第 12 回 続高橋セミナー

層別因子を含む探索的な回帰分析入門

2024 年 1 月 19 日

第 2 章　デザイン行列を活用した 1 因子実験データの解析

平方和の分解，および，分散の加法性に基づく各種の実験データの解析法は，実験計画法を発展させてきた．伝統的な実験計画法は，多くの要因を含む多彩な実験データに対して手計算でも対処できるように，長年の創意工夫の集大成とも言うべき優れた方法である．ただし，量的変数に対しては，幾つかの水準を設定し質的変数として扱うことを前提にしなければならない問題を内在している．先進的な統計ソフトでは，質的変数を量的変数（ダミー変数）に置き換え，量的変数に対する線形モデルによる解析が行われている．だが，その効果的な使い方について十分に知られていない．第 1 章では，伝統的な平方和の分解では対応できない各種の事例について身近にある Excel の回帰分析に加え，デザイン行列 $\boldsymbol{X}$ を用いた Excel の行列計算を用いることにより，きめ細かな解析ができることを示した．本章では，1 因子実験データについて，平方和の分解による解析法と対比しつつ，ダミー変数を用いた線形モデルによる解析方法を詳しく解説する．

第 2 章　　目　次

次ページに続く

連絡先:高橋 行雄, 〒105-0014 東京都 港区 芝 1―12－3－1005
takahashi.stat@nifty.com ，FAX : 03-342-8035

第 12 回 続・高橋セミナー
層別因子を含む探索的な回帰分析入門

目 次 （全章）

2. デザイン行列を活用した 1 因子実験データの解析

平方和の分解，および，分散の加法性に基づく各種の実験データの解析法は，実験計画法を発展させてきた．伝統的な実験計画法は，多くの要因を含む多彩な実験データに対して手計算でも対処できるように，長年の創意工夫の集大成とも言うべき優れた方法である．ただし，量的変数に対しては，幾つかの水準を設定し質的変数として扱うことを前提にしなければならない問題を内在している．先進的な統計ソフトでは，質的変数を量的変数（ダミー変数）に置き換え，量的変数に対する線形モデルによる解析が行われている．だが，その効果的な使い方について十分に知られていない．第 1 章では，伝統的な平方和の分解では対応できない各種の事例について身近にある Excel の回帰分析に加え，デザイン行列 $\boldsymbol{X}$ を用いた Excel の行列計算を用いることにより，きめ細かな解析ができることを示した．本章では，1 因子実験データについて，平方和の分解による解析法と対比しつつ，ダミー変数を用いた線形モデルによる解析方法を詳しく解説する．

2.1. 繰り返しが等しい 1 因子実験データ

芳賀（2014），「医薬品開発のための統計解析 第 2 部 実験計画法 改訂版」，の「第 1 章 質的因子の 1 因子実験」に示されている繰り返しが 5 の 1 因子実験データを表 2.1 に示す．実験は，4 種類の薬剤を取り上げ，薬効の違いを調べるため，各薬剤を 5 匹の動物に投与し，薬効を評価した結果であり，各薬剤について平均値と標準偏差が示されている．

表 2.1　データと平均・標準偏差［芳賀（2014），表示 1.1.1］

	繰り返し						
水準	1	2	3	4	5	平均	標準偏差
A_1	10.8	9.9	9.7	10.4	10.7	10.30	0.48
A_2	10.7	10.6	11.0	10.8	10.9	10.80	0.16
A_3	11.4	10.7	10.9	11.3	11.7	11.20	0.40
A_4	11.9	11.2	11.0	11.1	11.3	11.30	0.35
Excelの関数，平均：Avarage (データの範囲)，標準偏差：StDev.S (データの範囲)							

データが得られたら，何らかのグラフを作成し概観するのが解析の最初の第一歩である．表 2.2 に示すのは，表 2.1 の（4 行×5 列）の Excel シート上の矩形データを行方向に（20 行

×1 列）に並べ替え，因子 A の水準番号を i として含めている．これは，Excel の散布図の X 軸の変数として使うためである．散布図は，水準番号 i とデータ y を選択して基本の「散布図」を作成し，散布図上のディフォルトの「●印」を「データ系列の書式設定」で「〇印」に置き換え，軸の目盛りを整えた結果である．散布図の X 軸の元の目盛りは（1　2　3　4）なので，（A_1　A_2　A_3　A_4）を図形の「テキストボックス」を用いた重ね書きした結果である．各水準の平均を，「データの選択」により散布図上に重ね書きし，●印を「マーカのオプション」で「種類 ー」に変更し，実線を点線に変更し書式を整えている．Excel の「箱ひげ図」は，因子 A とデータ y を選択して作成し，目盛りなどを整えた結果である．

表 2.2　行方向へのデータの並べ替え

因子 A	番号 i	反応 y
A_1	1	10.8
A_1	1	9.9
A_1	1	9.7
A_1	1	10.4
A_1	1	10.7
A_2	2	10.7
A_2	2	10.6
A_2	2	11.0
A_2	2	10.8
A_2	2	10.9
A_3	3	11.4
A_3	3	10.7
A_3	3	10.9
A_3	3	11.3
A_3	3	11.7
A_4	4	11.9
A_4	4	11.2
A_4	4	11.0
A_4	4	11.1
A_4	4	11.3

水準 A	番号 i	平均 $\bar{y}_{i\cdot}$
A_1	1	10.3
A_2	2	10.8
A_3	3	11.2
A_4	4	11.3

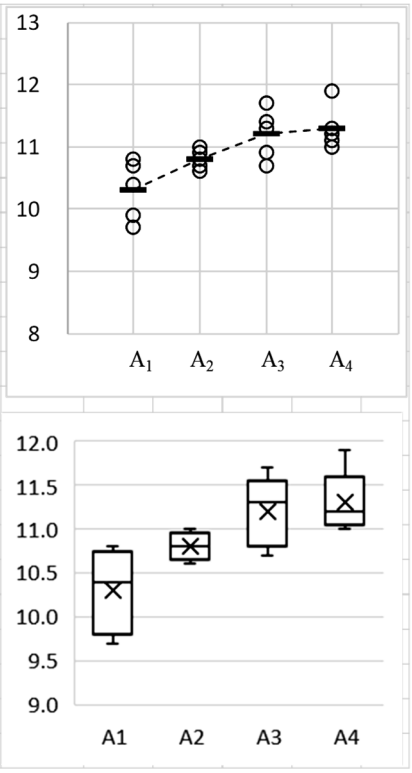

図 2.1 Excel による平均値の折れ線入り散布図および箱ひげ図

各種の平方和の分解による分散分析表

表 2.3 に示すのは，各種の平方和の計算のための Exeel シートである．デザイン行列 $\boldsymbol{X}$ を用いた回帰分析を適用する際には，全てのデータが，行方向に並んでいることを前提にしている．それに合わせ，行方向のデータの並びに対して平方和の分解による解析を行う．データの構造式

$$\left.\begin{aligned} y_{ij} &= \mu + \alpha_i + \varepsilon_{ij} \\ &= \bar{y}_{\cdot\cdot} + (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}) + (y_{ij} - \bar{y}_{i\cdot}) \\ \text{ただし，}\ & \textstyle\sum_i \alpha_i = 0, \quad \sum_i \sum_j \varepsilon_{ij} = 0 \end{aligned}\right\} \tag{2.1}$$

（μ:$\bar{y}_{i\cdot}$の平均，α_i:因子Aの効果，ε_{ij}:誤差）であり，それらに対応する各種の平方和

$$\sum_i \sum_j (y_{ij} - \bar{y}_{\cdot\cdot})^2 = \sum_i \sum_j (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 + \sum_i \sum_j (y_{ij} - \bar{y}_{i\cdot})^2 \qquad (2.2)$$

$$S_{\mathrm{T}} \quad = \quad S_{\mathrm{A}} \quad + \quad S_e$$

を計算する．

表 2.3 Excel による各種の平方和の計算

			反応	1: μ^			2: α^$_i$	3: ε^$_{ij}$	1+2+3	分散分析表		
A	i	j	y_{ij}	$\bar{y}_{\cdot\cdot}$	$y_{ij} - \bar{y}_{\cdot\cdot}$	$\bar{y}_{i\cdot}$	$\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}$	$y_{ij} - \bar{y}_{i\cdot}$	y_{ij}	要因	自由度	平方和
A_1	1	1	10.80	10.90	-0.10	10.30	-0.60	0.50	10.80	A	3	3.10
	1	2	9.90	10.90	-1.00	10.30	-0.60	-0.40	9.90	e	16	2.18
	1	3	9.70	10.90	-1.20	10.30	-0.60	-0.60	9.70	T	19	5.28
	1	4	10.40	10.90	-0.50	10.30	-0.60	0.10	10.40			
	1	5	10.70	10.90	-0.20	10.30	-0.60	0.40	10.70			
A_2	2	1	10.70	10.90	-0.20	10.80	-0.10	-0.10	10.70			
	2	2	10.60	10.90	-0.30	10.80	-0.10	-0.20	10.60			
	2	3	11.00	10.90	0.10	10.80	-0.10	0.20	11.00			
	2	4	10.80	10.90	-0.10	10.80	-0.10	0.00	10.80			
	2	5	10.90	10.90	0.00	10.80	-0.10	0.10	10.90			
A_3	3	1	11.40	10.90	0.50	11.20	0.30	0.20	11.40			
	3	2	10.70	10.90	-0.20	11.20	0.30	-0.50	10.70			
	3	3	10.90	10.90	0.00	11.20	0.30	-0.30	10.90			
	3	4	11.30	10.90	0.40	11.20	0.30	0.10	11.30			
	3	5	11.70	10.90	0.80	11.20	0.30	0.50	11.70			
A_4	4	1	11.90	10.90	1.00	11.30	0.40	0.60	11.90			
	4	2	11.20	10.90	0.30	11.30	0.40	-0.10	11.20			
	4	3	11.00	10.90	0.10	11.30	0.40	-0.30	11.00			
	4	4	11.10	10.90	0.20	11.30	0.40	-0.20	11.10			
	4	5	11.30	10.90	0.40	11.30	0.40	0.00	11.30			
	平均		10.90	平方和	**5.28**		**3.10**	**2.18**				
					S_{T}		S_{A}	S_e				

表 2.3 に示した各平方和の計算は，以下の手順で行っている．まず，総平均$\bar{y}_{\cdot\cdot}$の計算は，Excel の Average () 関数を用いて

$$\left.\begin{aligned} \bar{y}_{\cdot\cdot} &= \sum_i \sum_j y_{ij} / N, \quad N:\text{データ数} \\ &= \text{Average}\ (y_{ij}\text{の範囲}) \ = \ 10.90 \end{aligned}\right\} \qquad (2.3)$$

として計算し「μ」の推定値としている．総平方和S_{T}は，反応y_{ij}と総平均$\bar{y}_{\cdot\cdot}$の差$(y_{ij} - \bar{y}_{\cdot\cdot})$の平方和であり，して SumSq () 関数により，

$$\left.\begin{aligned} S_{\mathrm{T}} &= \sum_i \sum_j (y_{ij} - \bar{y}_{\cdot\cdot})^2 \\ &= \text{SumSq}((y_{ij} - \bar{y}_{\cdot\cdot})\text{の範囲}) \ = \ 5.28 \end{aligned}\right\} \qquad (2.4)$$

として求めている．なお，S_{T}は，DevSq (y_{ij}の範囲) 関数で偏差平方和を直接求められるが，元のシグマを用いた式に沿った計算法としている．因子 A の各水準の平均値は，水準ごとに Average () 関数を適用し，水準平均$\bar{y}_{i\cdot}$を求めている．

$$\left.\begin{aligned}\bar{y}_{i}. &= \textstyle\sum_j (y_{ij})/n_i, \qquad n_i : \text{各水準のデータ数} \\ &= \text{Average}(\bar{y}_{i}.\text{の範囲})\end{aligned}\right\} \tag{2.5}$$

因子 A の平方和 S_{A} は，水準平均 $\bar{y}_{i}.$ と総平均 $\bar{y}..$ との差 $(\bar{y}_{i}. - \bar{y}..)$ を計算し，SumSq () 関数で

$$\left.\begin{aligned}S_{\mathrm{A}} &= \textstyle\sum_i \sum_j (\bar{y}_{i}. - \bar{y}..)^2 \\ &= \text{SumSq}((\bar{y}_{i}. - \bar{y}..)\ \text{の範囲}) \ = \ 3.10\end{aligned}\right\} \tag{2.6}$$

を求めている．残差平方和 S_e は，反応 y_{ij} と因子 A の水準平均 $\bar{y}_{i}.$ との差 $(y_{ij} - \bar{y}_{i}.)$ を計算し，SumSq () 関数で

$$\left.\begin{aligned}S_e &= \textstyle\sum_i \sum_j (y_{ij} - \bar{y}_{i}.)^2 \\ &= \text{SumSq}((y_{ij} - \bar{y}_{i}.)\text{の範囲}) \ = \ 2.18\end{aligned}\right\} \tag{2.7}$$

を求めている．総平方和 S_{T} が，

$$\left.\begin{aligned}S_{\mathrm{T}} &= S_{\mathrm{A}} + S_e \\ 5.28 &= 3.10 + 2.18\end{aligned}\right\} \tag{2.8}$$

のように S_{A} と S_e の和に分解できるも表 2.3 右上段の分散分析表の形式で確認できる．また，式 (2.1) のデータの構造式 $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ に対し，

$$\left.\begin{aligned}y_{ij} &= \ \hat{\mu} \ + \ \hat{\alpha}_i \ + \ \hat{\varepsilon}_{ij} \\ y_{1,1} \quad 10.80 &= 10.90 + (-0.60) + 0.50 \\ y_{1,2} \quad 9.90 &= 10.90 + (-0.60) + (-0.40) \\ &\vdots \\ y_{4,5} \quad 11.30 &= 10.90 + 0.40 + 0.00\end{aligned}\right\} \tag{2.9}$$

のように成り立っていることも表 2.3 右端の（1+2+3）の列で確認できる．

ダミー変数を用いた線形モデル

平方和の分解による分散分析表の作成は，取り上げる因子（変数）が全て質的変数であることを前提にしている．線形モデルによる解析では，質的変数を量的変数，いわゆるダミー変数に変換する必要がある．2 水準の（なし，あり）を（0，1）と置き換えてダミー変数とすると，線形モデルで推定されたパラメータが，「なし」に対する「あり」の差の推定値となり，結果の解釈がしやすい．そのために，ダミー変数といえば「(0，1）に変換する方法である」と断定する人が大多数である．

2 水準の（1：男，2：女）or（1：Female，2：Male）とコード化されている場合に，(0，1）型か，(1，0）型かどちらの型を選ぶかは，悩ましい問題でもある．そのまま（1，2）型ダミー変数としたらどうなのであろうか．実際に試してみると，パラメータの推定値は（0，1）型と同じであり，切片の推定値が異なるだけである．(1，0）型とすると，パラメータの推定値の符号が変化する．

ただし，多水準の因子，複数の因子を同時に取り扱い，交互作用なども考慮するような場合には，(0，1）型ダミー変数は，各種の推定値とその 95%信頼区間の計算に際し，不都合な側面が出てくる．これは，式 (2.1) で示したデータの構造式では，効果 α_i について

$$\sum_i \alpha_i = 0 \tag{2.10}$$

という制約条件が付いており，2 水準であれば，

$$\left.\begin{aligned} \alpha_1 + \alpha_2 = 0 \\ \alpha_2 = -\alpha_1 \end{aligned}\right\} \tag{2.11}$$

