

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 2 月 7 日

第 7 章 共変量を含む 3 因子実験データの探索的解析

これまで取り上げてきたほとんど事例は、1 つの質的変数と 1 つの量的変数の場合であった。現実には、複数の質的変数と複数の量的変数を同時に取り上げ、一気に問題解決を図りたい。多くの実験研究あるいは観察研究の場で、対象分野の特質により交互作用よりも主効果に重きを置く分野、常に交互作用を念頭に置く分野があり、実験計画の考え方が異なり、解析の考え方も異なる。宮川 (2008), 「問題発見の解決の科学—SQC の基本—, 第 6.3 節 計画的に採取した 3 元配置データの解析事例」で取り上げられている「冷蔵庫の塗装不良の改善」の実験データを取り上げる。原書で実践されているのは、層別散布図に回帰直線を重ね書きした図を主体にしたアプローチである。そこで、Excel による格子状の層別散布図の作成方法を示すと共に、Excel による探索的な回帰分析と結果のグラフ化による問題解決法を例示する。さらに、JMP による交互作用も考慮した変数選択機能を活用した探索的な解決法も合わせて示す。

第 7 章 目 次

7.	共変量を含む 3 因子実験データの探索的解析 -----	219
7.1.	Excel の散布図を主体にした問題解決-----	219
	2×2 格子状の層別散布図に回帰直線の重ね書き, 共変量 X の分布の 均一性の確認, 2×2 格子状の層別散布図の併合	
7.2.	回帰直線に対する個別データの 95%信頼区間 -----	224
	交互作用を含む線形モデル, 個別データの 95%信頼区間, 色 B 別の個別データの 95%信頼区間	
	次ページ続く	

7.3. Excel の回帰分析による交互作用を含む変数減少法-----	230
ダミー変数, 交互作用を含む線形モデル, 変数選択の手順, 推定されたパラメータを用いた層別回帰直線の推定	
7.4. JMP のステップワイズ法による変数選択 -----	236
手作業での変数選択, 自動的な変数選択, 質的変数と量的変数の 交互作用を含む変数選択, 変数選択の結果をさらに深掘する	
文献索引, 索引, 解析ファイル一覧 -----	(243)

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)

はじめに -----	1
1. 層別因子を含む各種の回帰分析の実際 -----	7
2. デザイン行列を活用した 1 因子実験データの解析 -----	55
3. 繰り返しが不揃いの 2 因子実験データの解析 -----	89
4. 欠測値がある直交表の線形モデルによる解析 -----	121
5. デザイン行列を用いた回帰分析の基礎 -----	155
6. 伝統的な共分散分析からの脱却 -----	195
7. 共変量を含む 3 因子実験データの探索的解析 -----	219
8. 交絡変数と共変量を含む 2 群比較 -----	243
9. 前後差データの群間比較に潜む前値の影 -----	281
10. 層別因子を含むロジスティック曲線のあてはめ -----	327
11. 各種のシグモイド曲線を用いた逆推定 -----	367
12. ミカエリス・メンテン式をめぐる新たな統計解析 -----	405
13. ロジスティック曲線のさらなる活用 -----	455
文献, 文献索引, 索引, 解析用ファイル一覧 -----	489

7. 共変量を含む 3 因子実験データの探索的解析

これまで取り上げてきたほとんど事例は、1 つの質的変数と 1 つの量的変数の場合であった。現実には、複数の質的変数と複数の量的変数を同時に取り上げ、一気に問題解決を図りたい。多くの実験研究あるいは観察研究の場で、対象分野の特質により交互作用よりも主効果に重きを置く分野、常に交互作用を念頭に置く分野があり、実験計画の考え方が異なり、解析の考え方も異なる。宮川（2008）、「問題発見の解決の科学—SQC の基本—、第 6.3 節 計画的に採取した 3 元配置データの解析事例」で取り上げられている「冷蔵庫の塗装不良の改善」の実験データを取り上げる。原書で実践されているのは、層別散布図に回帰直線を重ね書きした図を主体にしたアプローチである。そこで、Excel による格子状の層別散布図の作成方法を示すと共に、Excel による探索的な回帰分析と結果のグラフ化による問題解決法を例示する。さらに、JMP による交互作用も考慮した変数選択機能を活用した探索的な解決法も合わせて示す。

7.1. Excel の散布図を主体にした問題解決

表 7.1 に示すのは、ある大手の家電メーカーでの現業実習で遭遇した事例であり、冷蔵庫の塗装不良の改善のために、取り上げられた因子は、(シンナーの種類, 塗装の色, 冷蔵庫の機種) であり、塗装の中心膜厚 y を測定し、共変量としてガンスプレーの吐出量 x が測定された。因子は各 2 水準で全ての組み合わせに対し 10 回の塗装が反復して行われた。それぞれの因子の組合せで 10 回の反復の平均と SD を示してある。中心膜厚の管理限界は、(28~36 単位) であり、管理限界を超えた場合には不良品となる。シンナー1 の場合には、40 個のデータ中 3 件の管理下限を下回る不良品が発生し、シンナー2 の場合には、40 個のデータ中 3 件の管理上限を上回る不良品が発生していることが示されている。

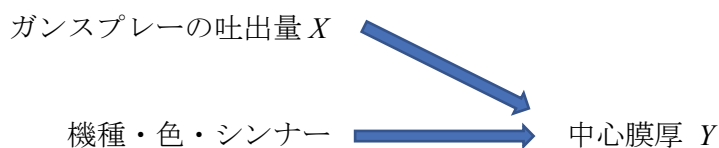
共変量がなければ、各セルの 10 個中の不良品の割合に注目し、不良の割合に対する解析で済ませたくなる。シンナー1・色 2 の場合の機種（国内・国外）に関わらず管理限界の下側を超える不良が発生し、シンナー2・色 1 の場合には逆に機種に関わらず管理限界の上側を超え

表 7.1 冷蔵庫の塗装データ [宮川 (2008), 表 6.1]

	機種: 国内					機種: 国外				
	色1		色2		注	色1		注	色2	
シンナー	x	y	x	y		x	y		x	y
1	91.3	33.5	88.6	29.2		89.0	31.1		90.9	30.7
	89.3	32.7	88.4	27.6	▼	90.1	30.7		88.0	28.5
	88.9	31.6	89.7	28.8		90.3	33.4		90.7	29.3
	90.4	32.6	90.4	29.7		89.6	32.1		92.9	31.2
	88.5	30.3	91.2	28.8		90.7	32.0		92.6	30.4
	92.6	33.0	89.3	28.0	▼	92.2	32.2		89.0	29.9
	91.3	31.4	91.7	31.0		88.3	33.2		89.3	29.2
	91.5	32.4	92.3	29.8		90.4	31.3		88.7	28.4
	87.7	31.2	89.7	30.3		92.2	33.7		90.0	28.0
	89.7	31.5	91.2	30.1		87.5	30.2		91.5	29.4
平均	90.12	32.02	90.25	29.33		90.03	31.99		90.36	29.50
SD	1.55	0.98	1.32	1.06		1.52	1.18		1.65	1.05
2	89.6	32.3	90.7	31.2		91.0	33.3		89.7	29.7
	90.8	35.0	91.0	32.4		91.6	34.0		92.0	31.6
	88.7	32.8	89.3	31.0		89.7	35.0		89.6	31.7
	88.8	31.7	92.4	34.3		92.4	36.7	△	91.3	33.4
	90.6	35.6	92.7	33.7		90.4	34.2		91.6	32.2
	89.3	33.1	89.7	29.1		91.2	36.1	△	91.8	30.8
	92.4	35.4	89.2	30.2		90.4	32.6		90.3	30.1
	89.1	33.9	90.4	30.1		91.0	34.3		91.5	30.3
	89.7	33.7	90.1	32.1		89.3	33.0		90.1	30.9
	92.2	36.1	88.9	28.9		90.2	33.5		91.0	31.5
平均	90.12	33.96	90.44	31.30		90.72	34.27		90.89	31.22
SD	1.34	1.51	1.30	1.82		0.91	1.33		0.89	1.10

注) △:流れ不良 (y が36以上), ▼:ウス不良 (y が28以下)

る不良が発生している. このような不良の発生原因の探求し, 不良品の発生を防止するための方法を見いだしたい. 中心膜厚 Y の平均値も不良の発生に関連して上下することが確認でき, 不良発生の代用特性であることが推測される.

