第 IX 部　母平均の差の検定

目次

7 F 分布

7.1 はじめに

- データのばらつきに差があるかを判定する方法を紹介します。
- ばらつきの指標は χ^2 分布を使います。
- 二つの χ^2 分布の比で差の有無を判断します。

7.1.1 ポイント

- χ^2 分布の自由度
- F 分布
- F 分布の自由度
- F 分布の臨界値

7.1.2 標本平均

標本 A に属する要素の数を n とし、各要素を x_i であらわす。標本 A に属する要素の総和を $\sum_{x_i \in A}^{n} x_i$ であらわす。

標本 A に属する要素の総和を n で割った値を『**標本平均**』といい $\bar{X}$ であらわす。

$$\bar{X} = \frac{1}{n} \sum_{x_i \in A}^{n} x_i \tag{1}$$

7.1.3 標本分散

要素の数が n の標本 A に属する要素 x_i の分散を『**標本分散**』といい、$V_{\bar{X}}$ であらわす。

$$V_{\bar{X}} = \frac{1}{n} \sum_{x_i \in A}^{n} \left(x_i - \bar{X}\right)^2 \tag{2}$$

7.1.4　不偏分散

要素の数が n である標本 A に属する要素 x_i の『「標本平均 $\bar{X}$ からの偏差」の二乗』の総和を $n-1$ で割った値を『**不偏分散**』といい s^2 であらわす。

$$s^2 = \frac{1}{n-1}\sum_{x_i \in A}^{n}\left(x_i - \bar{X}\right)^2 \tag{3}$$

この時の $n-1$ を『**自由度**』という。

7.2　標準正規分布

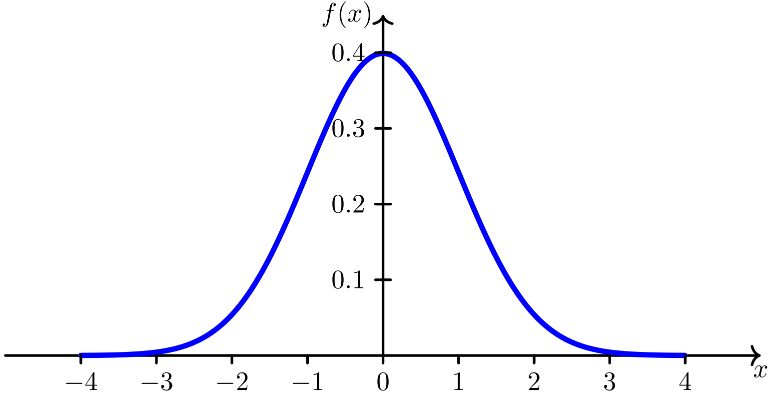
正規曲線

$$f(z) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\right) \tag{4}$$

を確率密度関数に持つ確率分布を標準正規分布といい、$N(0,\ 1)$ であらわす。

標準正規分布 $N(0,\ 1)$ は、平均 0, 分散 1 である。

図 1 標準正規分布の確率密度関数



7.3 χ^2 分布

標準正規分布 $N(0,\ 1)$ に従う n 個の確率変数 $z_1,\ z_2,\ \cdots,\ z_n$ の平方和から統計量を作り $\chi^2_{(0)}$ とあらわす。

$$
\begin{aligned}
\chi^2_{(0)} &= {z_1}^2 + {z_2}^2 + \cdots + {z_n}^2 && (5) \\
&= \sum_{x_i \in A}^{n} {z_i}^2 && (6)
\end{aligned}
$$

このとき、n を『**自由度**』といい ν をつかってあらわす。$\chi^2_{(0)}$ が従う分布を『**自由度 $\boldsymbol{\nu = n}$ の $\boldsymbol{\chi^2}$ 分布**』という。χ^2 分布は自由度によってその形状を大きく変える。

