

第9回 続高橋セミナー 最尤法によるポアソン回帰分析入門

2020年7月14日

第13章 最小2乗平均の謎を予測プロファイルで解く

最小2乗法は、統計的用語として広く使われているが、「最小2乗平均」は、竹内ら(1989)、「統計学辞典」の索引にも載っていない方言みたいなものである。「最小2乗平均」は、SAS ユーザであれば、Lsmeans で出力される「調整平均」と広く認識されており、SAS を使った解析結果には、方言であることの認識なく広く使われている。高橋・大橋・芳賀(1989)、「SAS による実験データの解析」の第15章では、最小2乗平均について丁寧に説明している。残念ながら、SAS ユーザの方言のまま現在に至っている。JMP では、「予測プロファイル」が新たな方言として加わっており、探索的な解析結果を可視化するために斬新な統計的な方法であることをこれまで示してきた。「最小2乗平均」は、「予測プロファイル」に包含される統計量の一種であり、「予測プロファイル」は、探索的なデータ解析に際し、データの内部構造を描き出すことに優れていることを示してきた。

第13章 目次

13. 最小2乗平均の謎を予測プロファイルで解く	421
13.1. 最小2乗平均(Lsmeans)とは	421
13.2. 交互作用を考慮した共分散分析	423

共分散分析の拡張、データのグラフ表示、伝統的な共分散分析の考え方、質的変数と量的変数を含む重回帰分析における交互作用解、統計ソフト JMP を用いた共分散分析、予測プロファイル、対比による水準間の差の推定、Excel による交互作用を含む解析、4本の回帰直線の推定、分散分析表、Excel による予測プロファイル、水準間の差の予測プロファイル、洗浄水の温度に関する予測プロファイル、回収液の濃度の差についての予測プロファイル、最小2乗平均(Lsmeans)、Excel による探索的な交互作用解析

続く

13.3. 共変量が 2 変量の場合の探索的な共分散分析 -----	440
共変量の効き方, Excel による 2 変量の共分散分析, 最小 2 乗平均, 水準間の差の推定, デザイン変数の活用, 共変量の影響の可視化	
13.4. 繰返しが不揃いな 2 因子の共分散分析 -----	449
対比型デザイン変数, SAS の GLM プロシジャでのデザイン変数, R 言語などでの最初的水準を基準にする場合のデザイン変数	
13.5. ポアソン回帰における最小 2 乗平均(Lsmeans) -----	457
文献索引, 索引, 解析用ファイル一覧 -----	461

第 9 回 続高橋セミナー 最尤法によるポアソン回帰分析入門

第 9 回 続高橋セミナー「最尤法によるポアソン回帰分析入門」は、ページ数が多いので章ごとに公開する。全体の章立てを次に示す。

目 次

はじめに -----	1
1. ポアソン分布に従う各種のカウント・データ-----	7
2. ニュートン・ラフソン法によるポアソン回帰 -----	63
3. 尤度比検定のためのデザイン行列-----	95
4. デザイン行列を用いた回帰分析入門-----	135
5. 反復重み付き最尤法によるポアソン回帰 -----	175
6. 過分散・ゼロ過剰への対応 -----	207
7. 過分散がある場合の探索的ポアソン解析 -----	237
8. 2 本の回帰直線の比較-----	269
9. 花数を共変量とした種子数の探索的ポアソン回帰-----	293
10. オフセットを含む探索的ポアソン回帰-----	321
11. デビアンス・逸脱度・残差・テコ比-----	359
12. パラメータの共分散行列の活用 -----	383
13. 最小 2 乗平均の謎を予測プロファイルで解く -----	421
文献, 文献索引, 索引, 解析用ファイル一覧 -----	461

13. 最小 2 乗平均の謎を予測プロファイルで解く

最小 2 乗法は、統計的用語として広く使われているが、「最小 2 乗平均」は、竹内ら (1989)、「統計学辞典」の索引にも載っていない方言みたいなものである。「最小 2 乗平均」は、SAS ユーザであれば、Lsmeans で出力される「調整平均」と広く認識されており、SAS を使った解析結果には、方言であることの認識なく広く使われている。高橋・大橋・芳賀 (1989)、「SAS による実験データの解析」の第 15 章では、最小 2 乗平均について丁寧に説明している。残念ながら、SAS ユーザの方言のまま現在に至っている。JMP では、「予測プロファイル」が新たな方言として加わっており、探索的な解析結果を可視化するために斬新な統計的な方法であることをこれまで示してきた。「最小 2 乗平均」は、「予測プロファイル」に包含される統計量の一種であり、「予測プロファイル」は、探索的データ解析に際し、データの内部構造を描き出すことに優れていることを示してきた。

13.1. 最小 2 乗平均 (Lsmeans) とは

予測プロファイルについては、第 7.2 節「カブトガニのサテライト数に対する探索的解析」で、メインの解析方法として例示してきた。この事例は、ポアソン回帰での適用であるが、予測プロファイルの算出方法は、共分散分析、重回帰分析などの場合も同様に適用できる汎用的な方法である。特に、質的因子と量的因子が混在し、交互作用が無視できないような探索的な解析に際し、結果の考察に役に立つ。

JMP の「予測プロファイル」は、「最小 2 乗平均」の概念を拡張し、GUI を兼ね備えた動的なグラフ表示となっていて、JMP ユーザにとって身近な存在である。ただ、どのような計算方法なのか、多くのユーザにとってブラック・ボックス的な存在でもある。使い勝手の良い方法であっても、その統計的な性質が適切に示され Excel などでも計算が誰にでも再現できなければ、しよせん「方言」の域を出ない。第 7.2 節では、JMP で作成した予測プロファイルを Excel により JMP と同様のグラフとして再現した。その際に必要なのは、パラメータの共分散行列であり、その活用方法について繰り返し言及してきた。

重回帰分析では、多重共線性の把握にパラメータの相関行列が広く活用されているが、パラメータの共分散行列は軽視されている。たとえば、JMP の「モデルのあてはめ」においてデフォルトの最小 2 乗法による解析を選択した場合には、パラメータの相関行列は得るこ

とができるが、パラメータの共分散行列を得ることができない。これは、多くの統計の教科書に「パラメータの共分散行列」の活用事例が見いだされないことの反映でもある。

JMP の「予測プロファイル」は、最小 2 乗平均の考え方を拡張し、その結果を GUI による探索的な要素を持った動的なグラフとして提供している。これにより、従来は、パラメータの推定値ベースによる隔靴搔痒的な結果の解釈に対し、直接的に複雑なデータの構造を切り出し、データの内部構造の可視化に貢献している。

とはいえ、多くの JMP ユーザにとって「予測プロファイル」が、ブラック・ボックス的であれば、しょせん方言のままとなる。幸い、現代の算盤である Excel を用いて自己完結的に「予測プロファイル」を作成できることを示してきた。ただし、主体はポアソン回帰についてであり、通常の回帰については断片的な取り扱いであった。

そこで、豊富で魅力的な事例に富んだ奥野ら（1981）で取り上げられている重回帰分析の事例に着目した。第 12.3 節では、2 変量の重回帰の事例として「材料、工数と生産量の関係」について奥野ら（1981）で用いられている「偏差平方和ベース」に対し「デザイン行列ベース」の解析アプローチの違いを比較検討するために用いた。さらに、奥野ら（1981）から共変量が 2 変数の共分散分析の事例、守屋ら（2018）の繰り返しが不揃いな 2 因子の共分散分析のデータについて最小 2 乗平均（Lsmeans）の謎を Excel によって解き明かす。

13.2. 交互作用を考慮した共分散分析

奥野ら(1981)の事例7.1は、4水準の層別因子を含む回帰分析の事例である。表13.1に示すように、亜硫酸ガス回収塔の洗浄水の温度 x と、回収液の濃度 y との関係を調べたい。しかし、測定回数が少ないので、長期間にわたってとられたデータを使わなければならない。その間に季節の変化ばかりでなく、操業条件も変わっているため、データを4つ季節に分けてみることにした。

共分散分析の拡張

狭義の意味での共分散分析は、1元配置型の実験で、制御することはできないが、結果に影響を与える量的変数を解析モデルに共変量として含める解析方法である。伝統的な解析法は、1元配置型の分散分析に回帰分析を複合的に付け加えており、一般的には難解な方法として認識されている。

奥野ら(1981)のタイトルは、「層別因子を含む回帰分析」であり、回帰分析を主体するが結果に影響がある質的因子（層別因子）を付加的に解析モデルに組み込もうとしている。層別因子として操業月を4区分の季節としている。解析モデル的には、共分散分析と同じであるが、狭義の意味での共分散分析ではなく、回帰分析を主体にした探索的な解析である。

表 13.1 季節ごとの洗浄水の温度 x と回収液の濃度 y [奥野(1981), 表 7.2]

No.	A ₁ (7月・8月)		A ₂ (9月・10月)		A ₃ (11月・12月)		A ₄ (1月・2月)	
	x	y	x	y	x	y	x	y
1	30	8.0	34	9.1	27	15.0	16	23.4
2	28	10.5	19	19.4	20	21.0	16	28.2
3	30	8.2	22	20.5	14	24.2	20	29.5
4	29	13.0	25	14.2	18	15.3	16	22.2
5	28	10.1	22	11.0	12	17.3	6	40.2
6	17	16.2	25	19.1	18	24.7	16	36.6
7	24	16.1	22	16.0	13	23.9	12	35.5
8	22	13.4	23	11.1	21	20.2	11	42.1
9	30	13.0	28	12.0	24	13.6	9	42.5
10	30	7.3	24	17.1			22	25.8
11	25	14.0	30	13.3				
12			24	8.9				
平均	26.6364	11.8000	24.8333	14.3083	18.5556	19.4667	14.4000	32.6000
						総平均	21.4762	19.1119

データのグラフ表示

得られたデータの関連を概観するために、洗浄水の温度 x と回収液の濃度 y の散布図上に季節別の回帰直線を上書きした結果を図13.1に示す。季節が A₄ (1月・2月) の場合に多重

比較などをするまでもなく明らかに他の季節に比べ洗浄水の温度が低い方に分布し、他の季節と共通の洗浄水の温度 x が 20 度で比べた場合に、回収液の濃度 y が高めになっている。季節に関連する操作条件などの他の因子の影響があるのかも知れない。

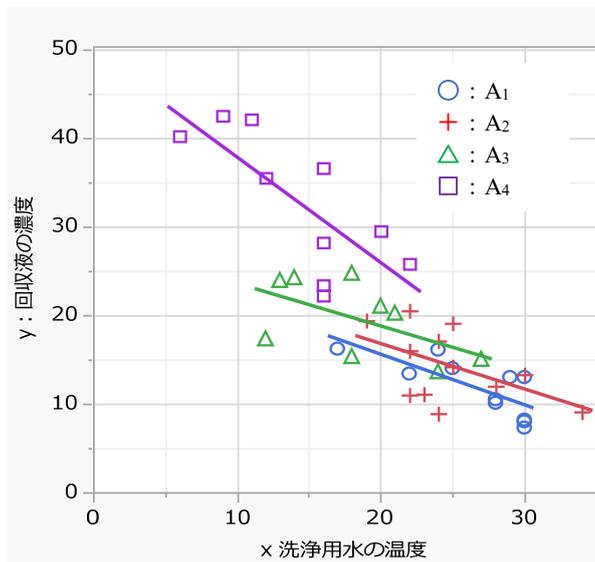


図 13.1 JMP のグラフ・ビルダーによる 4 本の回帰直線のあてはめ

伝統的な共分散分析の考え方

奥野ら(1981)には、共分散分析における伝統的な交互作用の解析方法が示されている。生物統計の名著であるスネデガー・コ克蘭(1972)でも、医療統計の名著であるアーミテージら(2001)でも、同様の解析方法が示されていることもあり、解析法として定着している。

1 元配置モデルでの解析に際し、反応 y に影響を与えることがはっきりしている変数があるが、実験に際し一定に保つことがどうしてもできない。この変数を「共変量」として、共変量の影響を除去して、因子 A の水準間の比較をしたいとの要望に対処するための方法として共分散分析が定着している。そのために、共変量の影響が均一であることが条件であり、均一でなければ共分散分析が適用できないとの共通認識となっている。

この節で取り上げる観察データに対する「層別因子を含む回帰分析」は、データの構造および解析の考え方は、狭義の意味での「共分散分析」と全く同一である。ただし、後付けで設定した層ごとの共変量 x が同一の範囲に入り、反応 y に対する回帰直線の傾きが層ごとに共通であるとは限らない。季節ごとの共変量 x の平均値には、統計的方法を用いなくても散布図から明らかと言い切ったのであるが、 p 値による判断を好む人たちもいるので、共変量 x について季節で 1 元配置分散分析をすると $F=16.2716$, $p=0.00000058$ ($df_1=3$, $df_2=38$) となる。し

たがって、「交互作用を考慮した共分散分析」というのではなく、「質的変数と量的変数を含む重回帰分析における交互作用解析」というべきであろう。

質的変数と量的変数を含む重回帰分析における交互作用解析

あえて「交互作用を考慮した共分散分析」としたのは、狭義の意味での共分散分析の解析の考え方が普及していて、交互作用があったら共分散分析ができないとの頑なな迷信が蔓延している。これら事柄については、高橋 (2019)、「投与前値がある場合の解析のレビュー」に歴史的な経緯が示されている。このために、共分散分析と同様のデータ構造をもつ観察データの探索的な解析に際しても「迷信」とらわれてしまいがちになる。そこで、あえて、狭義の共分散分析の適用がためられる事例について、交互作用が否定できない場合の探索的な解析方法を示すことにした。交互作用の探索的な解析には、最小 2 乗平均および予測プロファイルによるアプローチが必要不可欠であることも示す。

伝統的な共分散分析の解析方法での交互作用の検出方法は、因子 A の水準ごとに反応 y に対する共変量 x を説明変数とする単回帰分析を行い、偏差平方和 ($S_{xy}^{(1)}, S_{xy}^{(2)}, \dots, S_{xy}^{(k)}$), ($S_{xx}^{(1)}, S_{xx}^{(2)}, \dots, S_{xx}^{(k)}$) を求めて、共通の切片 $\hat{\beta}_1^{(\text{共通})}$ として、

$$\hat{\beta}_1^{(\text{共通})} = \frac{\sum_{i=1}^k S_{xy}^{(k)}}{\sum_{i=1}^k S_{xx}^{(k)}}$$

を計算し、別々の直線をあてはめた場合の平方和、共通の傾きを持つ回帰直線をあてはめた場合の誤差平方和を計算し、それらの誤差平方和の差から、交互作用の検討を行っている。

この方法は、コンピュータによる重回帰分析が手軽にできない時代の計算手順であり、「共分散分析」というと、この伝統的な解析手順にとらわれてしまいがちになる。正確には、質的変数と量的変数、さらにそれらの交互作用を含むデザイン行列を活用した探索的な重回帰分析と認識すべきある。反応がカウント・データデータであれば、ポアソン回帰を使った解析となり、2 値反応ならば、ロジスティック回帰を使った解析となる。第 3.5 節の「2 本の回帰直線に対する各種のデザイン行列」は、ポアソン回帰での結果を示しているが、重回帰分析の場合でも、まったく共通する考え方である。

統計ソフト JMP を用いた共分散分析

JMP の最小 2 乗法による回帰分析を用い、層別因子 A と洗浄水の温度 x との交互作用 $A \times x$ を含む重回帰分析を適用し、予測プロファイルを用いて比較検討する。解析モデルは、4 水準の季節 A、洗浄水の温度 x 、それらの交互作用 $A \times x$ を含める。表 13.2 に示すような分散分析表が得られる。

表 13.2 季節 A と洗浄水の温度 x の交互作用を含む分散分析表

分散分析					
要因	自由度	平方和	平均平方	F値	
モデル	7	3145.7754	449.3965	30.0158	
誤差	34	509.0487	14.9720		p値(Prob>F)
全体(修正済み)	41	3654.8240			<.0001*
効果の検定					
要因	自由度	平方和	平均平方	F値	p値(Prob>F)
A	3	182.8749	60.9583	4.0715	0.0142*
x	1	369.0901	369.0901	24.6520	<.0001*
A*x	3	71.2462	23.7487	1.5862	0.2107

交互作用 $A \times x$ が有意ではないので、解析モデルから除いて、季節 A に共通の傾きをあてはめたい。だが、季節により洗浄水の温度 x の範囲が明らかに異なるので、ためらわれる。このような観察データでは、狭義の意味での共分散分析では起こりえないような状況が発生する。したがって、交互作用を含めた“共分散分析”とし、季節に共通の洗浄水の温度である 20 度での季節間の回収液の温度 y の推定値の比較を試みよう。

予測プロファイル

JMP の予測プロファイルを用い、図 13.2 に示すように、洗浄水の温度 x を 20 度に固定し回収液の濃度 y を季節 A₄ (1 月・2 月) と季節 A₃ (11 月・12 月) の推定値と 95%信頼区間を求める。季節 A₄ の回収液の濃度の推定値 \hat{y} は、25.9338 で 95%信頼区間は (22.0423, 29.8254) 季節 A₃ の場合は、18.7719 (16.0328, 21.5109) であり、95%信頼区間が互いに重ならないので有意な差があると判断される。