図 7.1 吐出量 X などが中心膜厚 Y に与える影響

2×2 格子状の層別散布図に回帰直線の重ね書き

塗装に用いるガンスプレーの吐出量 X が, 中心膜厚 Y に対する共変量であることの確認は, X と Y の散布図を描き, 回帰直線を描くことにより確認できる. 表 7.1 のデータの並びに対応し, Excel で作成した 2×2 の格子状の層別散布図を図 7.2 に示す. この図の作成は, まず, 左上段に 機種: 国内・色 1・シンナー 1 の散布図を作成, シンナー 2 を重ね書きし, マーカの種類と色を整え, 「近似曲線の追加」により回帰直線を引く. これをコピーし 2×2 の格子状に

並べ、コピーしたそれぞれの散布図に対して、散布図の点をマウスで選択すると（ x の欄， y の欄）の最初の参照先が Excel シート上で色別し表示されるので、それらに対応する位置にマウスで移動する．すると自動的に散布図の点が新たに表示され、回帰直線と回帰式も更新される．

管理限界の上限および下限を示す点線は、Excel シート上に始点と終点の（ x ， y ）の座標を与え、「データの選択」により追加で重ね書きし点線に変更している．各々の図に対する説明は、Excel シートのセルの幅を図の大きさに調整して説明を加える．ワードに張り付ける際には、 2×2 格子状の図を含むセル全体をコピーすることで全体のバランスを保つことができる．

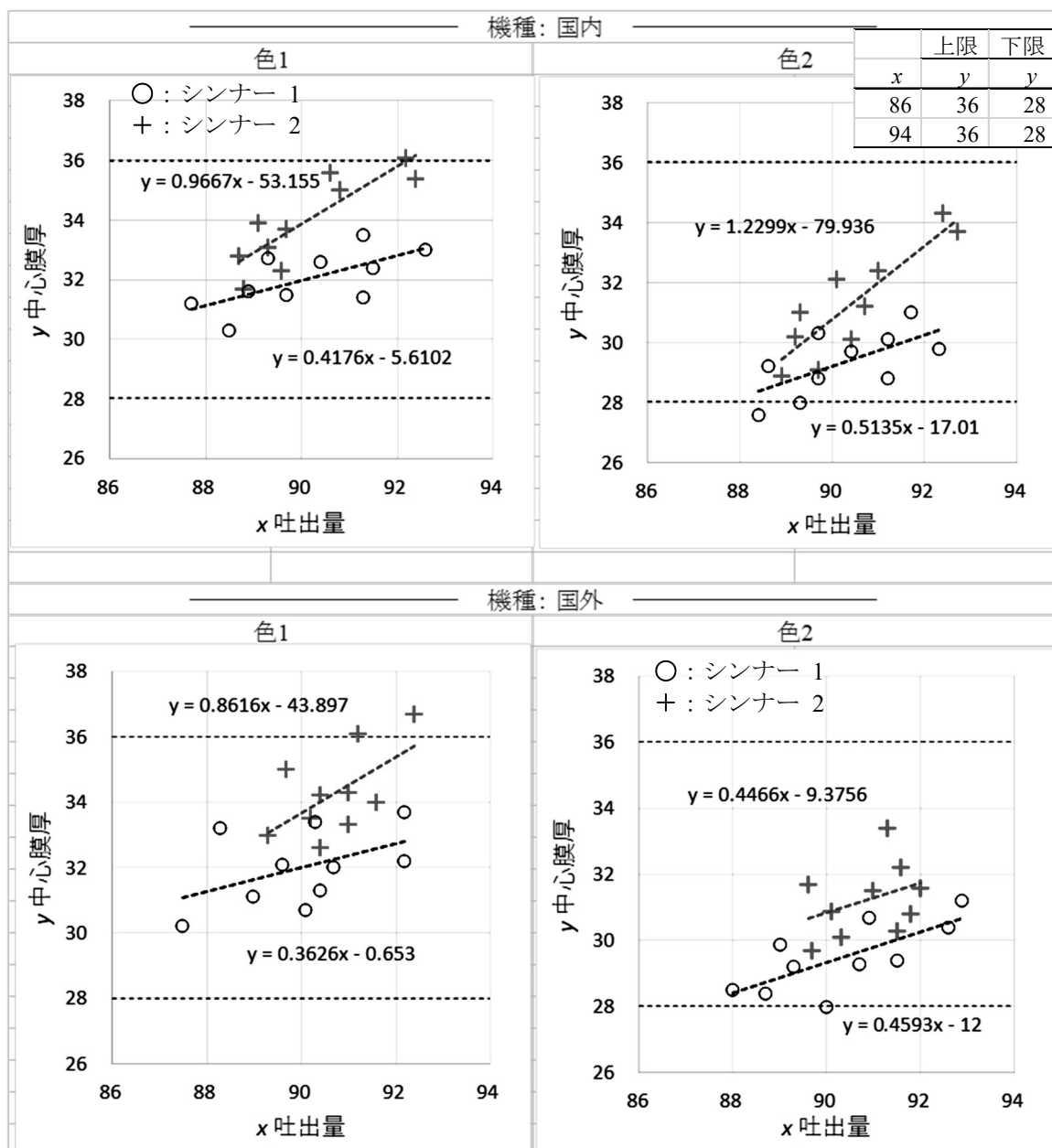


図 7.2 Excel による格子状の層別散布図行列に回帰直線の重ね書き

図7.2から、吐出量 X と中心膜厚 Y の関係は、直線的であることが確認できる。また、色1は、色2に対して中心膜厚 Y が厚めで、シンナー2の場合に中心膜厚が36を超える流れ不良が発生している。色2でシンナー1の場合には、中心膜厚が28を下回るウス不良が発生している。シンナー1は、シンナー2に対して散布図上で回帰直線の傾きが小さ目である。機種（国内、国外）を色とシンナーで対応づけて比較すると、ほぼ同様の散布図であることも確認される。

これらの視察から、色2でシンナー1の場合は、吐出量を多めに、色1でシンナー2の場合には、吐出量を少なめにすることが、不良を出にくくするために必要と思われる。これらの格子状の層別散布図から得られた結果から、3因子間の交互作用はないと判断されるが、シンナーと吐出量 X には、交互作用が統計的にあることを確認する必要がある。

共変量 X の分布の均一性の確認

ガンスプレーの吐出量 X は、中心膜厚 Y に対して影響を与える共変量であることが、図7.2で確認することができた。また、各実験条件の下で吐出量 X の範囲が微妙に異なるようでもある。そこで、図7.3に $2 \times 2 \times 2$ の各実験条件の下での吐出量 X を散布図にし、平均 \pm SDを重ね書きした結果を示す。なお、この図は、表7.1のデータを行方向に並べ直し、 $2 \times 2 \times 2$ の因子の組み合わせに対して1, 2, ..., 8の組み合わせ番号を付与し散布図を作成している。さらに、平均とSDを計算し、平均 \pm SDを求めて散布図に重ね書きし、デフォルトのマーカーを「データ系列の書式設定」によって「+」などに変更し、全体を点線で結んだ後に一部を消している。

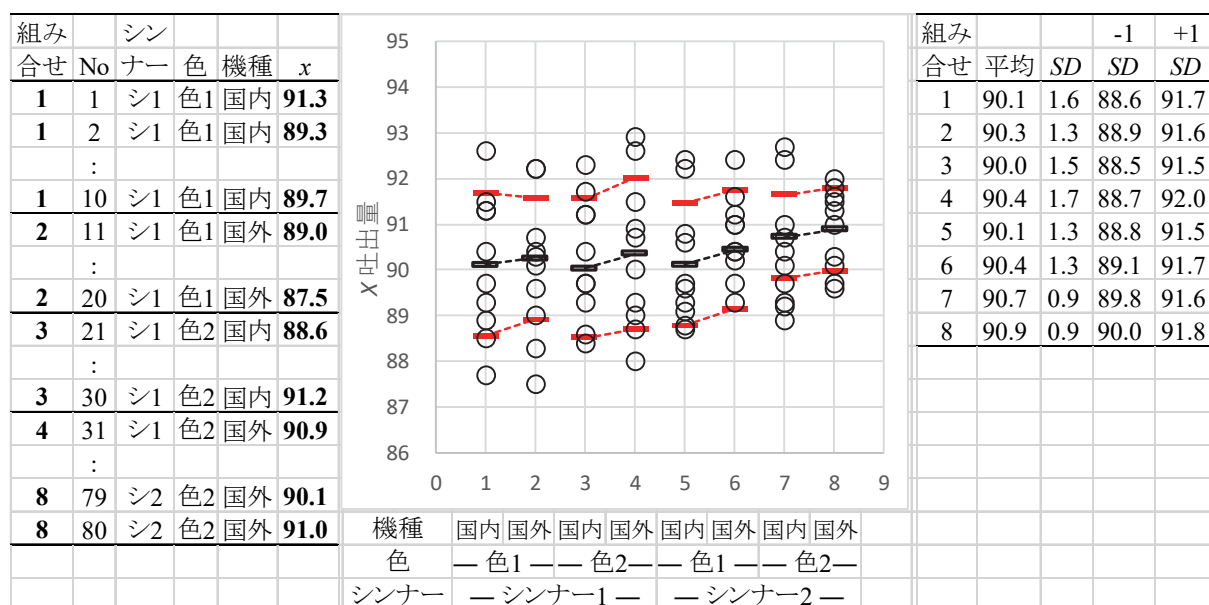


図7.3 ガンスプレーの吐出量 X のExcelの散布図に平均 \pm SDを重ね書き

2×2 格子状の層別散布図の併合

闇雲に散布図を作るのではなく、どのような散布図が問題解決の糸口になるか深い洞察力が必要である。図 7.2 から、冷蔵庫の機種（国内、国外）による差異が相対的に小さいので、表 7.1 を表 7.2 のように、行方向に機種（国内、国外）を展開し、それらを併合した 2×1 格子状の層別散布図を図 7.4 示す。