図 2 自由度 $\nu \leqq 2$ の χ^2 分布の確率密度関数

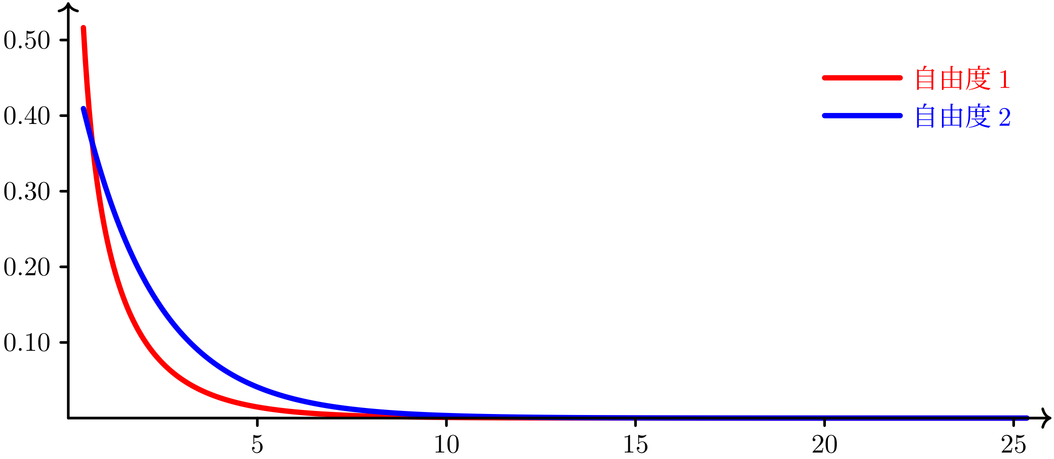
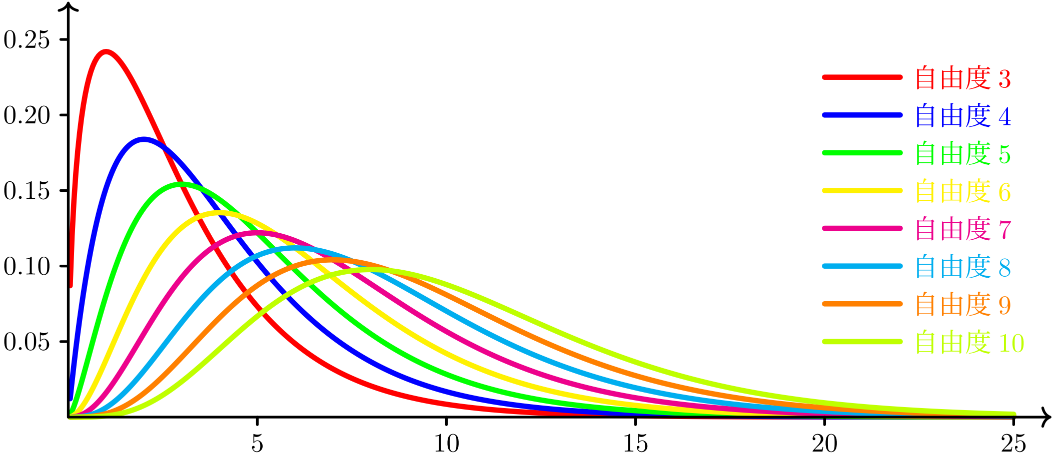


図 3 自由度 $\nu \geqq 3$ の χ^2 分布の確率密度関数



7.3.1 母平均 μ が既知のときの χ^2 分布

母平均 μ が既知であるとする。既知の母平均を μ^* とあらわす。正規母集団 $N\left(\mu^*,\ \sigma^2\right)$ から、サンプル・サイズ n の標本 A を抽出し、それぞれの要素を x_i であらわす。このとき、

$$\chi^2_{(\mu^*)} = \sum_{x_i \in A}^{n} \left(\frac{x_i - \mu^*}{\sigma}\right)^2 \tag{7}$$

は、自由度 $\nu = n$ の χ^2 分布に従う。

7.3.2　母平均 μ が未知のときの χ^2 分布

正規母集団 $N\left(\mu,\ \sigma^2\right)$ から、サンプル・サイズ n の標本を抽出し、それぞれの要素を x_i であらわす。母平均が未知なので、母平均の代わりに標本平均 $\bar{X}$ を用いた統計量

$$\chi^2_{(\bar{X})} = \sum_{x_i \in A}^{n} \left(\frac{x_i - \bar{X}}{\sigma}\right)^2 \tag{8}$$