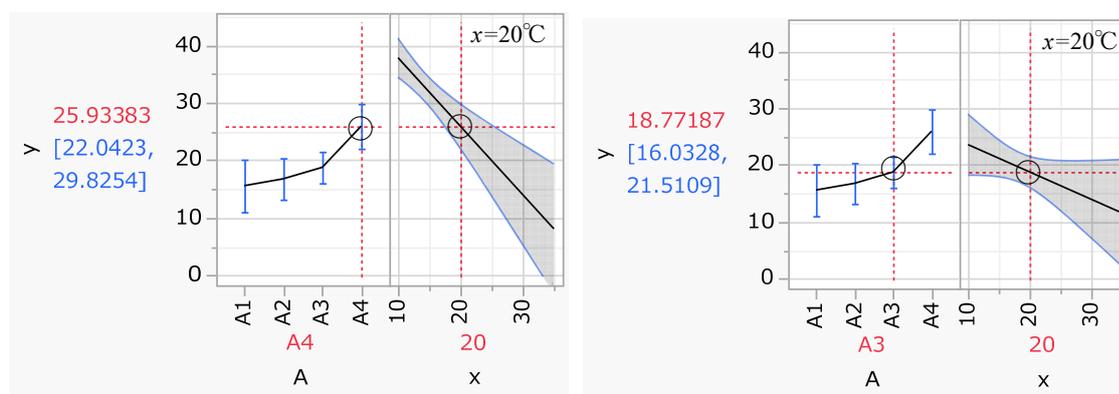


図 13.2 A₄および A₃とした場合の予測プロファイル

予測プロファイルには、季節を A₄ (1月・2月) とした場合、洗浄水の温度 x に対する回収液の濃度 y について回帰直線および 95%信頼区間が示されている。マウスで「季節 A₃」を選択すると図 13.2 右の傾きが異なる回帰直線が示される。これは、解析モデルに交互作用を含めたからである。洗浄水の温度 x を 20 度に設定してあるが、デフォルトでは、総平均の 21.4762 度である。他の温度を自由に設定することができ、図 13.3 に示すように温度の設定を 10 度と 30 度に設定すると、交互作用を含めているために季節ごとの水準平均と 95%信頼区間ががらりと変わることが示されている。

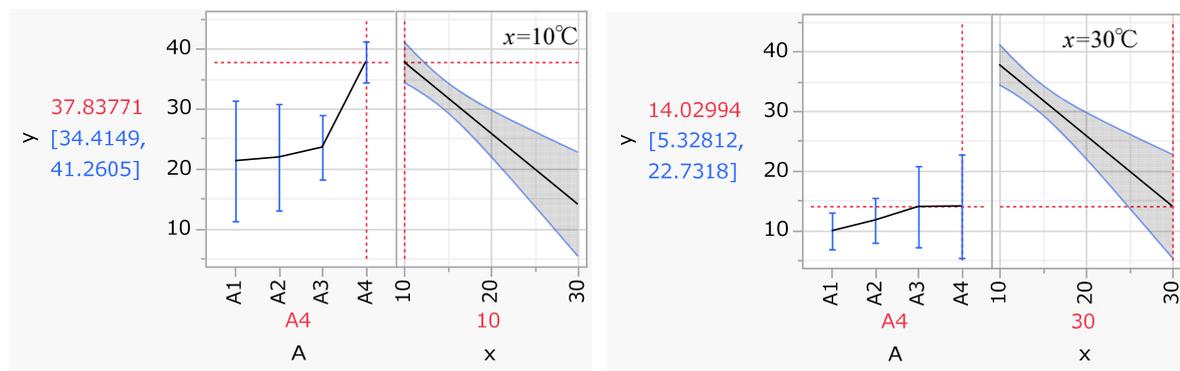


図 13.3 洗浄水の温度を変えた場合の季節 A の各水準の推定値と 95%信頼区間

対比による水準間の差の推定

JMP には、対比の機能を使って任意の水準間の差について統計量を算出する機能が備わっている。表 13.3 に示すように、洗浄水の温度を 20 度に固定し、季節 A₄ を -1 季節 A₃ を +1 とする対比の指定により、図 13.2 に示した A₃ の推定値の 18.7719 と A₄ の推定値 25.9338 との差が -7.1620 と定され、その標準誤差が 2.3417 と推定されている。95%信頼区間の出力はな

表 13.3 洗浄水の温度 20 度の場合の季節 A₄ と季節 A₃ の差の統計量

対比		検定の詳細		パラメータ関数	
対比の指定		A1	0.0000	パラメータ	
A 温度を 20°C に設定		A2	0.0000	切片	0
A1	0	A3	1.0000	A[A1]	1
A2	0	A4	-1.0000	A[A2]	1
A3	1	推定値	-7.1620	A[A3]	2
A4	-1	標準誤差	2.3417	x	0
+または-をクリックして対比值を作成。		t値	-3.0585	x*A[A1]	20
		p値(Prob> t)	0.0043	x*A[A2]	20
		平方和	140.053	x*A[A3]	40
		平方和	140.053	分子自由度	1.0000
				分母自由度	34.0000
				F値	9.3543
				p値(Prob>F)	0.0043
		x	20		

いが、 $t=3.0585$ と有意な差であることが分かる。JMP での対比の設定は、直観的で季節 ($+A_3-A_4$) となっているが、JMP の内部での設定は、表 13.3 右の「パラメータ関数」で示すように ($+A_1+A_2+2A_3$) が使われている。これは、JMP が対比型のデザイン変数をデフォルトにしていることに関係している。表 13.4 に示す対比型デザイン変数から明らかなように、($+A_3-A_4$) は、内部では ($+a_1+a_2+2a_3$) と置き換えられている。

表 13.4 対比型デザイン変数 ($+A_3-A_4$)

A	a_1	a_2	a_3
A ₁	1	0	0
A ₂	0	1	0
A ₃	0	0	1
A ₄	-1	-1	-1
A ₃ -A ₄	1	1	2

パラメータ関数	
パラメータ	
切片	0
A[A1]	1
A[A2]	1
A[A3]	2

なお、SAS/GLM プロシジャなどの場合は、Estimate ステートメントあるいは Contrast ステートメントで同様の推定を行うことができる。一般化線形モデルに対する SAS/GENMOD プロシジャでも同様の推定ができる。

Excel による交互作用を含む解析

JMP の予測プロファイルは、探索的な解析のための有用なツールであることを多くの事例で示したきたが、さらに、交互作用を含む事例について Excel で予測プロファイルを再現することにより、さらに理解を深めたい。交互作用を含むので表 13.6 に示すようにデザイン行列 X は、 (42×8) と大きくなるが、これまでと同様に Excel の分析ツールの回帰分析を適用することができる。

季節 A は 4 水準なので、表 13.5 に示すように対比型デザイン変数 (a_1, a_2, a_3) を与える。洗浄水の温度 x との交互作用は、(a_1x, a_2x, a_3x) のように積で与える。

表 13.5 対比型デザインでの交互作用に対するデザイン変数

A	a_1	a_2	a_3
A ₁	1	0	0
A ₂	0	1	0
A ₃	0	0	1
A ₄	-1	-1	-1

A	a_1x	a_2x	a_3x
A ₁	x	0	0
A ₂	0	x	0
A ₃	0	0	x
A ₄	$-x$	$-x$	$-x$

表 13.6 に交互作用モデルに対するデザイン行列に対して、Excel の回帰分析による分散分析表およびパラメータの推定結果を示す。その下に、Excel の行列関数を用いて計算したパラメータの共分散行列の計算結果を示す。

表 13.6 交互作用モデルに対する Excel によるパラメータの共分散行列の推定

No.	A	— デザイン行列 X —								Excelによる回帰分析 ($x_0 \sim a_{3x}$ 8列)								
		x_0	a_1	a_2	a_3	x	a_1x	a_2x	a_3x	y	分散分析表		「定数に0を使用」を on とする					
1	A ₁	1	1	0	0	30	30	0	0	8.0	自由度	変動	分散					
2		1	1	0	0	28	28	0	0	10.5	回帰	8	18487	2310.9				
3		1	1	0	0	30	30	0	0	8.2	残差	34	509.0	14.972 σ^2				
4		1	1	0	0	29	29	0	0	13.0	合計	42	18996					
5		1	1	0	0	28	28	0	0	10.1	係数	SE	t	P-値				
6		1	1	0	0	17	17	0	0	16.2	切片	0	#N/A	#N/A	#N/A			
7		1	1	0	0	24	24	0	0	16.1	X 値 1	33.04	3.11	10.61	0.00 β_0 x_0			
8		1	1	0	0	22	22	0	0	13.4	X 値 2	-6.03	6.33	-0.95	0.35 β_1 a_1			
9		1	1	0	0	30	30	0	0	13.0	X 値 3	-6.02	5.95	-1.01	0.32 β_2 a_2			
10		1	1	0	0	30	30	0	0	7.3	X 値 4	-4.65	4.81	-0.97	0.34 β_3 a_3			
11		1	1	0	0	25	25	0	0	14.0	X 値 5	-0.69	0.14	-4.97	0.00 β_4 x			
12	A ₂	1	0	1	0	34	0	34	0	9.1	X 値 6	0.12	0.25	0.47	0.64 β_5 a_1x			
13		1	0	1	0	19	0	19	0	19.4	X 値 7	0.18	0.24	0.72	0.48 β_6 a_2x			
14		1	0	1	0	22	0	22	0	20.5	X 値 8	0.21	0.24	0.88	0.39 β_7 a_3x			
15		1	0	1	0	25	0	25	0	14.2								
16		1	0	1	0	22	0	22	0	11.0	パラメータの共分散行列 $\Sigma(\hat{\beta}) = (X^T X)^{-1} \sigma^2$							
17		1	0	1	0	25	0	25	0	19.1	x_0	a_1	a_2	a_3	x	a_1x	a_2x	a_3x
18		1	0	1	0	22	0	22	0	16.0	9.70	5.52	3.18	-2.97	-0.41	-0.14	-0.09	0.07
19		1	0	1	0	23	0	23	0	11.1	5.52	40.12	-18.40	-12.25	-0.14	-1.53	0.65	0.49
20		1	0	1	0	28	0	28	0	12.0	3.18	-18.40	35.45	-9.91	-0.09	0.65	-1.43	0.43
21		1	0	1	0	24	0	24	0	17.1	-2.97	-12.25	-9.91	23.15	0.07	0.49	0.43	-1.09
22		1	0	1	0	30	0	30	0	13.3	-0.41	-0.14	-0.09	0.07	0.02	0.00	0.00	0.00
23		1	0	1	0	24	0	24	0	8.9	-0.14	-1.53	0.65	0.49	0.00	0.06	-0.02	-0.02
24	A ₃	1	0	0	1	27	0	0	27	15.0	-0.09	0.65	-1.43	0.43	0.00	-0.02	0.06	-0.02
25		1	0	0	1	20	0	0	20	21.0	0.07	0.49	0.43	-1.09	0.00	-0.02	-0.02	0.06
26		1	0	0	1	14	0	0	14	24.2								
27		1	0	0	1	18	0	0	18	15.3	対比型デザイン変数							
28		1	0	0	1	12	0	0	12	17.3	A	a_1	a_2	a_3				
29		1	0	0	1	18	0	0	18	24.7	A ₁	1	0	0				
30		1	0	0	1	13	0	0	13	23.9	A ₂	0	1	0				
31		1	0	0	1	21	0	0	21	20.2	A ₃	0	0	1				
32		1	0	0	1	24	0	0	24	13.6	A ₄	-1	-1	-1				
33	A ₄	1	-1	-1	-1	16	-16	-16	-16	23.4	交互作用のデザイン変数							
34		1	-1	-1	-1	16	-16	-16	-16	28.2	A	a_1x	a_2x	a_3x				
35		1	-1	-1	-1	20	-20	-20	-20	29.5	A ₁	x	0	0				
36		1	-1	-1	-1	16	-16	-16	-16	22.2	A ₂	0	x	0				
37		1	-1	-1	-1	6	-6	-6	-6	40.2	A ₃	0	0	x				
38		1	-1	-1	-1	16	-16	-16	-16	36.6	A ₄	-x	-x	-x				
39		1	-1	-1	-1	12	-12	-12	-12	35.5								
40		1	-1	-1	-1	11	-11	-11	-11	42.1								
41		1	-1	-1	-1	9	-9	-9	-9	42.5	$t(0.05, 34) = $ 2.0322							
42		1	-1	-1	-1	22	-22	-22	-22	25.8	=T.Inv.2T(0.05, 34)							
総平均						21.4762				19.1119								

パラメータに関する共分散行列 $\Sigma(\hat{\beta})$ は、デザイン行列と分散分析表の誤差分散=14.972 を用いて

$$\Sigma(\hat{\beta}) = (X^T X)^{-1} \sigma^2$$

$$= \text{Minverse}(\text{Mmult}(\text{Transpose}(X\text{の範囲}), X\text{の範囲})) * 14.972$$

$$=$$

x_0	a_1	a_2	a_3	x	a_1x	a_2x	a_3x
9.70	5.52	3.18	-2.97	-0.41	-0.14	-0.09	0.07
5.52	40.12	-18.40	-12.25	-0.14	-1.53	0.65	0.49
3.18	-18.40	35.45	-9.91	-0.09	0.65	-1.43	0.43
-2.97	-12.25	-9.91	23.15	0.07	0.49	0.43	-1.09
-0.41	-0.14	-0.09	0.07	0.02	0.00	0.00	0.00
-0.14	-1.53	0.65	0.49	0.00	0.06	-0.02	-0.02
-0.09	0.65	-1.43	0.43	0.00	-0.02	0.06	-0.02
0.07	0.49	0.43	-1.09	0.00	-0.02	-0.02	0.06

のように計算されている。なお、これら行列計算の詳細は、第4章を参照のこと。

4本の回帰直線の推定

季節 A₁ の回帰直線は、対比型のデザインであることから、次のように求めることができる。図 13.1 を参照して、推定結果が適切であることが確認できる。

$$A_1 : \begin{cases} y_{A1} = (\hat{\beta}_0 + \hat{\beta}_1) \times 1 + (\hat{\beta}_4 + \hat{\beta}_5)x \\ = 33.04 - 6.03 + (-0.69 + 0.12)x \\ = 27.02 - 0.57x \end{cases}$$

$$A_2 : \begin{cases} y_{A2} = (\hat{\beta}_0 + \hat{\beta}_2) \times 1 + (\hat{\beta}_4 + \hat{\beta}_6)x \\ = 27.03 - 0.51x \end{cases}$$

$$A_3 : \begin{cases} y_{A3} = (\hat{\beta}_0 + \hat{\beta}_3) \times 1 + (\hat{\beta}_4 + \hat{\beta}_7)x \\ = 28.39 - 0.48x \end{cases}$$

$$A_4 : \begin{cases} y_{A4} = [\hat{\beta}_0 - (\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3)] \times 1 + [\hat{\beta}_4 - (\hat{\beta}_5 + \hat{\beta}_6 + \hat{\beta}_7)]x \\ = 27.02 - (-16.70) + (-0.69 - 0.50)x \\ = 49.74 - 1.19x \end{cases}$$

分散分析表

表 13.6 で示した Excel での分散分析表は、切片を含むデザイン変数すべてを用いたので、「回帰」の自由度が 8 で（回帰の変動=18487）となり、表 13.2 に示した JMP の結果（モデルの平方和=3145.7754）とは異なるが、以後の計算に必要な（残差の分散=14.972）は、JMP の（誤差の平均平方=14.9720）と同じである。Excel による回帰分析では、因子 A、温度 x、交互作用 A×x に関する平方和の分解がないので、別途対応する必要がある。

交互作用 A×x の効果は、表 13.7(右) に示すように交互作用を含まない主効果モデルでの分散分析表を作成し（S_{A+x}：回帰の変動=3047.5291）を得る。表 13.7(左) に示すように交互

作用を含めたモデルで ($S_{A+x+(A \times x)}$: 回帰の変動=3145.7754) を得る. 交互作用 $S_{(A \times x)}$ の変動 (平方和) は, それらの差

$$\begin{aligned} S_{(A \times x)} &= S_{A+x+(A \times x)} - S_{A+x} \\ &= 3145.7754 - 3074.5291 \\ &= 71.2462 \end{aligned}$$

となり, 表 13.2 に示した JMP での結果 71.2462 が得られる.

表 13.7 交互作用モデル vs 主効果モデル

交互作用モデル: $y = A + x + (A \times x)$				主効果モデル: $y = A + x$			
	自由度	変動	分散		自由度	変動	分散
回帰	7	3145.7754	449.3965	回帰	4	3074.5291	768.6323
残差	34	509.0487	14.9720	残差	37	580.2949	15.6836
合計	41	3654.8240		合計	41	3654.8240	

$a_1 \ a_2 \ a_3 \ x \ a_1x \ a_2x \ a_3x$ を選択 $a_1 \ a_2 \ a_3 \ x$ を選択
 「定数に 0 を使用」を off に設定

表 13.8 に示すように, 主効果 A および x の平方和は, Type III の平方和と言われて, 因子 A ごとの x の平均が等しいと仮定するような平方和で, Excel の回帰分析では計算することができない. 表 13.2 の交互作用を含む場合に対して, 因子 A の平方和は 128.8749 → 551.4944 と大幅に増加し, 因子 x の平方和は 369.0901 → 389.1143 と微妙に異なる. 表 13.8 に示した平方和は, Type II の平方和と言われていて, Excel の回帰分析で求めることができる.