を満たす必要がある．このような制約条件を実現するためのダミー変数は，足して 0 となる（1，–1）対比型ダミー変数が必要となる．因子 A が 4 水準の場合には，表 2.4 左に示すように第 4 水準が –1 となるような（1，–1）対比型ダミー変数とすることにより，データの構造式の制約条件を満たすことができる．もちろん，表 2.4 右に示すように（0，1）型ダミー変数とすることも可能である．ダミー変数名を（a_2，a_3，a_4）としたのは，ダミー変数が「1」となる因子 A の水準に対対させるための配慮である．

表 2.4 （1，–1）対比型ダミー変数 vs.（0，1）型ダミー変数

	(1, –1) 対比型				(0, 1) 型		
	a_1	a_2	a_3		a_2	a_3	a_4
A_1	1	0	0	A_1	0	0	0
A_2	0	1	0	A_2	1	0	0
A_3	0	0	1	A_3	0	1	0
A_4	-1	-1	-1	A_4	0	0	1
和	0	0	0	和	1	1	1
平均	0	0	0	平均	1/4	1/4	1/4

表 2.5 左に示すのは，切片に $x_0 = 1$ を明示的に加え，因子 A を（1，–1）対比型ダミーとし（20 行×4 列）のデザイン行列 $\boldsymbol{X}$ を設定している．線形モデルは，切片の推定のための変数 x_0 を含めた式を

$$\left.\begin{aligned} y_{ij} = \beta_0 x_0 + \beta_1 a_{1,i} + \beta_2 a_{2,i} + \beta_3 a_{3,i} + \varepsilon_{ij} \\ \text{or} \\ y_{ij} = \theta_0 x_0 + \theta_1 a_{1,i} + \theta_2 a_{2,i} + \theta_3 a_{3,i} + \varepsilon_{ij} \end{aligned}\right\} \tag{2.12}$$

とする．通常の回帰分析では，切片のための変数は内部で補ってくれるが，Excel により各種の推定値を求め，その 95%信頼区間求めるためには，パラメータの分散と共分散が必要となる．それらを求めるためには，切片 $x_0 = 1$ を含めた（20 行×4 列）のデザイン行列 $\boldsymbol{X}$ が必要不可欠である．

表 2.5 右上段に Excel の「分析ツール」の「回帰分析」を適用した結果を示す．デザイン行列には，「切片 x_0」が含まれているが，「入力 X 範囲」の設定で「定数に 0 を使用」をオフとし，ダミー変数（a_1，a_2，a_3）を選択した結果が示されている．ここでは，Excel の「回帰分析」の結果の出力の左側 5 列分に限定し示している．

表 2.5　対比型ダミー変数によるデザイン行列 $\boldsymbol{X}$ を用いた Excel の回帰分析

				デザイン行列 $\boldsymbol{X}$			
A	i	j	y	x_0	a_1	a_2	a_3
A_1	1	1	10.8	1	1	0	0
	1	2	9.9	1	1	0	0
	1	3	9.7	1	1	0	0
	1	4	10.4	1	1	0	0
	1	5	10.7	1	1	0	0
A_2	2	1	10.7	1	0	1	0
	2	2	10.6	1	0	1	0
	2	3	11.0	1	0	1	0
	2	4	10.8	1	0	1	0
	2	5	10.9	1	0	1	0
A_3	3	1	11.4	1	0	0	1
	3	2	10.7	1	0	0	1
	3	3	10.9	1	0	0	1
	3	4	11.3	1	0	0	1
	3	5	11.7	1	0	0	1
A_4	4	1	11.9	1	-1	-1	-1
	4	2	11.2	1	-1	-1	-1
	4	3	11.0	1	-1	-1	-1
	4	4	11.1	1	-1	-1	-1
	4	5	11.3	1	-1	-1	-1

分散分析表,「定数に 0 を使用」off

	自由度	変動	分散	分散比
回帰	3	3.1000	1.0333	7.5841
残差	16	2.1800	**0.1363**	$=\sigma$^^{2}
合計	19	5.2800		

		係数	標準誤差	t	P-値
β^$_0$	切片x_0	10.9000	0.0825	132.0606	0.0000
β^$_1$	a_1	-0.6000	0.1430	-4.1970	0.0007
β^$_2$	a_2	-0.1000	0.1430	-0.6995	0.4943
β^$_3$	a_3	0.3000	0.1430	2.0985	0.0521

パラメータの共分散行列 $\boldsymbol{\Sigma}$ ($\boldsymbol{\beta}$^)=$(\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma$^^{2}

	x_0	a_1	a_2	a_3
β^$_0$	**0.0068**	0.0000	0.0000	0.0000
β^$_1$	0.0000	**0.0204**	-0.0068	-0.0068
β^$_2$	0.0000	-0.0068	**0.0204**	-0.0068
β^$_3$	0.0000	-0.0068	-0.0068	**0.0204**

=Minverse (Mmult (Transpose ($\boldsymbol{X}$の範囲), $\boldsymbol{X}$の範囲))*σ^^{2}

Excel の「回帰分析」は，データを変更した場合などに手動による再計算を行う必要があるので，自動計算の機能がある線形モデル（回帰分析）のための LinEst () 関数を使用することも可能である．ただし，LinEst () 関数の結果を活用するためには，あらかじめ表頭・表側に結果の解釈ができるように書式を整えておく必要があること，さらにパラメータの表示形式が逆順であることなどの理由で，Excel の「回帰分析」を使っている．ここに示した回帰分析の結果は，別シートに出力された結果の一部を選択・コピーし，デザイン行列の横にペーストして書式を整えた結果である．再計算が必要な場合には，新たなシート上に出力された結果の数値のみをコピーし「値のみ」をペーストすることにより，あらかじめ整えた書式が維持される．

表 2.5 右上段に示した分散分析表の「回帰」の行の変動が，表 2.3 に示した平方和の計算による $S_{\mathrm{A}}=3.10$ となり，残差の変動が $S_e=2.18$，合計が $S_{\mathrm{T}}=5.28$ に対応する．切片の推定値 $\hat{\beta}_0=11.90$ は，一般平均 μ の推定値であり $\bar{y}_{\cdot\cdot}$ に等しい．ダミー変数 a_1 のパラメータの推定値 $\hat{\beta}_1=-0.60$ が，データの構造式における効果 α_1 に対応し $p=0.0007$ と有意な差であり，他は有意な差でないことが読み取れる．表 2.5 右下段が（4 行×4 列）のパラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\hat{\sigma}^2$ であり，その対角要素が，回帰パラメータ（係数）の分散の推定値であり，平方根を取ると標準誤差 $SE(\hat{\beta}_1)$ となる．2 行 2 列目の $Var(\hat{\beta}_1)=0.0204$ の平方根が，$\sqrt{Var(\hat{\beta}_1)}=0.1430$ となり標準誤差 $SE(\hat{\beta}_1)$ に等しいことが確かめられる．なお，行列計算に不慣れな場合は，第 5 章を参照のこと．

デザイン行列 $\boldsymbol{X}$ を用いた計算の実際

行列による回帰分析の計算式は，統計の理論を重視する成書で多用されているが，実際のデータの解析方法を例示する成書では，極まれに飾り程度に示されている．Excel の行列関数による計算は，従来のスクリプト型の行列計算言語にない特徴を持っていて，四則演算のごとく誰にでも簡単に Excel シート上に表示された行列計算の結果を直接見ることができる．表 2.5 右下段に示したパラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ の計算は，計算結果が出力される（4 行×4 列）のセルを最初に選択し，

$$\left.\begin{aligned}\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) &= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\hat{\sigma}^2 \\ &= \mathrm{Minverse}(\mathrm{Mmult}(\mathrm{Transpose}(\boldsymbol{X}\text{の範囲}), \boldsymbol{X}\text{の範囲}))*\hat{\sigma}^2\end{aligned}\right\} \qquad (2.13)$$

のように行列関数による計算式を入力した結果が，（4 行×4 列）のセルに直接表示されている．ただし，周辺の変数名などは，Excel シート上に別途書き込んだものである．このように，元のデザイン行列 $\boldsymbol{X}$ と並列して $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ の計算結果を示せることが Excel の特徴であり，良さでもある．なお，従来のスクリプト型の行列計算言語を使った場合には，結果がテキストとして出力されるのが常であり，表 2.5 に示したようなコンパクトな結果に表示するためには，表示形式を整える努力を必要とする．

実験データに限らず観察データに対する解析結果の解釈には，各種の推定値とその 95%信頼区間のグラフ表示が欠かせない．と，言うのはやさしいが，それを実践するのは難儀である．Excel では，パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ が同一シート上に得られ，それを活用することにより，どのような推定値に対しても分散を手軽に求めることができる．推定値の分散・共分散が計算できれば，95%信頼区間は簡単に求めることができ，Excel シート上にある推定値に 95%信頼区間を重ね書きした予測プロファイルの作成も容易にできる．理論的には周知のことであり，新規性はまったくないのであるが，行列計算に引き続き，ストレスなく予測プロファイルの作成ができるようになったことは，画期的である．

伝統的な実験計画法の解析方法は，分散分析表の作成のための計算方法を示すことを目的にしており，各種の推定値の分散を求める際には，「分散の加法性」を注意深く適用した手計算を前提としてきた．パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ を使った分散の計算は，線形モデルの理論で示されているものの，行列計算を前提にしているために難儀なことであった．そこで，「手計算」の方法と対比しつつ，パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ を使った計算法を対比して示す．

データの構造式における効果の推定

式 (2.1) に示したデータの構造式に基づき，表 2.5 左に示したように，デザイン行列 $\boldsymbol{X}$ を設定し，変数（x_0， a_1， a_2， a_3）を用いた回帰分析により得た推定値（$\hat{\beta}_0$， $\hat{\beta}_1$， $\hat{\beta}_2$， $\hat{\beta}_3$）

を表 2.5 右に示した．式 (2.1) のデータの構造式に対応づけると，$\hat{\mu}=\hat{\beta}_0 x_0=\hat{\beta}_0$，$\hat{\alpha}_1=\hat{\beta}_1 a_1=\hat{\beta}_1$，$\hat{\alpha}_2=\hat{\beta}_2 a_2=\hat{\beta}_2$，$\hat{\alpha}_3=\hat{\beta}_3 a_3=\hat{\beta}_3$ が得られる．A4 に対応する効果 $\hat{\alpha}_4$ は，直接推定されてないので，表 2.4 の（1．-1）対比型ダミー変数の設定により，

$$\left.\begin{aligned}\hat{\alpha}_4&=\hat{\beta}_1(-a_1)+\hat{\beta}_2(-a_2)+\hat{\beta}_3(-a_3)\\&=-\hat{\beta}_1-\hat{\beta}_2-\hat{\beta}_3\end{aligned}\right\} \qquad (2.14)$$

で求められる．

表 2.6 に示すのは，パラメータの推定値 $\hat{\boldsymbol{\beta}}$ を用いた各種の線形和 L の推定と 95%信頼区間の計算結果である．$L^{(0)}$ に対し行ベクトル $\boldsymbol{l}^{(0)}=[1\ \ 0\ \ 0\ \ 0]$ が設定され，$L^{(0)}=\hat{\mu}=\boldsymbol{l}^{(0)}\hat{\boldsymbol{\beta}}$ によって一般平均 $\hat{\mu}=10.900$ が推定されている．同様に行ベクトル $\boldsymbol{l}^{(1)}=[0\ \ 1\ \ 0\ \ 0]$ を用いて $L^{(1)}=\boldsymbol{l}^{(1)}\hat{\boldsymbol{\beta}}=\hat{\beta}_1 a_1=\hat{\alpha}_1=-0.60$ が推定されて，行ベクトル $\boldsymbol{l}^{(4)}=[0\ -1\ -1\ -1]$ に対しては，$\hat{\alpha}_4=0.40$ が推定されている．これらの線形和 $L^{(i)}=\boldsymbol{l}^{(i)}\hat{\boldsymbol{\beta}}$ の分散は，$Var(\boldsymbol{l}^{(i)}\hat{\boldsymbol{\beta}})=\boldsymbol{l}^{(i)}\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})(\boldsymbol{l}^{(i)})^T$ によって求められ，この平方根が表 2.5 右中段のパラメータの推定値の標準誤差 SE と一致することが確かめられる．

表 2.6　データの構造式の効果 α_i の推定と 95%信頼区間

		l_0	l_1	l_2	l_3		推定値	分散	標準誤差	幅	95%信頼区間	
	L	x_0	a_1	a_2	a_3	$\hat{\boldsymbol{\beta}}$	$\boldsymbol{l}\hat{\boldsymbol{\beta}}$	$Var(\boldsymbol{l}\hat{\boldsymbol{\beta}})$	SE	$t_{0.05}\times SE$	L95%	U95%
μ	$L^{(0)}$	1	0	0	0	10.900	10.9000	0.0068	0.0825	0.1750	10.7250	11.0750
α_1	$L^{(1)}$	0	1	0	0	-0.600	-0.6000	0.0204	0.1430	0.3031	-0.9031	-0.2969
α_2	$L^{(2)}$	0	0	1	0	-0.100	-0.1000	0.0204	0.1430	0.3031	-0.4031	0.2031
α_3	$L^{(3)}$	0	0	0	1	0.300	0.3000	0.0204	0.1430	0.3031	-0.0031	0.6031
α_4	$L^{(4)}$	0	-1	-1	-1		0.4000	0.0204	0.1430	0.3031	0.0969	0.7031
			推定値 $\boldsymbol{l}\hat{\boldsymbol{\beta}}$			=Mmult ($\boldsymbol{l}$範囲, $\boldsymbol{\beta}$^の範囲)		$t_{0.05}(20-4)$=	2.1199			
			分散 $Var(\boldsymbol{l}\hat{\boldsymbol{\beta}})$			=Mmult (Mmult ($\boldsymbol{l}$の範囲, $\boldsymbol{\Sigma}(\boldsymbol{\beta}$^)の範囲), Transpose ($\boldsymbol{l}$の範囲))						

表 2.6 でデータの構造式に対応する $\hat{\alpha}_i$ の推定値と 95%信頼区間が得られたので，Excel の折れ線グラフを用いて結果の表示を行なう．図 2.2 に示すように，1 因子実験での効果 $\hat{\alpha}_i$ と 95%信頼区間が示されている．水準効果 $\hat{\alpha}_i$ は，定義により全て加えると 0.0 となるので，効果 $\hat{\alpha}_i$ は，0.0 を挟む形になる．

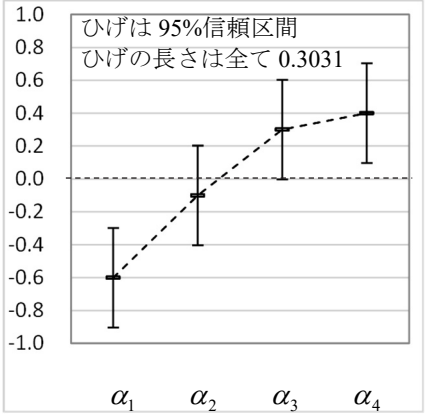

図 2.2　折れ線グラフによる効果 $\hat{\alpha}_i$ に対する予測プロファイル

水準平均の 95%信頼区間

表 2.5 右に示したパラメータの推定値 $\hat{\boldsymbol{\beta}}$ を用い，因子 A の A_i 水準平均 $\bar{y}_{i\cdot}$，および，その 95%信頼区間を求める．表 2.7 に示すように，行ベクトル $\boldsymbol{l}^{(5)}$ を用いて $L^{(5)}=\boldsymbol{l}^{(5)}\hat{\boldsymbol{\beta}}=\hat{\beta}_0+\hat{\beta}_1$ =10.30 であり，A_1 水準の算術平均と一致する．その分散は，$Var(\boldsymbol{l}^{(5)}\hat{\boldsymbol{\beta}})=\boldsymbol{l}^{(5)}\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})(\boldsymbol{l}^{(5)})^T$ であり，$\hat{\beta}_0$ と $\hat{\beta}_1$ の共分散は表 2.5 から 0.0 なので

$$\left.\begin{aligned} Var(\hat{\beta}_0+\hat{\beta}_1) &= Var(\hat{\beta}_0)+2Cov(\hat{\beta}_0,\hat{\beta}_1)+Var(\hat{\beta}_1) \\ &= 0.0068+0.0+0.0204=0.0273 \end{aligned}\right\} \tag{2.15}$$