は、自由度 $\nu = n - 1$ の χ^2 分布に従うことが知られている。

7.4　$\boldsymbol{F}$ 分布

自由度 $\nu = n$ の χ^2 分布に従う確率変数を χ_x^2 とする。自由度 $\nu = m$ の χ^2 分布に従う確率変数を χ_y^2 とする。この二つの χ^2 分布に従う変数から、統計量

$$F_{(n,m)} = \frac{\left(\frac{\chi_x^2}{n}\right)}{\left(\frac{\chi_y^2}{m}\right)} \tag{9}$$

を作る。このとき統計量 $F_{(n,m)}$ は、『**自由度 $\boldsymbol{(n, m)}$ の $\boldsymbol{F}$ 分布**』に従う。

図 4　$n = m$; 5 ∼ 20 の F 分布

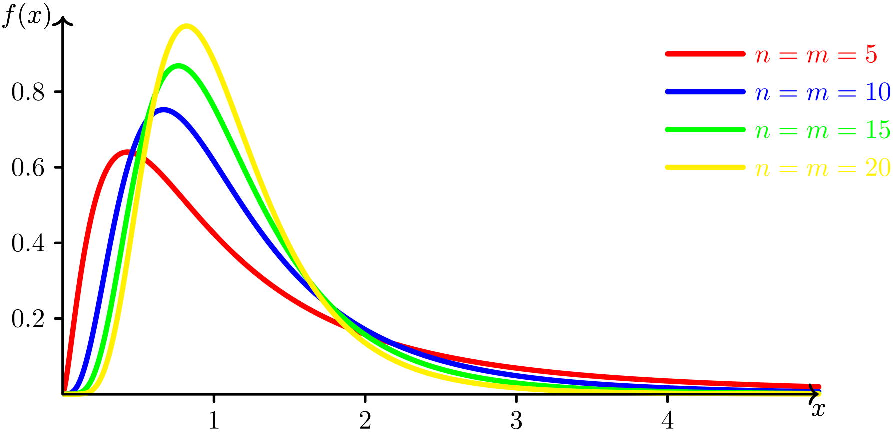


図 5 $n=m$;10 ∼ 40 の F 分布

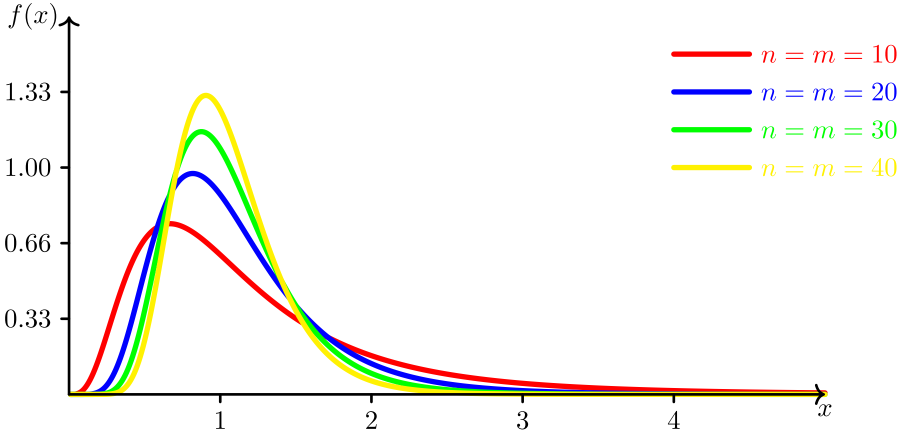


図 6 $\boldsymbol{n = m}$;100 ∼ 400 の $\boldsymbol{F}$ 分布

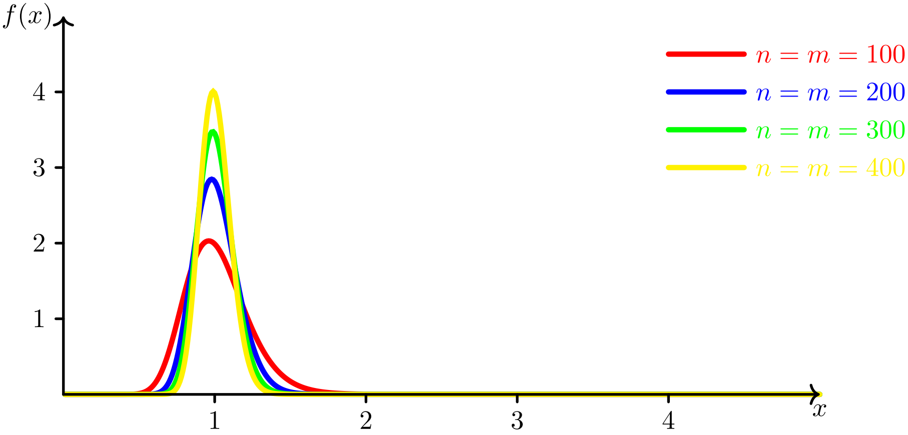
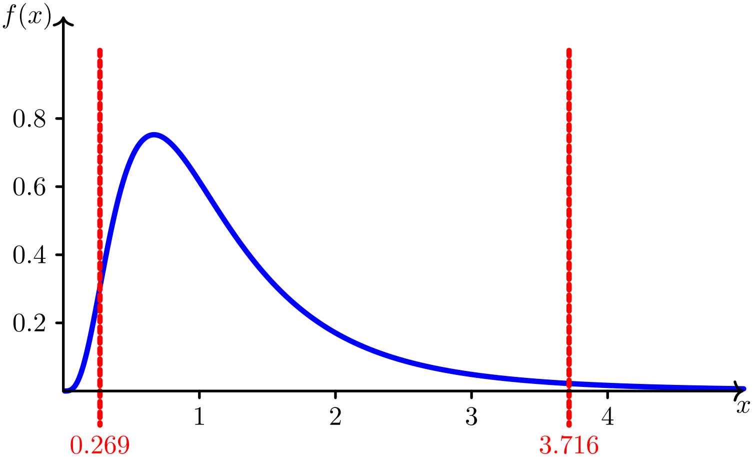


図 7 $n = 10$, $m = 10$ の F 分布の $\alpha = 0.05$ 臨界値



7.4.1 χ^2 分布を自由度で割った値

標準正規分布 $N(0,\ 1)$ の平均は 0 である。従って、

$$z \sim N(0,\ 1) \tag{10}$$

である z は、平均からの偏差

$$z = z - 0 \tag{11}$$