表 13.8 JMP での主効果モデルでの効果の検定

分散分析					
要因	自由度	平方和	平均平方	F値	p値(Prob>F)
モデル	4	3074.5291	768.6323	49.0085	
誤差	37	580.2949	15.6836		p値(Prob>F)
全体(修正済み)	41	3654.8240			<.0001*

効果の検定					
要因	自由度	平方和	平均平方	F値	p値(Prob>F)
A	3	551.4944	183.8315	11.7212	<.0001*
x	1	389.1143	389.1143	24.8102	<.0001*

表 13.9 に示すよう ($y = A+x$) モデルでの平方和は $S_{A+x} = 3074.5291$, 因子 A のみの ($y = A$) モデルでの平方和は $S_A = 2685.4149$ であり, この差

$$\begin{aligned} S_x &= S_{A+x} - S_A \\ &= 3074.5291 - 2685.4149 = 389.1143 \end{aligned}$$

が, 因子 A の存在下で因子 x を加えた場合の平方和の増分となっている. 同様に, 因子 x の

みの $(y=x)$ モデルでの平方和は $S_x = 2523.0347$ であり、この差

$$S_A = S_{A+x} - S_x = 3074.5291 - 2523.0347 = 551.4944$$

が、因子 x の存在下で因子 A を加えた場合の平方和の増分となっている。

表 13.9 3種の主効果モデル

主効果モデル: $y = A + x$			主効果モデル: $y = A$			主効果モデル: $y = x$		
	自由度	変動		自由度	変動		自由度	変動
回帰	4	3074.5291	回帰	3	2685.4149	回帰	1	2523.0347
残差	37	580.2949	残差	38	969.4092	残差	40	1131.7893
合計	41	3654.8240	合計	41	3654.8240	合計	41	3654.8240

$a_1 \ a_2 \ a_3 \ x$ を選択
 $a_1 \ a_2 \ a_3$ を選択
 x を選択

交互作用がモデルに含まれている場合の主効果の平方和の計算方法には、幾つかの考え方があり、そもそも、交互作用がある場合の主効果の検定にどんな意味があるのかとも言われている。Excel で追試ができる範囲の Type II の平方和を基準にして対応することを勧める。これらの平方和の扱いについては、芳賀 (2009), 「医薬品開発のための統計解析 第2部 実験計画法 初版」の第4章「共分散分析」に詳しく説明されている。また、同書について「じっくり勉強すれば身につく統計入門」シリーズ第3回で、杉本・橋田 (2011) が「共分散分析の基礎・医薬品開発における共分散分析の例」で丁寧に解析しているので参照してもらいたい。

Excel による予測プロファイル

JMP の予測プロファイルを用いた推定を Excel で行う。洗浄水の温度 x を季節に共通の 20 度に固定し回収液の濃度の推定値 \hat{y}_i と分散 $Var(\hat{y}_i)$ を推定する。表 13.10 に示すように、デザイン行列の変数をセットし、

$$\hat{y}_i = \mathbf{x}_i \hat{\boldsymbol{\beta}}$$

$$Var(\hat{y}_i) = \mathbf{x}_i \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) \mathbf{x}_i^T$$

$$(L95\%, U95\%) = \hat{y}_i \pm t(0.05, 34) \sqrt{Var(\hat{y}_i)}$$

による計算を行う。季節 A_1 の場合は、

	\mathbf{x}_1								$\hat{\boldsymbol{\beta}}$	
$y_{A_1}^{\wedge} =$	1	1	0	0	20	20	0	0	33.0450	= 15.5912
									-6.0281	
									-6.0157	
									-4.6528	
									-0.6887	
									0.1175	
									0.1765	
									0.2077	

$Var(\hat{y}_{A_1})=$		$\Sigma(\hat{\beta})$								x_1^T								
1	1	0	0	20	20	0	0	9.70	5.52	3.18	-2.97	-0.41	-0.14	-0.09	0.07	1	=	5.0542
								5.52	40.12	-18.40	-12.25	-0.14	-1.53	0.65	0.49	1		
								3.18	-18.40	35.45	-9.91	-0.09	0.65	-1.43	0.43	0		
								-2.97	-12.25	-9.91	23.15	0.07	0.49	0.43	-1.09	0		
								-0.41	-0.14	-0.09	0.07	0.02	0.00	0.00	0.00	20		
								-0.14	-1.53	0.65	0.49	0.00	0.06	-0.02	-0.02	20		
								-0.09	0.65	-1.43	0.43	0.00	-0.02	0.06	-0.02	0		
								0.07	0.49	0.43	-1.09	0.00	-0.02	-0.02	0.06	0		

$$\begin{aligned}
 (L95\%, U95\%) &= \hat{y}_1 \pm t(0.05, 34) \sqrt{Var(\hat{y}_1)} \\
 &= 15.5912 \pm 2.0322 \sqrt{5.0542} \\
 &= (11.0224, 20.1600)
 \end{aligned}$$

が得られる。他の季節に対しても同様に計算する。季節 A₄ の推定値 \hat{y}_{A_4} は、25.9338 で 95%信頼区間は (22.0423, 29.8254) であり、季節 A₃ の場合は、18.7719 (16.0328, 21.5109) となり図 13.2 に示した予測プロファイルの推定値に一致する。

表 13.10 交互作用を考慮した季節別の推定値と 95%信頼区間

No.	A	x_0	a_1	a_2	a_3	x	a_1x	a_2x	a_3x	y^\wedge	$Var(y^\wedge)$	SE	L95%	U95%
1	A ₁	1	1	0	0	20	20	0	0	15.5912	5.0542	2.2482	11.0224	20.1600
2	A ₂	1	0	1	0	20	0	20	0	16.7842	3.1520	1.7754	13.1762	20.3922
3	A ₃	1	0	0	1	20	0	0	20	18.7719	1.8165	1.3478	16.0328	21.5109
4	A ₄	1	-1	-1	-1	20	-20	-20	-20	25.9338	3.6669	1.9149	22.0423	29.8254

この結果を Excel の「折れ線」グラフで作成した結果を図 13.4 に示す。Excel での信頼区間の幅を付ける際には、推定値からの長さでの設定になっているので $SE \times t(0.05, df)$ による髭の長さを別途計算する必要がある。

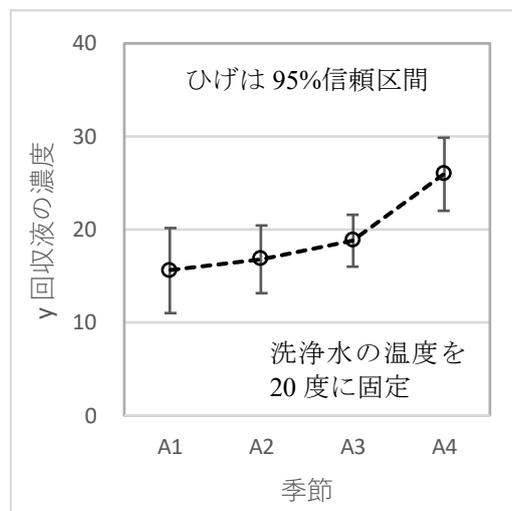


図 13.4 Excel の「折れ線」による予測プロファイル

Excel での「折れ線グラフ」の髭の長さの設定は、かなり深いところにあり、

グラフの要素 → 誤差範囲 → その他のオプション → ユーザの設定 →
 値の設定 → 正の誤差の値（範囲） → 負の誤差の値（範囲）

のように設定する。

水準間の差の予測プロファイル

回収液の濃度が最も高い季節 A_4 を基準とし、温度が 20 度の場合の他の季節との差の推定値と 95%信頼区間を求めたい。季節 A_3 と A_4 の差の推定値は、表 13.11 に示すように、デザイン変数の差を計算する。推定値は、

$$\begin{aligned}\hat{y}_{(A_3-A_4)} &= \hat{y}_{A_3} - \hat{y}_{A_4} \\ &= 18.7719 - 25.9338 = -7.1620\end{aligned}$$

として求められる。この差に対する分散の推定のためのベクトル $\mathbf{d}_{(A_3-A_4)}$ は、季節 A_3 のベクトル \mathbf{x}_{A_3} と季節 A_4 のベクトル \mathbf{x}_{A_4} の差

$$\begin{aligned}\mathbf{x}_{A_3} &= [1 \quad 0 \quad 0 \quad 1 \quad 20 \quad 0 \quad 0 \quad 20] \\ \mathbf{x}_{A_4} &= [1 \quad -1 \quad -1 \quad -1 \quad 20 \quad -20 \quad -20 \quad -20] \\ \mathbf{d}_{(A_3-A_4)} &= \mathbf{x}_{A_3} - \mathbf{x}_{A_4} = [0 \quad 1 \quad 1 \quad 2 \quad 0 \quad 20 \quad 20 \quad 40]\end{aligned}$$

であり、 $\hat{y}_{(A_3-A_4)}$ の分散 $Var[\hat{y}_{(A_3-A_4)}]$ は、

$$\begin{aligned}Var[\hat{y}_{(A_3-A_4)}] &= \mathbf{d}_{(A_3-A_4)} \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) \mathbf{d}_{(A_3-A_4)}^T \\ &= 5.4834\end{aligned}$$

として求められ。この結果は、表 13.3 に示した「パラメータ関数」に示された結果に一致する。差 -7.1620 の 95%信頼区間が $(-11.9208, -2.4031)$ と推定されゼロを含まないので、有意な差であることが示されている。

表 13.11 2 水準間の差のデザイン変数と推定値

No.	A	x_0	a_1	a_2	a_3	x	a_1x	a_2x	a_3x	\hat{y}	$Var(\hat{y})$	SE	L95%	U95%
3	A_3	1	0	0	1	20	0	0	20	18.7719				
4	A_4	1	-1	-1	-1	20	-20	-20	-20	25.9338				
差	d	0	1	1	2	0	20	20	40	-7.1620	5.4834	2.3417	-11.9208	-2.4031

同様に季節 A_4 を基準とし、他の季節の差を表 13.12 に示す。差のデザイン行列については、季節 A_3 と季節 A_4 の場合と同様にする。差の分散についてもそれぞれの差のデザイン行列の場合と同様に推定し、95%信頼区間を計算する。JMP にも表 13.3 で示したように、ある温度設定に対し、因子 A の水準間の差の推定ができる。

表 13.12 季節 A₄ を基準とした 2 水準間の差の推定

No.	A	x_0	a_1	a_2	a_3	x	a_1x	a_2x	a_3x	y^\wedge	$Var(y^\wedge)$	SE	L95%	U95%
5	A ₁ -A ₄	0	2	1	1	0	40	20	20	-10.3426	8.7211	2.9532	-16.3441	-4.3411
6	A ₂ -A ₄	0	1	2	1	0	20	40	20	-9.1496	6.8189	2.6113	-14.4564	-3.8428
7	A ₃ -A ₄	0	1	1	2	0	20	20	40	-7.1620	5.4834	2.3417	-11.9208	-2.4031
8	A ₄ -A ₄	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0000

図 13.5 に季節 A₄ を基準とした差の推定値および 95% 信頼区間の予測プロフィールを示す。この図は、洗浄水の温度を 20 度に固定した場合であり、温度が低くなれば季節 A₄ との差は広がり、温度が高くなれば差は縮まる。

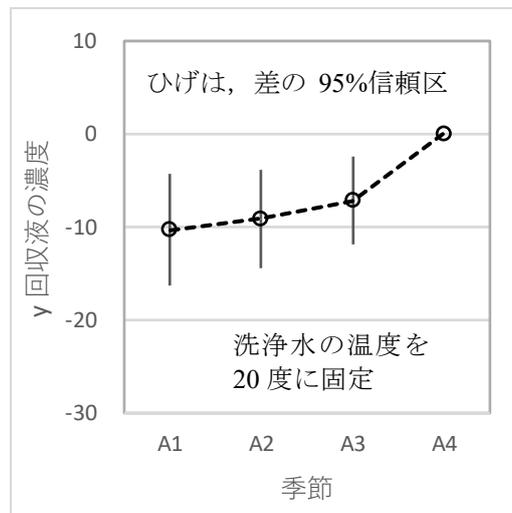


図 13.5 Excel の「折れ線」による差の予測プロフィール

洗浄水の温度に関する予測プロフィール

季節 A₄ と季節 A₃ について洗浄水の温度を変化させた場合の予測プロフィールを作成する。表 13.13 に示すように、季節 A₄ のデザイン行列変数のベクトル $[-1 \ -1 \ -1]$ 、洗浄水の温度 x を 5 度から 25 度まで変化させ、さらに $[-1 \ -1 \ -1]$ と x の交互作用を加えたベクトルとする。

季節 A₃ のベクトル $[0 \ 0 \ 1]$ に対しても同様に、洗浄水の温度 x を 5 度から 25 度まで変化させ、さらに $[0 \ 0 \ 1]$ と x の交互作用を加えたベクトルを生成する。これらのベクトルに対し、パラメータの推定値 $\hat{\beta}$ およびパラメータの共分散行列 $\Sigma(\hat{\beta})$ を用いて予測値 \hat{y} と分散を推定し、95% 信頼区間を求める。

表 13.13 洗浄水の温度を変えた場合の回収液の濃度

No.	A	x_0	a_1	a_2	a_3	x	a_1x	a_2x	a_3x	y^\wedge	$Var(y^\wedge)$	SE	$L95\%$	$U95\%$
9	A ₄	1	-1	-1	-1	5	-5	-5	-5	43.7896	7.6105	2.7587	38.1833	49.3960
10	A ₄	1	-1	-1	-1	10	-10	-10	-10	37.8377	2.8367	1.6842	34.4149	41.2605
11	A ₄	1	-1	-1	-1	15	-15	-15	-15	31.8858	1.5221	1.2337	29.3785	34.3930
12	A ₄	1	-1	-1	-1	20	-20	-20	-20	25.9338	3.6669	1.9149	22.0423	29.8254
13	A ₄	1	-1	-1	-1	25	-25	-25	-25	19.9819	9.2710	3.0448	13.7940	26.1697
14	A ₃	1	0	0	1	10	0	0	10	23.5820	7.0298	2.6514	18.1937	28.9703
15	A ₃	1	0	0	1	15	0	0	15	21.1769	2.5904	1.6095	17.9061	24.4478
16	A ₃	1	0	0	1	20	0	0	20	18.7719	1.8165	1.3478	16.0328	21.5109
17	A ₃	1	0	0	1	25	0	0	25	16.3668	4.7083	2.1699	11.9571	20.7765
18	A ₃	1	0	0	1	30	0	0	30	13.9618	11.2657	3.3564	7.1407	20.7829

表 13.13 で計算された季節 A₄ と季節 A₃ の予測値と 95%信頼区間を図 13.6 に示すような予測プロファイルを作成する。季節 A₄ の洗浄水の温度 x が 10 度の場合、回収液の濃度 \hat{y} は、37.8377 と季節 A₃ の 23.5820 と 14.2557 の差がある。洗浄水の温度 x が 10 度の場合、差が 7.1620 と縮小するが、95%信頼区間が互いに重なっていないので統計的な差があると判断される。

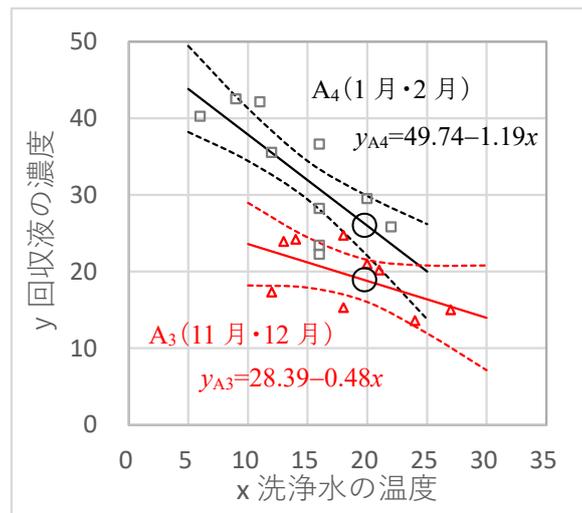


図 13.6 洗浄水の温度に対する季節 A₄ と季節 A₃ の回収液の濃度

回収液の濃度の差についての予測プロファイル

洗浄水の温度を変化させた場合に季節 A₄ と季節 A₃ の回収液の濃度の差の予測値と 95%信頼区間について定量的に検討したい。季節 A₃ のデザイン変数のベクトル \mathbf{x}_{A_3} と季節 A₄ のベクトル \mathbf{x}_{A_4} の差は、

$$\begin{aligned}
 \mathbf{x}_{A_3} &= [1 \quad 0 \quad 0 \quad 1 \quad x \quad 0 \quad 0 \quad x] \\
 \mathbf{x}_{A_4} &= [1 \quad -1 \quad -1 \quad -1 \quad x \quad -x \quad -x \quad -x] \\
 \mathbf{d}_{(A_3-A_4)} &= \mathbf{x}_{A_3} - \mathbf{x}_{A_4} = [0 \quad 1 \quad 1 \quad 2 \quad 0 \quad x \quad x \quad 2x]
 \end{aligned}$$