と計算される．他の A_i 水準についても同様に求められている．

表 2.7 水準平均の 95%信頼区間

			l_0	l_1	l_2	l_3		推定値	分散	標準誤差	幅	95%信頼区間	
	$\bar{y}_{i\cdot}$	L	x_0	a_1	a_2	a_3	$\hat{\boldsymbol{\beta}}$	$\boldsymbol{l}\hat{\boldsymbol{\beta}}$	$Var(\boldsymbol{l}\hat{\boldsymbol{\beta}})$	SE	$t_{0.05}\times SE$	L 95%	U 95%
A_1	$\bar{y}_{1\cdot}$	$L^{(5)}$	1	1	0	0	10.900	10.3000	0.0273	0.1651	0.3499	9.9501	10.6499
A_2	$\bar{y}_{2\cdot}$	$L^{(6)}$	1	0	1	0	-0.600	10.8000	0.0273	0.1651	0.3499	10.4501	11.1499
A_3	$\bar{y}_{3\cdot}$	$L^{(7)}$	1	0	0	1	-0.100	11.2000	0.0273	0.1651	0.3499	10.8501	11.5499
A_4	$\bar{y}_{4\cdot}$	$L^{(8)}$	1	-1	-1	-1	0.300	11.3000	0.0273	0.1651	0.3499	10.9501	11.6499

この結果を表 2.6 に示したデータの構造式の $\hat{\alpha}_i$ の $SE(\hat{\alpha}_i)=0.1430$ と比べると，$SE(\bar{y}_{i\cdot})$ =0.1651 と大きくなっている．これは，水準平均に $\bar{y}_{i\cdot}=\hat{\mu}+\hat{\alpha}_i$ のように一般平均の推定値 $\hat{\mu}$ が含まれているためである．

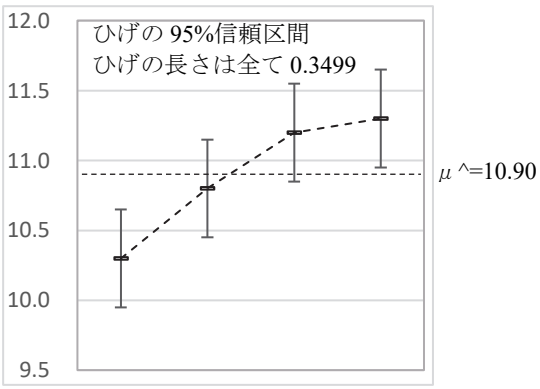

図 2.3 因子 A の水準平均と 95%信頼区間

水準間の差と 95%信頼区間

A_1 水準を基準として他の水準の差および 95%信頼区間を求めてみよう．表 2.8 に示すように「(A_2 - A_1) $L^{(9)}$」は，A_2 水準の推定のための行ベクトル $\boldsymbol{l}^{(6)}$ から A_1 水準の推定のために行ベクトル $\boldsymbol{l}^{(5)}$ を引いたものである．表には示していないが，［(A_3＋A_4)/ 2 - A_1］などのように，有意ではない A_3 水準と A_4 水準の平均値と A_1 水準との差なども同様に設定することができる．

表 2.8　A_2 と A_1 の差の推定のためのベクトル

				l_0	l_1	l_2	l_3
			L	x_0	a_1	a_2	a_3
		A_2	$L^{(5)}$	1	0	1	0
−)		A_1	$L^{(6)}$	1	1	0	0
A_2	-	A_1	$L^{(9)}$	0	-1	1	0

表 2.9 に各種の水準間の差について計算した結果を示す．（推定値，分散，SE，t 値，$t_{0.05}\times SE$，$L95\%$，$U95\%$）の計算式は，表 2.6 に示した効果 $\hat{\alpha}_i$ に対す場合と全く同じで，線形和 $L^{(i)}=\boldsymbol{l}^{(i)}\hat{\boldsymbol{\beta}}$ の行ベクトル $\boldsymbol{l}^{(i)}$ を推定目的に合わせて設定している．また，これらの計算方法は，繰り返しが不揃いの場合でも，1 因子実験の場合に限らず，あらゆる実験モデルの解析で共通な汎用的な方法である．

表 2.9　因子 A の水準間の比較と 95%信頼区間

				l_0	l_1	l_2	l_3		推定値	分散	標準誤差	幅	95%信頼区間	
			L	x_0	a_1	a_2	a_3	$\hat{\boldsymbol{\beta}}$	$\boldsymbol{l}\hat{\boldsymbol{\beta}}$	$Var(\boldsymbol{l}\hat{\boldsymbol{\beta}})$	SE	$t_{0.05}\times SE$	$L95\%$	$U95\%$
A_2	- A_1	$L^{(9)}$	0	-1	1	0	10.900		0.5000	0.0545	0.2335	0.4949	0.0051	0.9949
A_3	- A_1	$L^{(10)}$	0	-1	0	1	-0.600		0.9000	0.0545	0.2335	0.4949	0.4051	1.3949
A_4	- A_1	$L^{(11)}$	0	-2	-1	-1	-0.100		1.0000	0.0545	0.2335	0.4949	0.5051	1.4949
A_3	- A_2	$L^{(12)}$	0	0	-1	1	0.300		0.4000	0.0545	0.2335	0.4949	-0.0949	0.8949
A_4	- A_2	$L^{(13)}$	0	-1	-2	-1			0.5000	0.0545	0.2335	0.4949	0.0051	0.9949
A_4	- A_3	$L^{(14)}$	0	-1	-1	-2			0.1000	0.0545	0.2335	0.4949	-0.3949	0.5949

表 2.9 で推定された因子 A の水準間の差と 95%信頼区間について，図 2.4 に Excel の折れ線グラフを用いて描いた結果を示す．

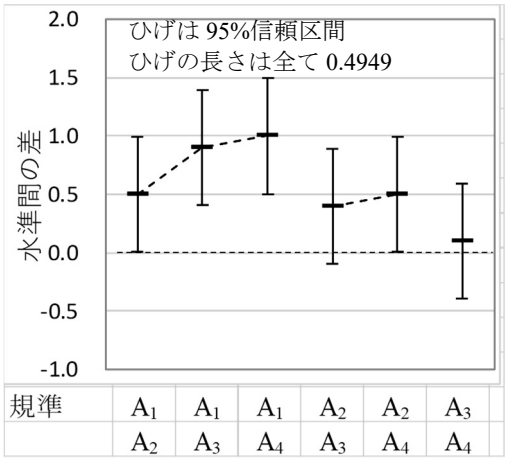

図 2.4　因子 A の水準間の差と 95%信頼区間

これまでも各種の推定値と 95%信頼区間の折れ線グラフを示してきたのであるが，作図する際の幾つかのヒントを示す．1）推定値のマーカは，組み込みオプションで - を選択しサイズを拡大する．2）水準間のつなぎ線は，適当な太さで適切な線種を選択する．3）上下の

ひげの幅を$t_{0.05}(df)\times SE$で前もって計算し，ユーザ設定の誤差範囲とする．4）水準間のつなぎ線の消去は，右側の点のみを選択し「線なし」にする．5）込み入った軸ラベルは，Excelシート上に設定し，全て込みでコピーし，「図（拡張メタファイル)」として貼り付ける．

水準平均に対する伝統的な分散の推定

伝統的な実験データの解析では，各種の偏差平方和を計算し，それらを分散分析表としてまとめ，結果を吟味することを主体にしており，各種の推定値と95%信頼区間の計算については，統計の基礎である分散の加法性による対応が必要となる．

Excelの回帰分析などで作成された分散分析表には，「残差の分散」，「誤差の平均平方」などの表記で推定された誤差分散の推定値$\hat{\sigma}^2$が必ず示されている．表2.5に示した分散分析表から$\hat{\sigma}^2=0.1363$が得られている．この$\hat{\sigma}^2$を使い，A_1水準の平均10.30は，5個のデータy_{ij}の平均であり，それぞれの誤差ε_{ij}は，互いに独立なので分散の加法性を用い，A_1水準平均の分散$Var(\bar{A}_1)$は，

$$\left.\begin{aligned} Var(\bar{A}_1) &= Var\left(\frac{y_{1,1}+y_{1,2}+\cdots+y_{1,5}}{5}\right) \\ &= \frac{Var(y_{1,1})}{5^2}+\frac{Var(y_{1,2})}{5^2}+\cdots+\frac{Var(y_{1,5})}{5^2} \\ &= \frac{5\hat{\sigma}^2}{5^2}=\frac{0.1363}{5}=0.0273 \end{aligned}\right\} \tag{2.16}$$

との計算結果を得る．他のA_i水準も同様に計算することができ，表2.7に示した分散に一致する．

水準平均の差とその分散の推定

A_2水準とA_1水準の差の分散は，A_2の平均とA_1の平均が互いに独立なので，

$$\left.\begin{aligned} Var(\bar{A}_2-\bar{A}_1) &= Var(\bar{A}_2)+(-1)^2Var(\bar{A}_1) \\ &= 2\frac{\hat{\sigma}^2}{5}=2\times 0.0273=0.0545 \end{aligned}\right\} \tag{2.17}$$

との計算結果を得る．このように，どのような推定なのかを見極めて分散の計算を別々行う必要がある．繰り返しが全て同じなので，因子Aの全ての水準の組合せの差についても同じ分散となり，表2.9に示した結果に一致する．

このように，「水準平均に対する分散の計算」と「水準間の差の分散の計算」が分散分析表の誤差分散の推定値$\hat{\sigma}^2$を用いて簡単に求めることができる事は素晴らしいのであるが，第1章で示した量的変数を含む実験データの解析に応用することができない．

水準効果の分散の推定

厄介なのは，因子 A の効果としての（$\hat{\alpha}_1$，$\hat{\alpha}_2$，$\hat{\alpha}_3$，$\hat{\alpha}_4$）の推定値の分散の導出である．効果の推定値は，各水準の平均値から総平均（$\overline{y}_{i\cdot}$ の平均）を差し引くことで，

$$\hat{\alpha}_i = \overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot} \tag{2.18}$$

簡単に求められる．因子 A の各水準の分散は，5 個のデータの平均なので，分散分析表の誤差分散の推定値 $\hat{\sigma}^2 = 0.1363$ を使って，$Var(\overline{\mathrm{A}}_i) = \hat{\sigma}^2 / 5 = 0.0273$ として求められた．ただし，因子 A の効果 $\hat{\alpha}_i$ は，

$$Var(\hat{\alpha}_i) = Var(\overline{\mathrm{A}}_i - \hat{\mu}) = Var(\overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot}) \tag{2.19}$$

で定義される．ここでは，$\overline{y}_{\cdot\cdot}$ の中に $\overline{y}_{i\cdot}$ も含まれているので，互いに独立ではなく，分散の加法性が成り立たない．

デザイン行列 $\boldsymbol{X}$ を用いた場合に $Var(\hat{\alpha}_i) = Var(\hat{\beta}_i)$ は，表 2.5 に示したようパラメータの推定値の標準誤差 SE が直接推定されているので，SE の 2 乗が分散として得られる．しかし，伝統的な分散分析による誤差分散 $\hat{\sigma}^2$ を用いた導出に際しては，元のデータ y_{ij} に戻って分散を計算する必要がある．実際に計算すると

$$\left.\begin{aligned}
Var(\hat{\alpha}_1) &= Var(\overline{y}_{1\cdot} - \overline{y}_{\cdot\cdot}) \\
&= Var\left(\frac{y_{1,1} + y_{1,1} + \cdots + y_{1,5}}{5} - \frac{y_{1,1} + y_{1,1} + \cdots + y_{2,1} + y_{2,1} + \cdots + y_{4,5}}{20} \right) \\
&= Var\left(\frac{3y_{1,1} + 3y_{1,1} + \cdots + 3y_{1,5}}{20} - \frac{y_{2,1} + y_{2,1} + \cdots + y_{4,5}}{20} \right) \\
&= \frac{(3^2 \times 5)\hat{\sigma}^2}{20^2} + \frac{15\hat{\sigma}^2}{20^2} \\
&= \frac{60\hat{\sigma}^2}{400} = \frac{0.1363}{6.6667} = 0.0204
\end{aligned}\right\} \tag{2.20}$$

のように，$\hat{\sigma}^2 = 0.1363$ を有効反復数 $n_e = 6.6667$ で割って分散 $Var(\hat{\alpha}_1) = 0.0204$ が得られる．この分散の平方根が，表 2.5 の回帰パラメータの $SE(\hat{\beta}_1) = 0.1430$ に一致する．

このような場合に伝統的な解析法では，水準効果の分散の推定に「伊奈の式」を使った有効反復数 n_e を算出して分散を計算する方法が知られている．他にも，「田口の式」による有効反復数 n_e の計算法などがあり，秘伝的な香りに満ちあふれている．しかも，繰り返し数が等しい場合のみ適用でき，繰り返し数が不揃いの場合には適用できないこので，発展性に欠ける．手計算時代には，素晴らしい解決方法であったことは間違いないが，新たな方法の発展・普及を阻害しているがごとくである．なお，「伊奈の式」については，芳賀（2014）の「第 1.6 節 補遺 （1）伊奈の法則」に詳細に示されている．

見方を変えれば，$\bar{A}_1$ の推定値は $\bar{A}_1 = \hat{\mu} + \hat{\beta}_1$ であり，表 2.5 右下段から共分散が $Cov(\hat{\mu}, \hat{\beta}_1) = 0$ なので，分散の加法性が成り立ち

$$Var(\bar{A}_1) = Var(\hat{\mu}) + Var(\hat{\beta}_1) \tag{2.21}$$

のように成り立つ．したがって，効果の分散は，分散の加法性ではなく，分散の引き算

$$\left.\begin{aligned} Var(\hat{\beta}_1) &= Var(\bar{A}_1) - Var(\hat{\mu}) \\ &= Var(\bar{y}_{i\cdot}) - Var(\bar{y}_{\cdot\cdot}) \\ &= \frac{\hat{\sigma}^2}{5} - \frac{\hat{\sigma}^2}{20} \\ &= 0.0273 - 0.0068 = 0.0204 \end{aligned}\right\} \tag{2.22}$$

で求められる．このような各種の推定値の分散を誤差分散 $\hat{\sigma}^2$ から推定することは，統計の基礎的な素養を高めるために必要ではあるが，繰り返しが等しい場合に限定され，汎用性に欠ける．パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ を使う方法は，線形和 $L^{(i)}$ を計算するためにベクトル $\boldsymbol{l}^{(i)}$ を設定する必要があるが，同じ式で計算できる汎用的な方法である．

伝統的な「平方和の分解」と「分散の加法性」による 1 因子実験による解析では，各種の推定値の分散を求める際に，手作業的な計算を必要としている．このこと自体は，統計の理論の実践には欠かせないことも理解できるが，手計算の方法を強いる方法だけでいいのだろうか．現代の算盤である Excel を活用した線形モデルによる 1 元配置の解析方法も合わせて示すことが，実践のための応用力を付けるため不可欠のように思えてならない．線形モデルに対してデザイン行列 $\boldsymbol{X}$ を用いた Excel による一連の解析方法は，多くの統計ソフトと同様の解析方法であり，統計ソフトを使用する前の入門として適している．逆説的には，ブラックボックス的な統計ソフトによる解析方法を理解し，更なる応用力を付けるために，Excel による一連の解析方法の習得は，学習の助けになる．