である。そして二乗は平均からの偏差の二乗

$$z^2 = (z-0)^2 \tag{12}$$

である。これを n 個足した値が、

$$\chi^2_{(z)} = \sum_{x_i \in A}^{n} (z_i - 0)^2 \tag{13}$$

である。このときの n が自由度 $\nu = n$ である。この $\chi^2_{(z)}$ を自由度 $\nu = n$ で割ると、

$$\frac{\chi^2_{(z)}}{n} = \frac{1}{n} \sum_{x_i \in A}^{n} (z_i - 0)^2 \tag{14}$$

を得るが、これは Z の分散 V_z である。

7.5　等分散の検定

独立な 2 つの確率変数 $X = x_i,\ Y = y_i$ がそれぞれ

$$x_i \sim N\left(\mu_x,\ {\sigma_x}^2\right) \quad ; i = 1,\ 2,\ \cdots,\ n \tag{15}$$

$$y_j \sim N\left(\mu_y,\ {\sigma_y}^2\right) \quad ; j = 1,\ 2,\ \cdots,\ m \tag{16}$$

であるとする。この 2 つの確率変数の母分散が等しいかどうかを検定することを考える。

$X,\ Y$ の不偏分散をそれぞれ ${s_x}^2,\ {s_y}^2$ とすると、

$${s_x}^2 = \frac{1}{n-1}\sum_{x_i \in A}^{n}\left(x_i - \bar{X}\right)^2 \tag{17}$$

$${s_y}^2 = \frac{1}{m-1}\sum_{y_j \in B}^{m}\left(y_j - \bar{Y}\right)^2 \tag{18}$$