となるので、表 13.14 に洗浄水の温度 x_0 を変化させたときの予測値と 95%信頼区間を計算した結果を示す。図 13.7 には、差の予測値と 95%信頼区間を示したものである。95%信頼区間がゼロを横切るのは、おおよそ 22 度と推測される。正確には、95%上側信頼区間が 0.0 となるように温度 x をソルバーで変化させると 21.99 度が得られる。同等に、推定値が同じになるのは、30.10 度と推定される。

表 13.14 A_4 と季節 A_3 の回収液の濃度の差の 95%信頼区間

No.	A	x_0	a_1	a_2	a_3	x	a_1x	a_2x	a_3x	y^\wedge	$Var(y^\wedge)$	SE	L95%	U95%
19	A_3-A_4	0	1	1	2	0	5	5	10	-17.8026	22.7455	4.7692	-27.4948	-8.1104
20	"	0	1	1	2	0	10	10	20	-14.2557	9.8665	3.1411	-20.6392	-7.8722
21	"	0	1	1	2	0	15	15	30	-10.7088	4.1125	2.0279	-14.8301	-6.5876
22	"	0	1	1	2	0	20	20	40	-7.1620	5.4834	2.3417	-11.9208	-2.4031
23	"	0	1	1	2	0	25	25	50	-3.6151	13.9793	3.7389	-11.2134	3.9833
24	"	0	1	1	2	0	30	30	60	-0.0682	29.6002	5.4406	-11.1248	10.9884
	ソルバー	0	1	1	2	0	21.99	21.99	43.98	-5.7513	8.0090	2.8300	-11.5026	0.0000
	ソルバー	0	1	1	2	0	30.10	30.10	60.19	0.0000	29.9703	5.4745	-11.1256	11.1256

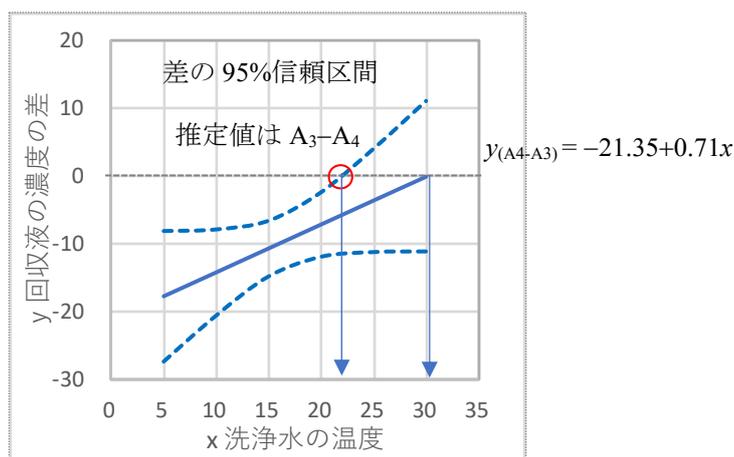


図 13.7 洗浄水の温度に対する季節 A_4 と季節 A_3 の回収液の濃度の差

最小 2 乗平均 (Lsmeans)

表 13.2 に示した JMP による「分散分析表・効果の検定」, に引き続き表 13.3 で季節 A の水準間の差の検討結果を示したが、季節 A についての最小 2 乗平均の結果を示さなかった。これは、簡単に説明しにくいからで、予測プロファイルのある特定の場合として、最小 2 乗平均を説明することが理解しやすいと思われる。

表 13.15 に示すように、季節 A について最小 2 乗平均、標準誤差、95%信頼区間、平均が出力されている。平均は表 13.1 に示した回収液の濃度の季節 A の各種水準の濃度の算術平均である。季節 A_4 の濃度 y の最小 2 乗平均は、24.1766、 A_3 の場合は、18.0618 と表示されている。

これは、洗浄水の温度 x の総平均 21.4762 に固定した場合の季節 A の各種水準の濃度 y の推定値で、図 13.6 で示した洗浄水の温度を $x=20.0$ から $x=21.4762$ にずらした場合の濃度 y の推定値に相当する。

表 13.15 JMP による季節 A の各水準に対する最小 2 乗平均と 95%信頼区間

最小2乗平均表					
水準	最小2乗平均	標準誤差	下側95%	上側95%	平均
A1	14.7479	1.8958	10.8953	18.6006	11.8000
A2	16.0280	1.4719	13.0368	19.0192	14.3083
A3	18.0618	1.5129	14.9872	21.1364	19.4667
A4	24.1766	2.2275	19.6499	28.7033	32.6000

JMP の予測プロファイルを使って確認する。図 13.8 に示すように、洗浄水の温度を $x=21.4762$ とした場合の、季節 A₄ の濃度 y の推定値は、24.1766 (19.6498, 28.7033)、季節 A₃ の場合は、18.0618 (14.9872, 19.4667) のように、表 13.15 に示す最小 2 乗平均に一致する。



図 13.8 洗浄水の温度 21.476 に対する季節 A₄ および季節 A₃ の予測プロファイル

なぜ、洗浄水温度 x の総平均を各季節の予測値の計算のために使用したのだろうか。これは、最小 2 乗法による回帰分析に際し、 y の推定値の分散が最小になるのが x の総平均であるためと理解される。このような、設定方法であると理解しても、各水準の標準誤差、95%信頼区間は、どのようにして求めているかの説明は、簡単ではない。説明のためには、これまで繰り返し使ってきたパラメータの共分散行列を使わずにはできない。

表 13.10 は、洗浄水の温度を $x=20.0$ に固定した場合の季節 A の各水準での濃度 y の推定値であるが、総平均 $x=21.4762$ に置き換えれば、「最小 2 乗平均」を推定することができる。表

13.16 に Excel の行列計算による「最小 2 乗平均」の推定と「95%信頼区間」の推定結果を示す。JMP で出力した表 13.15 と完全に一致していることが確かめられる。

表 13.16 季節 A の各水準に対する最小 2 乗平均と 95%信頼区間

No.	A	x_0	a_1	a_2	a_3	x	a_1x	a_2x	a_3x	y^{\wedge}	$Var(y^{\wedge})$	L 95%	U 95%
1	A ₁	1	1	0	0	21.4762	21.48	0	0	14.7479	3.5939	10.8952	18.6006
2	A ₂	1	0	1	0	21.4762	0	21.48	0	16.0280	2.1664	13.0368	19.0192
3	A ₃	1	0	0	1	21.4762	0	0	21.48	18.0618	2.2889	14.9872	21.1364
4	A ₄	1	-1	-1	-1	21.4762	-21.48	-21.48	-21.48	24.1766	4.9616	19.6498	28.7033

Excel による探索的な交互作用解析

Excel を主体にし、交互作用を考慮した解析方法を詳細に示し、JMP および SAS/GLM プロシジャで同様の解析を試みたのであるが、なかなか大変である。表 13.14 で示したような、共変量 x がどのくらいから有意な差となるかとの推定は、実現できなかった。本質的に交互作用がある場合の“共分散分析”の解析方法として、図 13.7 に示した結果の表示が容易にできたことは、Excel による探索的な解析の賜物であり、また、驚きでもあった。

交互作用があるか否かは、伝統的に分散分析表によつての判断することで対応がなされてきた。交互作用があった場合には、私自身も「適切なグラフ作成して考察する」ことで対応してきた。これは、95%信頼区間が含まれないような図 13.1 の回帰直線のみで「適当」に考察するような態度であり、「統計的」な判断を放棄したことに等しい。少なくとも図 13.6 に示すように、2 本の回帰直線に 95%信頼区間を上書きすることにより「統計的」判断を行うことができるようになる。

先進的な JMP でも、図 13.6 のように 2 本の回帰直線に 95%信頼区間を上書きしようと試みたことはある。ただし、細かなファイル操作などを経て、JMP でのグラフ作成機能では、図 13.6 と同等なもの作成は、やってできないことはないが、生産性が著しく低く、まったく薦められない。これは、Excel の「散布図」のきめ細かさが JMP のスペックより勝っているためである。

13.3. 共変量が2変量の場合の探索的な共分散分析

共分散分析は、質的因子を対象にした1元配置による実験に際し、反応変数 y に何らかの影響を与えることが事前に分かっている測定値を得ることができる量的因子があるが、実験に際して制御し難い変数を「共変量」として解析モデルに組み込む方法である。したがって、共分散分析は、「共分散・分析」ではなく「共(変量を含む)・分散分析」の意味である。

表13.17に示すのは、奥野ら(1981)の第7章の事例7.2であり、ある電気部品の製造工程で、熱処理前の部品寸法 x_1 , x_2 が事前に測定され、4つの炉で熱処理をして電気特性 y を計測した結果である。このデータは、4つの炉で熱処理した電気部品の電気特性 y に対し、炉によるバイアスが疑われたのであろう。そこで、幾つかの試験片を作成し、炉による熱処理を行い、電気特性 y を測定したと思われる。質的因子である4つの炉に対し、さまざま寸法の電気部品を集めて熱処理後の電気特性 y に対し、部品寸法 x_1 , x_2 が、特徴的な共変量として認識される。

表13.17 電気部品と電気特性 [奥野(1981), 表7.6]

i	1号炉			2号炉			3号炉			4号炉		
	x_1	x_2	y									
1	31	24	14	29	15	14	11	36	0	16	15	15
2	25	31	24	29	34	21	26	32	13	44	21	29
3	37	22	23	36	25	22	26	26	16	29	41	22
4	20	35	9	16	34	3	18	13	4	18	49	22
5	11	28	0	35	16	12	32	34	9	21	33	23
6	21	39	2	30	26	16	31	26	20	31	7	17
7	9	28	1	28	21	19	30	49	19	21	37	10
8				29	27	19	19	40	7	29	12	22
9				21	11	14				21	52	18
10				22	37	17						
平均	22.0	29.6	10.4	27.5	24.6	15.7	24.1	32.0	11.0	25.6	29.7	19.8
									総平均	25.06	28.71	14.59

共変量の効き方

通常共分散分析は、制御できる質的因子に対して1つの共変量に対する解析法として知られている。この事例は、2つの共変量があるので、探索的な解析が必要となる。まず、部品寸法 x_1 , x_2 のどちらが、共変量として効いているのか、両方が効いているとした場合に、互いに独立なのか、互いに補完し合っているのか、などを最初に明らかにする必要がある。

図13.9にJMPの「二変量の関係」で作成した $(x_1$ と $y)$, $(x_2$ と $y)$ の散布図上に炉ごとに50%の確率楕円をあてはめた結果を示す。このような層別確率楕円図は、炉ごとの説明変数

と反応変数 y の関係を一目で把握することができる。なお、探索的な解析の場合に、経験的に 50% 程度の緩い確率楕円が、探索的な解析では見通しがよいようである。

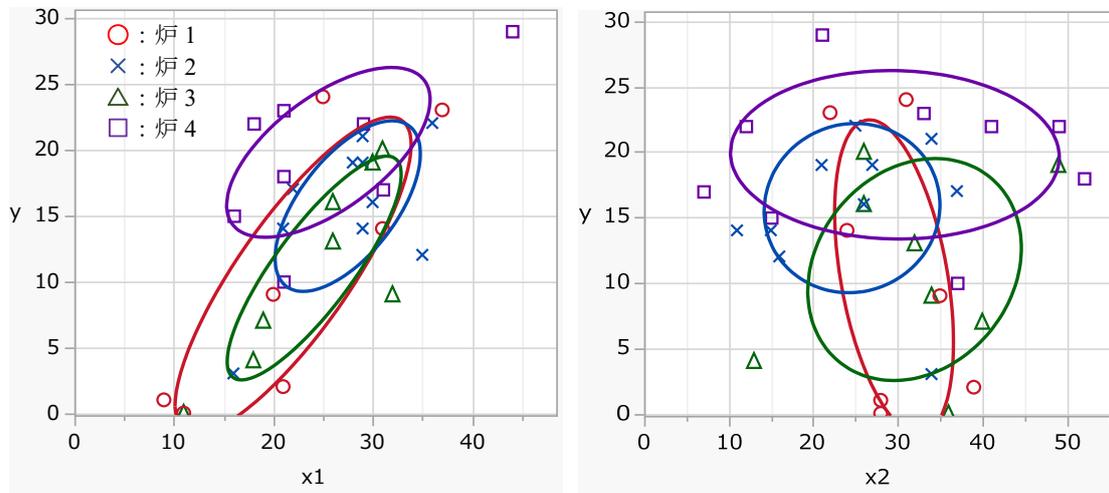


図 13.9 寸法 x_1 と x_2 別の電気特性 y について炉ごとの 50% 確率楕円

寸法 x_1 と y の関係は、どの炉でも同様の正の関連がみられる。炉の 1~3 は重なっており、炉 4 が他に比べて高い電気特性を持っているようである。寸法 x_2 と y の関係は、ほとんど関連がなく、電氣的性能に及ぼす影響が単独ではないと判断される。

2 つの特性が互いに関係しているかを調べるためには、交互作用をモデルに入れることで検討できる。また、共変量と y の関係が 1 次的であることの確認のために、解析モデルに x_1^2 、 x_2^2 および $x_1 \times x_2$ を加えた応答局面法による解析も有益である。ここでは、詳細は省くが JMP の応答局面プロファイルの結果を図 13.10 に示す。

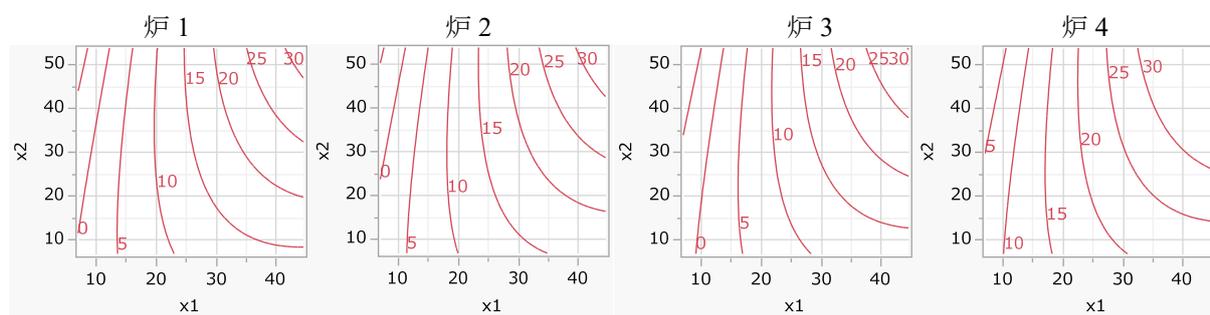


図 13.10 JMP の応答局面解析で作成した等高線プロファイル

応答局面解析の結果から寸法 x_1 、 x_2 間で明確な交互作用は見出されず、寸法 x_1 の 1 次成分のみが明確な縦縞の等高線として得られている。

Excel による 2 変量の共分散分析

実際の解析では、寸法 x_1 のみを共変量とした共分散分析の問題に帰着するが、寸法 x_1 , x_2 が互いに独立に、電気特性 y に対し直線関係であるとして、2 つの共変量とする共分散分析の事例とする。表 13.18 に Excel シート上に展開したデザイン行列、分析ツールの回帰分析で

表 13.18 分析ツールの回帰分析による結果およびパラメータの共分散行列

— デザイン行列 —									Excelによる回帰分析					
i	炉	— 炉 —			— 寸法 —		反応	分散分析表						
		x_0	a_1	a_2	a_3	x_1		x_2	y	自由度	変動	分散	分散比	有意 F
1	A1	1	1	0	0	31	24	14	回帰	6	8554.77	1425.80	60.38	0.0000
2		1	1	0	0	25	31	24	残差	28	661.23	23.62	σ^2	
3		1	1	0	0	37	22	23	合計	34	9216			
4		1	1	0	0	20	35	9						
5		1	1	0	0	11	28	0		係数	標準誤差	t	P-値	
6		1	1	0	0	21	39	2	切片	0	#N/A	#N/A	#N/A	
7		1	1	0	0	9	28	1	X 値 1	-5.4994	4.1387	-1.3288	0.1946	x_0
8	A2	1	0	1	0	29	15	14	X 値 2	-1.9702	1.5777	-1.2487	0.2221	a_1
9		1	0	1	0	29	34	21	X 値 3	0.0893	1.4326	0.0623	0.9507	a_2
10		1	0	1	0	36	25	22	X 値 4	-3.0839	1.4964	-2.0609	0.0487	a_3
11		1	0	1	0	16	34	3	X 値 5	0.6763	0.1134	5.9624	0.0000	x_1
12		1	0	1	0	35	16	12	X 値 6	0.1021	0.0801	1.2747	0.2129	x_2
13		1	0	1	0	30	26	16						
14		1	0	1	0	28	21	19	パラメータの共分散行列 $\Sigma(\hat{\beta})=(X^T X)^{-1}\sigma^2$					
15		1	0	1	0	29	27	19	x_0	a_1	a_2	a_3	x_1	x_2
16		1	0	1	0	21	11	14	17.1288	-0.7848	-0.1288	0.5044	-0.3821	-0.2399
17		1	0	1	0	22	37	17	-0.7848	2.4892	-0.8111	-0.8578	0.0346	0.0022
18	A3	1	0	0	1	11	36	0	-0.1288	-0.8111	2.0524	-0.7055	-0.0253	0.0221
19		1	0	0	1	26	32	13	0.5044	-0.8578	-0.7055	2.2391	0.0020	-0.0181
20		1	0	0	1	26	26	16	-0.3821	0.0346	-0.0253	0.0020	0.0129	0.0022
21		1	0	0	1	18	13	4	-0.2399	0.0022	0.0221	-0.0181	0.0022	0.0064
22		1	0	0	1	32	34	9						
23		1	0	0	1	31	26	20						
24		1	0	0	1	30	49	19	対比型デザイン変数					
25		1	0	0	1	19	40	7	A	a_1	a_2	a_3		
26	A4	1	-1	-1	-1	16	15	15	A ₁	1	0	0		
27		1	-1	-1	-1	44	21	29	A ₂	0	1	0		
28		1	-1	-1	-1	29	41	22	A ₃	0	0	1		
29		1	-1	-1	-1	18	49	22	A ₄	-1	-1	-1		
30		1	-1	-1	-1	21	33	23						
31		1	-1	-1	-1	31	7	17	$t(0.08,28)=$	2.0484				
32		1	-1	-1	-1	21	37	10						
33		1	-1	-1	-1	29	12	22						
34		1	-1	-1	-1	21	52	18						