分散の加法性が成り立たない効果の差

分散の加法性が成り立つことを前提に実験計画法は，平方和の分解を使った分散分析表の作成が，基本的な解析方法として定着し，ほとんど全ての「実験計画法」に関する書物で踏襲されて続けている．嘆かわしいとしか言いようのないことであり，量的な変数が含まれる場合，欠測値が生じた場合，繰り返しが不揃いの場合などへの適用などの応用に際し，融通の利かない方法として自らの発展を阻害している．

デザイン行列 $\boldsymbol{X}$ を用いた回帰分析の結果を表 2.5 に示したのであるが，パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ から（$\hat{\beta}_1$，$\hat{\beta}_2$，$\hat{\beta}_3$）の間には，共通の共分散

$$Cov(\beta_i, \beta_{i'}) = -0.0068$$

が存在していることが示されている．そのために，A_2 水準の平均と A_1 水準の平均の差の推定値は

$$\begin{aligned}\bar{A}_2 - \bar{A}_1 &= (\hat{\mu} + \hat{\beta}_2) - (\hat{\mu} + \hat{\beta}_1) \\ &= \hat{\beta}_2 - \hat{\beta}_1 \\ &= -0.60 - (-0.10) = 0.50\end{aligned}$$

として求まるが，効果の差の分散を求める際に「分散の加法性」が成り立たないために

$$\begin{aligned}Var(\hat{\beta}_2 - \hat{\beta}_1) &= Var(\hat{\beta}_1) - 2Cov(\hat{\beta}_2, \hat{\beta}_1) + Var(\hat{\beta}_2) \\ &= 0.0204 - 2 \times (-0.0068) + 0.0204 = 0.0545\end{aligned}$$

のようにパラメータの共分散を考慮しなければならない．

見方を変えて，$\bar{A}_2$ の分散と $\bar{A}_1$ の分散を使った場合には，互いに独立なので，分散の加法性が成り立ち，

$$\begin{aligned}Var(\bar{A}_2 - \bar{A}_1) &= Var(\bar{A}_2) + Var(\bar{A}_1) \\ &= 0.0273 + 0.0273 = 0.0545\end{aligned}$$

のように，共分散を含めずに分散を求めることができる．

セル平均モデル

1 因子実験のデータの構造式を式 (2.1) では，因子 A の効果を α_i とし，（1） $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ としたのであるが，$\mu + \alpha_i$ を $\mu_i = \mu + \alpha_i$ とし，

（2）　$y_{ij} = \mu_i + \varepsilon_{ij}$　　(2.23)

のようなデータの構造式として表すことも可能である．式 (2.23) は，セル平均モデル（cell means model）と言われ，このモデルに対応するためには，表 2.10 に示すように（1，1）標示型ダミー変数を用いてデザイン行列 $\boldsymbol{X}$ を生成する必要がある．Excel の回帰分析では，「定数に 0 を設定」オプションをオンとすることにより，因子 A の各水準の平均値，*SE* などを直接推定することができる．

表 2.10　（1，1）標示型ダミー変数

		(1, 1) 型		
	a_1	a_2	a_3	a_4
A_1	1	0	0	0
A_2	0	1	0	0
A_3	0	0	1	0
A_4	0	0	0	1

2.2. 繰り返しが不揃いな 1 因子実験データ

繰り返しが等しい 1 因子実験の場合には，平方和の分解による従来の解析方法とデザイン行列 $\boldsymbol{X}$ を用いた線形モデルに解析方法が一致することを前節で示した．繰り返しが不揃いの 1 因子実験の場合は，伝統的な平方和の分解による方法とデザイン行列 $\boldsymbol{X}$ を用いた方法には，微妙な食い違いが発生する．芳賀（2014）の第 1.2 節で示されているデータを表 2.11 に示す．

実験は，A_0 を対照群とし，処置群 A_1～A_4 と比較したい．その際，基準となる対照群の繰り返し数を増やすことにより，A_0 と A_1～A_4 群の差の標準誤差が小さくなることが期待される．表 2.11 に示すように全データ 18 個から求めた総平均は $\bar{y}_{..}=50.50$ であり，水準平均の平均は，$\hat{\mu}=51.40$ と異なる．

表 2.11 繰り返しが不揃いの 1 因子実験データ［芳賀（2014），表示 1.2.1］

			繰り返し					
A	*n*	平均	1	2	3	4	5	6
A_0	6	46.00	43	45	42	47	49	50
A_1	3	49.00	47	51	49			
A_2	3	53.00	54	48	57			
A_3	3	58.00	55	58	61			
A_4	3	51.00	52	48	53			
総平均	18	50.50	$=\bar{y}_{..}$					
水準平均の平均		51.40	$=\mu$^					

図 2.5 に示すのは，前節と同様に表 2.11 のデータを行方向に並べ直し，の散布図に平均値を上書きした結果である．さらに，全データに対する総平均 $\bar{y}_{..}$，水準平均の平均 $\hat{\mu}$ につても重ね書きしてある．

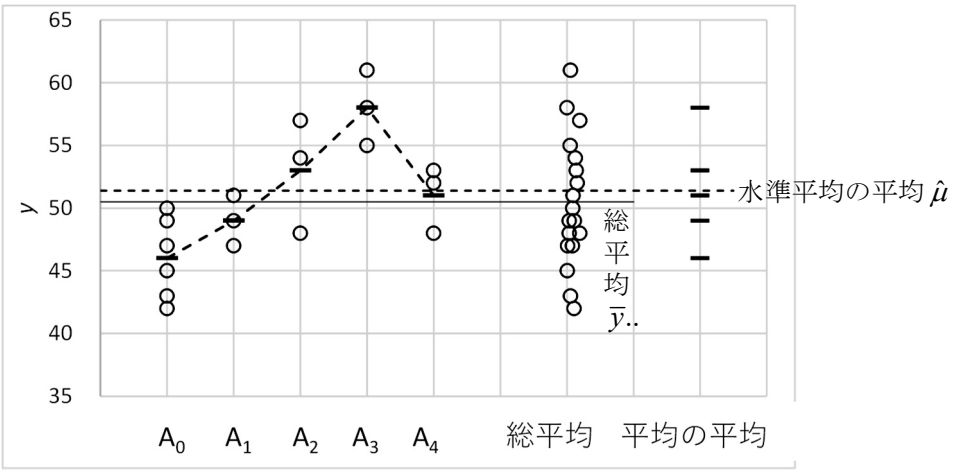

図 2.5 平均の折れ線入りの散布図

平方和の分解

表 2.12 に示すように，平方和の計算を前節の表 2.3 に準じて行う．平方和（S_{T}，S_{A}，S_e）の計算結果が示され，

$$\left.\begin{aligned} S_{\mathrm{T}} &= S_{\mathrm{A}} + S_e \\ &= 316.50 + 134.00 = 450.50 \end{aligned}\right\} \tag{2.24}$$

のように $S_{\mathrm{T}} = S_{\mathrm{A}} + S_e$ となることから，平方和の分解が成り立っていることが確認される．総平方和 S_{T} は，データの総平均 $\bar{y}.. = 50.50$ からの偏差平方和で計算している．計算量を減らすために，S_{A} を計算しなくても $S_{\mathrm{A}} = S_{\mathrm{T}} - S_e$ として求めることもできるが，平方和の定義に基づいた計算が，各平方和の分解の意味を理解するために適している．

表 2.12　各種の平方和の計算

				1: μ^′			2: α^$_i$′	3: ε^$_{ij}$′	1+2+3	分散分析表		
A	i	j	y_{ij}	$\bar{y}..$	$y_{ij} - \bar{y}..$	$\bar{y}_i.$	$\bar{y}_i. - \bar{y}..$	$y_{ij} - \bar{y}_i.$	y_{ij}	要因	自由度	平方和
A_0	0	1	43	50.50	-7.50	46.00	-4.50	-3.00	43.00	A	4	316.50
		2	45	50.50	-5.50	46.00	-4.50	-1.00	45.00	e	13	134.00
		3	42	50.50	-8.50	46.00	-4.50	-4.00	42.00	T	17	450.50
		4	47	50.50	-3.50	46.00	-4.50	1.00	47.00			
		5	49	50.50	-1.50	46.00	-4.50	3.00	49.00			
		6	50	50.50	-0.50	46.00	-4.50	4.00	50.00			
A_1	1	1	47	50.50	-3.50	49.00	-1.50	-2.00	47.00			
		2	51	50.50	0.50	49.00	-1.50	2.00	51.00			
		3	49	50.50	-1.50	49.00	-1.50	0.00	49.00			
A_2	2	1	54	50.50	3.50	53.00	2.50	1.00	54.00			
		2	48	50.50	-2.50	53.00	2.50	-5.00	48.00			
		3	57	50.50	6.50	53.00	2.50	4.00	57.00			
A_3	3	1	55	50.50	4.50	58.00	7.50	-3.00	55.00			
		2	58	50.50	7.50	58.00	7.50	0.00	58.00			
		3	61	50.50	10.50	58.00	7.50	3.00	61.00			
A_4	4	1	52	50.50	1.50	51.00	0.50	1.00	52.00			
		2	48	50.50	-2.50	51.00	0.50	-3.00	48.00			
		3	53	50.50	2.50	51.00	0.50	2.00	53.00			
平均			50.50	平方和	**450.50**		**316.50**	**134.00**				
					S_{T}		S_{A}	S_{e}				

因子 A の平方和 S_{A} は，各群の平均 $\bar{y}_i.$ と総平均 $\bar{y}..$ の差の平方和を算している．その差を効果 α_i'

$$\alpha_i' = \bar{y}_i. - \bar{y}.. \tag{2.25}$$

として表わす．ただし，効果 α_i' の合計は，表 2.13 に示すように

$$\sum_{i=0}^{4} \alpha_i' = 4.50, \qquad \sum_{i=0}^{4} n_i \alpha_i' = 0 \tag{2.26}$$

となり，0 とはならないが，それぞれの群のデータ数 n_i での重み付き合計とした場合には，0 となる．

表 2.13 総平均 $\bar{y}_{..}$ からの差の効果 $\hat{\alpha}_i'$ および水準平均の平均 $\hat{\mu}$ からの効果 $\hat{\alpha}_i$

	繰返し数	水準平均	総平均	$\bar{y}_{i.}-\bar{y}_{..}$	重み付き	$\bar{y}_{i.}$ の平均	$\bar{y}_{i.}-\hat{\mu}$
A	n_i	$\bar{A}_i$: $\bar{y}_{i.}$	$\bar{y}_{..}$	$\hat{\alpha}_i'$	$n_i\hat{\alpha}_i'$	$\hat{\mu}$	$\hat{\alpha}_i$
A_0	6	46.00	50.50	-4.50	-27.00	51.40	-5.40
A_1	3	49.00	50.50	-1.50	-4.50	51.40	-2.40
A_2	3	53.00	50.50	2.50	7.50	51.40	1.60
A_3	3	58.00	50.50	7.50	22.50	51.40	6.60
A_4	3	51.00	50.50	0.50	1.50	51.40	-0.40
	$\bar{y}_{i.}$ の平均	51.40	計	4.50	0.00	計	0.00

因子の効果 α_i を因子 A の水準平均の平均 $\hat{\mu}=51.40$ からの差とすれば,

$$\hat{\alpha}_i=\bar{y}_{i.}-(\text{水準平均の平均})=\bar{y}_{i.}-\hat{\mu} \tag{2.27}$$

$\hat{\alpha}_i$ の合計は，表 2.13 に示すように 0.0 となる.

デザイン行列 $\boldsymbol{X}$ を用いた回帰分析

表 2.14 に示すのは，前節と同様に因子 A に対し（1，–1）対比型ダミー変数で設定された（18 行×5 列）のデザイン行列 $\boldsymbol{X}$ を用いて Excel の回帰分析を適用した結果である．分散分析表は，平方和の分解で求めた結果に一致する．ただし，切片の推定値 $\hat{\beta}_{00}=51.40$ は，表 2.13 に示した水準平均の平均 $\hat{\mu}$ となり，総平均 $\bar{y}_{..}$ とは異なり，因子 A の各水準の推定値 $\hat{\beta}_i$ は，各水準の平均値 $\bar{y}_{i.}$ と $\hat{\mu}$ との差となる効果 $\hat{\alpha}_i$ が推定されている．デザイン行列 $\boldsymbol{X}$ を用いた場合の分散分析表は，どのようにして求められているのであろうか.

表 2.14 （1，–1）対比型のダミー変数を用いた Excel による回帰分析

	番号		$\boldsymbol{Y}$	—デザイン行列 $\boldsymbol{X}$—										
群	i	j	y	x_0	a_0	a_1	a_2	a_3		分散分析表,「定数に 0 を使用」off				
A_0	0	1	43	1	1	0	0	0			自由度	変動	分散	分散比
	0	2	45	1	1	0	0	0		回帰	4	316.50	79.1250	7.6763
	0	3	42	1	1	0	0	0		残差	13	134.00	**10.3077**	$=\sigma^{\wedge 2}$
	0	4	47	1	1	0	0	0		合計	17	450.50		
	0	5	49	1	1	0	0	0						
	0	6	50	1	1	0	0	0			係数	標準誤差	t	P-値
A_1	1	1	47	1	0	1	0	0	$\beta^{\wedge}_{00}$	切片 x_0	51.400	0.7864	65.3592	0.0000
	1	2	51	1	0	1	0	0	$\beta^{\wedge}_{0}$	a_0	-5.400	1.2842	-4.2049	0.0010
	1	3	49	1	0	1	0	0	$\beta^{\wedge}_{1}$	a_1	-2.400	1.6371	-1.4660	0.1664
A_2	2	1	54	1	0	0	1	0	$\beta^{\wedge}_{2}$	a_2	1.600	1.6371	0.9774	0.3462
	2	2	48	1	0	0	1	0	$\beta^{\wedge}_{3}$	a_3	6.600	1.6371	4.0316	0.0014
	2	3	57	1	0	0	1	0						
A_3	3	1	55	1	0	0	0	1		パラメータの共分散行列 $\boldsymbol{\Sigma}(\boldsymbol{\beta}^{\wedge})=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma^{\wedge 2}$				
	3	2	58	1	0	0	0	1	$\beta^{\wedge}_{00}$	**0.6185**	-0.2749	0.0687	0.0687	0.0687
	3	3	61	1	0	0	0	1	$\beta^{\wedge}_{0}$	-0.2749	**1.6492**	-0.4123	-0.4123	-0.4123
A_4	4	1	52	1	-1	-1	-1	-1	$\beta^{\wedge}_{1}$	0.0687	-0.4123	**2.6800**	-0.7559	-0.7559
	4	2	48	1	-1	-1	-1	-1	$\beta^{\wedge}_{2}$	0.0687	-0.4123	-0.7559	**2.6800**	-0.7559
	4	3	53	1	-1	-1	-1	-1	$\beta^{\wedge}_{3}$	0.0687	-0.4123	-0.7559	-0.7559	**2.6800**
										x_0	a_0	a_1	a_2	a_3
										=Minverse (Mmult (Transpose ($\boldsymbol{X}$の範囲), $\boldsymbol{X}$の範囲))*$\sigma^{\wedge 2}$				

