

分散分析と多変量分散分析 ANOVA and MANOVA

樋口さぶろお <https://hig3.net>

龍谷大学大学院 先端理工学研究科 数理・情報科学コース

計算科学特論 L15(2026-07-27 Mon)

最終更新: Time-stamp: "2026-08-08 Sat 09:10 JST hig"

今日の目標

- 対応のない1元分散分析ができる
- 多変量分散分析の考え方を説明できる



L14-Q1

TA Prob and Sol: 母平均値の差の区間推定 (母分散未知)

ドーナツ製造マシン 1 号, 2 号が製造するドーナツの重さ X, Y g は, 未知の母平均値 μ_1, μ_2 の独立同分布にしたがう確率変数である. 母分散も未知だが, 1 号と 2 号で等しいことがわかっている.

1 号, 2 号で製造したドーナツの重さは次のようだった.

1 号: 51g, 52g, 47g, 50g.

2 号: 55g, 56g, 51g, 52g, 56g, 54g.

- ① $\bar{X}, \bar{Y}$, 合併した不偏標本分散 S^2 を求めよう.
- ② 母平均値の差 $\mu_1 - \mu_2$ を, 信頼係数 $1 - \alpha = 0.99$ で区間推定しよう.

略解

$$\textcircled{1} \quad \bar{X} = 50, \bar{Y} = 54.$$

$$S^2 = \frac{1}{4+6-2}[(51-50)^2 + \dots + (55-54)^2 + \dots] = \frac{1}{8}[14 + 22] = \frac{9}{2}.$$

2

$$\begin{aligned}
 (50 - 54) - t_{0.005}(8) \sqrt{\frac{9}{2} \cdot \left(\frac{1}{4} + \frac{1}{6}\right)} \\
 < \mu_1 - \mu_2 < (50 - 54) + t_{0.005}(8) \sqrt{\frac{9}{2} \cdot \left(\frac{1}{4} + \frac{1}{6}\right)}
 \end{aligned}$$

L14-Q2

TA Prob and Sol: 両側 2 標本 t 検定

ドーナツ製造マシン 1 号, 2 号が製造するドーナツの重さ X_i, Y_j (g) は, 未知の母平均値 μ_1, μ_2 , 未知の共通の母分散の独立な正規分布にしたがう確率変数である.

- 1 号の標本は, サイズ 4, 標本平均値 50g, 不偏標本分散 $\frac{14}{3}g^2$,
- 2 号の標本は, サイズ 6, 標本平均値 54g, 不偏標本分散 $\frac{22}{5}g^2$ だった.

2 つのドーナツ製造マシンの製造するドーナツの重さの母平均値に差があるか知りたい. 帰無仮説を H_0 : 差はない $\mu_1 - \mu_2 = 0$, として, 有意水準 $\alpha = 0.01$ で両側 2 標本 t 検定をしよう.

略解

- ① 有意水準 $\alpha = 0.01$ で,
- ② 母平均値の差の両側 2 標本 t 検定を行う
- ③ 帰無仮説 H_0 を, 「…ドーナツの重さの母平均値は等しい:
 $\mu_1 - \mu_2 = 0$ 」とする. すなわち, 対立仮説 H_1 を, $\mu_1 \neq \mu_2$ とする.

- ④ サイズ m, n の標本の, 標本平均値を $\bar{X}, \bar{Y}$, プールした不偏標本分散を S^2 とすると, 量 $T = \frac{\bar{X} - \bar{Y}}{\sqrt{S^2 \cdot (\frac{1}{n} + \frac{1}{m})}}$ は, 帰無仮説のもとで自由度 $n + m - 2$ の t 分布に従う. この量を検定統計量として用いる.
- ⑤ この標本に対して $T = \frac{-4}{\sqrt{\frac{1}{4+6-2}((4-1)\frac{14}{3} + (6-1)\frac{22}{5}) \cdot (\frac{1}{4} + \frac{1}{6})}} = -2.9218$.
- ⑥ t 分布表より, p 値 $P(|T| > 2.9218)$ は $\alpha = 0.01$ よりも大きい ($0.01 = P(|T| > 3.355)$ だから. あるいは, $2.9218 < t_{0.005}(8)$ だから (といっても同じこと). よって帰無仮説は棄却されない. 母平均値が異なると有意水準 0.01 では結論できない.

