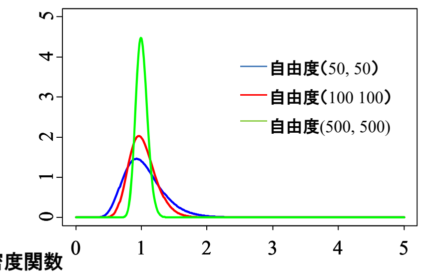
- 自由度 (n, m) の F 分布
  - V は自由度 n, W は自由度 m の χ^2 確率変数で互いに独立とするとき, $\frac{V/n}{W/m}$ の従う分布
    - n, m が大きくなると1の周辺に密度関数が集まる

check!

- $F = \frac{S_x^2/\sigma_x^2}{S_y^2/\sigma_y^2}$ は自由度 $(n-1, m-1)$ の F 分布に従う



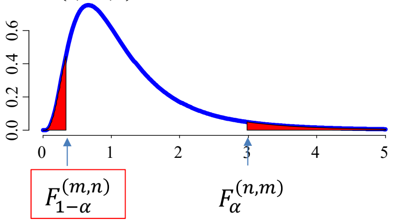
確率密度関数



F 分布 #2

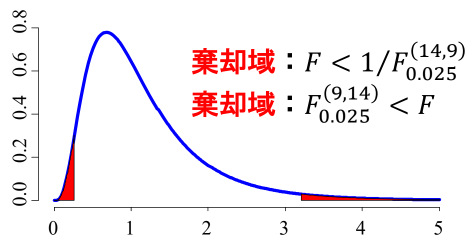
- F 分布の逆数の分布　check!
  - F が自由度 (n, m) の F 分布に従うとき, $1/F$ は自由度(m, n) の F 分布に従う
- F 分布のパーセント点
  - $F_\alpha^{(n,m)}$　自由度 (n, m) の F 分布の上側$100 \times \alpha$%点
    - $F_\alpha^{(n,m)}$の値は F 分布表より利用可
  - 自由度 (n, m) の F 分布の下側$100 \times \alpha$%点
    $F_{1-\alpha}^{(n,m)} = 1/F_\alpha^{(m,n)}$　check!

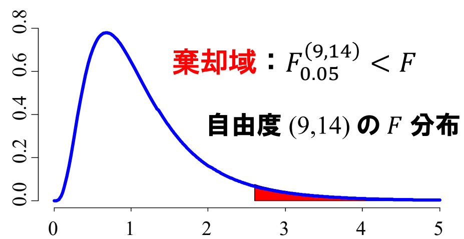
確率密度関数 $f(x) = \frac{1}{Beta(n/2, m/2)} \left(\frac{nx}{nx+m}\right)^{n/2} \left(1 - \frac{nx}{nx+m}\right)^{m/2} x^{-1}$