デザイン行列 $\boldsymbol{X}$ を用いた場合の分散分析表と平方和の分解による分散分析表が完全に一致するのは，なぜなのだろうか．分散分析表は，デザイン行列 $\boldsymbol{X}$ を用いて得られたパラメータから推定される各水準の推定値をベースに組み立てられている．対照群 A_0 の場合では，

$$\left.\begin{aligned} &\text{平方和の分解：}\quad \bar{A}_0' = \bar{y}_{..} + \alpha_0' = 50.50 - 4.50 = 46.00 \\ &\text{デザイン行列：}\quad \bar{A}_0 = \hat{\mu} + \hat{\alpha}_0 \;\; = 51.40 - 5.40 = 46.00 \end{aligned}\right\} \tag{2.28}$$

と水準平均が一致するので，因子 A の平方和 S_A も一致する．また，回帰分析の場合でも総平方和 S_T は，平方和の分解と同様に y_{ij} と $\bar{y}_{..}$ の偏差平方和で計算されており，回帰パラメータとして推定された $\hat{\mu} = 51.4$ からの平方和でない．

Excel の回帰分析の出力から，因子 A の全体（回帰）に対し $F = 7.6763$ と有意な差であること，ダミー変数 a_0 のパラメータの推定値が $\hat{\beta}_0 = -5.40$，$p = 0.0010$ と有意な差，ダミー変数 a_3 のパラメータの推定値が $\hat{\beta}_3 = 6.60$，$p = 0.0014$ と有意な差であり，他は有意な差でないことが読み取れる．各種の推定値の標準誤差 SE の算出には，表 2.14 右下段のパラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ が必要で，ちなみに $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ の（1 行・1 列）目は，切片 x_0 の分散で，標準誤差の 2 乗 $Var(\hat{\beta}_1) = 0.7864^2 = 0.6185$ に一致する．

パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ の計算は，(5行×18列) の転置行列 $\boldsymbol{X}^T$ と (18行×5列) の $\boldsymbol{X}$ の積和行列 $(\boldsymbol{X}^T\boldsymbol{X})$ は (5行×5列) の正方行列となり，$(\boldsymbol{X}^T\boldsymbol{X})$ の逆行列 $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ も同じ (5行×5列) の正方行列となる．Excel シート上では，(5行×5列) のセルを最初に選択し，Excel の行列関数による計算式

$$\left.\begin{aligned} \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) &= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\hat{\sigma}^2 \\ &= \text{Minverse}(\text{Mmult}(\text{Transpose}(\boldsymbol{X}\text{の範囲}), \boldsymbol{X}\text{の範囲}))*\hat{\sigma}^2 \end{aligned}\right\} \tag{2.29}$$

により求めることができる．もちろん，繰り返しが等しい場合でも，計算式は全く同じである．

データの構造式における効果

表 2.13 に示したように繰り返し不揃いの 1 因子実験のデータの構造式は，繰り返しが等しい場合の式 (2.1) とまったく同じ，

$$\left.\begin{aligned} &y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \\ &\text{ただし，} \sum\nolimits_i \alpha_i = 0,\quad \sum\nolimits_i \sum\nolimits_j \varepsilon_{ij} = 0 \end{aligned}\right\} \tag{2.30}$$

であり，表 2.14 で示したパラメータ推定値に示すように，切片 x_0 のパラメータの推定値として，$\hat{\mu} = 51.40$ が推定されている．効果 $\hat{\alpha}_0$ は，$\hat{\alpha}_0 = \hat{\beta}_0 = -5.40$ であり，同様に $\hat{\alpha}_1 = -2.40$，$\hat{\alpha}_2 = 1.60$，$\hat{\alpha}_3 = 6.60$ が求められている．効果 $\hat{\alpha}_4$ に対応する変数がないので，$\hat{\alpha}_4 = -\hat{\beta}_0 - \hat{\beta}_1 - \hat{\beta}_2 - \hat{\beta}_3$ で計算する必要がある．表 2.14 には，$\hat{\alpha}_3$ までの効果についての SE が計算されているが，効果 $\hat{\alpha}_4$ の SE は，どのように求めたらよいのであろうか．

表 2.15 に示すように効果 $\hat{\alpha}_4$ は，線形和

$$\hat{\alpha}_4 \ : \ L^{(6)} = \boldsymbol{l}^{(6)}\hat{\boldsymbol{\beta}} = [0 \ \ -1 \ \ -1 \ \ -1 \ \ -1] \begin{bmatrix} 51.40 \\ -5.40 \\ -2.40 \\ 1.60 \\ 6.60 \end{bmatrix} = -0.400 \qquad (2.31)$$

として計算している． *SE* は，表 2.14 右下段で計算されているパラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ を用いて Excel シートのセルに埋め込まれた計算式によって

$$SE(\hat{\alpha}_4) = \sqrt{Var(\hat{\alpha}_4)} = \sqrt{\boldsymbol{l}^{(6)}\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})(\boldsymbol{l}^{(6)})^T} = 1.6371 \qquad (2.32)$$

として計算できる．結果として，効果 $\hat{\alpha}_4$ の *SE* は，繰り返しが 3 の他の $\hat{\alpha}_i$ の *SE* と等しいことが分かる．表 2.13 で省略した 95%信頼区間についても表 2.15 で改めて計算した結果を示す．

表 2.15 効果 α_i の推定と 95%信頼区間

				l_{00}	l_0	l_1	l_2	l_3		推定値	分散	標準誤差	幅	95%信頼区間	
			L	x_0	a_0	a_1	a_2	a_3	$\hat{\boldsymbol{\beta}}$	$\boldsymbol{l}\hat{\boldsymbol{\beta}}$	$Var(\boldsymbol{l}\hat{\boldsymbol{\beta}})$	SE	$t_{0.05}\times SE$	$L95\%$	$U95\%$
	μ	$L^{(1)}$	1	0	0	0	0	51.400	51.40	0.6185	0.7864	1.6990	49.70	53.10
効果	α_0	$L^{(2)}$	0	1	0	0	0	-5.400	-5.40	1.6492	1.2842	2.7744	-8.17	-2.63
	α_1	$L^{(3)}$	0	0	1	0	0	-2.400	-2.40	2.6800	1.6371	3.5367	-5.94	1.14
	α_2	$L^{(4)}$	0	0	0	1	0	1.600	1.60	2.6800	1.6371	3.5367	-1.94	5.14
	α_3	$L^{(5)}$	0	0	0	0	1	6.600	6.60	2.6800	1.6371	3.5367	3.06	10.14
	α_4	$L^{(6)}$	0	-1	-1	-1	-1		**-0.40**	2.6800	**1.6371**	3.5367	-3.94	3.14
											$t_{0.05}(18\text{-}5)=$	2.1604		

図 2.6 に，表 2.15 で計算された，効果 $\hat{\alpha}_i$ の推定値と 95%信頼区間を Excel の折れ線グラフで作図した結果を示す．効果 $\hat{\alpha}_0$ の 95%信頼区間が他よりも短くなっているのは，繰り返し数が 6 で，他の群よりも大きいためである．

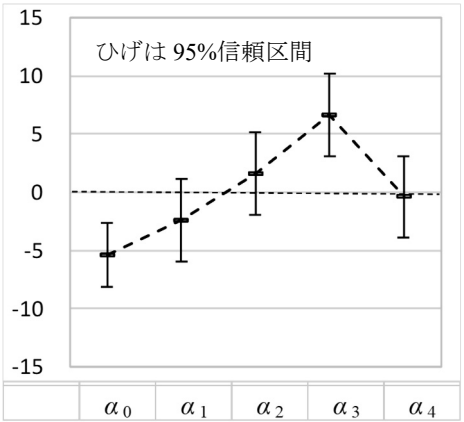

図 2.6 折れ線グラフによる効果に対する予測プロファイル

水準平均に対する 95%信頼区間

水準平均は，単純な算術平均であり，表 2.11 に示したごとくであり，その分散も表 2.14 左上段に示した分散分析表の誤差分散 $\hat{\sigma}^2=10.3077$ を用い，式 (2.16) と同様に個々の y_{ij} に戻ると分散の加法性が成り立つので，反復数が 6 の A_0 の $\bar{y}_{0.}$ の場合であれば，

$$Var(\bar{y}_{0.})=\frac{\hat{\sigma}^2}{6}=\frac{10.3077}{6}=1.7179$$

であり，反復数が 3 の A_1 場合は

$$Var(\bar{y}_{1.})=\frac{\hat{\sigma}^2}{3}=\frac{10.3077}{3}=3.4359$$

と計算することができる．表 2.16 に示すように行列関数を用いて大げさに計算する必要はないのであるが，自明である事例について，線形和による推定値，行ベクトル $\boldsymbol{l}^{(i)}$ の 2 次形式による分散の計算で結果が一致することが確認できる．

表 2.16　水準平均の推定と 95%信頼区間

				l_{00}	l_0	l_1	l_2	l_3		推定値	分散	標準誤差	幅	95%信頼区間	
			L	x_0	a_0	a_1	a_2	a_3	$\hat{\boldsymbol{\beta}}$	$\boldsymbol{l}\hat{\boldsymbol{\beta}}$	$Var(\boldsymbol{l}\hat{\boldsymbol{\beta}})$	SE	$t_{0.05}\times SE$	L95%	U95%
平均	A_0	$L^{(7)}$	1	1	0	0	0	51.400		46.00	1.7179	1.3107	2.8316	43.17	48.83
	A_1	$L^{(8)}$	1	0	1	0	0	-5.400		49.00	3.4359	1.8536	4.0045	45.00	53.00
	A_2	$L^{(9)}$	1	0	0	1	0	-2.400		53.00	3.4359	1.8536	4.0045	49.00	57.00
	A_3	$L^{(10)}$	1	0	0	0	1	1.600		58.00	3.4359	1.8536	4.0045	54.00	62.00
	A_4	$L^{(11)}$	1	-1	-1	-1	-1	6.600		51.00	3.4359	1.8536	4.0045	47.00	55.00
												$t_{0.05}(18\text{-}5)=$	2.1604		

水準平均の差に対する 95%信頼区間

対照群と設定された A_0 群と他の群との差およびその 95%信頼区間を求めたい．表 2.16 に示した A_0 水準の平均の推定値を求めるために $\boldsymbol{l}^{(7)}=[1\quad 1\quad 0\quad 0\quad 0]$ が設定され，同様に A_1 水準では，$\boldsymbol{l}^{(8)}=[1\quad 0\quad 1\quad 0\quad 0]$ なので，その差は，表 2.17 に示すように $\boldsymbol{l}^{(12)}=\boldsymbol{l}^{(8)}-\boldsymbol{l}^{(7)}$ となる．

表 2.17　水準間の差の推定のためのベクトル $\boldsymbol{x}'_{12}$ の生成

				l_{00}	l_0	l_1	l_2	l_3
			L	x_0	a_0	a_1	a_2	a_3
		A_1	$L^{(8)}$	1	0	1	0	0
−)		A_0	$L^{(7)}$	1	1	0	0	0
A_1	-	A_0	$L^{(12)}$	0	-1	1	0	0

他の群に対しても同様に差のベクトルのように行ベクトル $\boldsymbol{l}^{(i)}$ を求め，表 2.18 に示すように各種の差の推定値と 95%信頼区間が計算されている．

表 2.18 パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ を用いた各種の推定

				l_{00}	l_0	l_1	l_2	l_3		推定値	分散	標準誤差	幅	95%信頼区間	
			L	x_0	a_0	a_1	a_2	a_3	$\hat{\boldsymbol{\beta}}$	$\boldsymbol{l}\hat{\boldsymbol{\beta}}$	$Var(\boldsymbol{l}\hat{\boldsymbol{\beta}})$	SE	$t_{0.05}\times SE$	L95%	U95%
A_1	-	A_0	$L^{(12)}$	0	-1	1	0	0	51.400	3.00	5.1538	2.2702	4.9045	-1.90	7.90
A_2	-	A_0	$L^{(13)}$	0	-1	0	1	0	-5.400	7.00	5.1538	2.2702	4.9045	2.10	11.90
A_3	-	A_0	$L^{(14)}$	0	-1	0	0	1	-2.400	12.00	5.1538	2.2702	4.9045	7.10	16.90
A_4	-	A_0	$L^{(15)}$	0	-2	-1	-1	-1	1.600	5.00	5.1538	2.2702	4.9045	0.10	9.90
A_2	-	A_1	$L^{(16)}$	0	0	-1	1	0	6.600	4.00	6.8718	2.6214	5.6632	-1.66	9.66
:		:													
A_4	-	A_3	$L^{(17)}$	0	-1	-1	-1	-2		-7.00	6.8718	-2.6703	5.6632	-12.66	-1.34
												$t_{0.05}(18\text{-}5)=$	2.1604		

これらの推定値と 95%信頼区間は，全て何らかのグラフで表示することが，解析結果を理解し伝えるために必須である．図 2.6 に示した「効果」の予測プロファイルは，有効性の量的な判断のためには有益であるが，図 2.7 左に示すように伝統的な水準平均の予測プロファイルで表わすことも理解を深めるために有益である．また，図 2.7 右に示すように水準間の差の予測プロファイルは，統計的な判定を直接確認でき結果の解釈に役立つ．

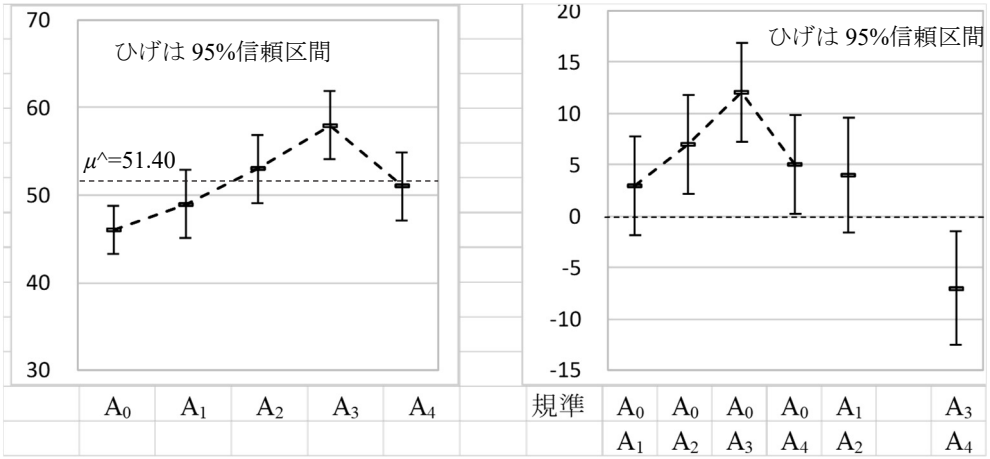

図 2.7 折れ線グラフによる水準平均と水準間の差の予測プロファイル

分散の加法性での対応と限界

パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ を使わない場合の 95%信頼区間の算出には，分散分析表の誤差分散の推定値 $\hat{\sigma}^2=10.3077$ を使って，どのような推定なのかを見極めて計算式を設定する必要があり，画一的な計算式で表わすことができない．A_0 群の平均 46.00 は，6 個のデータ y_{ij} の平均なので，分散の加法性を用い

$$\left.\begin{aligned} Var(\bar{A}_0) &= Var\left(\frac{y_{1,1}+y_{1,2}+\cdots+y_{1,6}}{6}\right) \\ &= \frac{Var(y_{1,1})}{6^2}+\frac{Var(y_{1,2})}{6^2}+\cdots+\frac{Var(y_{1,6})}{6^2} \\ &= \frac{6\hat{\sigma}^2}{6^2}=\frac{10.3077}{6}=1.7179 \end{aligned}\right\} \qquad (2.33)$$