L14-Q3

TA Prob and Sol: 両側 2 標本 t2 検定

各 30 人の 2 組で、 $p = 2$ 科目の試験を行った。

X, Y 組の 2 科目の成績の標本平均値ベクトルはそれぞれ $\begin{pmatrix} 65 \\ 70 \end{pmatrix}, \begin{pmatrix} 72 \\ 74 \end{pmatrix}$

だった。プールされた不偏共分散行列は $S = \begin{pmatrix} 100 & 60 \\ 60 & 90 \end{pmatrix}$ だった。

2 組の成績は、同じ共分散行列の正規分布にしたがうと仮定して、有意水準 $\alpha = 0.05$ で、成績の母平均値ベクトルに差があるか検定しよう。

略解

- ① 有意水準 $\alpha = 0.05$ で Hotelling の T^2 検定を行う。
- ② 帰無仮説を、「2 組の母平均値ベクトルは等しい」、対立仮説を「等しくない」とする。

- ③ $\frac{58-2+1}{2 \cdot 58} T^2 = {}^t(\mathbf{X} - \mathbf{Y}) \frac{1}{\frac{1}{30} + \frac{1}{30}} (S^2)^{-1} (\mathbf{X} - \mathbf{Y})$ は, 自由度 2, 57 の F

分布にしたがう.

- ④ F 検定統計量の実現値は

$$T^2 = (-7 \ -4) \frac{1}{\frac{1}{30} + \frac{1}{30}} \begin{pmatrix} 100 & 60 \\ 60 & 90 \end{pmatrix}^{-1} \begin{pmatrix} -7 \\ -4 \end{pmatrix} = 13.36. \text{ F}(2, 57) \text{ にしたがう F 統計量は, } f = \frac{58-2+1}{2 \cdot 58} T^2 = 6.56.$$

- ⑤ $\alpha = 0.05$ の, F 検定統計量の棄却限界は, $F(2, 57)$ の累積分布関数を F と書いたとき, $F^{-1}(1 - 0.05) = 3.159$. 棄却域は $f > 3.159$. よって, 棄却される.

- ⑥ 2 組の母平均値ベクトルは異なると結論する.

L14-Q4

TA Prob and Sol: 両側 2 標本 t² 検定

各 40 人の 2 組で、 $p = 2$ 科目の試験を行った。

X, Y 組の 2 科目の成績の標本平均値ベクトルはそれぞれ $\begin{pmatrix} 60 \\ 65 \end{pmatrix}, \begin{pmatrix} 66 \\ 73 \end{pmatrix}$

だった。プールされた不偏共分散行列は $S = \begin{pmatrix} 100 & 40 \\ 40 & 80 \end{pmatrix}$ だった。

2 組の成績は、同じ共分散行列の正規分布にしたがうと仮定して、有意水準 $\alpha = 0.05$ で、成績の母平均値ベクトルに差があるか検定しよう。

略解

$$T^2 = (-6 \ -8) \frac{1}{\frac{1}{30} + \frac{1}{30}} \begin{pmatrix} 100 & 40 \\ 40 & 80 \end{pmatrix}^{-1} \begin{pmatrix} -7 \\ -4 \end{pmatrix} = 22.00.$$

$F(2, 77)$ にしたがう F 統計量としてみると、

$$F = \frac{78-2+1}{2 \cdot 78} T^2 = 10.86.$$

棄却限界を越え、棄却される。

ここまで来たよ

14 2 標本 t 検定と Hotelling の T^2 検定

15 分散分析と多変量分散分析 ANOVA and MANOVA

- 母分散の検定
- 分散分析
- 多変量分散分析

母分散の両側カイ二乗検定

母平均値母分散未知の正規分布からの標本に基づき、母分散が σ_0^2 かどうか判定したい! (σ_0^2 でないと言いたい)

$X_i \sim N(\mu_0, \sigma^2)$, i.i.d.