である。

$s_x{}^2$, $s_y{}^2$ の大きい方を分子に、小さい方を分母にして統計量 F_0 をつくる。

$$s_x{}^2 > s_y{}^2 \tag{19}$$

とすると、

$$F_0 = \frac{s_x{}^2}{s_y{}^2} \tag{20}$$

は、自由度 $(n-1,\ m-1)$ の F 分布に従う。

ここで帰無仮説 H_0 と対立仮説 H_1 は

$$H_0 : \sigma_x{}^2 = \sigma_y{}^2 \tag{21}$$

$$H_1 : \sigma_x{}^2 \neq \sigma_y{}^2 \tag{22}$$

として、$\alpha = 0.05$ とすると、

$$F_0 \geqq F\,(n-1,\ m-1,\ 0.025) \tag{23}$$

のとき帰無仮説 H_0 を棄却する。

例題

2 群の分散 ${\sigma_A}^2$ と ${\sigma_B}^2$ に差があるかを検定しなさい。有意水準は $\alpha = 5\%$ とする。

表 1 数値例 1

A 群	10.5	9.8	10.9	9.7	10.6
B 群	11.1	11.7	10.8	11.5	10.9

解法

$$H_0 : {\sigma_A}^2 = {\sigma_B}^2$$

$$H_1 : {\sigma_A}^2 \neq {\sigma_B}^2$$

標本平均をもとめる。

$$\bar{X}_A = \frac{10.5 + 9.8 + 10.9 + 9.7 + 10.6}{5} = 10.3 \tag{24}$$

$$\bar{X}_B = \frac{11.1 + 11.7 + 10.8 + 11.5 + 10.9}{5} = 11.2 \tag{25}$$

不偏分散は

$$
{s_A}^2 = \frac{(10.5-10.3)^2+(9.8-10.3)^2+(10.9-10.3)^2+(9.7-10.3)^2+(10.6-10.3)^2}{5-1} \tag{26}
$$

$$
= 0.275 \tag{27}
$$

$$
{s_B}^2 = \frac{(11.1-11.2)^2+(11.7-11.2)^2+(10.8-11.2)^2+(11.5-11.2)^2+(10.9-11.2)^2}{5-1} \tag{28}
$$

$$
= 0.15 \tag{29}
$$

である。

$$
{s_A}^2 > {s_B}^2 \tag{30}
$$

なので、${s_B}^2$ を分母、${s_A}^2$ を分子として統計量 F_0 を求める。

$$F_0 = \frac{{s_A}^2}{{s_B}^2} = \frac{0.275}{0.15} = 1.833\cdots \tag{31}$$

自由度は

$$\nu_A = 5 - 1 = 4 \tag{32}$$

$$\nu_B = 5 - 1 = 4 \tag{33}$$

なので有意水準 5% の臨界値は

$$F(4, 4, 0.025) = 9.604529885 \tag{34}$$

である[1)]。

$$F_0 < F(4, 4, 0.025) \tag{35}$$

なので、H_0 は棄却されず、「${\sigma_A}^2$ と ${\sigma_B}^2$ は異なる」とは言えない。

1）Excel で F.INV.RT 関数で算出している。両側併せて 5%を取っているので上側 2.5%の臨界値を求めている。

問題 IX − 7 − 1

2 群の分散 ${\sigma_A}^2$ と ${\sigma_B}^2$ に差があるかを検定しなさい。有意水準は $\alpha = 5\%$ とする。

$F_{(11,11,0.025)}$ の上側の臨界値を 3.473 699 051、下側の臨界値を 0.287 877 558 とする。

A 群	117.3	117.0	123.0	122.6	129.0	129.0	134.5	133.9	140.6	140.1	146.7	146.0
B 群	116.1	115.9	122.4	122.3	128.6	128.0	134.6	134.0	142.0	140.7	148.3	148.0