との計算結果を得る．もちろん，表 2.16 に示す $L^{(7)}$ の分散に一致する．表 2.18 に示す $L^{(12)}$ の分散は，A_1 群の平均と A_0 群の平均が互いに独立なので，

$$\left.\begin{aligned} Var(\bar{A}_1-\bar{A}_0) &= Var(\bar{A}_1)+(-1)^2 Var(\bar{A}_0) \\ &= \frac{\hat{\sigma}^2}{3}+\frac{\hat{\sigma}^2}{6} \\ &= \frac{\hat{\sigma}^2}{2}=\frac{10.3077}{2}=5.1538 \end{aligned}\right\} \tag{2.34}$$

との計算結果に一致する．このように，どのような推定なのかを見極めて分散の計算を別々行う必要がある．

因子 A の各水準の分散は，6 個または 3 個のデータの平均なので，分散分析表の誤差分散 $\hat{\sigma}^2=10.3077$ を使って，

$$\left.\begin{aligned} Var(\bar{A}_0) &= \frac{\hat{\sigma}^2}{6}=\frac{10.3077}{6}=1.7179 \\ Var(\bar{A}_1) &= \frac{\hat{\sigma}^2}{3}=\frac{10.3077}{3}=3.4359 \end{aligned}\right\} \tag{2.35}$$

として求められる．ただし，厄介なのは，因子 A の効果としての α_i（β_i に対応）の分散の導出であり，$Var(\hat{\beta}_i)$ は，

$$Var(\hat{\beta}_i)=Var(\bar{A}_i-\hat{\mu})=Var(\bar{y}_i.-\frac{\bar{y}_0.+\bar{y}_1.+\bar{y}_2.+\bar{y}_3.+\bar{y}_4.}{5}) \tag{2.36}$$

となり，$\hat{\mu}$ の計算式の中に $\bar{y}_i.$ が含まれているので，互いに独立ではなく，分散の加法性が成り立たない．繰り返しが等しい場合は，$\hat{\mu}$ からの差ではなく $Var(\bar{A}_i-\bar{y}..)$ のように $\bar{y}..$ からの差であったので，式 (2.20) に示したように元の y_{ij} の分散に戻り，再構成することにより有効反復数 n_e を求めることができたが，繰り返し不揃いの場合は，計算することは容易ではない．（1，-1）対比型ダミー変数を用いた回帰分析を適用すれば，パラメータ $\hat{\beta}_i$ の *SE* が求められているので，有効反復数 n_e を計算する必要はない．

式 (2.30) に示した 1 因子実験に関するデータの構造式は，水準平均の平均 μ，水準の効果 α_i，誤差 ε_{ij} の和として定義されている．繰り返しが不揃いな 1 因子実験データの解析において（1，–1）対比型ダミー変数によるデザイン行列 $\boldsymbol{X}$ を設定することにより，回帰分析のパラメータの推定値として，μ および α_i が得られ，それらの標準誤差 *SE* も同時に得られることを示した．伝統的な平方和の分解による方法は，分散分析表を作成，表 2.13 に示したように μ および α_i の推定はできるのであるが，統計的な考察に必要な標準誤差 *SE* を求めるすべがない．

2.3. (0, 1)型ダミー変数を用いた 1 因子実験

繰り返しが等しい 1 因子実験データに対し第 2.1 節では，（1，-1）対比型のダミー変数を用いたデザイン行列 $\boldsymbol{X}$ による線形モデルを取り上げた．その際に（0，1）型ダミー変数については，線形モデルを念頭にした場合に不向きであることを示唆しただけで，理由を示さなかったので，第 2.2 節で取り上げた繰り返しが不揃いの事例を用い，実際に適用した結果により明らかにする．

表 2.19 右に示すように，（0，1）型ダミー変数は，質的変数の最初の水準 A_0 に対して（0，0，0，0）を与え，2 番目の水準 A_1 に（1，0，0，0）を与え，最後の水準 A_4 に（0，0，0，1）を与える．ダミー変数名としては，1 と設定されている水準に対応付けることにすると，因子 A が（A_0，A_1, . . . ，A_4）となっているので，ダミー変数名とし（a_1，a_2，a_3，a_4）を付与している．

表 2.19 （1，-1）対比型 vs.（0，1）型ダミー変数

	切片	(1, -1) 対比型			
	x_0	a_0	a_1	a_2	a_3
A_0	1	1	0	0	0
A_1	1	0	1	0	0
A_2	1	0	0	1	0
A_3	1	0	0	0	1
A_4	1	-1	-1	-1	-1
和	5	0	0	0	0
平均	1	0	0	0	0

	切片	(0, 1) 型			
	x_0	a_1	a_2	a_3	a_4
A_0	1	0	0	0	0
A_1	1	1	0	0	0
A_2	1	0	1	0	0
A_3	1	0	0	1	0
A_4	1	0	0	0	1
和	5	1	1	1	1
平均	1	1/5	1/5	1/5	1/5

(0, 1)型ダミー変数によるデザイン行列 $\boldsymbol{X}$ を用いた場合

表 2.20 に示すのは，（0，1）型ダミー変数のデザイン行列 $\boldsymbol{X}$ を用いた回帰分析の結果である．表 2.14 に示した（1，-1）対比型ダミー変数と比べて大きく異なるのは，最後の水準 A_4 が（–1，–1，–1，–1）から（0，0，0，1）となり，すっきりとしている．このことが，（0，1）型ダミー変数が好まれる理由でもある．

表 2.20 右上段に示す分散分析表は，表 2.14 右上の分散分析表と完全に一致しているが，パラメータの推定値は，まったく異なる．どのように理解したら良いのであろうか．切片 x_0 の推定値が，$\hat{\beta}_0 = 46.00$ となっている． A_0 のダミー変数は，$a_1 = 0$, . . . ，$a_4 = 0$ であり，切片が $x_0 = 1$ となっているので，切片のパラメータ $\hat{\beta}_0 = 46.00$ は，最初の水準 A_0 の推定値になる．

表 2.20　（0，1）型のデザイン行列 $\boldsymbol{X}$ に対する回帰分析

	番号		$\boldsymbol{Y}$	———$\boldsymbol{X}$———				
群	i	j	y	x_0	a_1	a_2	a_3	a_4
A_0	0	1	43	1	0	0	0	0
	0	2	45	1	0	0	0	0
	0	3	42	1	0	0	0	0
	0	4	47	1	0	0	0	0
	0	5	49	1	0	0	0	0
	0	6	50	1	0	0	0	0
A_1	1	1	47	1	1	0	0	0
	1	2	51	1	1	0	0	0
	1	3	49	1	1	0	0	0
A_2	2	1	54	1	0	1	0	0
	2	2	48	1	0	1	0	0
	2	3	57	1	0	1	0	0
A_3	3	1	55	1	0	0	1	0
	3	2	58	1	0	0	1	0
	3	3	61	1	0	0	1	0
A_4	4	1	52	1	0	0	0	1
	4	2	48	1	0	0	0	1
	4	3	53	1	0	0	0	1

分散分析表,「定数に 0 を使用」off

	自由度	変動	分散	分散比
回帰	4	316.50	79.1250	7.6763
残差	13	134.00	10.3077	
合計	17	450.50		

		係数	標準誤差	t	P-値
$\beta\hat{}_0$	切片 x_0	46.0000	1.3107	35.0956	0.0000
$\beta\hat{}_1$	a_1	3.0000	2.2702	1.3215	0.2091
$\beta\hat{}_2$	a_2	7.0000	2.2702	3.0834	0.0087
$\beta\hat{}_3$	a_3	12.0000	2.2702	5.2859	0.0001
$\beta\hat{}_4$	a_4	5.0000	2.2702	2.2024	0.0463

パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\hat{\sigma}^2$

		x_0	a_1	a_2	a_3	a_4
$\beta\hat{}_0$		**1.7179**	-1.7179	-1.7179	-1.7179	-1.7179
$\beta\hat{}_1$		-1.7179	**5.1538**	1.7179	1.7179	1.7179
$\beta\hat{}_2$		-1.7179	1.7179	**5.1538**	1.7179	1.7179
$\beta\hat{}_3$		-1.7179	1.7179	1.7179	**5.1538**	1.7179
$\beta\hat{}_4$		-1.7179	1.7179	1.7179	1.7179	**5.1538**

=Minverse (Mmult (Transpose ($\boldsymbol{X}$の範囲), $\boldsymbol{X}$の範囲))*$\hat{\sigma}^2$

第 2 の水準 A_1 の推定値は，表 2.19 の（0，1）型ダミー変数から

$$\left.\begin{aligned}\bar{A}_1 &= \hat{\beta}_0 x_0 + \hat{\beta}_1 a_1 \\ &= \hat{\beta}_0 + \hat{\beta}_1 \\ &= 46.0 + 3.0 = 49.0\end{aligned}\right\} \qquad (2.37)$$

として求められる．この関係から $\hat{\beta}_1 = 3.00$ は，$\bar{A}_1$ は，$\bar{A}_0$ からの増分（差分）であることが分かる．

表 2.21 に示すように，（0，1）型ダミー変数を用いた場合の推定値は，最初の水準の推定値からの差分となると理解される．対照群 A_0 を基準とし A_1～A_4 との差分だけに興味がある場合に，（0，1）型ダミー変数を用いた線形モデルを適用すれば，対照群 A_0 との差分に関する t 値および p 値が得られる．表 2.21 には含めていないが，95%信頼区間も得られるので，対照群 A_0 との差の検定を目的とする場合に最も適している．

表 2.21　（0，1）型ダミー変数のデザイン行列 $\boldsymbol{X}$ を用いたパラメータ推定

		推定値		和	水準平均
A	n	$\hat{\boldsymbol{\beta}}$		$\beta\hat{}_0+\beta\hat{}_i$	$\bar{y}_{i\cdot}$
A_0	6	$\beta\hat{}_0$	46.00	—	46.00
A_1	3	$\beta\hat{}_1$	3.00	49.00	49.00
A_2	3	$\beta\hat{}_2$	7.00	53.00	53.00
A_3	3	$\beta\hat{}_3$	12.00	58.00	58.00
A_4	3	$\beta\hat{}_4$	5.00	51.00	51.00

(0, 1)型ダミー変数の場合の各水準の 95%信頼区間

表 2.22 に示すようにダミー変数の型が異なったとしても，因子 A の各水準の推定値は，得られたパラメータから，

$$\bar{A}_0 = 46.0,\ \bar{A}_1 = 46.0 + 3.0 = 49.0, \ldots,\ \bar{A}_4 = 46.0 + 5.0 = 51.0$$

となり，（1，-1）対比型ダミー変数の場合の推定値は，

$$\bar{A}_0 = 51.4 - 5.4 = 46.0,\ \bar{A}_1 = 51.4 - 2.4 = 49.0, \ldots,\ \bar{A}_4 = 51.4 + 5.4 + 2.4 - 1.6 - 6.6 = 51.0$$

と，全く同じになる．

表 2.22　パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ を用いた平均値の 95%信頼区間

				l_0	l_1	l_2	l_3	l_4			推定値	分散		幅	95%信頼区間	
			L	x_0	a_1	a_2	a_3	a_4		$\boldsymbol{\beta}\hat{}$		$\boldsymbol{l\beta}\hat{}$	$Var(\boldsymbol{l\beta}\hat{})$	SE	$t_{0.05}\times SE$	L95%	U95%
平均	A_0	$L^{(1)}$		1	0	0	0	0		46.00	=	46.00	1.7179	1.3107	2.8316	43.17	48.83
	A_1	$L^{(2)}$		1	1	0	0	0		3.00		49.00	3.4359	1.8536	4.0045	45.00	53.00
	A_2	$L^{(3)}$		1	0	1	0	0		7.00		53.00	3.4359	1.8536	4.0045	49.00	57.00
	A_3	$L^{(4)}$		1	0	0	1	0		12.00		58.00	3.4359	1.8536	4.0045	54.00	62.00
	A_4	$L^{(5)}$		1	0	0	0	1		5.00		51.00	3.4359	1.8536	4.0045	47.00	55.00
														$t_{0.05}(18\text{-}5)$=	2.1604		

因子 A の各水準の 95%信頼区間も報告書に掲載したい．そのためには，表 2.20 に示したパラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ を用いた各種の推定が必要となる． A_1 水準の平均値は，

$$\bar{A}_1 = \hat{\beta}_0 + \hat{\beta}_1 = 46.0 + 3.0 = 49.0$$

であり，その分散は，

$$\begin{aligned} Var(\hat{\beta}_0 + \hat{\beta}_1) &= Var(\hat{\beta}_0) + 2Cov(\hat{\beta}_0, \hat{\beta}_1) + Var(\hat{\beta}_1) \\ &= 1.7179 + 2\times(-1.7179) + 5.1538 = 3.4359 \end{aligned}$$

として求められる．このように，水準平均の 95%信頼区間が必要な場合は，表 2.10 に示したように（1，1）標示型ダミー変数を用いたデザイン行列 $\boldsymbol{X}$ に対し，「定数に 0 を使用」オプションをオフとした回帰分析を適用すれば，簡単に得られる．

選択したダミー変数によってパラメータの推定結果が全く異なるが，相互に互換性があることを示した．解析の目的に即したダミー変数を選択することにより，スピーディで効果的な解析が行える．それらの結果を「予測プロファイル」としてまとめることにより，結果の解釈に役立つことを示した．

2.4. 1 因子実験の量的変数に対する多項式回帰

量的変数に対して幾つかの水準を設定して質的変数として解析する方法が広く普及している．手計算時代には，量的変数のまま回帰分析を適用すること自体が難儀であった．そのために，量的変数を質的変数として扱うことにより平方和の分解による分散分析表を作成し，F 検定による統計的な考察が行えることは，画期的な方法であり，実験データの解析法として合理的な方法であったともいえる．さて，F 検定で何らかの水準間に差が統計的にあった場合に，さらに，直線的な関係なのか，あるいは曲線的な関係なのか明らかにしたい．

量的変数を質的変数に置き換えた分散分析では，直線関係か否かの判定に LOF（Lack Of Fit，あてはまりの悪さ）解析が知られている．芳賀（2014）の第 2 章「量的因子の 1 因子実験」に LOF 解析が展開されているが，伝統的な平方和の分解による分散分析と回帰分析を併合した解析法となっていて煩雑で見通しが悪い．そこで，デザイン行列 $\boldsymbol{X}$ を用いた多項式回帰の適用により，見通しの良い解析法を提示する．表 2.23 に示すのは，対照群 0 mg/kg に対し，10，20，30 mg/kg を 5 匹の動物に対し投与し，ある薬理反応を得た結果である．

表 2.23　ある薬理作用に対する用量反応実験データ［芳賀（2014），表示 2.2.1］

	用量 *dose*	繰り返し						
A	mg/kg	1	2	3	4	5	平均	標準偏差
A_1	0	10.5	9.6	10.4	10.2	9.4	10.020	0.4919
A_2	10	10.8	10.7	11.1	10.9	11.0	10.900	0.1581
A_3	20	11.4	10.7	10.9	11.3	11.7	11.200	0.4000
A_4	30	11.9	11.2	11.0	11.1	11.3	11.300	0.3536
	全体						10.855	

この実験の目的は，どの程度の用量から反応が対照群に対して統計的な差（有意な差）があるのかを確認するのが主たる目的である．A_2 の 10 mg/kg 群の反応は，A_1 の 0 mg/kg 対照群の最大値である 10.5 よりも（10.7〜11.1）と全て大きく，統計的な方法によらずとも明らかな反応の差があると判断される．さらに，用量反応関係が直線的なのか，何らかの曲線関係が示唆されるのかを統計的に調べたい．