- 対立仮説 H_1 母分散 $\sigma \neq \sigma_0$.
- 帰無仮説 H_0 母分散 $\sigma = \sigma_0$.

帰無仮説のもとで、サイズ n の標本の不偏標本分散 S^2 は $(n-1) \times S^2 / \sigma_0^2$ は自由度 $n-1$ のカイ二乗分布にしたがう。

定義 ($\chi_{k,\alpha}^2$)

$$P(W > \chi_{k,\alpha}^2) = \alpha$$

$$P(\chi_{n-1, 1-\frac{\alpha}{2}}^2 < (n-1) \times \frac{S^2}{\sigma_0^2} < \chi_{n-1, \frac{\alpha}{2}}^2) = 1 - \alpha.$$

一般の統計的仮説検定の、レポートや論文での書き方

母集団を決める. 母集団の分布タイプを仮定する. 使う検定を決める/決まる (将来は自作できる).

- ① 「有意水準 $\alpha = \dots$ で」「 $\dots$ 検定を行う」(2,3 を名前で予告する)
- ② 「帰無仮説を $\dots$ とする」「対立仮説を $\dots$ とする」
- ③ 「帰無仮説のもとでナントカ検定統計量 Y は $\dots$ 分布にしたがう」
- ④ 「この標本に対してナントカ検定統計量 Y の実現値は $y = \dots$ である」
- ⑤ (棄却限界を計算して)「棄却域は $\dots$ 」
- ⑥ 「 $\dots$ より帰無仮説を棄却する/受容する」「よって (問に即した対立仮説の内容) と結論する/とは結論できない」

L15-Q1

TA Prob and Sol: 母分散の両側カイ二乗検定

あるファーストフードチェーンのポテトフライ S の重さは、母分散が 4g^2 であることが定められているという。

トレーニング中のアルバイトの人に、ポテトフライ S サイズを 9 個作ってもらったところ、重さは下のようだった (単位は g)。

76, 76, 76, 76, 80, 84, 84, 84, 84.

このアルバイトの作るポテトフライ S の重さの母分散 σ^2 は、 2^2 と異なるか？ 重さが正規分布にしたがうと仮定し、有意水準 $\alpha = 0.05$ で、母分散のカイ二乗検定で判定しよう。

略解

- ① 有意水準 $\alpha = 0.05$ で、母分散の両側カイ二乗検定を行う。

- ② 帰無仮説を, 「アルバイトの…重さの正規分布の母分散 σ^2 は, 2^2 に等しい」とする
- ③ サイズ n の標本の不偏標本分散を s^2 とすると, 量 $Y = (n - 1) \times \frac{s^2}{2^2}$ は, 自由度 $n - 1$ のカイ二乗分布に従う. この量を検定統計量として用いる.
- ④ この標本に対して $Y = (n - 1) \times \frac{s^2}{2^2} = (9 - 1) \cdot \frac{16}{2^2} = 32$.
- ⑤ カイ二乗分布表より, この値は, 棄却域 $Y < \chi_{n-1, 1-\frac{\alpha}{2}}^2 = 2.180$, $Y > \chi_{n-1, \frac{\alpha}{2}}^2 = 17.53$ に含まれるので帰無仮説を棄却する. 母分散は 2^2 と異なると結論する.