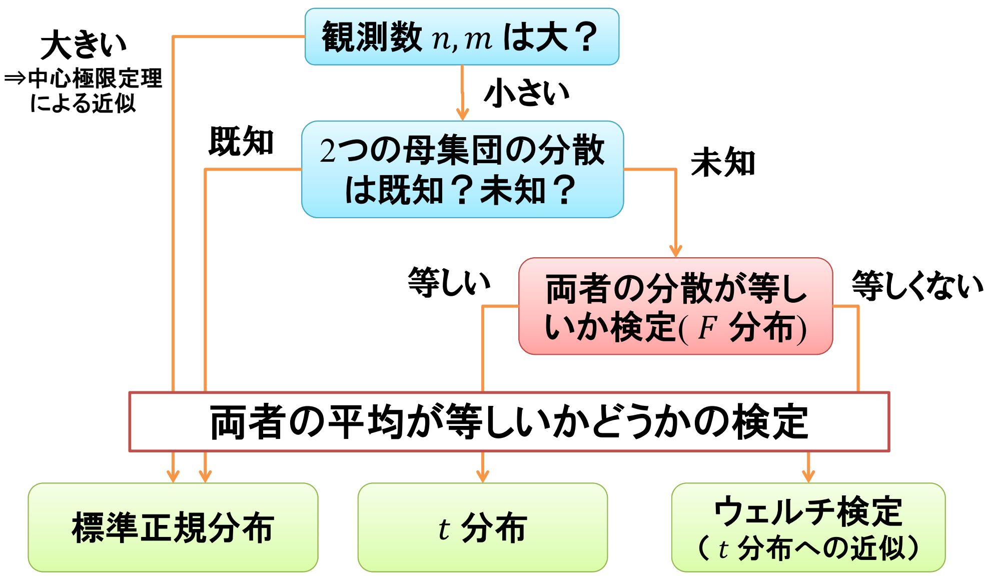
② 正規母集団の等分散性の検定

- 検定統計量　$F = S_x^2/S_y^2$
  - 帰無仮説　$\sigma_x^2 = \sigma_y^2$
  - 例 $n = 10, m = 15$での5%水準の両側検定 (対立仮説 $\sigma_x^2 \neq \sigma_y^2$)
    - 境界値 $R_+ = F_{0.025}^{(9,14)} = 3.21$ (F 分布表より)
    - 境界値 $R_- = F_{0.975}^{(9,14)} = 1/F_{0.025}^{(14,9)} = 0.31$ ($F_{0.025}^{(14,9)}$の値は F 分布表より)
  - 例 $n = 10, m = 15$での5%水準の右側検定 (対立仮説 $\sigma_x^2 > \sigma_y^2$)
    - 境界値 $R = F_{0.05}^{(9,14)} = 2.65$ (F 分布表より)





検定統計量の選択のフローチャート



演習問題

- **問題**
  - ある工場では重さ10.0gの薬品を大量生産している。あるとき, 製造ライン X から7個, 製造ライン Y から9個のサンプルを無作為に抜き打ち検査したところ $\bar{X} = 10.1, \bar{Y} = 9.8, S_x^2 = (0.2)^2, S_y^2 = (0.3)^2$ であった。

1. このとき、両ラインで製造される製品の母分散は等しいか5%有意水準で両側検定しなさい
2. 母平均に差があるか検定したい。このとき, 検定統計量 Z (標準正規分布), 検定統計量 T (t 分布), Welchの t 検定のうちどの検定方法を用いるべきか議論しなさい

$F_{0.025}^{(6,8)} = 4.65, F_{0.025}^{(8,6)} = 5.60, F_{0.025}^{(7,9)} = 4.20, F_{0.025}^{(9,7)} = 4.82,$
$F_{0.05}^{(6,8)} = 3.58, F_{0.05}^{(8,6)} = 4.15, F_{0.05}^{(7,9)} = 3.29, F_{0.05}^{(9,7)} = 3.68$

③ 比率に関する検定

- **母比率がある値と異なるかどうかの検定**
  - 帰無仮説 $p = p_0$
    - p: 母比率, p_0: 帰無仮説の下での比率
  - 検定統計量 $Z = \frac{\bar{X} - p_0}{\sqrt{p_0(1-p_0)/n}}$ (標準正規分布)
    - n が大きいとき, 中心極限定理により正規近似
- **2つの母比率の差の検定**
  - 帰無仮説 $p_x = p_y$
    - p_x, p_y: 2つの母集団（集団XとY）の母比率
  - 検定統計量 $Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\tilde{p}(1-\tilde{p})/\{(1/n)+(1/m)\}}}, \tilde{p} = \frac{n\bar{X} + m\bar{Y}}{n+m}$
    - n が大きいとき, 中心極限定理により正規近似

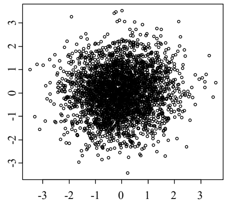
演習問題

- **問題**
  - A社とB社がそれぞれ内閣の支持率調査を行った。
  - A社　　回答者900人、支持者495人。$\bar{X} = 0.55$
  - B社　　回答者100人、支持者55人。$\bar{X} = 0.55$
  - 以前の調査での内閣支持率は50%であった。このとき“内閣の支持率は上昇した”という対立仮説をA社とB社のそれぞれの調査結果に対して5%有意水準で片側検定しなさい

$Z_{0.025} = 1.96, Z_{0.05} = 1.65$

④ 相関係数の検定

- **相関係数がゼロか否かの検定**
  - **帰無仮説** $\rho_{xy} = 0$, **対立仮説** $\rho_{xy} \neq 0$
    - **母相関係数** ρ_{xy}, **標本相関係数** r_{xy}
- **検定統計量** $T = r_{xy}\sqrt{\frac{n-2}{1-r_{xy}^2}}$
  - T **は自由度** $n-2$ **の** t **分布に従う**



例 $n = 5{,}000$, $r_{xy} = 0.054$, p値:0.0002

  - **実用上の注意すべき例**
    - $\rho_{xy} = 0.05$ (社会科学分野では相関無し)**の分布から各サンプルサイズで**1000**回乱数発生させたときの**5%**水準で有意差ありの回数**(シミュレーション)
    - **検定では"効果の大きさ"は測っていないことが分かる例**

サンプルサイズ	$n = 100$	$n = 1{,}000$	$n = 5{,}000$	$n = 10{,}000$
有意差あり	66回	356回	950回	1000回