デザイン行列 $\boldsymbol{X}$ を用いた多項式回帰

これまで示してきた質的因子 A の場合のデザイン行列 $\boldsymbol{X}$ と同様に，1 次式，2 次式, . . . をあてはめる回帰分析を効率よく行うためにデザイン行列 $\boldsymbol{X}$ を前もって設定する．表 2.24 右に示すように用量 *dose* は 2 桁なので，冪乗すると桁数のインフレーションが起きるので，10 で割って 1 桁とし，2 乗および 3 乗とした“多項式型ダミー変数”とする．

表 2.24 多項式回帰のためのダミー変数

	(0, 1) 型				*dose*	*x*	多項式型			
	a_2	a_3	a_4		mg/kg	*dose*/10	x^0	x^1	x^2	x^3
A_1	0	0	0	A_1	0	0	1	0	0	0
A_2	1	0	0	A_2	10	1	1	1	1	1
A_3	0	1	0	A_3	20	2	1	2	4	8
A_4	0	0	1	A_4	30	3	1	3	9	27

3 次式のあてはめ

表 2.25 右上段に示すのは，3 次の多項式型ダミー変数を（20 行×4 列）のデザイン行列 $\boldsymbol{X}$ とした回帰分析の結果である．分散分析表の分散比 F 値は，12.2742 と十分に大きく 3 次式がよくあてはまっているとの結果であるが，パラメータ（係数）の変数 x^3 の p 値からは，統計的に有意ではないとの結果である．これは，低次の（x^0， x^1， x^2）の変数が回帰モデルにすでに含まれている条件下でさらに x^3 を加えた時の x^3 自身の寄与分を表していて，2 次式のあてはめに加えて 3 次式にする必要がないと解釈される．同様に 1 次の係数 $\hat{\beta}_1^{(3)}=1.2967$ の場合は，（x^0， x^2， x^3）の変数が回帰モデルにすでに含まれている条件下でさらに x^1 を加えて時の x^1 自身の寄与分を表していて，単純に 1 次式をあてはめた結果ではない．

表 2.25 デザイン行列を用いた 3 次式に対する回帰分析

	dose		— デザイン行列 $\boldsymbol{X}$ —				$\boldsymbol{X\beta}$^
A	mg/kg	*y*	x^0	x^1	x^2	x^3	*y*^
A_1	0	10.5	1	0	0	0	10.02
	0	9.6	1	0	0	0	10.02
	0	10.4	1	0	0	0	10.02
	0	10.2	1	0	0	0	10.02
	0	9.4	1	0	0	0	10.02
A_2	10	10.8	1	1	1	1	10.90
	10	10.7	1	1	1	1	10.90
	10	11.1	1	1	1	1	10.90
	10	10.9	1	1	1	1	10.90
	10	11.0	1	1	1	1	10.90
A_3	20	11.4	1	2	4	8	11.20
	20	10.7	1	2	4	8	11.20
	20	10.9	1	2	4	8	11.20
	20	11.3	1	2	4	8	11.20
	20	11.7	1	2	4	8	11.20
A_4	30	11.9	1	3	9	27	11.30
	30	11.2	1	3	9	27	11.30
	30	11.0	1	3	9	27	11.30
	30	11.1	1	3	9	27	11.30
	30	11.3	1	3	9	27	11.30

分散分析表,「定数に 0 使用」off

	自由度	変動	分散	分散比
回帰	3	5.0815	1.6938	12.2742
残差	16	2.2080	**0.1380**	
合計	19	7.2895		

	係数	標準誤差	*t*	*P*-値
$\beta\hat{}_0^{(3)}$ 切片 x^0	10.0200	0.1661	60.3133	0.0000
$\beta\hat{}_1^{(3)}$ x^1	1.2967	0.6374	2.0342	0.0589
$\beta\hat{}_2^{(3)}$ x^2	-0.4800	0.5634	-0.8520	0.4068
$\beta\hat{}_3^{(3)}$ x^3	0.0633	0.1238	0.5115	**0.6160**

パラメータの共分散行列 $\boldsymbol{\Sigma}(\boldsymbol{\beta}\hat{}^{(3)})=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma\hat{}^2$

	x_0	x^1	x^2	x^3
$\beta\hat{}_0^{(3)}$	**0.0276**	-0.0506	0.0276	-0.0046
$\beta\hat{}_1^{(3)}$	-0.0506	**0.4063**	-0.3450	0.0721
$\beta\hat{}_2^{(3)}$	0.0276	-0.3450	**0.3174**	-0.0690
$\beta\hat{}_3^{(3)}$	-0.0046	0.0721	-0.0690	**0.0153**

=Minverse (Mmult (Transpose ($\boldsymbol{X}$の範囲), $\boldsymbol{X}$の範囲))*$\sigma\hat{}^2$

パラメータの推定値からは，3 次式のあてはめに関して否定的な結果であったが，表 2.26 に示すように 3 次曲線のあてはめを試みる．3 次式は，

$$\begin{aligned}\hat{y}^{(3)} &= \hat{\beta}_0^{(3)}x^0 + \hat{\beta}_1^{(3)}x^1 + \hat{\beta}_2^{(3)}x^2 + \hat{\beta}_3^{(3)}x^3 \\ &= 10.0200x^0 + 1.2967x^1 - 0.4800x^2 + 0.0633x^3\end{aligned} \tag{2.38}$$

であり，その分散 $Var(\hat{y}^{(3)})$ は，これまで用いてきたの線形和の係数ベクトル $\boldsymbol{l}$ を，［$l_0^{(3)} = x^0 = 1$，$l_1^{(3)} = x^1$，$l_2^{(3)} = x^2$，$l_3^{(3)} = x^3$］としたときに，これまでと同様にパラメータの共分散行列に間するベクトル $\boldsymbol{l}$ の 2 次形式により，

$$Var(\hat{y}^{(3)}) = \boldsymbol{l}^{(3)}\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^{(3)})(\boldsymbol{l}^{(3)})^T \tag{2.39}$$

として分散の推定ができる．

表 2.26 に示すように 3 次曲線を滑らかに描くために，用量 *dose* を -10 から 5 刻みに 40 まで与え，式 (2.38) に代えて

$$\hat{y}^{(3)} = \hat{\beta}_0^{(3)}l_0^{(3)} + \hat{\beta}_1^{(3)}l_1^{(3)} + \hat{\beta}_2^{(3)}l_2^{(3)} + \hat{\beta}_3^{(3)}l_3^{(3)} \tag{2.40}$$

により推定し，分散 $Var(\hat{y}^{(3)})$ は，式 (2.39) によって求めている．推定値結果は，当然のことながら，それぞれの用量 *dose* の平均値に完全にフィットしている．図 2.8 に示すように 95%信頼区間は，6 次式となり複雑な曲線となっている．統計的には，より少ないパラメータでの推定が，より安定した回帰式となることが知られている．この 95%信頼区間から，得られたデータの範囲を超えた場合に上下に跳ね上がっていることから，オーバーフィットによる予測式の不安定さが現われている．

表 2.26　3 次曲線をあてはめた予測プロファイル

dose	l_0	l_1	l_2	l_3	推定値	分散	幅	95%信頼区間	
mg/kg	x^0	x^1	x^2	x^3	y^	$Var(y$^$)$	$t_{0.05}\times SE$	L95%	U95%
-10	1	-1.0	1.0	-1.0	8.18	1.904	2.93	5.25	11.11
-5	1	-0.5	0.3	-0.1	9.24	0.314	1.19	8.06	10.43
0	1	0.0	0.0	0.0	**10.02**	0.028	0.35	9.67	10.37
5	1	0.5	0.3	0.1	10.56	0.030	0.37	10.19	10.92
10	1	1.0	1.0	1.0	**10.90**	0.028	0.35	10.55	11.25
15	1	1.5	2.3	3.4	11.10	0.018	0.28	10.82	11.38
20	1	2.0	4.0	8.0	**11.20**	0.028	0.35	10.85	11.55
25	1	2.5	6.3	15.6	11.25	0.030	0.37	10.89	11.62
30	1	3.0	9.0	27.0	**11.30**	0.028	0.35	10.95	11.65
35	1	3.5	12.3	42.9	11.39	0.314	1.19	10.21	12.58
40	1	4.0	16.0	64.0	11.58	1.904	2.93	8.65	14.51
					$t_{0.05}(20-4)$=		2.1199		

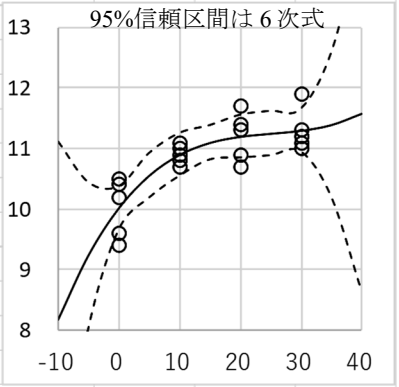

図 2.8　3 次曲線のあてはめ

2 次式のあてはめ

2 次式のあてはめは，デザイン行列 $\boldsymbol{X}$ の（x^0，x^1，x^2）について回帰分析を行なう．表 2.27 に示すように，2 次式をあてはめた場合に，x^2 変数の係数は，3 次式のあてはめで p

表 2.27 デザイン行列を用いた 2 次式に対する回帰分析

	dose		— デザイン行列 $\boldsymbol{X}$ —				$\boldsymbol{X\beta}$^
A	mg/kg	y	x^0	x^1	x^2	x^3	y^
A_1	0	10.5	1	0	0	0	10.04
	0	9.6	1	0	0	0	10.04
	0	10.4	1	0	0	0	10.04
	0	10.2	1	0	0	0	10.04
	0	9.4	1	0	0	0	10.04
A_2	10	10.8	1	1	1	1	10.84
	10	10.7	1	1	1	1	10.84
	10	11.1	1	1	1	1	10.84
	10	10.9	1	1	1	1	10.84
	10	11.0	1	1	1	1	10.84
A_3	20	11.4	1	2	4	8	11.26
	20	10.7	1	2	4	8	11.26
	20	10.9	1	2	4	8	11.26
	20	11.3	1	2	4	8	11.26
	20	11.7	1	2	4	8	11.26
A_4	30	11.9	1	3	9	27	11.28
	30	11.2	1	3	9	27	11.28
	30	11.0	1	3	9	27	11.28
	30	11.1	1	3	9	27	11.28
	30	11.3	1	3	9	27	11.28

分散分析表,「定数に 0 使用」off

	自由度	変動	分散	分散比
回帰	2	5.0454	2.5227	19.1105
残差	17	2.2441	**0.1320**	
合計	19	7.2895		

	係数	標準誤差	t	P-値
切片 x^0	10.0390	0.1584	63.3895	0.0000
x^1	0.9990	0.2543	3.9280	0.0011
x^2	-0.1950	0.0812	-2.4002	**0.0281**

パラメータの共分散行列 $\boldsymbol{\Sigma}(\boldsymbol{\beta}\text{^}^{(2)})=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma\text{^}^2$

	x_0	x^1	x^2
$\beta\text{^}_0^{(2)}$	**0.0251**	-0.0277	0.0066
$\beta\text{^}_1^{(2)}$	-0.0277	**0.0647**	-0.0198
$\beta\text{^}_2^{(2)}$	0.0066	-0.0198	**0.0066**

=Minverse (Mmult (Transpose ($\boldsymbol{X}$の範囲), $\boldsymbol{X}$の範囲))*$\sigma\text{^}^2$

$=0.4068$ と有意でなかったのであるが，$\hat{\beta}_2^{(2)}=-0.1950$，$p=0.0281$ と有意な差に変化する．この結果から，1 次式ではなく 2 次式のあてはめが支持されることになる．

表 2.28 に推定された 2 次式の回帰パラメータ $\hat{\beta}_1^{(2)}$ およびパラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^{(2)})$ を用いて計算した推定値 $\hat{y}$ および 95%信頼区間を散布図上に重ね書きした予測プロファイル

表 2.28 2 次曲線をあてはめた予測プロファイル

dose	l_0	l_1	l_2	l_3	推定値	分散	幅	95%信頼区間	
mg/kg	x^0	x^1	x^2	x^3	y^	$Var(y\text{^})$	$t_{0.05}\times SE$	L95%	U95%
-10	1	-1.0	1.0	-1.0	8.85	0.205	0.95	7.89	9.80
-5	1	-0.5	0.3	-0.1	9.49	0.078	0.59	8.90	10.08
0	1	0.0	0.0	0.0	**10.04**	0.025	0.33	9.70	10.37
5	1	0.5	0.3	0.1	10.49	0.012	0.23	10.26	10.72
10	1	1.0	1.0	1.0	**10.84**	0.015	0.25	10.59	11.10
15	1	1.5	2.3	3.4	11.10	0.017	0.27	10.82	11.37
20	1	2.0	4.0	8.0	**11.26**	0.015	0.25	11.00	11.51
25	1	2.5	6.3	15.6	11.32	0.012	0.23	11.08	11.55
30	1	3.0	9.0	27.0	**11.28**	0.025	0.33	10.95	11.62
35	1	3.5	12.3	42.9	11.15	0.078	0.59	10.56	11.73
40	1	4.0	16.0	64.0	10.92	0.205	0.95	9.96	11.87
					$t_{0.05}(20-3)=$		2.1098		

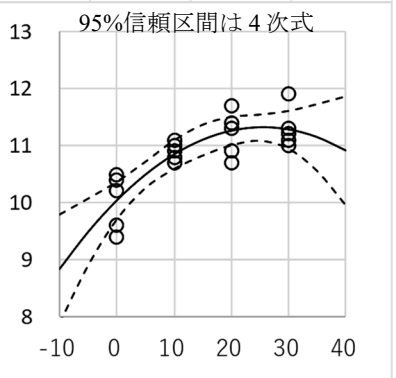

図 2.9 2 次曲線のあてはめ

を示す．統計ソフトでは，2 次式に限らず多項式回帰の 95%信頼区間を図示できるものもあるが，その計算方法はブラックボックス的である．

多くの統計入門書で，1 次式の回帰直線の場合の 95%信頼区間の計算式は，平方和を用いた式で示されているが，2 次式以上の回帰曲線の 95%信頼区間の計算式を見いだすことは不可能に近い．この理由は，1 次式の回帰直線のパラメータ推定に際し，平方和に基づく計算式が示されているが，2 次式への拡張性が全くない式となっているためである．デザイン行列 $\boldsymbol{X}$ に基づいた計算法では，これまでも示してきたように，線形和 $L=\boldsymbol{l}\hat{\boldsymbol{\beta}}$ の係数ベクトル $\boldsymbol{l}$ を用いた 2 次形式

$$Var(\boldsymbol{l}\hat{\boldsymbol{\beta}})=Var(\hat{y})=\boldsymbol{l}\,\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})\,\boldsymbol{l}^{T} \tag{2.41}$$

によって，推定値 $\hat{y}$ の分散 $Var(\hat{y})$ を計算し

$$95\%CL(\hat{y})=\hat{y}\pm t_{0.05}(df_e)\sqrt{Var(\hat{y})} \tag{2.42}$$

によって 95%信頼区間の計算をしてきた．ここでは，2 次式の推定値の分散 $Var(\hat{y})$ の計算過程を詳しく示す．

外挿となるが，$dose=40$ mg/kg の場合を示す．表 2.28 の係数ベクトル $\boldsymbol{l}^{(2)}_{dose=40}$ は，

$$\boldsymbol{l}^{(2)}_{dose=40}=[1\quad 4\quad 16]$$

である．推定値 $\hat{y}^{(2)}_{dose=40}$ は，

$$\begin{aligned}\hat{y}^{(2)}_{dose=40}&=\boldsymbol{l}^{(2)}_{dose=40}\hat{\boldsymbol{\beta}}^{(2)}\\&=[1\quad 4\quad 16]\begin{bmatrix}10.0390\\0.9990\\-0.1950\end{bmatrix}\\&=1\times 10.0390+4\times 0.9990+16\times(-0.1950)=10.92\end{aligned}$$