棄却域に含まれなかったときの書き方は, 「棄却域に含まれないので帰無仮説は棄却できない. 母分散と 2^2 が異なるとは結論できない.」

母平均値と母分散

	母平均値	母分散
推定量の分布	標本平均値は正規分布	不偏標本分散はカイ二乗分布
区間推定	両端は、標本平均値から和で修正	両端は、不偏標本分散から積で修正
検定 (2 標本の比較)	母平均値が異なるとき、標本平均値の差は正規分布。母分散未知のとき不偏標本分散を使うと、差は t 分布	母分散が異なるとき、不偏標本分散の比は F 分布

	母平均値	母分散
検定 (ある値と標本との比較)	t 検定	カイ二乗検定
検定 (2 標本の比較)	2 標本 t 検定	F 検定
	差	比

×

(片, 両)

母分散の比の F 検定

共通の正規分布からの 2 標本

- $X_1, \dots, X_{n_1} \sim N(\mu, \sigma^2)$, サイズ $n_1 \rightsquigarrow$ 不偏標本分散 S_1^2
- $Y_1, \dots, Y_{n_2} \sim N(\mu, \sigma^2)$, サイズ $n_2 \rightsquigarrow$ 不偏標本分散 S_2^2

$W_{n_i-1} = (n_i - 1) \times S_i^2 / \sigma^2$ は自由度 $n_i - 1$ のカイ二乗分布しがたう.

$$\text{不偏標本分散の比} \quad \frac{S_1^2}{S_2^2} = \frac{W_{n_1-1} \cdot \sigma^2 / (n_1 - 1)}{W_{n_2-1} \cdot \sigma^2 / (n_2 - 1)} = \frac{W_{n_1-1} / (n_1 - 1)}{W_{n_2-1} / (n_2 - 1)}$$

は自由度 $(n_1 - 1, n_2 - 1)$ の F 分布にしたがう (1 に近いはず). これが大小に極端な値をとったら, 母分散が共通の値 σ^2 であることを疑おう.

L15-Q2

Quiz(F 検定)

同じチェーン店のドーナツ屋さんの、各支店で製造されるオールドファッションドーナツの重さは正規分布にしたがうという (母平均値や母分散は支店によってことなるかも).

支店1でオールドファッションドーナツのサイズ10の標本を得たところ、不偏標本分散は $28g^2$ だった.

支店2でオールドファッションドーナツのサイズ5の標本を得たところ、不偏標本分散は $4g^2$ だった.

2つの支店で、分布の母分散が異なるかどうか、有意水準 $\alpha = 0.05$ で検定しよう.

- ① 有意水準 $\alpha = 0.05$ で, 母分散の比の両側 F 検定を行う
- ② 帰無仮説 H_0 を, 「…ドーナツの重さの母分散は等しい: $\sigma_1^2/\sigma_2^2 = 1$ 」とする. すなわち, 対立仮説 H_1 を, $\sigma_1^2/\sigma_2^2 \neq 1$ とする.
- ③ 標本サイズを n_1, n_2 , 不偏標本分散を S_1^2, S_2^2 とすると, 量
- ④ この標本に対して
- ⑤ F 分布表より,

ここまで来たよ

14 2 標本 t 検定と Hotelling の T^2 検定

15 分散分析と多変量分散分析 ANOVA and MANOVA

- 母分散の検定
- 分散分析
- 多変量分散分析

分散分析=ANOVA=Analysis of Variance とは

学級 カテゴリ	点数 量的
A	79
A	80
B	78
⋮	⋮
C	86
⋮	⋮

- 多群 (2 群以上何群でも) の間に平均値の差があるかを判定する.
- (カテゴリ変数, 量的変数) のデータで, 量的変数がカテゴリ変数に依存してるかどうかを判定する.

この2つは同じこと.

比較

- 2群の間に母平均値の差があるかを判定する $\leadsto$ 2群のt検定
- (カテゴリ変数, 量的変数) のデータで, 量的変数がカテゴリ変数 (値2個) に依存してるかどうかを判定する $\leadsto$ カテゴリ変数をダミー変数 $X = 0, 1$ と思って単回帰分析
- (量的変数1, 量的変数2) のデータで, 量的変数2が1に依存してるかどうかを判定する $\leadsto$ 単回帰分析

例

問「ドーナツの重さの母平均値は支店に依存しない」か?

i	支店	データ	個数	標本平均値	不偏標本分散
1	瀬田	79,80,80,81	4	80	$\frac{1}{4-1}[(79-80)^2 + \dots]$
2	石山	78,86,81,83	4	82	
3	草津	81,81,80,82	4	81	
計			12	81	

仮定 第 i 群, j 番目の実現値 $Y_{ij} \sim N(\mu_i, \sigma^2)$, i.i.d.