の結果，行列関数で計算したパラメータの共分散行列を示す．寸法 x_2 の p 値は，0.2129 と有意でないことが示されている．

回帰パラメータの推定値を用いた回帰式は，

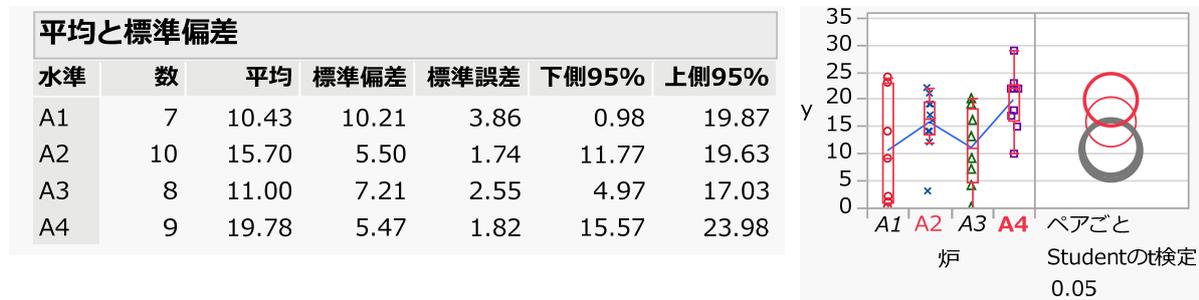
$$\begin{aligned}\hat{y} &= \hat{\beta}_0 x_0 + \hat{\beta}_1 a_1 + \hat{\beta}_2 a_2 + \hat{\beta}_3 a_3 + \hat{\beta}_4 x_1 + \hat{\beta}_5 x_2 \\ &= -5.4994 - 1.9702 a_1 + 0.0893 a_2 - 3.0839 a_3 + 0.6763 x_1 + 0.1021 x_2\end{aligned}$$

である． $\hat{\beta}_0 = -5.4994$ は， $a_1 = a_2 = a_3 = x_1 = x_2 = 0$ の場合でなので，寸法 x_1 と x_2 の共通の Y 切片であり，また，炉 A に関しては効果の推定値となっている． $\hat{\beta}_0 + \hat{\beta}_1 = -5.4994 - 1.9702$ は，寸法 x_1 と x_2 の切片における A_1 の推定値であり， $\hat{\beta}_4 = 0.6763$ は，寸法 x_1 が切片 0 から 1 単位増加した時の増分である．このような回帰パラメータの推定値の解釈を行なうのだが，実際の結果を説明する際に，現実から遊離していき説明しづらい．

最小 2 乗平均

まず，表 13.19 に示すように炉別の基本統計量を示し，ボックス・プロットなどを用いて，共変量抜きでの比較を行う．炉 A_4 が熱処理後の電気特性に対する性能が高く， A_2 は平均が 2 番目に高いが，飛び離れ値の存在が気になり，炉 A_1 と炉 A_3 が，低めになっている．

表 13.19 炉の番号別の基本統計量



炉によって，試験片の寸法が不揃いであり，図 13.9 に示したように，寸法 x_1 が共変量なので，共分散分析によって，試験片の平均値を統計的に調整し，調整済み平均値と 95%信頼区間を示したい．さらに，ある炉を基準とした調整済み平均値の 95%信頼区間も示し，総合的に評価を行いたい．

SAS および JMP ユーザならば，迷わず最小 2 乗平均 `Lsmeans` を算出して評価をするに違いない．JMP を使えば，表 13.20 に示すように炉の番号別の最小 2 乗平均と 95%信頼区間，さらにグラフが標示される．さて，さまざまな関係者から「最小 2 乗平均」とは何ですか，ど

のように計算されたものですか、との質問にどのように答えたら良いのでしょうか。表 13.20 に炉の番号別の最小 2 乗平均と 95%信頼区間を示す。表の右端に「算術平均」が示されており、最小 2 乗平均とは明らかに異なる。

表 13.20 炉 A の番号別の最小 2 乗平均と 95%信頼区間



回帰パラメータの推定値を用いた推論の場合には、 $x_1=0$ 、 $x_2=0$ の場合の共通の Y 切片上での炉の番号別の推定値となり、現実データと遊離しているので、説明がややこしくなる。そのために、「最小 2 乗平均」は、炉の推定値の分散が最小になると期待される x_1 の総平均 20.05 と x_2 の総平均 28.71 としたときの炉 A の水準ごとの反応 y の推定値を「最小 2 乗平均」としている。炉 A₁ の場合は、

$$\hat{y}_{A_1} = \hat{\beta}_0 x_0 + \hat{\beta}_1 a_1 + \hat{\beta}_2 a_2 + \hat{\beta}_3 a_3 + \hat{\beta}_4 x_1 + \hat{\beta}_5 x_2$$

	x_0	a_1	a_2	a_3	x_1	x_2	$\hat{\beta}$	
$y^{\wedge}_{A_1} =$	1	1	0	0	25.0588	28.7059	-5.4994	= 12.4088
							-1.9702	
							0.0893	
							-3.0839	
							0.6763	
							0.1021	

と推定される。表 13.21 に A₁ 以外の炉について最小 2 乗平均を推定した結果を示す。JMP で求めた表 13.20 と完全に一致していることが確認される。

表 13.21 Excel による炉 A の番号別の最小 2 乗平均と 95%信頼区間

炉	水準	— 炉 —			寸法		推定値	分散	— 誤差 —		95%信頼区間		
		x_0	a_1	a_2	a_3	x_1	x_2	y^{\wedge}	$Var(y^{\wedge})$	SE	$t_{\alpha}SE$	L95%	U95%
						25.06	28.71						
1	A ₁	1	1	0	0	25.06	28.71	12.41	3.49	1.87	3.83	8.58	16.23
2	A ₂	1	0	1	0	25.06	28.71	14.47	2.50	1.58	3.24	11.23	17.71
3	A ₃	1	0	0	1	25.06	28.71	11.30	3.02	1.74	3.56	7.74	14.85
4	A ₄	1	-1	-1	-1	25.06	28.71	19.34	2.64	1.62	3.33	16.02	22.67

注) $t_{\alpha}SE$ は、Excel で 95%信頼区間のひげを付けるために計算している。

JMP および SAS での「最小 2 乗平均」は、合理的ではあるが、出力された結果について、手計算で計算結果を検証することは、なかなか困難である。ここに示したように Excel の行列計算を用いることにより、「最小 2 乗平均」の内部を可視化することができる。表 13.21 で、 $x_1 = 25$, $x_2 = 30$ のように Excel でデータを置き換えることにより、切れの良い数字での推定値が得られ、結果の説明も容易になる。

水準間の差の推定

炉の水準ごとの推定値に対し、炉 A₄ を基準にし、他の炉との差および 95% 信頼区間を推定したい。また、炉 A₁, A₂, A₃ 水準は、互いに 95% 信頼区間が推定値に互いにかぶり合い、統計的には差がないことが明らかであるが、炉 A₄ との差も考察のために求めたいとする。水準間の差とその 95% 信頼区間を求めたい場合には、比較したい水準のデザイン変数の差を計算すればよい、炉 A₃ と炉 A₄ の差の場合であれば、

$$\begin{array}{r}
 \begin{array}{|c|c|c|c|c|c|c|}
 \hline
 \text{炉} & x_0 & a_1 & a_2 & a_3 & x_1 & x_2 \\
 \hline
 A_3 & 1 & 0 & 0 & 1 & 25.0 & 30.0 \\
 \hline
 \end{array} \\
 -) \\
 \hline
 \begin{array}{|c|c|c|c|c|c|c|}
 \hline
 A_4 & 1 & -1 & -1 & -1 & 25.0 & 30.0 \\
 \hline
 \end{array} \\
 = \\
 \hline
 \begin{array}{|c|c|c|c|c|c|c|}
 \hline
 A_3-A_4 & 0 & 1 & 1 & 2 & 0 & 0 \\
 \hline
 \end{array}
 \end{array}$$

とする。推定値も分散の計算も、表 13.22 に示すように、これまでと同様に計算すればよい。炉 A の水準間の推定に際しては、共変量の大きさが反映されるが、交互作用がない主効果モ

表 13.22 Excel による炉 A₄ との差の最小 2 乗平均と 95% 信頼区間

	炉	x_0	a_1	a_2	a_3	x_1	x_2	推定値 $y_{\text{差}}^{\wedge}$	分散 $Var(y_{\text{差}}^{\wedge})$	— 誤差 — SE	$t_{\alpha} SE$	95% 信頼区間 L95% U95%		
	A ₄	1	-1	-1	-1	25.0	30.0							
	A ₁	A ₄	0	2	1	1	0.0	0.0	-6.93	6.16	2.48	5.08	-12.02	-1.85
	A ₂	A ₄	0	1	2	1	0.0	0.0	-4.88	5.16	2.27	4.65	-9.53	-0.22
	A ₃	A ₄	0	1	1	2	0.0	0.0	-8.05	5.62	2.37	4.86	-12.91	-3.19
	A ₄	A ₄	0	0	0	0	0.0	0.0	0.00	0.00	0.00	0.00	0.00	0.00

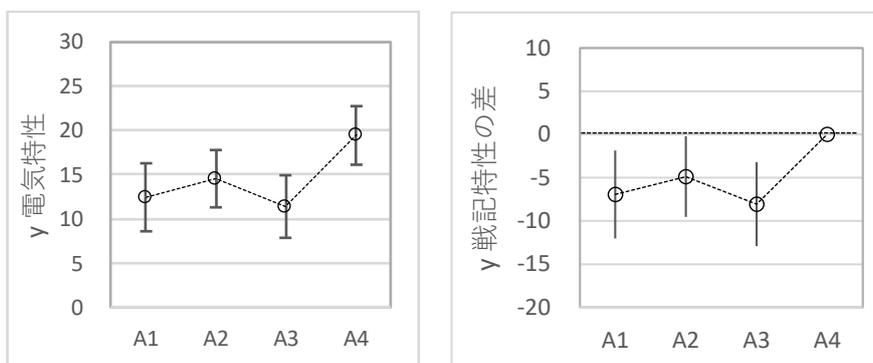


図 13.11 炉 A の水準間の最小 2 乗平均および炉 A₄ との差の 95% 信頼区間

デルの場合での水準間の差における共変量の大きさはゼロとなっており、共変量 x_1 および x_2 の大きさに関わらず差の推定値も差の分散も同じである。

デザイン変数の活用

ここに示した共分散分析は、共変量が2つ、炉 A が4水準で、複雑な共分散分析の事例となっている。一般的な共分散分析は、共変量が一つで、2群に対して平行な直線をあてはめる場面設定であり、通常の解析手順は、分散分析表を主体にした解析結果が示されている。推定された2本の回帰直線間の統計的な判断は、分散分析表の群間差の p 値に基づいている。

推定された2つの Y 切片の差の95%信頼区間を示し、結果を考察するような手順とはなっていないが、デザイン変数を、工夫することにより表 13.22 に示したと同等の結果を得ることができる。Excel の回帰分析で、表 13.23 に示すような炉 A₄ を基準とする (1, 0) 型のデザイン変数にすることにより、表 13.24 に示すように A₁ と A₄ の差、A₂ と A₄ の差、A₃ と A₄ の差についての SE および 95%信頼区間も同時推定ができる。

表 13.23 炉 A₄ を基準とするデザイン変数

A	x_0	a_1	a_2	a_3	推定値
A ₁	1	1	0	0	A ₁ -A ₄
A ₂	1	0	1	0	A ₂ -A ₄
A ₃	1	0	0	1	A ₃ -A ₄
A ₄	1	0	0	0	基準

表 13.24 に示した a_3 の行の「係数」欄には、炉 A₃ と炉 A₄ の差の推定値 -8.0468 (-12.9057 , -3.1914) が推定されていて、表 13.22 の結果に一致する。

表 13.24 Excel の回帰分析による炉 A₄ を基準とする他の水準との差と 95%信頼区間

		自由度	変動	分散	分散比	有意 F	
回帰		6	8554.77	1425.80	60.38	0.0000	
残差		28	661.23	23.62			
合計		34	9216				
		係数	標準誤差	t	P-値	下限 95%	上限 95%
切片		0	#N/A	#N/A	#N/A	#N/A	#N/A
x_0	X 値 1	-0.5347	4.4698	-0.1196	0.9056	-9.6907	8.6213
a_1	X 値 2	-6.9349	2.4823	-2.7938	0.0093	-12.0196	-1.8502
a_2	X 値 3	-4.8754	2.2707	-2.1471	0.0406	-9.5266	-0.2242
a_3	X 値 4	-8.0486	2.3712	-3.3943	0.0021	-12.9057	-3.1914
x_1	X 値 5	0.6763	0.1134	5.9624	0.0000	0.4439	0.9086
x_2	X 値 6	0.1021	0.0801	1.2747	0.2129	-0.0620	0.2662

共変量の影響の可視化

共変量の効き方は、表 13.18 の回帰係数の推定値と標準誤差によって評価できるが、共変量の変化によって電気特性 y に対しどのような状況かを可視化したい。図 13.12 に JMP の予測プロフィールで（炉 A=A4, $x_1=25$, $x_2=30$ ）を選択した結果を示す。

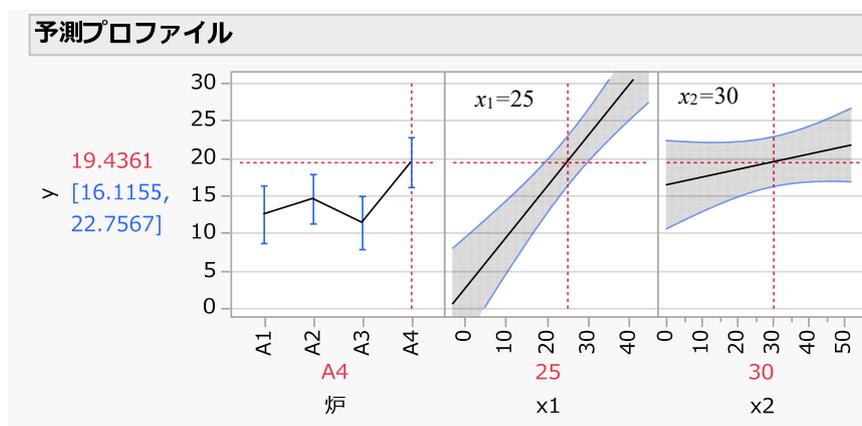


図 13.12 炉 A4 の場合の共変量の影響

Excel を用いて内部での計算方法を表 13.25 に示す。基本は、表 13.21 と同様であり、表 13.24 に示すように炉 A4 と $x_2=30$ に固定し x_1 を (0, 10, ..., 50) と変化させた場合の炉 A4 と $x_1=25$ を固定し x_2 を (0, 10, ..., 50) と変化させた場合に電気特性 y に対する推定値、分散、SE、および、95%信頼区間の計算をしている。

表 13.25 炉 A4 に対する共変量の変化に対する電気特性 y に対する影響

	炉	— 炉 —			寸法		推定値	分散	誤差	95%信頼区間		
		x_0	a_1	a_2	a_3	x_1				x_2	y^{\wedge}	$Var(y^{\wedge})$
$x_1=25$	A1	1	1	0	0	25.0	30.0	12.50	3.50	1.87	8.67	16.33
$x_2=30$	A2	1	0	1	0	25.0	30.0	14.56	2.57	1.60	11.28	17.84
	A3	1	0	0	1	25.0	30.0	11.39	2.98	1.73	7.85	14.92
	A4	1	-1	-1	-1	25.0	30.0	19.44	2.63	1.62	16.12	22.76
$x_2=30$	A4	1	-1	-1	-1	0.00	30.00	2.53	10.99	3.32	-4.26	9.32
	A4	1	-1	-1	-1	10.00	30.00	9.29	5.72	2.39	4.39	14.19
	A4	1	-1	-1	-1	25.00	30.00	19.44	2.63	1.62	16.12	22.76
	A4	1	-1	-1	-1	30.00	30.00	22.82	2.89	1.70	19.34	26.30
	A4	1	-1	-1	-1	40.00	30.00	29.58	5.33	2.31	24.85	34.31
	A4	1	-1	-1	-1	50.00	30.00	36.34	10.35	3.22	29.75	42.93
$x_1=25$	A4	1	-1	-1	-1	25.00	0.00	16.37	8.35	2.89	10.45	22.29
	A4	1	-1	-1	-1	25.00	10.00	17.39	5.16	2.27	12.74	22.05
	A4	1	-1	-1	-1	25.00	20.00	18.41	3.25	1.80	14.72	22.11
	A4	1	-1	-1	-1	25.00	30.00	19.44	2.63	1.62	16.12	22.76
	A4	1	-1	-1	-1	25.00	40.00	20.46	3.29	1.81	16.74	24.17
	A4	1	-1	-1	-1	25.00	50.00	21.48	5.23	2.29	16.79	26.16