表 7.2 機種（国内、国外）を行方向に並べ替え

	シンナー:1				シンナー:2			
	色1		色2		色1		色2	
機種	x	y	x	y	x	y	x	y
国内	91.3	33.5	88.6	29.2	89.6	32.3	90.7	31.2
	89.3	32.7	88.4	27.6	90.8	35.0	91.0	32.4
	⋮				⋮			
	89.7	31.5	91.2	30.1	92.2	36.1	88.9	28.9
国外	89.0	31.1	90.9	30.7	91.0	33.3	89.7	29.7
	⋮				⋮			
	92.2	33.7	90.0	28.0	89.3	33.0	90.1	30.9
	87.5	30.2	91.5	29.4	90.2	33.5	91.0	31.5

図 7.4 から色 1 に対するシンナー 2 の場合に中心膜厚が 36 を超えて「流れ不良」発生し、色 2 に対するシンナー 1 の場合に中心膜厚が 28 以下になる「ウス不良」発生していることが鮮明になり、不良を発生しないような解決方法の糸口が見いだされる。Excel では、散布図に対する回帰直線の重ね書きは容易にできるのであるが、残念ながら回帰直線の 95%信頼区間を描く機能はない。そのためには、きちっとした統計解析のスキルが必要であり、グラフ表示のための作図のスキルも必要となる。

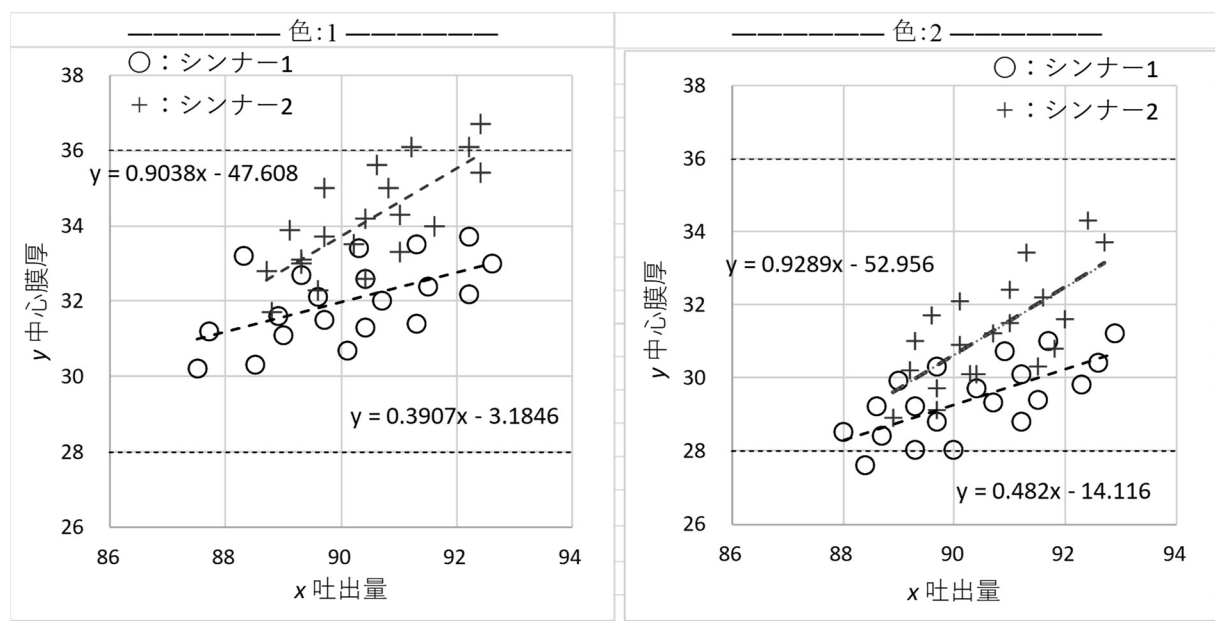


図 7.4 機種（国内、国外）を合わせた色別のシンナーに関する層別散布図

7.2. 回帰直線に対する個別データの95%信頼区間

管理限界がある場合の統計的な判断をするためには、それぞれの回帰直線に対する個別データの95%信頼区間の重ね書きが必要である。図7.2および図7.4でExcelの近似曲線の機能を用いて回帰直線を重ね書きしたのであるが、さらに個別データの95%信頼区間を重ね書きしたい。Excelによる回帰直線の95%信頼区間の重ね書きによる作図は、第1.2節の図1.9に2本の傾きが異なる直線に対する事例、第5.5節の図5.1による個別データの95%信頼区間の事例、第6.2節の図6.6により平行な直線に対する事例を示してきた。図7.4は、層別因子としての「色別」に2種類の「シンナー」に対する回帰直線が重ね書きされ、全体で4本の回帰直線となっている。この4本の回帰直線についての個別データについての個別データの同時95%信頼区間を重ね書きし、統計的な考察をしたい。

交互作用を含む線形モデル

前節の格子状の層別散布図に対する回帰直線の重ね書き図による検討から、線形モデルで取り上げる変数を、シンナーA（シンナーA₁: $a=0$, シンナーA₂: $a=1$ ）、色B（色B₁: $b=0$, 色B₂: $b=1$ ）のように（0, 1）型ダミー変数で与え、ガンスプレーの吐出量 x 、シンナーの種類 a と吐出量 x の交互作用のための変数を ax とする。線形モデルを、

$$y_i = \theta_0 + \theta_1 a_i + \theta_2 b_i + \theta_3 x_i + \theta_4 (ax)_i + \varepsilon_i \quad (7.1)$$

とし、表7.3に解析用のデータを示す。

表 7.3 交互作用を含む解析用のデザイン行列 X

No.	シンナー A	色 B	機種 C	膜厚 y	デザイン行列 X				
					x_0	a	b	x	ax
1	1	1	国内	33.5	1	0	0	91.3	0.0
⋮									
20	1	1	国外	30.2	1	0	0	87.5	0.0
21	1	2	国内	29.2	1	0	1	88.6	0.0
⋮									
40	1	2	国外	29.4	1	0	1	91.5	0.0
41	2	1	国内	32.3	1	1	0	89.6	89.6
⋮									
60	2	1	国外	33.5	1	1	0	90.2	90.2
61	2	2	国内	31.2	1	1	1	90.7	90.7
⋮									
80	2	2	国外	31.5	1	1	1	91.0	91.0

デザイン行列 X の設定は、最初の1行目で切片の変数 x_0 に対し「1」とし、変数 a は、Excelのif()関数を用い「if((シンナーA=1), 0, 1)」によって（0, 1）型ダミー変数とし、変数

b も同様に「if ((色 B=1) , 0, 1) 」と (0, 1) 型ダミー変数を生成する. 交互作用 ax は, 「 $a*x$ 」と積で与える. 2行目から80行目は, 変数 x を除いてフィルハンドルで数式のコピー機能でセットする. このような操作により, (80行×5列) のデザイン行列 X を作成する. 表7.4左にExcelの回帰分析による分散分析表とパラメータの推定値を示す.

表7.4 デザイン行列 X に対する回帰分析の結果

	シ	色	機		デザイン行列 X						分散分析表 (切片に 0 を使用) off				
No	A	B	種	y	x_0	a	b	x	ax			自由度	変動	分散	分散比
1	1	1	内	33.5	1	0	0	91.3	0.0		回帰	4	282.7236	70.6809	84.3497
:											残差	75	62.8463	0.8380	
20	1	1	外	30.2	1	0	0	87.5	0.0		合計	79	345.5699		
21	1	2	内	29.2	1	0	1	88.6	0.0						
:															
40	1	2	外	29.4	1	0	1	91.5	0.0	θ^{\wedge}_0	切片	-7.5835	9.0466	-0.8383	0.4045
41	2	1	内	32.3	1	1	0	89.6	89.6	θ^{\wedge}_1	a	-40.3310	14.8599	-2.7141	0.0082
:										θ^{\wedge}_2	b	-2.8842	0.2056	-14.0253	0.0000
60	2	1	外	33.5	1	1	0	90.2	90.2	θ^{\wedge}_3	x	0.4406	0.1004	4.3904	0.0000
61	2	2	内	31.2	1	1	1	90.7	90.7	θ^{\wedge}_4	ax	0.4656	0.1643	2.8329	0.0059
:															
80	2	2	外	31.5	1	1	1	91.0	91.0						