$Y_{ij} = \mu_i + \epsilon_{ij}, \epsilon_{ij} \sim N(0, \sigma^2)$, i.i.d. も同じこと.

そもそも論

瀬田=石山を帰無仮説として2標本t検定，石山=草津を帰無仮説として2標本t検定，草津=瀬田を帰無仮説として2標本t検定，と繰り返せばように思える（どこの間に差があるかまでわかる）．

しかし，多重検定による有意水準の調整が必要になる（検出力が低くなる）．1度の検定で帰無仮説瀬田=石山=草津を棄却して，対立仮説「重さは店による」を主張できないか？

内緒でどの支店間に差があるかを知ってから，それを対立仮説にするのは禁じ手．HARKing (HARK=Hypothesize After Results are Known) の例．

分散分析の用語と記号

問「群内平均値は「群」(=「級」or「水準」, class, level)に依存する」か?

$\mu_1 = \dots = \mu_\ell$ か?

群	データ	個数	群内平均	残差平方和
1	$y_{11}, y_{12}, \dots, y_{1r}$	r	$\bar{y}_{1\bullet}$	$\sum_j (y_{1j} - \bar{y}_{1\bullet})^2$
2	$y_{21}, y_{22}, \dots, y_{2r}$	r	$\bar{y}_{2\bullet}$	$\sum_j (y_{2j} - \bar{y}_{2\bullet})^2$
$\vdots$				
ℓ	$y_{\ell 1}, y_{\ell 2}, \dots, y_{\ell r}$	r	$\bar{y}_{\ell\bullet}$	$\sum_j (y_{\ell j} - \bar{y}_{\ell\bullet})^2$
計		$r\ell$	$\bar{y}_{\bullet\bullet}$	

●はその添字で平均したという意味.

群内平均値 $\bar{y}_{i\bullet} = \frac{1}{r} \sum_{j=1}^r y_{ij}$.

全平均値 $\bar{y}_{\bullet\bullet} = \frac{1}{r\ell} \sum_{i=1}^{\ell} \sum_{j=1}^r y_{ij}$.

分散 (= ちらばり) の比較に言い換え

「横の箱内のちらばりと、縦 (箱の間) のちらばりは同じ」か?

$\mu_1 = \dots = \mu_\ell$ でないなら縦によけいにちらばるはず.

横方向のちらばりの合計: ϵ_{ij} の効果 = 残差平方和

$$S_E = \sum_{i=1}^{\ell} \sum_{j=1}^r (y_{ij} - \bar{y}_{i\bullet})^2 \sim \chi_{(r-1)-(\ell-1)}^2$$

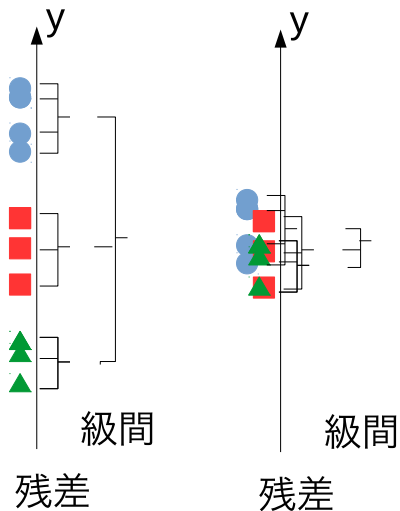
縦方向のちらばりの合計: μ_i の差の効果 = 群間平方和

$$S_A = \sum_{i=1}^{\ell} \sum_{j=1}^r (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2 \sim \chi_{\ell-1}^2$$

すべてのちらばりの合計 = 全平方和

$$S_T = \sum_{i=1}^{\ell} \sum_{j=1}^r (y_{ij} - \bar{y}_{\bullet\bullet})^2$$

実は $S_A + S_E = S_T$.