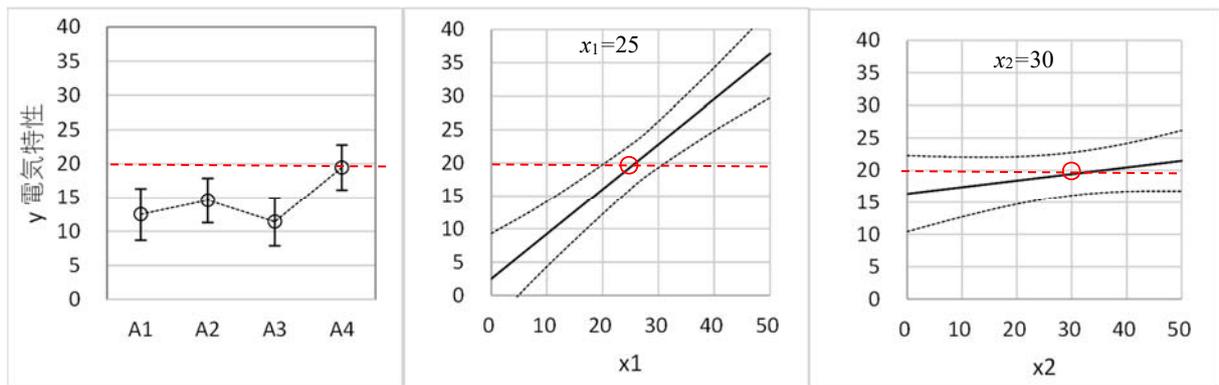


図 13.13 Excel による炉 A₄ に対応する共変量の予測プロファイル

炉 A の予測プロファイルは、部品寸法を ($x_1 = 25$, $x_2 = 30$) に固定した場合、部品寸法 x_1 の予測プロファイルは、(炉 A=A₄, $x_2 = 30$) に固定した場合、部品寸法 x_2 の予測プロファイルは、(炉 A=A₄, $x_1 = 25$) に固定した場合の推定値であり、互いに関連している。炉 A=A₄, $x_1 = 25$, $x_2 = 30$ とそれぞれが一致する場合の予測値は 19.44 と揃っていることが確認される。

JMP が使えるならば、Excel でわざわざ作成する必要はないが、「予測プロファイル」は統計の一般用語ではないことを認識し、他者に対する資料として作成する場合には、何らかの解説を付ける必要がある。

Excel による作図のヒントは、本章以外でも第 7.2 節、第 9.2 節、第 10.3 節、第 12.3 節で例示しているので参考にしてもらいたい。

13.4. 繰返しが不揃いな2因子の共分散分析

検索エンジンを使って、「最小2乗平均」について書かれている論文を探したところ、守屋・広岡（2018）の「Rパッケージを用いた最小2乗分散分析と最小2乗平均値の算出」が見いだされた。彼らは、データ数が不揃いの実験データの解析のために SAS/GLM プロシジャに代えて、Rのlsmeansパッケージの使用法について論じている。

守屋らは、「各要因のグループ（水準）内のデータ数が等しい釣り合い型データ（balanced data）に対しては通常の生物統計学の教科書で紹介されている分散分析法が適用でき、EXCELのような広く普及している表計算ソフトでも対応できる。一方、データ数が等しくない不釣り合い型データ（unbalanced data）や分析に共変量を含むケースは通常の分散分析の手法ではなく最小2乗分散分析法を用いる必要がある」ことを提示している。

そのために、「SASのGLMプロシジャで実行されることが多かった。しかし、SASは有料でしかも高価であったため、実行できる環境は非常に限定されていた」と述べ、最近、R環境においてもlsmeansパッケージが提供されたので、SASのGLMプロシジャでのLsmeansと比較検討した結果を報告している。論文には、例示として、（因子A、因子B、共変量 x 、反応変数 y ）についての15個のデータが示されている。

表 13.26 守屋らの2因子共分散データ

因子A	因子B	共変量 x	反応 y
A ₁	B ₁	350	970
		400	1000
		360	980
A ₁	B ₂	350	980
		340	970
A ₂	B ₁	390	990
		340	950
A ₂	B ₂	410	980
		430	990
		390	980
A ₃	B ₁	400	990
		320	940
A ₃	B ₂	330	930
		390	1000
		420	1000
	平均	374.67	976.67
反応 y :1日当たりの増体量			

解析モデルとして，交互作用を含むモデル

$$\text{反応 } y = A + B + A \times B + \text{共変量 } x$$

についての最小 2 乗平均を提示し，SAS/GLM プロシジャの Lsmeans の結果と一致することを報告している。

対比型デザイン変数を用いた場合の最小 2 乗平均

表 13.26 のデータに対して，対比型のデザイン変数を表 13.27 に示す．交互作用は，因子 A と因子 B の水準の組み合わせで， $A_i \times B_j$ に対し ($x_{4,(i,j)} = x_{1,i} \cdot x_{3,j}$, $x_{5,(i,j)} = x_{2,i} \cdot x_{3,j}$) のようにデザイン変数間の積で定義する．なお，共変量 x は， x_6 として定義し直す。

表 13.27 対比型のデザイン変数

A	x_1	x_2		B	x_3		A×B	x_4	x_5
A ₁	1	0		B ₁	1		A ₁ B ₁	1	0
A ₂	0	1		B ₂	-1		B ₂	-1	0
A ₃	-1	-1					A ₂ B ₁	0	1
							B ₂	0	-1
							A ₃ B ₁	-1	-1
							B ₂	1	1

第 13.2 節では，因子が 1 つで共変量が 1 つの場合であり，第 13.3 節では因子が 1 つで共変量が 2 つの場合であった．ここでは，因子が 2 つで交互作用を含め共変量が 1 つの場合であり，最小 2 乗平均はどのように求められているのだろうか．表 13.28 にデザイン変数を組み込み，Excel の回帰分析を適用した結果を示す。

表 13.28 対比型のデザイン変数を用いた Excel による回帰係数の推定

		—— デザイン行列 X ——												
		A		B	A×B		共変量	反応						
i	A B	x_0	x_1	x_2	x_3	x_4	x_5	x_6	y	分散分析表				
1	A ₁ B ₁	1	1	0	1	1	0	350	970	自由度	変動	分散		
2		1	1	0	1	1	0	400	1000	回帰	7	14314030	2044861	
3		1	1	0	1	1	0	360	980	残差	8	670.4128	83.8016 σ^2	
4	B ₂	1	1	0	-1	-1	0	350	980	合計	15	14314700		
5		1	1	0	-1	-1	0	340	970					
6	A ₂ B ₁	1	0	1	1	0	1	390	990	係数	標準誤差	t	P-値	
7		1	0	1	1	0	1	340	950	切片	0	#N/A	#N/A	#N/A
8	B ₂	1	0	1	-1	0	-1	410	980	X 値 1	718.117	32.6778	21.9757	0.0000 x_0
9		1	0	1	-1	0	-1	430	990	X 値 2	13.4238	3.6307	3.6973	0.0061 x_1
10		1	0	1	-1	0	-1	390	980	X 値 3	-9.8560	3.6833	-2.6759	0.0281 x_2
11	A ₃ B ₁	1	-1	-1	1	-1	-1	400	990	X 値 4	1.8400	2.4822	0.7413	0.4797 x_3
12		1	-1	-1	1	-1	-1	320	940	X 値 5	-6.3315	3.8031	-1.6648	0.1345 x_4
13	B ₂	1	-1	-1	-1	1	1	330	930	X 値 6	7.0782	3.6833	1.9217	0.0909 x_5
14		1	-1	-1	-1	1	1	390	1000	X 値 7	0.6927	0.0877	7.8996	0.0000 x_6
15		1	-1	-1	-1	1	1	420	1000					

パラメータの共分散行列 $\Sigma(\hat{\beta})$ は、デザイン行列 X と誤差分散 $\hat{\sigma}^2$ を用いて

$$\Sigma(\hat{\beta}) = (X^T X)^{-1} \hat{\sigma}^2$$

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

として表 13.29 に示すように計算される。

表 13.29 対比型デザイン行列に対するパラメータの共分散行列

	— A —			B	— A×B —		共変量
x_0	x_1	x_2	x_3	x_4	x_5	x_6	
1067.84	-40.4806	45.2430	-18.6617	53.2160	-44.4671	-2.8575	
-40.4806	13.1821	-7.5441	-0.8258	-2.8635	2.1125	0.1089	
45.2430	-7.5441	13.5665	-0.0356	2.7211	-0.3755	-0.1217	
-18.6617	-0.8258	-0.0356	6.1613	-0.9824	0.8115	0.0513	
53.2160	-2.8635	2.7211	-0.9824	14.4635	-8.1527	-0.1474	
-44.4671	2.1125	-0.3755	0.8115	-8.1527	13.5665	0.1217	
-2.8575	0.1089	-0.1217	0.0513	-0.1474	0.1217	0.0077	
パラメータの共分散行列 $\Sigma(\hat{\beta}) = (X^T X)^{-1} \hat{\sigma}^2$							

表 13.30 に対比型デザイン変数を用いた因子 A、因子 B および交互作用 A×B の各水準の Lsmeans を示す。この計算は、Lsmeans のデザイン変数を x_i ベクトルとし、推定されたパラメータ $\hat{\beta}$ としたときに、因子 A の第 1 水準 A_1 は、共変量 x_6 については、総平均 374.67 とし、因子 B の $x_3 = 0$ 、交互作用 A×B の ($x_4 = 0$, $x_5 = 0$) た場合の y_{A1} は、991.06 が Lsmeans となる。

$$\hat{y}_{A1} = x_{16} \hat{\beta} =$$

x_{16}							$\hat{\beta}$		
1	1	0	0	0	0	374.67	718.1167		
							13.4238		
x_0	x_1	x_2	x_3	x_4	x_5	x_6	-9.8560		y_{A1}
							1.8400		
							-6.3315		
							7.0782		
							0.6927		

表 13.30 最小 2 乗平均・Lsmeans の計算結果

	A	B	— デザイン行列 X —							共変量	反応	最小 2 乗		標準 誤差
			x_0	x_1	x_2	x_3	x_4	x_5	x_6			平均	分散	
i										y	y^{\wedge}	$Var(y^{\wedge})$	SE	
16	A₁		1	1	0	0	0	0	374.67	-	991.06	19.7243	4.4412	
17	A ₂		1	0	1	0	0	0	374.67	-	967.78	18.7249	4.3272	
18	A ₃		1	-1	-1	0	0	0	374.67	-	974.07	17.6261	4.1983	
19	B ₁		1	0	0	1	0	0	374.67	-	979.47	13.1335	3.6240	
20	B ₂		1	0	0	-1	0	0	374.67	-	975.79	10.9665	3.3116	
21	A ₁	B ₁	1	1	0	1	1	0	374.67	-	986.57	28.1013	5.3011	
22		B ₂	1	1	0	-1	-1	0	374.67	-	995.55	48.6673	6.9762	
23	A ₂	B ₁	1	0	1	1	0	1	374.67	-	976.70	42.6192	6.5283	
24		B ₂	1	0	1	-1	0	-1	374.67	-	958.86	37.5322	6.1264	
25	A ₃	B ₁	1	-1	-1	1	-1	-1	374.67	-	975.16	43.5546	6.5996	
26		B ₂	1	-1	-1	-1	1	1	374.67	-	972.97	28.1526	5.3059	

さらに分散と SE は、パラメータの共分散行列 $\Sigma(\hat{\beta})$ を用いて、

$$Var(\hat{y}_{A_1}) = \mathbf{x}_{16} \Sigma(\hat{\beta}) \mathbf{x}_{16}^T$$

=							$\Sigma(\hat{\beta})$							\mathbf{x}_{16}^T		$Var(A_1)$
1	1	0	0	0	0	374.67	1067.8	-40.48	45.24	-18.66	53.22	-44.47	-2.86	1	=	19.7243
							-40.48	13.18	-7.54	-0.83	-2.86	2.11	0.11	1		
x_0	x_1	x_2	x_3	x_4	x_5	x_6	45.24	-7.54	13.57	-0.04	2.72	-0.38	-0.12	0		SE
							-18.66	-0.83	-0.04	6.16	-0.98	0.81	0.05	0		4.4412
							53.22	-2.86	2.72	-0.98	14.46	-8.15	-0.15	0		
							-44.47	2.11	-0.38	0.81	-8.15	13.57	0.12	0		
							-2.86	0.11	-0.12	0.05	-0.15	0.12	0.01	375		
パラメータの共分散行列 $\Sigma(\hat{\beta})$																

として計算されている。表 13.31 に守屋らの R の lsmeans パッケージによる最小 2 乗平均の算出例を示す。Excel で求めた Lsmeans と SE が、一致することが確かめられる。

表 13.31 R の lsmeans パッケージによる最小 2 乗平均の算出例

FactorA	lsmean	SE	df	lower.CL	upper.CL	.group
A1: a	991.0573	4.441210	8	980.8159	1001.2988	2
A2: b	967.7775	4.327225	8	957.7989	977.7561	1
A3: c	974.0657	4.198345	8	964.3843	983.7472	12

守屋・広岡 (2018) の表 3 の一部を抜粋・編集

探索的な解析では、何らかの結果のグラフ表示が欠かせない。JMP の予測プロファイルによるグラフ表示、グラフ・ビルダーによる層別散布図を示そう。

JMP による 2 因子交互作用に共変量を含む最小 2 乗法での解析で「予測プロファイル」を選択すると、図 13.14 に示すような図が表示される。マウスで因子 A の A_1 を選択し、因子 B の B_1 を選択した結果が示されている。共変量 x は、最小 2 乗平均の推定のために 374.67 が自動的にセットされている。この条件下で反応 y の推定値が、986.67 (974.34, 998.79) と表示され、表 13.30 に示されている推定値 \hat{y} と分散 $Var(\hat{y})$ を使い 95% 信頼区間は

$$\begin{aligned} (L95\%, U95\%) &= \hat{y} \pm t(0.05, 8) \sqrt{Var(\hat{y})} \\ &= 986.67 \pm 2.3060 \sqrt{28.1013} \quad (L95\%, U95\%) \\ &= (974.34, 998.79) \end{aligned}$$

となり、図 13.14 の上段の結果に一致する。因子 A は A_1 水準のまま、因子 B を B_2 水準に変更したのが図 13.14 の下段である。因子 A のプロファイルが上段と下段で異なるのは、交互作用を解析モデルに含めたからである。

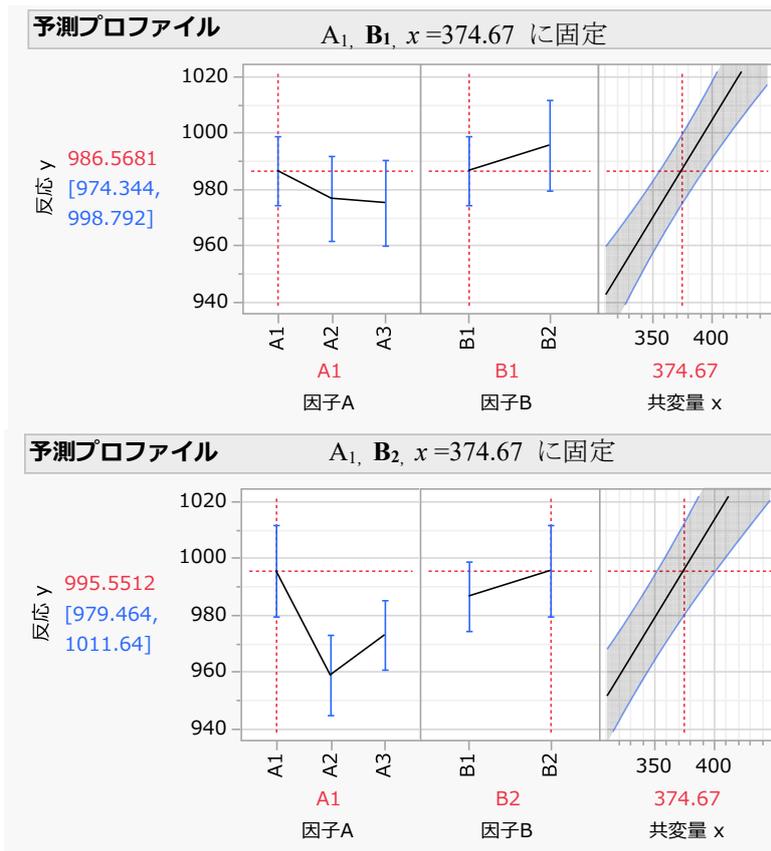


図 13.14 交互作用モデルに対する最小 2 乗平均による予測プロファイル

探索的な解析に際し、因子 A と因子 B を組み合わせた層別散布図により、データの全体像を把握することの必要性をこれまでも示してきた。図 13.15 に JMP のグラフ・ビルダーでの結果を示す。最小 2 乗平均の図 13.14 と比べることにより、探索的な解析結果についての理解が深まると思われる。