パラメータ (係数) の推定値から, 交互作用 ax と主効果 b が共に有意な差であることが確認される. なお, 主効果 a および x については, 交互作用 ax がモデルに含まれているので, 慎重に取り扱わなければならない. 特に主効果 a の推定値 $\hat{\theta}_1 = -40.3310$ は, 共変量が, $x=0$ における切片の差の推定値であることを認識する必要がある. この回帰分析の結果の解釈のためには, 推定されたパラメータを用いて, 回帰式を明示し図に重ね書きする必要がある. パラメータの推定値から, 因子Aと因子Bの組み合わせによる回帰式は,

$$\left. \begin{aligned}
 \text{シンナーA}_1, \text{色B}_1: \hat{y}^{(1,1)} &= \hat{\theta}_0 + \hat{\theta}_3 x = -7.5835 + 0.4656x \\
 \text{シンナーA}_1, \text{色B}_2: \hat{y}^{(1,2)} &= (\hat{\theta}_0 + \hat{\theta}_2) + \hat{\theta}_3 x = -10.4677 + 0.4656x \\
 \text{シンナーA}_2, \text{色B}_1: \hat{y}^{(2,1)} &= (\hat{\theta}_0 + \hat{\theta}_1) + (\hat{\theta}_3 + \hat{\theta}_4)x = -47.9145 + 0.9061x \\
 \text{シンナーA}_2, \text{色B}_2: \hat{y}^{(2,2)} &= (\hat{\theta}_0 + \hat{\theta}_1 + \hat{\theta}_2) + (\hat{\theta}_3 + \hat{\theta}_4)x = -50.7987 + 0.9061x
 \end{aligned} \right\} \quad (7.2)$$

として得られる. なお, これらの回帰式は, 図7.4で示した色別・シンナー別に求めた個別データから求めた回帰式とは, 微妙に異なる. これは, それぞれの組み合わせごとのデータによる回帰式となっているためである.

個別データの95%信頼区間

式(7.2)に示された4本の回帰式の個別データの95%信頼区間を描くためには, それぞれの回帰直線の推定値 \hat{y} に対する分散を推定する必要があり, 別々に回帰分析をすることも考えられるが, 交互作用の検討に用いた表7.4に示したパラメータの推定値を反映したい. その

ためには、表 7.4 に示した (80 行×5 列) のデザイン行列 X を用いたパラメータの共分散行列 $\Sigma(\hat{\theta}) = (X^T X)^{-1} \hat{\sigma}^2$ を必要とする。

表 7.5 に示すのは、転置行列 X^T とデザイン行列 X の積 $(X^T X)$ の計算方法である。転置行列 X^T は、Transpose (X の範囲) 関数を使い、行列の積 $(X^T X)$ は、Mmult (X^T の範囲, X の範囲) 関数を用いて行う。転置行列 X^T を Excel シート上に作成するのは難儀なので、Transpose () 関数を Mmult () 関数の引数とすることにより一括計算することができる。

表 7.5 転置行列 X^T とデザイン行列 X の積 $(X^T X)$

転置行列 $X^T = \text{Transpose}(X \text{ の範囲})$						デザイン行列 X					=Mmult ($X^T \text{ の範囲}, X \text{ の範囲}$) ($X^T X$)					
1	2	3	...	79	80	x_0	a	b	x	ax	=					
1	1	1	...	1	1	1	0	0	91.3	0.0		80	40	40	7229	3622
0	0	0	...	1	1	1	0	0	89.3	0.0		40	40	20	3622	3622
0	0	0	...	1	1	1	0	0	88.9	0.0		40	20	40	3619	1813
91.3	89.3	88.9	...	90.1	91.0	:	:		:	:		7229	3622	3619	653420	327967
0.0	0.0	0.0	...	90.1	91.0	1	1	1	90.1	90.1		3622	3622	1813	327967	327967
						1	1	1	91.0	91.0		x_0	a	b	x	ax
(5×80)						(80×5)						(5×5)				

表 7.6 に示すのは、 $(X^T X)$ について Minverse () 関数により逆行列 $(X^T X)^{-1}$ を求め、回帰分析の分散分析表の残差の分散 (誤差分散の推定値) $\hat{\sigma}^2 = 0.8380$ を掛けて、パラメータの共分散行列 $\Sigma(\hat{\theta})$ を求めた結果である。

表 7.6 パラメータの共分散行列 $\Sigma(\hat{\theta})$

逆行列 ($X^T X$) ⁻¹					=Minverse ($X^T X$ の範囲)					パラメータの共分散行列 $\Sigma(\hat{\theta}) = (X^T X)^{-1} \hat{\sigma}^2$				
							$\hat{\sigma}^2$							
97.67	-97.27	0.10	-1.08	1.08		0.8380	=	81.8408	-81.5062	0.0839	-0.9077	0.9035		
-97.27	263.52	0.10	1.08	-2.91				-81.5062	220.8166	0.0847	0.9030	-2.4419		
0.10	0.10	0.05	0.00	0.00				0.0839	0.0847	0.0423	-0.0012	-0.0009		
-1.08	1.08	0.00	0.01	-0.01				-0.9077	0.9030	-0.0012	0.0101	-0.0100		
1.08	-2.91	0.00	-0.01	0.03				0.9035	-2.4419	-0.0009	-0.0100	0.0270		
(5×5)					(5×5)					(5×5)				

表 7.7 に示すのは、パラメータの共分散行列 $\Sigma(\hat{\theta})$ の対角要素の平方根が表 7.4 に示した Excel の回帰分析の標準誤差 SE に一致することの確認である。

表 7.7 パラメータの共分散行列 $\Sigma(\hat{\theta})$ の対角要素の平方根

パラメータの共分散行列 $\Sigma(\hat{\theta}) = (X^T X)^{-1} \hat{\sigma}^2$					対角要素の平方根		表 7.4 回帰分析の結果		
					$\text{Var}(\hat{\theta})$	SE		係数	標準誤差
81.8408	-81.5062	0.0839	-0.9077	0.9035	81.8408	9.0466	切片	-7.5835	9.0466
-81.5062	220.8166	0.0847	0.9030	-2.4419	220.8166	14.8599	a	-40.3310	14.8599
0.0839	0.0847	0.0423	-0.0012	-0.0009	0.0423	0.2056	b	-2.8842	0.2056
-0.9077	0.9030	-0.0012	0.0101	-0.0100	0.0101	0.1004	x	0.4406	0.1004
0.9035	-2.4419	-0.0009	-0.0100	0.0270	0.0270	0.1643	ax	0.4656	0.1643

色B別の個別データの95%信頼区間

表 7.8 に示すのは、吐出量 x を $[87, 88, \dots, 93]^T$ のように設定し、推定値 \hat{y} および分散 $Var(\hat{y})$ を計算し、個別データの95%信頼区間の推定値である。変数名 (x_0, a, b, x, ax) に対して別名を $\mathbf{l} = (l_0, l_1, l_2, l_3, l_4)$ とし、 \mathbf{l} に推定目的に沿った係数を与え、推定されたパラメータ $\hat{\boldsymbol{\theta}}$ との積和（線形和） $L = \mathbf{l}\hat{\boldsymbol{\theta}}$ によって推定値 \hat{y} を計算し、分散を $Var(L) = Var(\mathbf{l}\hat{\boldsymbol{\theta}}) = \mathbf{l}\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})\mathbf{l}^T$ によって求めた結果である。

表 7.4 で推定された回帰パラメータは、

$$\hat{\boldsymbol{\theta}} = [-7.5835 \quad -40.3310 \quad -2.8842 \quad 0.4406 \quad 0.4656]^T$$

であり、表 7.8 の最初の行のベクトルを $\mathbf{l}^{(1)}$ としたときに、線形和

$$\hat{y}_{x=87}^{(1)} = \mathbf{l}^{(1)}\hat{\boldsymbol{\theta}}$$

=Mmult ($\mathbf{l}^{(1)}$ の範囲, $\hat{\boldsymbol{\theta}}$ の範囲)

	x_0	a	b	x	ax	$\hat{\boldsymbol{\theta}}$		\hat{y}
=	1	0	0	87	0	-7.5835	=	30.7466
						-40.3310		
						-2.8842		
						0.4406		
						0.4656		

は、色 B_1 ・シンナー A_1 について $x=87$ の場合の回帰直線の推定値が得られている。その分散 $Var(L^{(1)}) = Var(\mathbf{l}^{(1)}\hat{\boldsymbol{\theta}})$ は、表 7.7 に示したパラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ を用いて