分散分析 (ANOVA) or 分散分析の F 検定 の方針.

(帰無仮説) $\mu_1 = \dots = \mu_\ell$ のもとで,

S_E は自由度 $\phi_E = r\ell - \ell$ のカイ二乗分布,

S_A は自由度 $\phi_A = \ell - 1$ のカイ二乗分布にしたがう

よって, $F = \frac{S_A/(\ell-1)}{S_E/(r\ell-\ell)}$ は自由度 $(\ell - 1, r\ell - \ell)$ の F 分布にしたがう.

もし $\mu_1 = \dots = \mu_\ell$ でないなら, S_A が, したがって比 F が想定より大きくなる. そのとき, 上側片側検定で帰無仮説を棄却する.

1 元配置の分散分析表

変動要因	平方和	自由度	平均平方	F
群間	S_A	$\phi_A = \ell - 1$	$V_A = S_A / \phi_A$	V_A / V_E
残差	S_E	$\phi_E = (r\ell - 1) - (\ell - 1)$	$V_E = S_E / \phi_E$	
全	S_T	$\phi_T = r\ell - 1$		

L15-Q3

Quiz(分散分析)

次のデータに対して，1元配置の分散分析表を作ろう．有意水準 $\alpha = 0.05$ でF検定しよう．

水準

A_1	11	9	12	9	9
A_2	10	17	18	20	10
A_3	25	23	21	22	24

L15-Q4

Quiz(1 元分散分析)

次のデータに対して，1 元配置の分散分析表を作ろう．有意水準 $\alpha = 0.05$ で，F 分布を用いた水準間の差の検定を行おう．

水準

A_1	10	10	10	10	10
A_2	11	17	18	18	11
A_3	25	23	21	22	24

ここまで来たよ

14 2 標本 t 検定と Hotelling の T^2 検定

15 分散分析と多変量分散分析 ANOVA and MANOVA

- 母分散の検定
- 分散分析
- 多変量分散分析

多変量分散分析=MANOVA=Multivariate ANOVA

$\mathbf{Y}_{ij} \in \mathbb{R}^p$ に対して, 帰無仮説 $\mathbf{Y}_{ij} \sim N(\boldsymbol{\mu}_i, \Sigma)$, i.i.d. p 変量正規分布.

残差平方和に相当する共分散行列 $S_E = \frac{1}{n} \sum_i \sum_j (\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i\cdot})^t (\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i\cdot})$

群間平方和に相当する共分散行列 $S_A = \frac{1}{n} \sum_i \sum_j (\bar{\mathbf{y}}_{i\cdot} - \bar{\mathbf{y}}_{\cdot\cdot})^t (\bar{\mathbf{y}}_{i\cdot} - \bar{\mathbf{y}}_{\cdot\cdot})$

全平方和に相当する共分散行列 $S_T = \frac{1}{n} \sum_i \sum_j (\mathbf{y}_{ij} - \bar{\mathbf{y}}_{\cdot\cdot})^t (\mathbf{y}_{ij} - \bar{\mathbf{y}}_{\cdot\cdot})$

いずれも $p \times p$ 行列.

検定統計量 Wilks' Lambda

標本の

$$A = S_A S_E^{-1}$$

の k 番目の固有値を λ_k としたとき、
検定統計量 Wilks' Lambda を

$$\Lambda = \prod_{k=1}^p \frac{1}{1 + \lambda_k} = \det(I + A)^{-1} = \frac{\det(S_E)}{\det(S_E + S_A)}$$

で定める.

Y の正規性の仮定の下で, S は Wishart 分布に従い, Λ は Wilks のラムダ分布 $\Lambda(p, r-1, \ell-1)$ にしたがう. Λ は $[0, 1]$ に分布し, $\Lambda = 0$ に近いところに棄却域がある.

他にも何個かの検定統計量が提案されている.

検定統計量は自分で発明できる. 帰無仮説が正しい時の値がわかり, かつ検定力の大きい検定統計量を作る競争