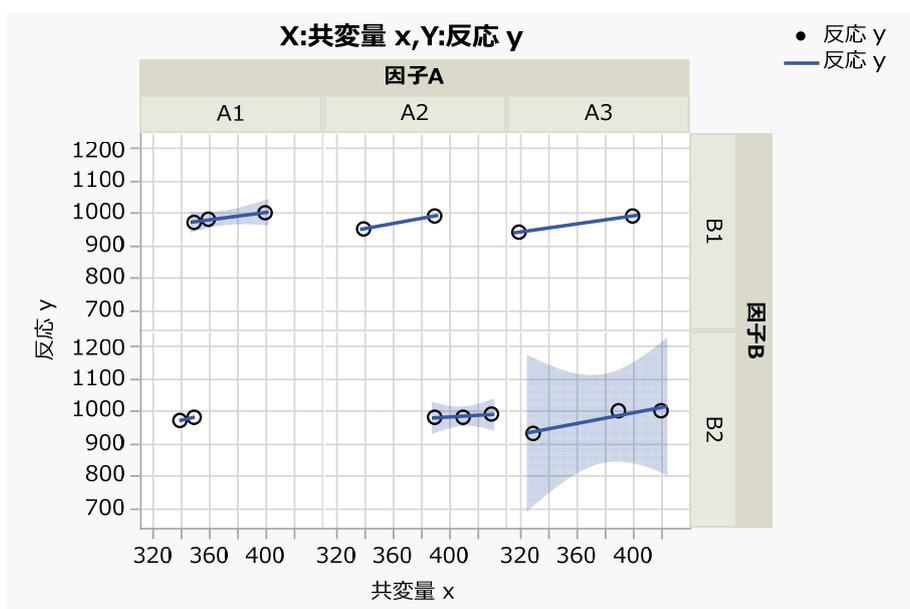


図 13.15 グラフ・ビルダーによる層別散布図によるデータの俯瞰

SAS の GLM プロシジャでのデザイン変数

これまででは、対比型のデザイン変数を用いた Lsmeans を Excel で再現してきたのであるが、SAS の GLM プロシジャでのデザイン変数は、各因子の最後の水準を基準としているので、表 13.32 で示す (1, 0) 型デザイン変数を使う。

表 13.32 最後の水準を基準としたデザイン変数

A	x_1	x_2		B	x_3		A×B		x_4	x_5
A ₁	1	0		B ₁	1		A ₁	B ₁	1	0
A ₂	0	1		B ₂	0			B ₂	0	0
A ₃	0	0		平均	1/2		A ₂	B ₁	0	1
平均	1/3	1/3						B ₂	0	0
							A ₃	B ₁	0	0
								B ₂	0	0

表 13.33 に示すように、因子 A の各水準の Lsmeans を求めるためには、因子 A のデザイン変数 x_1 と x_2 を使い、因子 B の x_3 には、デザイン変数の平均としての (1/2) を用いて、交互作用は、それらの積として、Lsmeans は、 $\hat{y}_{1.16} = \mathbf{x}_{1.16} \hat{\boldsymbol{\beta}}$ として計算されている。もちろん、表 13.30 に示した lsmeans パッケージの結果に一致している。因子 B の場合は、因子 A のデザイン変数 x_1 と x_2 には、3 水準の平均(1/3)を用いる。交互作用については、因子 A と因子 B のデザイン変数の組み合わせによる積のままでよい。

表 13.33 GLM 式デザイン行列に対する Lsmeans を求めるためのデザイン変数

----- デザイン変数 -----										最小2乗		標準					
— A —										B	A×B		共変量	反応	平均	分散	誤差
i	A	B	x_0	x_1	x_2	x_3	x_4	x_5	x_6	y	y^{\wedge}	$Var(y^{\wedge})$	SE				
1.16	A ₁		1	1	0	0.50	0.50	0	374.67	-	991.06	19.724	4.441				
1.17	A ₂		1	0	1	0.50	0	0.50	374.67	-	967.78	18.725	4.327				
1.18	A ₃		1	0	0	0.50	0	0	374.67	-	974.07	17.626	4.198				
1.19	B ₁		1	0.33	0.33	1	0.33	0.33	374.67	-	979.47	13.133	3.624				
1.20	B ₂		1	0.33	0.33	0	0	0	374.67	-	975.79	10.967	3.312				
1.21	A ₁	B ₁	1	1	0	1	1	0	374.67	-	986.57	28.101	5.301				
1.22		B ₂	1	1	0	0	0	0	374.67	-	995.55	48.667	6.976				
1.23	A ₂	B ₁	1	0	1	1	0	1	374.67	-	976.70	42.619	6.528				
1.24		B ₂	1	0	1	0	0	0	374.67	-	958.86	37.532	6.126				
1.25	A ₃	B ₁	1	0	0	1	0	0	374.67	-	975.16	43.555	6.600				
1.26		B ₂	1	0	0	0	0	0	374.67	-	972.97	28.153	5.306				
$\hat{\boldsymbol{\beta}}^T =$			$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$								
			713.4557	22.5765	-14.1131	2.1865	-11.1697	15.6498	0.6927								

表 13.27 に示した対比型のデザイン行列の場合は、それぞれのデザイン変数の合計および平均はゼロなので、表 13.32 に示した最後の水準を基準とした SAS/GLM プロシジャと同様の考え方が必要なのであるが、他の因子での調整は必要とない。R パッケージの場合は、最初

の水準を基準としているので、本質的には SAS/GLM プロシジャと同様のデザイン変数なので、Lsmeans の推定には、求めたい因子以外の因子のデザイン変数について、該当する因子の水準平均を用いる必要がある。

詳細は、高橋ら（1989）の 15 章「4 種の平方和と LSMESN」および第 16 章「GLM プロシジャの計算方式」、魚住（2014）の「LS-Means 再考—GLM と PLM によるモデル推定後のプロセス」、Littell ら（2002）の「SAS for Linear Medels, Chapter 6 Understanding Linear Models Concepts」を参照のこと。

R 言語などでの最初の水準を基準にする場合のデザイン変数

最初の水準を基準にする場合について、SAS/GLM プロシジャでの最後の水準を基準とする場合を参考にして、Lsmeans の計算のためのデザイン変数を表 13.34 に示し、表 13.35 にこれらを反映した (0, 1) 型デザイン行列を用いて Excel の回帰分析の結果を示す。

表 13.34 最後の水準を基準としたデザイン変数

A	x_1	x_2		B	x_3		A×B		x_4	x_5
A ₁	0	0		B ₁	0		A ₁	B ₁	0	0
A ₂	1	0		B ₂	1			B ₂	0	0
A ₃	0	1		平均	1/2		A ₂	B ₁	0	0
平均	1/3	1/3						B ₂	1	0
							A ₃	B ₁	0	0
								B ₂	0	1

表 13.35 最初の水準を基準とする Excel による回帰係数の推定

——— デザイン行列 ———																
		— A—B		A×B		共変量		反応								
<i>i</i>	A	B	x_0	x_1	x_2	x_3	x_4	x_5	x_6	<i>y</i>	分散分析表					
2.01	A ₁	B ₁	1	0	0	0	0	0	350	970	自由度	変動	分散			
2.02			1	0	0	0	0	0	400	1000	回帰	7	14314030	2044861		
2.03			1	0	0	0	0	0	360	980	残差	8	670.41	83.8016		
2.04		B ₂	1	0	0	1	0	0	350	980	合計	15	14314700			
2.05			1	0	0	1	0	0	340	970						
2.06	A ₂	B ₁	1	1	0	0	0	0	390	990	係数	標準誤差	t	P-値		
2.07			1	1	0	0	0	0	340	950	切片	0	#N/A	#N/A	#N/A	
2.08		B ₂	1	1	0	1	1	0	410	980	X 値 1	727.049	32.8702	22.1188	0.0000	x_0
2.09			1	1	0	1	1	0	430	990	X 値 2	-9.8700	8.3682	-1.1795	0.2721	x_1
2.10			1	1	0	1	1	0	390	980	X 値 3	-11.4067	8.4026	-1.3575	0.2117	x_2
2.11	A ₃	B ₁	1	0	1	0	0	0	400	990	X 値 4	8.9832	8.6394	1.0398	0.3288	x_3
2.12			1	0	1	0	0	0	320	940	X 値 5	-26.8196	13.3170	-2.0139	0.0788	x_4
2.13		B ₂	1	0	1	1	0	1	330	930	X 値 6	-11.1697	12.4595	-0.8965	0.3962	x_5
2.14			1	0	1	1	0	1	390	1000	X 値 7	0.6927	0.0877	7.8996	0.0000	x_6
2.15			1	0	1	1	0	1	420	1000						

デザイン行列は、表 13.28 と表 13.35 は異なるので、パラメータの推定値も異なる。表 13.36 で示すパラメータの共分散行列は、表 13.29 と比較すると共変量 x_6 の分散 0.0077 は一致するが、他は全く異なる。

表 13.36 最初の水準を基準とした場合のパラメータの共分散行列

切片	— A —		B	A×B		共変量
x_0	x_1	x_2	x_3	x_4	x_5	x_6
1080.4513	-42.1571	-56.3803	-99.0499	227.0588	155.9427	-2.8446
-42.1571	70.0269	28.3183	28.8949	-72.5255	-29.6637	0.0384
-56.3803	28.3183	70.6035	29.8559	-33.3156	-73.2944	0.0769
-99.0499	28.8949	29.8559	74.6398	-83.2891	-78.4839	0.1922
227.0588	-72.5255	-33.3156	-83.2891	177.3416	94.0526	-0.5382
155.9427	-29.6637	-73.2944	-78.4839	94.0526	155.2380	-0.3460
-2.8446	0.0384	0.0769	0.1922	-0.5382	-0.3460	0.0077
パラメータの共分散行列 $\Sigma(\hat{\beta})=(X^T X)^{-1}\sigma^2$						

表 13.37 に最初の水準を基準とした共分散分析に対して、Lsmeans を求めるためのデザイン変数を示す。因子 A の Lsmeans を求めるために他の因子 B にはデザイン変数 x_3 の平均とし、共変量 x_6 はデータの平均とする。交互作用 A×B は、デザイン変数の掛け算で求める。因子 A の各水準の Lsmeans は、 $\hat{y}_{A_1} = x_{2.16}\hat{\beta} = 991.06$ として計算され、表 13.31 に示した lsmeansn パッケージの結果に一致する。因子 A₁ の標準誤差は、 $Var(\hat{y}_{A_1}) = x_{2.16}\Sigma(\hat{\beta})x_{2.16}^T = 19.7243$ の平方根で 4.4412 と推定されていて、表 13.31 の結果に一致する。

表 13.37 最初の水準を基準とした場合の Lsmeans を求めるためのデザイン変数

————— デザイン変数 —————										最小2乗		標準	
	— A —		B	A×B		共変量	反応	平均	分散	誤差			
<i>i</i>	A	B	x_0	x_1	x_2	x_3	x_4	x_5	x_6	y	y^\wedge	$Var(y^\wedge)$	SE
2.16	A ₁		1	0	0	0.50	0	0	374.67	-	991.06	19.7243	4.4412
2.17	A ₂		1	1	0	0.50	0.50	0	374.67	-	967.78	18.7249	4.3272
2.18	A ₃		1	0	1	0.50	0	0.50	374.67	-	974.07	17.6261	4.1983
2.19	B ₁		1	0.33	0.33	0	0	0	374.67	-	979.47	13.1335	3.6240
2.20	B ₂		1	0.33	0.33	1	0.33	0.33	374.67	-	975.79	10.9665	3.3116
2.21	A ₁	B ₁	1	0	0	0	0	0	374.67	-	986.57	28.1013	5.3011
2.22		B ₂	1	0	0	1	0	0	374.67	-	995.55	48.6673	6.9762
2.23	A ₂	B ₁	1	1	0	0	0	0	374.67	-	976.70	42.6192	6.5283
2.24		B ₂	1	1	0	1	1	0	374.67	-	958.86	37.5322	6.1264
2.25	A ₃	B ₁	1	0	1	0	0	0	374.67	-	975.16	43.5546	6.5996
2.26		B ₂	1	0	1	1	0	1	374.67	-	972.97	28.1526	5.3059

因子 B の各水準の Lsmeans は、因子 A のデザイン変数 (x_1 , x_2) の平均値 $1/3=0.3333$ とし、交互作用 A×B は、デザイン変数の掛け算で求める。このように謎めいた Lsmeans も Excel の行列計算の活用により、身近な統計量として使われるようになることを期待したい。

13.5. ポアソン回帰における最小 2 乗平均 (Lsmeans)

これまでの探索的なポアソン回帰の例示では、「最小 2 乗平均」ではなく「予測プロファイル」によって結果を示してきた。これは、最小 2 乗平均は、説明変数に質的変数 A と量的変数 X が混合するような場合に、量的変数 X の平均を \bar{X} としたときの質的変数 A の水準平均を推定する方法であり、量的変数については適用されないからである。JMP の予測プロファイルは、質的変数 A のある水準にセットした場合に量的変数のプロファイルを図示する機能が含まれていて探索的な解析の結果の解釈に役立つことを示してきた。

第 1.13 節で取り上げた「雌のカブトガニに連結する雄の数」は、質的変数 2 因子、量的変数 2 変数の対数リンクでのポアソン回帰の事例で、過分散となっていることを示した [アグレスティ (2003)]。第 7.2 節では、「カブトガニのサテライト数に対する探索的解析」において予測プロファイルの使い方を詳細に示した。表 13.38 に示すように質的変数として (甲羅の色, 後体部の棘), 量的変数として (甲羅の幅, 体重) があり, 反応変数はサテライト数である。

表 13.38 カブトガニの一タのリスト

No	甲羅の色	後体部の棘	甲羅の幅	体重	サテライト数
1	2:中ぐらい	3:両方破損	28.3	3.05	8
2	3:やや暗い	3:両方破損	22.5	1.55	0
3	1:やや明るい	1:正常	26.0	2.30	9
4	3:やや暗い	3:両方破損	24.8	2.10	0
5	3:やや暗い	3:両方破損	26.0	2.60	4
:					
171	1:やや明るい	1:正常	28.0	2.63	0
172	4:暗い	3:両方破損	27.0	2.63	0
173	2:中ぐらい	2:一方破損	24.5	2.00	0
		平均	26.2988	2.4372	2.9191

カブトガニのデータは、観察研究で得られたデータなので、表 13.39 に示すように、2 つの質的変数のクロス表のセルは、不揃いである。

表 13.39 甲羅の色, 後体部の棘の状態のクロス表

甲羅の色	後体部の棘			計
	1:正常	2:一方破損	3:両方破損	
1:やや明るい	9	2	1	12
2:中ぐらい	24	8	63	95
3:やや暗い	3	4	37	44
4:暗い	1	1	20	22
計	37	15	121	173

2つの量的変数（甲羅の幅，体重）の内，体重がサテライト数に対してより説明力があることが結果として示されているので，サテライト数を反応変数とし，甲羅の色，後体部の棘，および，体重を説明変数としたポアソン回帰を行い，最小2乗平均を求める．デザイン変数は，表 13.40 に示すように JMP のデフォルトの対比型である．

表 13.40 対比型デザイン変数

甲羅の色	a_1	a_2	a_3	後体部の棘	b_1	b_2
1:やや明るい	1	0	0	1:正常	1	0
2:中ぐらい	0	1	0	2:一方破損	0	1
3:やや暗い	0	0	1	3:両方破損	-1	-1
4:暗い	-1	-1	-1			

JMP によるポアソン回帰の結果を表 13.41 に示す．適合度統計量から過分散であるが，ここでは，Lsmeans の例示として用いるので無視する．

表 13.41 JMP によるポアソン回帰（対数リンク，過分散なし）

モデル全体の検定				
モデル	(-1)*対数尤度	尤度比カイ2乗	自由度	p値(Prob>ChiSq)
差分	41.5446	83.0892	6	<.0001*
完全	452.5001			
縮小	494.0447			
適合度統計量		カイ2乗	自由度	p値(Prob>ChiSq)
Pearson		533.4818	166	<.0001*
デビアン		549.7025	166	<.0001*
AICc				
919.6789				
効果の検定				
要因	自由度	尤度比カイ2乗	p値	
甲羅の色	3	9.7851	0.0205*	
後体部の棘	2	2.1025	0.3495	
体重	1	52.4907	<.0001*	
パラメータ推定値				
項	推定値	標準誤差	尤度比カイ2乗	p値
切片	-0.3979	0.1990	4.0206	0.0449*
甲羅の色[1:やや明るい]	0.3321	0.1313	6.0106	0.0142*
甲羅の色[2:中ぐらい]	0.0644	0.0735	0.7721	0.3796
甲羅の色[3:やや暗い]	-0.1888	0.0961	3.9353	0.0473*
後体部の棘[1:正常]	0.0233	0.0945	0.0607	0.8054
後体部の棘[2:一方破損]	-0.1374	0.1292	1.1923	0.2749
体重	0.5476	0.0732	52.4907	<.0001*

予測プロファイルを図 13.16 に示す。甲羅の色を 1:明るい，後体部の棘を 1:正常，体重を平均値の 2.4372 kg としたときの推定値が，3.6396 (2.7211, 4.8681) と推定されている。