表 7.8 色 B_1 についてのシンナー A_1 と A_2 についての個別データの95%信頼区間

シ	色		l_0	l_1	l_2	l_3	l_4	$\mathbf{l}\hat{\boldsymbol{\theta}}$	分散	個別95%信頼区間	
A	B	L	x_0	a	b	x	ax	\hat{y}	$Var(\hat{y})$	個別L95%	個別U95%
1	1	$L^{(1)}$	1	0	0	87	0	30.7466	0.1303	28.7864	32.7068
1	1	$L^{(2)}$	1	0	0	88	0	31.1872	0.0773	29.2814	33.0930
1	1	$L^{(3)}$	1	0	0	89	0	31.6278	0.0444	29.7566	33.4990
1	1	$L^{(4)}$	1	0	0	90	0	32.0684	0.0317	30.2107	33.9261
1	1	$L^{(5)}$	1	0	0	91	0	32.5090	0.0391	30.6434	34.3745
1	1	$L^{(6)}$	1	0	0	92	0	32.9495	0.0666	31.0549	34.8442
1	1	$L^{(7)}$	1	0	0	93	0	33.3901	0.1143	31.4461	35.3341
シ	色		l_0	l_1	l_2	l_3	l_4	$\mathbf{l}\hat{\boldsymbol{\theta}}$	分散	個別95%信頼区間	
A	B	L	x_0	a	b	x	ax	\hat{y}	$Var(\hat{y})$	個別L95%	個別U95%
2	1	$L^{(8)}$	1	1	0	87	87	30.9196	0.2381	28.8531	32.9861
2	1	$L^{(9)}$	1	1	0	88	88	31.8257	0.1364	29.8593	33.7921
2	1	$L^{(10)}$	1	1	0	89	89	32.7319	0.0689	30.8349	34.6289
2	1	$L^{(11)}$	1	1	0	90	90	33.6380	0.0354	31.7763	35.4997
2	1	$L^{(12)}$	1	1	0	90.56	90.56	34.1424	0.0316	32.2848	36.0000 #
2	1	$L^{(13)}$	1	1	0	92	92	35.4503	0.0708	33.5512	37.3493
2	1	$L^{(14)}$	1	1	0	93	93	36.3564	0.1397	34.3867	38.3261

Excel のソルバーで個別U95%が、上限の36になるように x を変化させ 90.56 が得られている。

$$Var(L^{(1)}) = l^{(1)} \Sigma(\hat{\theta}) l^{(1)T}$$

	x_0	a	b	x	ax	$\Sigma(\hat{\theta}) = (X^T X)^{-1} \sigma^2$					x'^T	$Var(y^{\wedge})$
=	1	0	0	87	0	81.8408	-81.5062	0.0839	-0.9077	0.9035	1	= 0.1303
						-81.5062	220.8166	0.0847	0.9030	-2.4419	0	
						0.0839	0.0847	0.0423	-0.0012	-0.0009	0	
						-0.9077	0.9030	-0.0012	0.0101	-0.0100	87	
						0.9035	-2.4419	-0.0009	-0.0100	0.0270	0	

によって計算され、個別データの95%信頼区間は、

$$\begin{aligned} \text{個別 } 95\%CL &= L^{(1)} \pm t_{0.05}(80-5) \sqrt{\hat{\sigma}^2 + Var(L^{(1)})} \\ &= 30.7466 \pm 1.9921 \times \sqrt{0.8380 + 0.1303} \\ &= (28.7864, 32.7068) \end{aligned}$$

となる。個別データの95%信頼区間が管理限界の上限 $y=36$ を超える吐出量 x を求めるために、Excelのソルバーを用いて、表7.8の個別U95%が、管理限界の上限の $y=36$ となるように x を変化させると $\hat{x}=90.56$ を求めることができる。

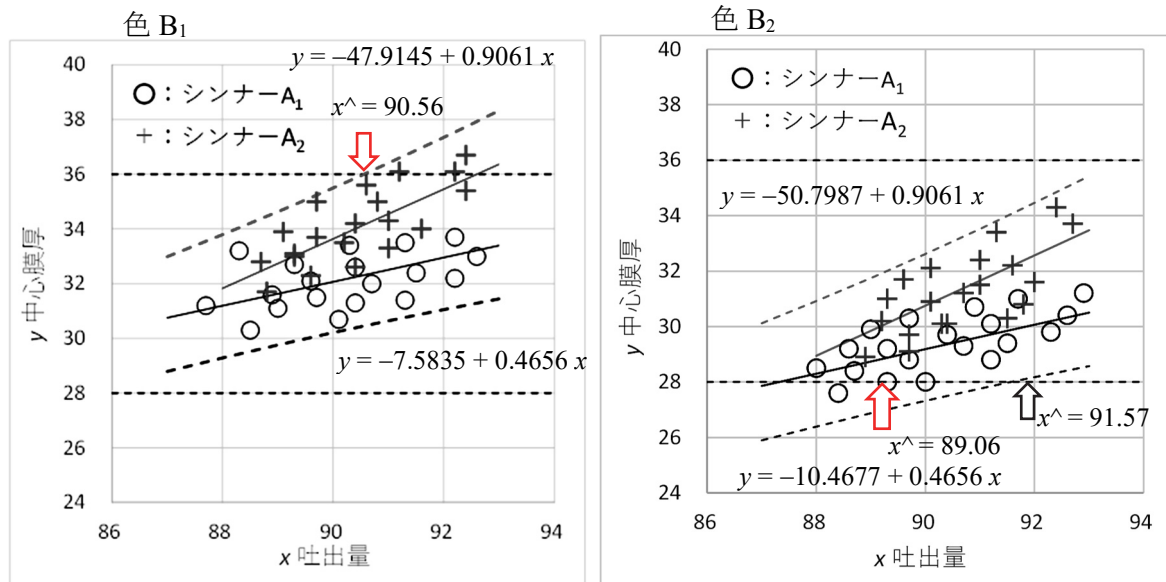
表7.9に示すは、色B₂に対する個別データの95%信頼区間の計算結果であり、図7.5に示すのは、表7.8および表7.9で求めた、回帰直線の推定値を重ね書きし、それらの個別データの95%信頼区間をさらに重ね書きした結果である。

表7.9 色B₂についてのシンナーA₁とA₂についての個別データの95%信頼区間

シ	色		l_0	l_1	l_2	l_3	l_4	$l\theta^{\wedge}$	分散	個別95%信頼区間	
A	B	L	x_0	a	b	x	ax	y^{\wedge}	$Var(y^{\wedge})$	個別L95%	個別U95%
1	2	$L^{(15)}$	1	0	1	87	0	27.8625	0.1377	25.8948	29.8302
1	2	$L^{(16)}$	1	0	1	88	0	28.3031	0.0824	26.3920	30.2141
1	2	$L^{(17)}$	1	0	1	89	0	28.7436	0.0472	26.8694	30.6178
1	2	$L^{(18)}$	1	0	1	90	0	29.1842	0.0321	27.3260	31.0424
1	2	$L^{(19)}$	1	0	1	91	0	29.6248	0.0372	27.7612	31.4884
1	2	$L^{(20)}$	1	0	1	91.57	0	29.8762	0.0491	28.0000	31.7525 #
1	2	$L^{(21)}$	1	0	1	93	0	30.5059	0.1078	28.5687	32.4432
シ	色		l_0	l_1	l_2	l_3	l_4	$l\theta^{\wedge}$	分散	個別95%信頼区間	
A	B	L	x_0	a	b	x	ax	y^{\wedge}	$Var(y^{\wedge})$	個別L95%	個別U95%
2	2	$L^{(22)}$	1	1	1	87	87	28.0354	0.2530	25.9547	30.1161
2	2	$L^{(23)}$	1	1	1	88	88	28.9416	0.1471	26.9644	30.9187
2	2	$L^{(24)}$	1	1	1	89.06	89	29.9005	0.0722	28.0000	31.8010 #
2	2	$L^{(25)}$	1	1	1	90	90	30.7538	0.0377	28.8897	32.6179
2	2	$L^{(26)}$	1	1	1	91	91	31.6600	0.0341	29.7996	33.5203
2	2	$L^{(27)}$	1	1	1	92	92	32.5661	0.0647	30.6735	34.4588
2	2	$L^{(28)}$	1	1	1	93	93	33.4723	0.1294	31.5130	35.4315

#ソルバーで個別L95%が、下限の28になるように x を変化させ 91.57 および 89.06 が得られている。

図7.5左に示す色B₁でシンナーA₁(+)で上側管理限界36を超える吐出量は、個別データの95%信頼区間から、目の子で吐出量 x が 90.5 ぐらいと読み、Excelのソルバーによる正確な

図 7.5 色 B₁ および B₂ についての回帰直線に対する個別データの 95%信頼区間の表示

推定では、 $\hat{x} = 90.56$ となる．吐出量の最大値が 90.56 を超えないようにすることで，不良率の軽減が図られる．また，交互作用 ax が統計的に有意であることから，シンナー-A₂ の傾きが，シンナー-A₁ に比べて急であるので不良が出やすくなった原因でもある．実験結果の報告に際し，このような統計的分析を行なうことにより，よりの確な数値に基づく対策が提言できる．