解例 IX − 7 − 1

$$H_0 : {\sigma_A}^2 = {\sigma_B}^2$$

$$H_1 : {\sigma_A}^2 \neq {\sigma_B}^2$$

s_A	108.039 015 2
s_B	130.288 106 1
ν_A	11
ν_B	11
F_0	1.205 935 706

$$F_0 < F(4, 4, 0.025) \tag{36}$$

なので、H_0 は棄却されず、「${\sigma_A}^2$ と ${\sigma_B}^2$ は異なる」とは言えない。

問題 IX − 7 − 2

2 群の分散 ${\sigma_A}^2$ と ${\sigma_B}^2$ に差があるかを検定しなさい。有意水準は $\alpha = 5\%$ とする。

$F_{(11,11,0.025)}$ の上側の臨界値を 3.473 699 051、下側の臨界値を 0.287 877 558 とする。

A 群	21.7	21.4	24.3	24.0	27.9	27.6	31.3	30.2	35.6	35.4	39.8	38.8
B 群	20.9	21.0	24.0	23.8	27.1	26.8	30.6	30.2	35.5	34.3	40.1	39.9

解例 IX − 7 − 2

$$H_0 : {\sigma_A}^2 = {\sigma_B}^2$$

$$H_1 : {\sigma_A}^2 \neq {\sigma_B}^2$$

s_A	41.536 969 70
s_B	45.750 606 06
ν_A	11
ν_B	11
F_0	1.101 443 037

$$F_0 < F(4, 4, 0.025) \tag{37}$$

なので、H_0 は棄却されず、「${\sigma_A}^2$ と ${\sigma_B}^2$ は異なる」とは言えない。

問題 IX − 7 − 3

2 群の分散 ${\sigma_A}^2$ と ${\sigma_B}^2$ に差があるかを検定しなさい。有意水準は $\alpha = 5\%$ とする。

$F_{(11,11,0.025)}$ の上側の臨界値を 3.473 699 051、下側の臨界値を 0.287 877 558 とする。

A 群	153.7	154.3	154.6	154.1	161.0	161.0	161.9	161.2	166.3	166.4	166.5	167.0
B 群	152.4	152.6	152.9	152.6	155.2	155.6	155.5	155.4	156.6	156.8	157.1	156.9

解例 IX − 7 − 3

$$H_0 : {\sigma_A}^2 = {\sigma_B}^2$$
$$H_1 : {\sigma_A}^2 \neq {\sigma_B}^2$$

s_A	28.160 606 06
s_B	3.391 515 152
ν_A	11
ν_B	11
F_0	8.303 252 323

$$F_0 > F(4, 4, 0.025) \tag{38}$$

なので、H_0 は棄却され、「${\sigma_A}^2$ と ${\sigma_B}^2$ は等しい」とは言えない。

問題 IX − 7 − 4

2 群の分散 ${\sigma_A}^2$ と ${\sigma_B}^2$ に差があるかを検定しなさい。有意水準は $\alpha = 5\%$ とする。
$F_{(11,11,0.025)}$ の上側の臨界値を 3.473 699 051、下側の臨界値を 0.287 877 558 とする。

A 群	44.7	45.3	45.6	45.1	49.7	51.0	50.9	49.8	55.5	55.2	55.0	54.5
B 群	44.5	44.7	44.3	43.8	47.7	47.4	47.0	47.2	49.9	49.9	49.3	48.6

解例 IX − 7 − 4

$$H_0 : {\sigma_A}^2 = {\sigma_B}^2$$

$$H_1 : {\sigma_A}^2 \neq {\sigma_B}^2$$

s_A	17.962 651 52
s_B	4.947 500 00
ν_A	11
ν_B	11
F_0	3.630 652 151

$$F_0 > F(4, 4, 0.025) \tag{39}$$

なので、H_0 は棄却され、「${\sigma_A}^2$ と ${\sigma_B}^2$ は等しい」とは言えない。

問題 IX − 7 − 5

2 群の分散 ${\sigma_A}^2$ と ${\sigma_B}^2$ に差があるかを検定しなさい。有意水準は $\alpha = 5\%$ とする。

$F_{(11,11,0.025)}$ の上側の臨界値を 3.473 699 051、下側の臨界値を 0.287 877 558 とする。

A 群	168.0	169.0	169.0	168.9	170.0	169.6	170.7	170.5	171.0	171.1	171.4	171.1
B 群	157.2	157.4	157.9	157.8	157.7	158.2	158.5	158.4	158.0	158.5	158.7	158.3