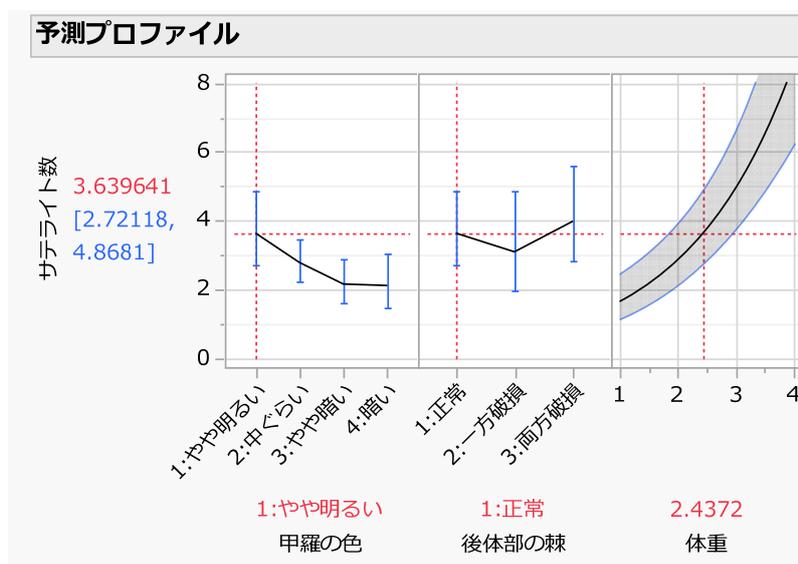


図 13.16 予測プロファイルによる結果の表示

表 13.42 に，JMP によって得られた推定値とパラメータの共分散行列を Excel に取り込んだ結果を示す。

表 13.42 JMP によるポアソン回帰の推定値とパラメータの共分散行列

項	推定値	共分散							
		切片	1:やや明るい	2:中ぐらい	3:やや暗い	1:正常	2:一方破損	体重	
切片	-0.3979	0.0396	-0.0005	0.0006	-0.0007	0.0031	0.0015	-0.0134	
甲羅の色									
1:やや明るい	0.3321	-0.0005	0.0172	-0.0028	-0.0056	-0.0034	-0.0001	0.0004	
2:中ぐらい	0.0644	0.0006	-0.0028	0.0054	0.0003	0.0006	-0.0001	-0.0012	
3:やや暗い	-0.1888	-0.0007	-0.0056	0.0003	0.0092	0.0018	-0.0007	0.0000	
後体部の棘									
1:正常	0.0233	0.0031	-0.0034	0.0006	0.0018	0.0089	-0.0095	-0.0022	
2:一方破損	-0.1374	0.0015	-0.0001	-0.0001	-0.0007	-0.0095	0.0167	0.0019	
体重	0.5476	-0.0134	0.0004	-0.0012	0.0000	-0.0022	0.0019	0.0054	

表 13.43 に，図 13.16 の予測プロファイルでの推定値と 95%信頼区間をされた表 13.42 の結果を元に計算した結果を示す。甲羅の色を 1:明るい，後体部の棘を 1:正常，体重を平均値の 2.4372 kg としたときの推定値が，3.6396 (2.7211, 4.8681) が得られている。JMP の予測プロファイルは，ここに示したように各因子の水準の組み合わせた場合を基本としている

表 13.43 因子の組み合わせ推定値と 95%信頼区間

	x_0	a_1	a_2	a_3	b_1	b_2	x	$\ln y^\wedge$	$Var(y^\wedge)$	y^\wedge	$L95\%$	$U95\%$
甲羅の色												
1:やや明るい	1	1	0	0	1	0	2.4372	1.2919	0.0220	3.6396	2.7212	4.8681
2:中ぐらい	1	0	1	0	1	0	2.4372	1.0242	0.0125	2.7849	2.2367	3.4674
3:やや暗い	1	0	0	1	1	0	2.4372	0.7710	0.0221	2.1620	1.6156	2.8931
4:暗い	1	-1	-1	-1	1	0	2.4372	0.7522	0.0342	2.1217	1.4768	3.0482
後体部の棘												
1:正常	1	1	0	0	1	0	2.4372	1.2919	0.0220	3.6396	2.7212	4.8681
2:一方破損	1	1	0	0	0	1	2.4372	1.1312	0.0530	3.0994	1.9739	4.8669
3:両方破損	1	1	0	0	-1	-1	2.4372	1.3827	0.0301	3.9858	2.8375	5.5989

表 13.44 に互いの因子についてセル平均ゼロとした最小 2 乗平均を示す。表 13.43 と大きく異なるのは、ほとんどの分散が小さくなることである。これは、名目のサンプル数が増加することによる影響である。JMP でのポアソン回帰の場合には、この「最小 2 乗平均」は、現在は求めることができないが、SAS/GENMOD プロシジャでは、Lsmeans ステートメントによって求めることができる。

表 13.44 甲羅の色および体部の棘についての最小 2 乗平均の推定値と 95%信頼区間

	x_0	a_1	a_2	a_3	b_1	b_2	x	$\ln y^\wedge$	$Var(y^\wedge)$	y^\wedge	$L95\%$	$U95\%$
甲羅の色												
1:やや明るい	1	1	0	0	0	0	2.4372	1.2686	0.0243	3.5559	2.6198	4.8265
2:中ぐらい	1	0	1	0	0	0	2.4372	1.0009	0.0068	2.7208	2.3140	3.1992
3:やや暗い	1	0	0	1	0	0	2.4372	0.7477	0.0140	2.1122	1.6754	2.6630
4:暗い	1	-1	-1	-1	0	0	2.4372	0.7290	0.0274	2.0729	1.4981	2.8682
体部の棘												
1:正常	1	0	0	0	1	0	2.4372	0.9598	0.0108	2.6112	2.1300	3.2012
2:一方破損	1	0	0	0	0	1	2.4372	0.7992	0.0351	2.2237	1.5406	3.2098
3:両方破損	1	0	0	0	-1	-1	2.4372	1.0507	0.0050	2.8596	2.4880	3.2867

いずれにしても、「最小 2 乗平均」は SAS および JMP ユーザの方言であり、「予測プロファイル」は、JMP ユーザの方言ではあるが、探索的な解析における結果の表示方法として優れた推定方法である。

これまでも示してきたように「予測プロファイル」は、「最小 2 乗平均」包含する概念であり、探索的な解析のための基本ツールとの認識し、JMP と同様の予測プロファイルを Excel で作成する手順を示してきた。

第13章 文献索引

アグレスティ著, 渡邊・菅波・吉田・角野・寒水・松永 訳(2003) - カテゴリカルデー解析入門	457
魚住(2014) - LS-Means再考 - GLMとPLMによるモデル推定後のプロセス	455
奥野ら(1981) - 多変量解析法 改訂版	423, 440
杉本・橋田(2011) - 共分散分析の基礎・医薬品開発における共分散分析の例	432
高橋(2019b) - 投与前値がある場合の解析のレビュー	425
高橋・大橋・芳賀(1989) - SASによる実験データの解析	421, 454
竹内ら(1989) - 統計学辞典	421
芳賀(2009) - 医薬品開発のための統計解析 第2部 実験計画法	432
守屋・広岡(2018) - Rパッケージを用いた最小2乗分散分析と最小2乗平均値の算出	449
Littleら(2002) - SAS for Linear Models	455

第13章 索引

<p>あ R - lsmeansパッケージ 449</p> <p>Rパッケージ - 守屋ら(2018) 449</p> <p>アグレスティ(2003) - カブトガニ 457</p> <p>at オプション - Lsmeansステートメント 439</p> <p>R言語 - 最初の水準を基準 455</p> <p style="padding-left: 20px;">- デザイン変数 455</p> <p>一般用語ではない - 予測プロファイル 448</p> <p>魚住(2014) - LS-Means再考 455</p> <p>Excel - Minverse() 関数 430</p> <p style="padding-left: 20px;">- 共分散分析 442</p> <p style="padding-left: 20px;">- 交互作用 429</p> <p style="padding-left: 20px;">- Transpose() 関数 430</p> <p style="padding-left: 20px;">- Mmult() 関数 430</p> <p style="padding-left: 20px;">- 予測プロファイル 421, 432, 459</p> <p>Estimateステートメント - GLMプロシジヤ 428</p> <p>Minverse() 関数 - Excel 430</p> <p>Lsmeans - 最初の水準を基準 456</p> <p style="padding-left: 20px;">- 最小2乗平均 421, 437</p> <p style="padding-left: 20px;">- 総平均 438</p> <p style="padding-left: 20px;">- ポアソン回帰 457</p> <p style="padding-left: 20px;">- 方言 421</p> <p style="padding-left: 20px;">- 予測プロファイル 438</p> <p>Lsmeansステートメント - at オプション 439</p> <p>Lsmeansの推定値 - デザイン変数 451</p> <p>lsmeansパッケージ - R 449</p> <p style="padding-left: 20px;">- 最小2乗平均 452</p> <p>LS-Means再考 - 魚住(2014) 455</p> <p>応答局面法 - JMP 441</p> <p style="padding-left: 20px;">- 等高線プロファイル 441</p> <p>奥野ら(1981) - 層別因子 423</p> <p style="padding-left: 20px;">- 電気特性 440</p> <p style="padding-left: 20px;">- 魅力的な事例 422</p>	<p>か 共分散行列 - 交互作用 429</p> <p style="padding-left: 20px;">- 重回帰分析 421</p> <p style="padding-left: 20px;">- パラメータ 430</p> <p>共分散分析 - アーミテージら(2001) 424</p> <p style="padding-left: 20px;">- Excel 442</p> <p style="padding-left: 20px;">- 共変量が2変量 440</p> <p style="padding-left: 20px;">- 狭義の意味での 425</p> <p style="padding-left: 20px;">- 交互作用 423, 426</p> <p style="padding-left: 20px;">- 杉本・橋田(2011) 432</p> <p style="padding-left: 20px;">- スネデガー・コ克蘭(1972) 424</p> <p style="padding-left: 20px;">- 伝統的 424</p> <p style="padding-left: 20px;">- 芳賀(2009) 432</p> <p>共変量 - 可視化 447</p> <p style="padding-left: 20px;">- 部品寸法 448</p> <p>共変量が2変量 - 共分散分析 440</p> <p>共変量の影響 - 電気特性 447</p> <p>狭義の意味での - 共分散分析 425</p> <p>分散分析表 - 誤差分散 430</p> <p>グラフ・ビルダー - 層別散布図 454</p> <p>繰返し不揃い - 2因子の共分散分析 449</p> <p>交互作用 - Excel 429</p> <p style="padding-left: 20px;">- 95%信頼区間 433</p> <p style="padding-left: 20px;">- 共分散行列 429</p> <p style="padding-left: 20px;">- 共分散分析 423, 426</p> <p style="padding-left: 20px;">- 推定値 433</p> <p style="padding-left: 20px;">- 積 428</p> <p style="padding-left: 20px;">- デザイン変数 428</p> <p style="padding-left: 20px;">- デザイン変数間の積 450</p> <p style="padding-left: 20px;">- 分散分析表 424, 430</p> <p>交互作用モデル - 主効果モデル 431</p> <p>甲羅の色 - 最小2乗平均 457</p> <p>誤差分散 - 分散分析表 430</p> <p>50%程度 - 二変量の関係 440</p> <p>50%程度の緩い - 確率楕円 441</p> <p>Contrustステートメント - GLMプロシジヤ 428</p>
<p>か 回帰分析 - 層別因子を含む 423</p> <p style="padding-left: 20px;">- 炉A4を基準 446</p> <p>回収液の濃度の差 - 予測プロファイル 436</p> <p>確率楕円 - 50%程度 441</p> <p>可視化 - 共変量 447</p> <p style="padding-left: 20px;">- 最小2乗平均 445</p> <p>カブトガニ - アグレスティ(2003) 457</p> <p>95%信頼区間 - 交互作用 433</p> <p style="padding-left: 20px;">- 差の推定値 435</p> <p>共分散分析 - 投与前値 425</p>	<p>さ 最後の水準 - デザイン変数 454</p> <p>最初の水準を基準 - R言語 455</p> <p style="padding-left: 20px;">- Lsmeans 456</p> <p>最小2乗分散分析法 - 不釣り合い型データ 449</p> <p>最小2乗平均 - Lsmeans 421, 437</p> <p style="padding-left: 20px;">- lsmeansパッケージ 452</p> <p style="padding-left: 20px;">- 可視化 445</p>

さ	- 甲羅の色	457	た	デザイン変数間の積 - 交互作用	450
	- SASユーザの方言	460		電気特性 - 共変量の影響	447
	- 算術平均	444		伝統的 - 共分散分析	424
	- JMPユーザの方言	460		電気特性 - 奥野ら(1981)	440
	- 高橋ら(1989)	421		電気特性対 - ボックス・プロット	443
	- 竹内ら(1989)	421		等高線プロファイル - 応答局面法	441
	- 調整済み平均	443		投与前値 - 共分散分析	425
	- ポアソン回帰	457		- 高橋(2019b)	425
	- 方言	421		動的なグラフ - 予測プロファイル	422
	- 守屋ら(2018)	422, 449		Transpose() 関数 - Excel	430
	- 予測プロファイル	421, 438, 453	な	2因子の共分散分析 - 繰返し不揃い	449
SASユーザの方言 - 最小2乗平均	460			2水準間の差 - デザイン変数	434
差の推定値 - 95%信頼区間	435			二変量の関係 - 50%の確率楕円	440
算術平均 - 最小2乗平均	444		は	芳賀(2009) - 共分散分析	432
GLMプロシジャ - Estimeteステートメント	428			パラメータ - 共分散行列	430
- Contrastステートメント	428			パラメータ関数 - 対比	428
- デザイン変数	454			不釣り合い型データ - 最小2乗分散分析法	449
杉本・橋田(2011) - 共分散分析	432			部品寸法 - 共変量	448
JMP - 応答局面法	441			分散分析表 - 交互作用	424
JMPユーザの方言 - 最小2乗平均	460			- 交互作用	430
- 予測プロファイル	460			ポアソン回帰 - Lsmeans	457
JMP - ポアソン回帰	458			- 最小2乗平均	457
重回帰分析 - 共分散行列	421			- JMP	458
- 相関行列	421			- 予測プロファイル	421
- 予測プロファイル	421			方言 - Lsmeans	421
推定値 - 交互作用	433			- 最小2乗平均	421
水準間の差 - デザイン変数	445			ボックス・プロット - 電気特性対	443
- 予測プロファイル	434		ま	Mmult() 関数 - Excel	430
水準間の差の推定 - 対比	427			魅力的な事例 - 奥野ら(1981)	422
スネデガー・コ克蘭(1972) - 共分散分析	424			守屋ら(2018) - Rパッケージ	449
積 - 交互作用	428			- 最小2乗平均	422, 449
洗浄水の温度 - 予測プロファイル	435		や	予測プロファイル - 一般用語ではない	448
層別因子 - 奥野ら(1981)	423			- Excel	421, 432, 459
層別因子を含む - 回帰分析	423			- Lsmeans	438
層別散布図 - グラフ・ビルダー	454			- 回収液の濃度の差	436
相関行列 - 重回帰分析	421			- 最小2乗平均	421, 438, 453
総平均 - Lsmeans	438			- JMPユーザの方言	460
Type I の平方和 - 逐次型	431			- 重回帰分析	421
た Type II の平方和 - 主効果モデル	431			- 水準間の差	434
Type III の平方和 - JMP	431			- 洗浄水の温度	435
対比 - 水準間の差の推定	427			- デフォルト	457
- パラメータ関数	428			- 動的なグラフ	422
対比型 - デザイン変数	428, 450, 458			- ポアソン回帰	421
高橋(2019b) - 投与前値	425			4種の平方和 - 高橋ら(1989)	455
高橋ら(1989) - 最小2乗平均	421			- Littleら(2002)	455
- 4種の平方和	455		ら	Littleら(2002) - 4種の平方和	455
竹内ら(1989) - 最小2乗平均	421			炉A4を基準 - 回帰分析	446
調整済み平均 - 最小2乗平均	443			- デザイン変数	446
デフォルト - 予測プロファイル	457				
デザイン変数 - R言語	455				
- Lsmeansの推定値	451				
- 交互作用	428				
- 最後の水準	454				
- GLMプロシジャ	454				
- 水準間の差	445				
- 対比型	428, 450, 458				
- 2水準間の差	434				
- 炉A4を基準	446				

第 13 章 Excel, JMP, SAS ファイル一覧

	9 KB	第13章02_層別共分散	JMP Data Table
	78 KB	第13章02_層別共分散	Microsoft Excel ワークシート
	14 KB	第13章03_2変量共分散	JMP Data Table
	58 KB	第13章03_2変量共分散	Microsoft Excel ワークシート
	8 KB	第13章04_守屋_2因子共変量	JMP Data Table
	60 KB	第13章04_守屋_2因子共変量	Microsoft Excel ワークシート
	17 KB	第13章05_カプトガニ	JMP Data Table
	26 KB	第13章05_カプトガニ_プロファイル	Microsoft Excel ワークシート

非売品, 無断複製を禁ずる

第 9 回 続高橋セミナー

最尤法によるポアソン回帰分析入門<第 13 章>

第 13 章 最小 2 乗平均の謎を予測プロファイルで解く

BioStat 研究所(株)

〒105-0014 東京都 港区 芝 1-12-3 の 1005

2020 年 7 月 14 日 高橋 行雄