図 7.5 右に示す色 B₂ の場合は，色 B₁ の場合に比べて，全体的に下限方向にシフトしている．表 7.4 のパラメータ b の推定値から，シフト量は $\hat{\theta}_2 = -2.8842$ であり， $p < 0.0000$ と高度な差となっている．その結果としてシンナー-A₁ の場合に下側管理限界 28 を下回る不良が発生する原因となっている．Excel のソルバーを用いて，個別データの下側 95%信頼区間が，下側管理限界 28 を超える吐出量を推定すると $\hat{x} = 91.57$ となり，吐出量 x の平均である 90.5 を超える値となっており，吐出量が少ないことが不良発生の原因である．また，シンナー-A₂ の場合には，表 7.9 から，吐出量が 89.06 で，下側管理限界 28 を下回ることも読み取れる．

以上の結果から，不良品を軽減する対策として短絡的には，吐出量 X の変化に対し勾配が大きいシンナー2 をやめてシンナー1 に統一し，色 2 に対しては，吐出量を 91.57 以上とすることにより中心被膜 y が，管理限界の下限以上となり，不良率の軽減となることが推測される．シンナー2 を使う別の理由があるならば，色 1 に対しては，吐出量を 90.56 以下とすることが望まれる．

7.3. Excel の回帰分析による交互作用を含む変数減少法

第7.1節で共変量を含む3因子実験データの解析に際し、適切な散布図を作成することにより、問題解決のための方法が見い出された。さらに、第7.2節でExcelの回帰分析に加え、Excelの行列関数を用いて回帰直線の個別データの95%信頼区間を重ね書きすることにより、きめ細かな統計解析を実現できることを示した。

本節では、表7.3に示したデータを用い、Excelの回帰分析による探索的な解析を試みる。共変量がなければ、繰り返しが10の3因子の分散分析による解析ができ、一般的な分散分析の手順により問題解決となる。だが、共変量が一つ含まれるだけで、従来の平方和の分解を主体にする解析法が適用できない。実験前に特定の値に設定することができない量的変数がある場合に、第6章に示した伝統的な共分散分析として解析手順が定式化されている。ただし、現実的には1因子実験の場合に限定されている。共変量を含む3因子実験データの解析は、どのように行ったらよいのだろうか。

ダミー変数

質的変数を量的変数に変換する方法として、ダミー変数が使われる。統計ソフトでは、質的変数をダミー変数に自動的に変換する機能があり、Rでは表7.10に示すように最初の水準を0とする(0, 1)型、SASでは最後の水準を0とする(1, 0)型、JMPでは表7.11に示すような水準ごとに足して0となるような(1, -1)対比型が使われている。

表 7.10 (0, 1) 型ダミー変数 (R の場合)

シナリー	<i>a</i>	色	<i>b</i>	機種	<i>c</i>	シナリー	色	<i>a</i>	<i>b</i>	<i>ab</i>
1	0	1	0	国内	0	1	1	0	0	0
2	1	2	1	国外	1		2	0	1	0
						2	1	1	0	0
							2	1	1	1

表 7.11 (1, -1) 対比型ダミー変数 (JMP の場合)

シナリー	<i>a</i>	色	<i>b</i>	機種	<i>c</i>	シナリー	色	<i>a</i>	<i>b</i>	<i>ab</i>
1	1	1	1	国内	1	1	1	1	1	1
2	-1	2	-1	国外	-1		2	1	-1	-1
						2	1	-1	1	-1
							2	-1	-1	1

Excelの回帰分析では、このような自動変換の機能がないので自ら設定する必要がある。どのような型のダミー変数とするかは自由なので、ここでは最初の水準を0とする(0, 1)型を用いる。

交互作用を含む線形モデル

経験的に交互作用を考慮する必要性がないような分野ならば、考慮すべき変数をすべて含む線形モデルを適用し、 p 値などを見ながらモデルに含める変数を適宜減らす方法が適している。ただし、Excel の回帰分析で、取り上げる変数は、すべて隣り合っていないければならないので、モデル化から外したい変数ベクトルを **Ctrl+X** で選択し、適当な列の前に同じサイズのセルを選択し、「切り取ったセルの挿入」することにより、デザイン行列 X の変数の入れ替えが、サイズを変更することなく容易にできる。

ここでは、経験的に交互作用があることを考慮にするような実験データの解析とする。量的変数 X の場合には、2 乗項 X^2 も考慮する必要もあるが、吐出量 X に関しては、直線のあてはめとし、質的変数（A シンナー、B 色、C 機種）について（0, 1）型ダミー変数とする。解析モデルとして、2 因子交互作用は、すべての組合せについて設定し、3 因子交互作用は、念のために $A \times B \times C$ を含めるが、吐出量 X との 3 因子交互作用は、無いものとみなす。

主効果： A シンナー、B 色、C 機種、吐出量 X ,

交互作用： $A \times X$, $B \times X$, $C \times X$,

$A \times B$, $A \times C$, $B \times C$,

$A \times B \times C$

表 7.12 に最大モデルでの回帰分析を行うために、（80 行×12 列）のデザイン行列 X を示す。交互作用の設定は、主効果の変数（ a , b , c , x ）に対し掛け算で与え、Excel のフィルハンドルを用いて計算式のコピー機能を用いることにより、ストレスなく作成できる。

表 7.12 （80 行×12 列）のデザイン行列 X

A	B	C	No.	x_0	a	b	c	x	ax	bx	cx	ab	ac	bc	abc	y
1	1	1	1	1	0	0	0	91.3	0.0	0.0	0.0	0	0	0	0	33.5
			:													:
		2	11	1	0	0	1	89.0	0.0	0.0	89.0	0	0	0	0	31.1
			:													:
	2	1	21	1	0	1	0	88.6	0.0	88.6	0.0	0	0	0	0	29.2
			:													:
		2	31	1	0	1	1	90.9	0.0	90.9	90.9	0	0	1	0	30.7
			:													:
2	1	1	41	1	1	0	0	89.6	89.6	0.0	0.0	0	0	0	0	32.3
			:													:
		2	51	1	1	0	1	91.0	91.0	0.0	91.0	0	1	0	0	33.3
			:													:
	2	1	61	1	1	1	0	90.7	90.7	90.7	0.0	1	0	0	0	31.2
			:													:
		2	71	1	1	1	1	89.7	89.7	89.7	89.7	1	1	1	1	29.7
			:													:
			80	1	1	1	1	91.0	91.0	91.0	91.0	1	1	1	1	31.5

直交表の解析を想定すると、すべての変数 (a, b, \dots, bc, abc) を対象に回帰分析を行い、有意でない変数の平方和をまとめて誤差項にプールすればと思われるかもしれない。しかし、 $(0, 1)$ 型ダミー変数としたこと、さらに、共変量 x が含まれることにより、変数間の直交性が失われているために、一つの変数の除いただけで、他のすべての変数の平方和が変化してしまい戸惑うことになる。

変数選択の手順

取り上げた要因を（主効果, 2 因子交互作用, 3 因子交互作用）のように区別する。そして、高次の交互作用が含まれている場合には、主効果も含め低次の交互作用の p 値は、完全に無視し、同じレベルの交互作用の要因のみの p 値を用いて、要因の取捨選択を前もって定めた ($p=0.15$) などの基準で行なう。

表 7.13 左に示すのは、すべての変数 (a, b, \dots, bc, abc) を含んだ 3 因子モデルでの回帰分析の結果で、3 因子交互作用 abc が有意ではないとの結論を得る。表 7.13 右に示すのは、交互作用 abc を除いた 2 因子モデルでの回帰分析の結果である。分散分析表を比較すると、3 因子モデルの回帰の自由度が 11 から 2 因子モデルで 10 に減少し変動（平方和）がわずかに減り、その分が残差の平方和に移行している。

表 7.13 3 因子モデル vs. 2 因子モデル

分散分析表(3因子モデル)						分散分析表(2因子モデル)					
		自由度	変動	分散	分散比		自由度	変動	分散	分散比	
	回帰	11	286.264	26.024	29.839		10	286.061	28.606	33.169	
	残差	68	59.305	0.872			69	59.508	0.862		
	合計	79	345.570				79	345.570			
		係数	標準誤差	t	P -値		係数	標準誤差	t	P -値	
θ^{\wedge}_0	切片	-12.077	14.082	-0.858	0.3941	切片	-12.552	13.969	-0.899	0.3720	
θ^{\wedge}_1	a	-41.352	15.865	-2.606	0.0112 *	a	-41.303	15.776	-2.618	0.0109 *	
θ^{\wedge}_2	b	-10.703	14.899	-0.718	0.4750 ?	b	-9.882	14.719	-0.671	0.5042 ?	
θ^{\wedge}_3	c	16.105	15.281	1.054	0.2956	c	16.142	15.195	1.062	0.2918	
θ^{\wedge}_4	x	0.489	0.156	3.132	0.0026 *	x	0.494	0.155	3.186	0.0022 *	
θ^{\wedge}_5	ax	0.480	0.176	2.730	0.0081 *	ax	0.481	0.175	2.748	0.0076 *	
θ^{\wedge}_6	bx	0.088	0.165	0.533	0.5955	bx	0.080	0.163	0.490	0.6256	
θ^{\wedge}_7	cx	-0.179	0.170	-1.054	0.2956	cx	-0.178	0.169	-1.056	0.2948	
θ^{\wedge}_8	ab	-0.233	0.593	-0.394	0.6949	ab	-0.433	0.421	-1.028	0.3074	
θ^{\wedge}_9	ac	-0.163	0.604	-0.269	0.7888	ac	-0.367	0.429	-0.856	0.3949	
θ^{\wedge}_{10}	bc	0.151	0.592	0.256	0.7989	bc	-0.049	0.419	-0.117	0.9069	
θ^{\wedge}_{11}	abc	-0.405	0.840	-0.482	0.6311						