であり，その分散は，

$$Var(\hat{y}^{(2)}_{dose=40})=\boldsymbol{l}^{(2)}_{dose=40}\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^{(2)})(\boldsymbol{l}^{(2)}_{dose=40})^{T}$$

=	1	4	16		**0.0251**	-0.0277	0.0066	1
					-0.0277	**0.0647**	-0.0198	4
					0.0066	-0.0198	**0.0066**	16
				=	0.0198	-0.0858	0.0330	1
								4
								16
				=	0.2046			

として計算されている．

シグマでの計算式では，パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^{(2)})$ の要素を $c_{ii'}$ とし，$\boldsymbol{l}^{(2)}_{dose=40}$ の係数を $l^{(2)}_{i}$ とすると

$$Var(\hat{y}^{(2)})=\sum_{i=0}^{2}\sum_{i'=0}^{2}\left[l^{(2)}_{i}\,l^{(2)}_{i'}\,c_{ii'}\right] \tag{2.43}$$

となる．この2次形式の式からも明らかなように，95%信頼区間の計算式は，4次式となり，これを四則演算式で示すこと自体が難儀なことであるし，実際に計算することも煩雑である．

表 2.28 で示した Excel の計算シートでの計算式は，1 行目で

$Var(\boldsymbol{l}^{(2)}\hat{\boldsymbol{\beta}}^{(2)})$=Mmult (Mmult ($\boldsymbol{l}^{(2)}_{dose=40}$ の範囲, $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^{(2)})$ の範囲), Transpose ($\boldsymbol{l}^{(2)}_{dose=40}$ の範囲)) = 0.205

$t_{0.05}\times SE$ = T.inv.2T (0.05, 20–3) * Sqrt ($Var(\boldsymbol{l}^{(2)}\hat{\boldsymbol{\beta}}^{(2)})$) = 0.95

$L95\%$ = 推定値 - $(t_{0.05}\times SE)$ = 7.89, $U95\%$ = 推定値 + $(t_{0.05}\times SE)$ = 9.80

として計算し，2 行目以後はフィルハンドルで計算式をコピーして求めている．これらの計算式は，汎用的であり線形モデルに限らずロジスティック回帰でも，ポアソン回帰でも共通である．

デザイン行列 $\boldsymbol{X}$ を用いた単回帰分析

一般的に，回帰分析というと直線をあてはめる単変数の場合が想定され，2 変数以上の場合に重回帰分析と区別されるが，本書では全て回帰分析として区別しないで使っている．ダミー変数を用いた場合に線形モデルとし，回帰モデルを区別しているが，どちらもデザイン行列を用いた解析方法としては，同一である．

これまでの流れで，1 変数の場合も切片を明示的に示したデザイン行列 $\boldsymbol{X}$ を用いた解析の流れで，表 2.29 にデザイン行列を用いた単回帰分析の結果を示す．分散分析表および回帰パラメータの推定は，Excel による回帰分析の結果であるが，その下の 2×2 のパラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^{(1)})$ は，切片を含めた（20 行×2 列）のデザイン行列 $\boldsymbol{X}$ から計算されている．

表 2.29 デザイン行列を用いた単回帰分析

	dose		— デザイン行列 $\boldsymbol{X}$ —				$\boldsymbol{X\beta}$^
A	mg/kg	y	x^0	x^1	x^2	x^3	y^
A_1	0	10.5	1	0	0	0	10.04
	:						
	0	9.4	1	0	0	0	10.04
A_2	10	10.8	1	1	1	1	10.84
	:						
	10	11.0	1	1	1	1	10.84
A_3	20	11.4	1	2	4	8	11.26
	:						
	20	11.7	1	2	4	8	11.26
A_4	30	11.9	1	3	9	27	11.28
	:						
	30	11.3	1	3	9	27	11.28

分散分析表,「定数に 0 使用」off

	自由度	変動	分散	分散比
回帰	1	4.2849	4.2849	25.6700
残差	18	3.0046	**0.1669**	
合計	19	7.2895		

	係数	標準誤差	t	P-値
切片 x^0	10.2340	0.1529	66.9460	0.0000
x^1	0.4140	0.0817	5.0666	0.0001

パラメータの共分散行列 $\boldsymbol{\Sigma}(\boldsymbol{\beta}\hat{}^{(1)})=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma\hat{}^2$

$\beta\hat{}_0^{(1)}$	**0.0234**	-0.0100
$\beta\hat{}_1^{(1)}$	-0.0100	**0.0067**
	x_0	x^1

=Minverse (Mmult (Transpose ($\boldsymbol{X}$の範囲), $\boldsymbol{X}$の範囲))*$\sigma\hat{}^2$

パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^{(1)})$ の活用は，平方和を用いる 95%信頼区間の計算方法からの脱却を図ることができる．Excel の単回帰分析を使っている人たちは，どのようにしたら 95%信頼区間を計算し，図示したいとの思いがあり，その際に頼りにするのは，平方和 S_{xx} などを用いた計算公式であり，元データを用いてシグマを用いた計算が別途必要となる．第 5 章で詳しく示すが，伝統的な計算公式は，過度に標準化されており，2 次式に対する計算式への拡張のヒントすら示されないのが最大の問題点である．まさに，応用力を発揮できないようにするための「ガラスの天井」がはめ込まれているがごとくである．

表 2.30 に示すように，回帰直線の 95%信頼区間を計算するために必要な推定値 $\hat{y}$ の分散 $Var(\hat{y}^{(1)})$ は，

$$\left.\begin{aligned} Var(\hat{y}^{(1)}) &= \boldsymbol{l}^{(1)}\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^{(1)})\boldsymbol{l}^{(1)T} \\ &= [x^0\ x^1]\begin{bmatrix} Ver(\beta_0^{(1)}) & Cov(\beta_0^{(1)},\beta_1^{(1)}) \\ Cov(\beta_0^{(1)},\beta_1^{(1)}) & Ver(\beta_1^{(1)}) \end{bmatrix}\begin{bmatrix} x^0 \\ x^1 \end{bmatrix} \\ &= (x^0)^2 Ver(\beta_0^{(1)}) + 2x^0x^1 Cov(\beta_0^{(1)},\beta_1^{(1)}) + (x^1)^2 Ver(\beta_1^{(1)}) \\ &= Ver(\beta_0^{(1)}) + 2xCov(\beta_0^{(1)},\beta_1^{(1)}) + x^2 Ver(\beta_1^{(1)}) \end{aligned}\right\} \quad (2.44)$$

のように合成分散の一般式で表すこともでき，$\boldsymbol{l}=[1\ \ -1.0]$ の場合であれば，

$$\left.\begin{aligned} Var(\hat{y}^{(1)}) &= = \boldsymbol{l}^{(1)}\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^{(1)})\boldsymbol{l}^{(1)T} \\ &= 0.0234 + 2\times(-1)\times(-0.0100) + (-1)^2\times 0.0067 = 0.0501 \end{aligned}\right\} \quad (2.45)$$

のように計算されている．この計算式が示されている成書を見いだすことができるが，偏差平方和 S_{xx} を用いた式へ誘導されるのが常である．平方和の分解と分散の加法性による解析法に対する先人たちの“こだわり”にほとんどの人達がからめとられている．

表 2.30 に示すように，パラメータの共分散 $Cov(\hat{\beta}_0,\hat{\beta}_1)$ を素直に使った計算により，回帰直線の分散を計算し，95%信頼区間を重ね書きした予測プロファイルを図 2.10 に示す．2 次式

表 2.30　単回帰式の回帰直線と 95%信頼区間の予測プロファイル

dose	l_0	l_1	l_2	l_3	推定値	分散	幅	95%信頼区間	
mg/kg	x^0	x^1	x^2	x^3	$\boldsymbol{l}\hat{\boldsymbol{\beta}}$	$Var(\boldsymbol{l}\hat{\boldsymbol{\beta}})$	$t_{0.05}\times SE$	L95%	U95%
-10	1	-1.0	1.0	-1.0	9.82	0.0501	0.47	9.35	10.29
-5	1	-0.5	0.3	-0.1	10.03	0.035	0.39	9.63	10.42
0	1	0.0	0.0	0.0	**10.23**	0.023	0.32	9.91	10.56
5	1	0.5	0.3	0.1	10.44	0.015	0.26	10.18	10.70
10	1	1.0	1.0	1.0	**10.65**	0.010	0.21	10.44	10.86
15	1	1.5	2.3	3.4	10.86	0.008	0.19	10.66	11.05
20	1	2.0	4.0	8.0	**11.06**	0.010	0.21	10.85	11.27
25	1	2.5	6.3	15.6	11.27	0.015	0.26	11.01	11.53
30	1	3.0	9.0	27.0	**11.48**	0.023	0.32	11.15	11.80
35	1	3.5	12.3	42.9	11.68	0.035	0.39	11.29	12.08
40	1	4.0	16.0	64.0	11.89	0.050	0.47	11.42	12.36
					$t_{0.05}(20-2)=$	1.9727			

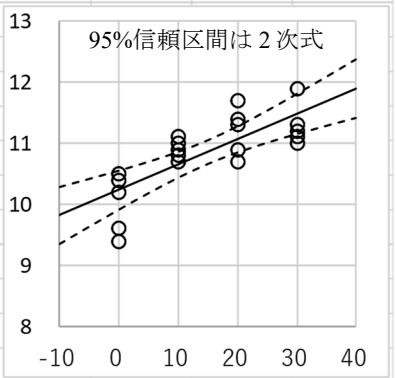

図 2.10　1 次式のあてはめ

のあてはまりの検討をしなければ，少々あてはまりは悪いようにも見えるが，直線があてはまっていると結論したくなるような予測プロファイルとなっている．

あてはまりの悪さLOF解析

量的変数に対する1因子実験データに多項式をあてはめ，次数を順次減らすことにより，2次式のあてはめが適していることを示した．平方和の分解による分散分析表の作成に引き続き，回帰直線をあてはめて分散分析表を作成し，有意な差が得られたのであるが，散布図に直線を引いてみると2次曲線らしき形状があり気になる．統計的に曲線の「あてはまり」は統計的に良くないと言えれば，直線のあてはめが正当化される．このような方法は，LOF（Lack OF Fit）解析として知られている．

表2.25に示した3次式のあてはめの分散分析表は，質的変数Aとしての分散分析表に一致する．表2.29に示した1次式のあてはめの分散分析表は，回帰分析そのものである．表2.31に示すように質的変数Aとしての自由度3の平方和は，$S_{\mathrm{A}}=5.0815$であり，回帰の平方和は，$S_{\text{回帰}}^{(1)}=4.2849$である．あてはまりの悪さLOFの平方和は，

$$\begin{aligned} S_{\mathrm{LOF}} &= S_{\mathrm{A}} - S_{\text{回帰}}^{(1)} \\ &= 5.0815 - 4.2849 = 0.7966 \end{aligned} \tag{2.46}$$

として計算される．

表2.31 質的変数Aとしての分散分析表および回帰分析の分散分析表

要因	*df*	平方和	平均平方	*F*値	*p*値
A	3	**5.0815**	1.6938	12.27	0.0001
e	16	2.2080	0.1380		
T	19	7.2895			

要因	*df*	平方和	平均平方	*F*値	*p*値
Reg	1	**4.2849**	4.2849	25.67	0.0001
e	18	3.0046	0.1669		
T	19	7.2895			

表2.32に示すのは，これらの分散分析表を統合してLOF解析のための分散分析表である．$S_{\mathrm{LOF}}=0.7966$なので，その自由度2で割った平均平方を，さらに誤差分散の推定値$\hat{\sigma}^2=0.1380$で割って求めた$F=2.8862$から$p=0.0851$が算出される．微妙な大きさのp値であるが，これを無視すれば，統計的にLOFが有意でないので，回帰直線のあてはまりに問題

表2.32 LOF解析のための分散分析表

要因	*df*		平方和	平均平方	*F*値	*p*値
A	3		5.0815	1.6938	12.2742	0.0001
Reg		1	4.2849	4.2849	31.0500	0.0000
LOF		2	0.7966	0.3983	2.8862	0.0851
e	16		2.2080	0.1380		
T	19		7.2895			

がないと言うことも可能である．別の見方をすれば，2 次式をあてはめ，更なる LOF 解析を行ってみる必要性もある．

LOF（あてはまりの悪さ）解析は，直線をあてはめていいのか，それで十分なのか，などの疑問に対する解析法であるが，込みいった解析法であるので，これまでの結果を整理し理解を深めたい．表 2.29 に示すように，1 次式をあてはめた場合に回帰係数 $\hat{\beta}_1^{(1)} = 0.4140$ $(p < 0.001)$ である．2 次式をあてはめた場合に 2 次の項 $\hat{\beta}_2^{(2)} = -0.1950$ $(p < 0.0281)$ と有意となっている．3 次式をあてはめた場合に 3 次の項 $\hat{\beta}_3^{(3)} = 0.0633$ (NS) となるので，2 次式のあてはめが支持される．

LOF 解析に代わる逐次平方和（タイプ I の平方和）

LOF 解析は，量的変数に対する 1 因子の分散分析表に回帰分析から得られた回帰の平方和を組み込む方法として知られているが，表 2.24 に示したように多項式型ダミー変数をデザイン行列 $\boldsymbol{X}$ とした用いた場合には，表 2.33 に示すように，1 次，2 次，3 次, . . . のように回帰分析を行ない，得られた回帰パラメータの p 値の変化を評価することにより平方和を主体にした LOF 解析の代わりとなる．

表 2.33　回帰パラメータに関する有意差検定

	1次式			2次式			3次式		
	係数	p		係数	p		係数	p	
x^0	10.2340	0.0000		10.0390	0.0000		10.0200	0.0000	
x^1	0.4140	0.0001	***	0.9990	0.0011	**	1.2967	0.0589	
x^2				-0.1950	0.0281	*	-0.4800	0.4068	
x^3							0.0633	0.6160	NS

JMP の「モデルのあてはめ」で表 2.34 に示すように 3 次式のあてはめを行い，「逐次検定」を選択することにより，多項式回帰の次数を一目で判定できる．

表 2.34 JMP の「モデルのあてはめ」による 3 次式の逐次（タイプ 1）の平方和

逐次(タイプ1)検定

要因	自由度	逐次平方和	F値	p値(Prob>F)
x	1	4.2849	31.0500	<.0001*
x*x	1	0.7605	5.5109	0.0321*
x*x*x	1	0.0361	0.2616	0.6160

x*x の $p = 0.0321$ と表 2.33 の $p = 0.0281$ が異なるのは，用いている誤差分散の違いによる．

第 2 章　文 献　索引

芳賀(2014) - 医薬品開発のための統計解析 第2部 実験計画法 改訂版. 第1章 質的因子の1因子実験　55
第2章 量的因子の1因子実験　80

第 2 章　索　引

第 2 章　解析用ファイル一覧

サイズ	名前	種類
152 KB	第02章_01_1因子_等しい	Microsoft Excel ワークシート
222 KB	第02章_02_1因子_不揃い	Microsoft Excel ワークシート
85 KB	第02章_03_1因子_不揃い01型	Microsoft Excel ワークシート
5 KB	第02章_04_多項式	JMP Data Table
67 KB	第02章_04_多項式	Microsoft Excel ワークシート

第 12 回 続高橋セミナー
層別因子を含む探索的な回帰分析入門

第 2 章　デザイン行列を活用した 1 因子実験データの解析

BioStat 研究所（株）
〒105-0014 東京都 港区 芝 1－12－3 の1005
2024 年 1 月 19 日　　高橋 行雄
takahashi.stat@nifty.com ， FAX : 03-342-8035