解例 IX − 7 − 5

$$H_0 : {\sigma_A}^2 = {\sigma_B}^2$$
$$H_1 : {\sigma_A}^2 \neq {\sigma_B}^2$$

s_A	1.225 681 818
s_B	0.217 272 727
ν_A	11
ν_B	11
F_0	5.641 213 389

$$F_0 > F(4, 4, 0.025) \tag{40}$$

なので、H_0 は棄却され、「${\sigma_A}^2$ と ${\sigma_B}^2$ は等しい」とは言えない。

問題 IX – 7 – 6

2 群の分散 ${\sigma_A}^2$ と ${\sigma_B}^2$ に差があるかを検定しなさい。有意水準は $\alpha = 5\%$ とする。
$F_{(11,11,0.025)}$ の上側の臨界値を 3.473 699 051、下側の臨界値を 0.287 877 558 とする。

A 群	58.9	58.7	58.4	57.9	61.2	60.0	60.6	59.9	62.0	61.8	62.4	61.7
B 群	51.6	51.5	50.9	51.3	52.0	52.8	52.0	51.3	53.1	53.3	52.3	51.7

解例 IX − 7 − 6

$$H_0 : {\sigma_A}^2 = {\sigma_B}^2$$
$$H_1 : {\sigma_A}^2 \neq {\sigma_B}^2$$

s_A	1.193 116 635
s_B	0.574 242 424
ν_A	11
ν_B	11
F_0	2.077 722 9

$$F_0 < F(4, 4, 0.025) \tag{41}$$

なので、H_0 は棄却されず、「${\sigma_A}^2$ と ${\sigma_B}^2$ は異なる」とは言えない。

7.6　まとめ

- 自由度 ν_A の χ^2 分布に従う統計量 $\chi^2_{(A)}$ と自由度 ν_B の χ^2 分布に従う統計量 $\chi^2_{(B)}$ の比 $\dfrac{\chi^2_{(A)}}{\chi^2_{(B)}}$ は自由度 $\nu_A,\ \nu_B$ の F 分布に従う。
- 2 つの χ^2 分布に従う統計量 $\chi^2_{(A)}$ と $\chi^2_{(B)}$ の比が

$$\frac{\chi^2_{(A)}}{\chi^2_{(B)}} > 1$$

ならば、上側 2.5 % で検定する。
- 2 つの χ^2 分布に従う統計量 $\chi^2_{(A)}$ と $\chi^2_{(B)}$ の比が

$$\frac{\chi^2_{(A)}}{\chi^2_{(B)}} < 1$$

ならば、下側 2.5 % で検定する。
- F 分布は自由度によって形状が大きく異なる。

参考文献

1. 『経済数学教室 1 巻』小山昭雄著「岩波書店」1994 年 5 月 30 日
2. 『新体系 大学数学 入門の教科書 上下』吉沢光雄著「講談社」2022 年 6 月 16 日
3. 『技術者のための高等数学 7 確率と統計（原書代 8 版）』E・クライツィグ著　田栗正章訳「培風館」2004 年 11 月 30 日第 8 版
4. 『データ解析のための数理統計入門』久保川達也著「共立出版」2023 年 10 月 15 日初版
5. 『統計学入門』東京大学教養学部統計学教室篇「東京大学出版会」1991 年 7 月 9 日初版
6. 『行動科学における統計解析法」芝祐順・南風原朝和著　「東京大学出版会」1990 年 3 月 20 日　初版
7. 『計量経済学の理論』A.S. ゴールドバーガー著, 福地崇生・森口親司訳「東洋経済新報社」

1970 年 9 月 10 日 第 1 刷

8.『計量経済学序説』R.J. ウォナコット・T.H. ウォナコット著, 国府田恒夫・田中一盛訳「培風館」 1975 年 6 月 10 日初版
9.『マーケティング・サイエンス』片平秀貴著　「東京大学出版会」1987 年 4 月 20 日初版
10.『回帰分析』佐和光男著 「朝倉書店」1979 年 4 月 20 日初版
11.『完全独習　統計学入門』小島寛之著「ダイヤモンド社」2006 年 9 月 28 日　初版
12.『分散分析の基礎』高橋敬子著「プレアデス出版」2009 年 10 月 1 日