回帰係数を比較するとわずかなではあるが、すべて異なる。伝統的な直交表の解析では、有意でない要因の平方和を残差平方和にプールし、誤差の自由度を大きくすることが手順とし

て確立して、交互作用も含めた全ての要因が直交しているので、平方和も回帰係数（パラメータ）も変化することはない。表 7.4 では、因子 B（変数 b ）の p 値が $p < 0.0000$ なのに対し、表 7.13 に示す 3 因子モデルで $p = 0.4750$ 、2 因子モデルで $p = 0.5042$ と全く異なる結果となっている。この原因は、因子 B の主効果を含む交互作用（ $B \times X$, $A \times B$, $B \times C$, $A \times B \times C$ ）が含まれていることに起因している。

表 7.14 左に示すのは、2 因子モデルで $p = 0.15$ 以上となる交互作用をモデルから外し、表 7.14 右に示すように、交互作用 ax のみを含めたモデルとする。因子 B の p 値が、 $p = 0.504$ から、 $p = 0.0000$ へ大変化を起こしている。このことを以ってしても、交互作用が含まれているときの主効果の p 値による判断は、全く意味をなさないことが示されている。推定された回帰パラメータも $\hat{\theta}_2' = -9.882$ から $\hat{\theta}_2'' = -2.886$ へと大きく変化している。これは、量的変数 x との交互作用がある場合に、 $x = 0$ における（ $b = 0$ および $b = 1$ ）に対する切片の差となるためである。

主効果＋交互作用モデルで、因子 C が有意ではないので、モデルから外すことにする。因子 A と共変量 X の主効果は有意な差なので、モデルから外す誘惑は起きないが、もしもどちらかが有意でない場合に外したくなる誘惑にかられるかもしれないが、絶対に外してはならない。

表 7.14 2 因子モデル vs. 主効果＋交互作用モデル

分散分析表(2因子モデル)						分散分析表(主効果＋交互作用モデル)						
		自由度	変動	分散	分散比		自由度	変動	分散	分散比		
	回帰	10	286.061	28.606	33.169		回帰	5	283.172	56.634	67.164	
	残差	69	59.508	0.862			残差	74	62.398	0.843		
	合計	79	345.570				合計	79	345.57			
		係数	標準誤差	t	P -値			係数	標準誤差	t	P -値	
θ^{\wedge}_0'	切片	-12.552	13.969	-0.899	0.3720		切片	-7.528	9.075	-0.830	0.4095	
θ^{\wedge}_1'	a	-41.303	15.776	-2.618	0.0109	*	a	-41.778	15.038	-2.778	0.0069	*
θ^{\wedge}_2'	b	-9.882	14.719	-0.671	0.5042	?	b	-2.886	0.206	-13.99	0.0000	*
θ^{\wedge}_3'	c	16.142	15.195	1.062	0.2918		c	-0.152	0.208	-0.729	0.4683	
θ^{\wedge}_4'	x	0.494	0.155	3.186	0.0022	*	x	0.441	0.101	4.379	0.0000	*
θ^{\wedge}_5'	ax	0.481	0.175	2.748	0.0076	*	ax	0.482	0.166	2.895	0.0050	*
θ^{\wedge}_6'	bx	0.080	0.163	0.490	0.6256							
θ^{\wedge}_7'	cx	-0.178	0.169	-1.056	0.2948							
θ^{\wedge}_8'	ab	-0.433	0.421	-1.028	0.3074							
θ^{\wedge}_9'	ac	-0.367	0.429	-0.856	0.3949							
θ^{\wedge}_{10}'	bc	-0.049	0.419	-0.117	0.9069							

表 7.15 に示すように、因子 C を外して最終モデルにようやく到達し、表 7.4 に示した層別散布図を使った探索的な結果にたどり着いた。推定されたパラメータの解釈を試みよう。パ

ラメータの表記として $\hat{\theta}_k'''$ としたいのであるが、煩雑なので $\hat{\theta}_k''' \rightarrow \hat{\theta}_k$ とする。切片の $\hat{\theta}_0 = -7.5835$ は、 $a=b=x=0$ の場合であり、図 7.5 左の（シンナーA₁・色 B₁）○印の切片である。因子 A の $\hat{\theta}_1 = -40.3310$ は、（シンナーA₁・色 B₁）の切片から（シンナーA₂・色 B₁）の切片の差となっている。したがって、表 7.15 右に示した因子 A の $p=0.0082$ は、図 7.5 右に示したデータの範囲での推論される p 値でないことに注意が必要である。このような不可解な推定値となるのは、交互作用 A×X が回帰モデルに含まれているためであり、交互作用がある場合の主効果の推定値および p 値に全く意味がないことに注意しなければならない。

表 7.15 （主効果＋交互作用）モデル vs. 最終モデル

分散分析表(主効果＋交互作用モデル)					分散分析表(最終モデル)				
	自由度	変動	分散	分散比		自由度	変動	分散	分散比
回帰	5	283.172	56.634	67.164	回帰	4	282.724	70.681	84.350
残差	74	62.398	0.843		残差	75	62.846	0.838	
合計	79	345.57			合計	79	345.57		
	係数	標準誤差	t	P -値		係数	標準誤差	t	P -値
θ_0'' 切片	-7.528	9.075	-0.830	0.4095	θ_0' 切片	-7.5835	9.0466	-0.8383	0.4045
θ_1'' a	-41.778	15.038	-2.778	0.0069 *	θ_1' a	-40.3310	14.8599	-2.7141	0.0082 *
θ_2'' b	-2.886	0.206	-13.99	0.0000 *	θ_2' b	-2.8842	0.2056	-14.0253	0.0000 *
θ_3'' c	-0.152	0.208	-0.729	0.4683	θ_3' x	0.4406	0.1004	4.3904	0.0000 *
θ_4'' x	0.441	0.101	4.379	0.0000 *	θ_4' ax	0.4656	0.1643	2.8329	0.0059 *
θ_5'' ax	0.482	0.166	2.895	0.0050 *					

推定されたパラメータを用いた回帰直線の推定

量的変数 x と 2 水準の質的変数をダミー変数とした回帰分析に際し、交互作用がなければ、推定されたパラメータの解釈は楽なのであるが、交互作用が入っていると説明に窮することになる。そのために、表 7.16 に示すように、量的変数 x の存在範囲を設定し、推定されたパラメータを用いて反応 y の推定値 \hat{y} を計算し、結果を考察することが必須である。因子 A の水準、因子 B の水準の組み合わせごとに、共変量 X の範囲を (0, 88, 93) と変化させた場合の推定値 \hat{y} の推定値を計算した結果である。なお、この推定された結果は、図 7.5 左および図 7.5 右に示した回帰直線に一致する。これらの 4 本の層別回帰直線の内、管理限界を超え

表 7.16 推定されたパラメータを用いた中心膜厚 y の推定値

シンナー	色						x	
A	B	切片		傾き		0	88	93
A ₁	B ₁ ($b=0$)	$\theta^{}_0=$	-7.5835	$\theta^{}_3=$	0.4406	-7.58	31.19	33.39
$a=0$	B ₂ ($b=1$)	$\theta^{}_0+ \theta^{}_2=$	-10.4677	$\theta^{}_3=$	0.4406	-10.47	28.31	30.51
A ₂	B ₁ ($b=0$)	$\theta^{}_0+\theta^{}_1=$	-47.9145	$\theta^{}_3+\theta^{}_4=$	0.9061	-47.91	31.82	36.35
$a=1$	B ₂ ($b=1$)	$\theta^{}_0+\theta^{}_1+ \theta^{}_2=$	-50.7987	$\theta^{}_3+\theta^{}_4=$	0.9061	-50.80	28.94	33.47

るデータが含まれる A_1B_2 と A_2B_1 の組み合わせについて散布図を作成し、2 本の回帰直線および個別データの 95%信頼区間を重ね書きする。

表 7.17 2 本の回帰直線の個別データの 95%信頼区間（表 7.8 および表 7.9 の要約）

B	A	印	X	切 片	a	b	x	ax	θ^{\wedge}	y^{\wedge}	$Var(y^{\wedge})$	L95%	U95%	個別 L95%	個別 U95%
B1	A2	●	87	1	1	0	87	87	-7.584	30.920	0.238	29.948	31.892	28.853	32.986
			89	1	1	0	89	89	-40.331	32.732	0.069	32.209	33.255	30.835	34.629
			91	1	1	0	91	91	-2.884	34.544	0.036	34.166	34.922	32.682	36.407
			93	1	1	0	93	93	0.441	36.356	0.140	35.612	37.101	34.387	38.326
B2	A1	△	87	1	0	1	87	0	0.466	27.862	0.138	27.123	28.602	25.895	29.830
			89	1	0	1	89	0		28.744	0.047	28.311	29.176	26.869	30.618
			91	1	0	1	91	0		29.625	0.037	29.241	30.009	27.761	31.488
			93	1	0	1	93	0		30.506	0.108	29.852	31.160	28.569	32.443
											$t_{0.05}(80-5)=$	1.9921			

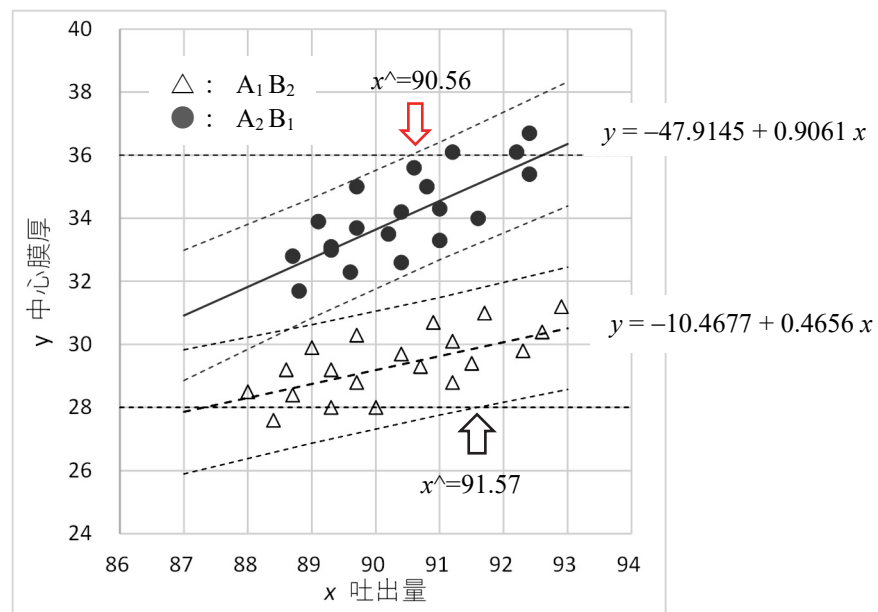


図 7.6 管理限界を超えるガンスプレーの吐出量 x の推定値

第 7.1 節では、各種の散布図に直線を重ね書きし、管理限界を超える事例について考察を行ってきたのであるが、図 7.6 に示すように、個別データの 95%信頼区間を重ね書きすることにより、統計的に管理限界を超える吐出量 x が、より低めであったり、より高めであったりが明確に数値として把握でき、現実の対策を考える際に役に立つ。

7.4. JMP のステップワイズ法による変数選択

表 7.13 に示した Excel による回帰分析によるモデル選択では、表 7.13 で「3 因子モデル vs. 2 因子モデル」、表 7.14 で「2 因子モデル vs. 主効果+交互作用モデル」、表 7.15 では、「主効果+交互作用モデル vs. 最終モデル」では、パラメータの p 値を見ながら手作業で段階的に変数選択を行った。

JMP では、「モデルのあてはめ」の中で「ステップワイズ法」がサポートされ、手動または自動的に変数を増加したり減少したりすることができる。連続変数のみならず質的変数も扱え、それらの交互作用も含めることができる。さらに、反応変数には、量的変数のみならず質的変数（名義尺度、順序尺度）も扱えるようになっている。また、変数の増減に際し、交互作用が対象となった場合には、主効果などの下位の変数も合わせて増減の判定を考慮する機能もある優れたものである。

手作業での変数選択

表 7.18 に示すのは、Excel での手作業での変数選択で用いた表 7.12 のデータを JMP にペーストして作成したデータである。

表 7.18 変数選択のための JMP データファイル

	No.	x0	a	b	c	x	ax	cx	bx	ab	ac	bc	abc	y
1	1	1	0	0	0	91.3	0.0	0.0	0.0	0	0	0	0	33.5
2	2	1	0	0	0	89.3	0.0	0.0	0.0	0	0	0	0	32.7
:														
9	9	1	0	0	0	87.7	0.0	0.0	0.0	0	0	0	0	31.2
10	10	1	0	0	0	89.7	0.0	0.0	0.0	0	0	0	0	31.5
11	11	1	0	0	1	89.0	0.0	89.0	0.0	0	0	0	0	31.1
:														
79	79	1	1	1	1	90.1	90.1	90.1	90.1	1	1	1	1	30.9
80	80	1	1	1	1	91.0	91.0	91.0	91.0	1	1	1	1	31.5

JMP の「ステップワイズ法」は、多機能で多彩であるが、ここでは手作業による変数選択の過程を示す。表 7.19 左（3 因子モデル）は、表 7.18 の全変数を追加した回帰分析の結果が示されている。3 因子の交互作用 abc の p 値が 0.631 なので、チェック・ボックスを off にし、2 因子交互作用モデルとする。表 7.19 中（2 因子モデル）の p 値を表 7.19 左（3 因子モデル）と比較すると p 値がかなり変化することが読み取れる。

Excel での変数選択は、変数の増減に伴い新たに回帰分析を行ない、結果をコピー＆ペーストし、書式を整える必要があり、煩雑であった。JMP の場合は、チェック・ボックスをクリックするだけで変数を増減した結果が即時に表示される。表 7.19 中（2 因子モデル）の 2 因子交互作用の p 値が大きいものから減少させると、交互作用 ax の p 値は揺れ動くが、小さいまま残り続ける。表 7.19 右（最終モデル）に示したのは、主効果 c が $p=0.468$ であり、交互作用項にも含まれていないので、減少させた結果であり表 7.15 右の Excel で求めた最終モデルに一致する。

表 7.19 JMP のステップワイズ法を用いた手作業による変数減少法
(3 因子モデル) (2 因子モデル) (最終モデル)

SSE	DFE	MSE	
59.305	68	0.872	
現在の推定値			
追加	パラメータ	推定値	p値
<input checked="" type="checkbox"/>	切片	-12.077	1.000
<input checked="" type="checkbox"/>	a	-41.352	0.011
<input checked="" type="checkbox"/>	b	-10.703	0.475
<input checked="" type="checkbox"/>	c	16.105	0.296
<input checked="" type="checkbox"/>	x	0.489	0.003
<input checked="" type="checkbox"/>	ax	0.480	0.008
<input checked="" type="checkbox"/>	cx	-0.179	0.296
<input checked="" type="checkbox"/>	bx	0.088	0.595
<input checked="" type="checkbox"/>	ab	-0.233	0.695
<input checked="" type="checkbox"/>	ac	-0.163	0.789
<input checked="" type="checkbox"/>	bc	0.151	0.799
<input checked="" type="checkbox"/>	abc	-0.405	0.631

SSE	DFE	MSE	
59.508	69	0.862	
現在の推定値			
追加	パラメータ	推定値	p値
<input checked="" type="checkbox"/>	切片	-12.552	1.000
<input checked="" type="checkbox"/>	a	-41.303	0.011
<input checked="" type="checkbox"/>	b	-9.882	0.504
<input checked="" type="checkbox"/>	c	16.142	0.292
<input checked="" type="checkbox"/>	x	0.494	0.002
<input checked="" type="checkbox"/>	ax	0.481	0.008
<input checked="" type="checkbox"/>	cx	-0.178	0.295
<input checked="" type="checkbox"/>	bx	0.080	0.626
<input checked="" type="checkbox"/>	ab	-0.433	0.307
<input checked="" type="checkbox"/>	ac	-0.367	0.395
<input checked="" type="checkbox"/>	bc	-0.049	0.907
<input type="checkbox"/>	abc	0.000	0.631

SSE	DFE	MSE	
62.846	75	0.838	
現在の推定値			
追加	パラメータ	推定値	p値
<input checked="" type="checkbox"/>	切片	-7.584	1.000
<input checked="" type="checkbox"/>	a	-40.331	0.008
<input checked="" type="checkbox"/>	b	-2.884	0.000
<input type="checkbox"/>	c	0.000	0.468
<input checked="" type="checkbox"/>	x	0.441	0.000
<input checked="" type="checkbox"/>	ax	0.466	0.006
<input type="checkbox"/>	cx	0.000	0.458
<input type="checkbox"/>	bx	0.000	0.777
<input type="checkbox"/>	ab	0.000	0.347
<input type="checkbox"/>	ac	0.000	0.202
<input type="checkbox"/>	bc	0.000	0.528
<input type="checkbox"/>	abc	0.000	0.133

注) 切片の P 値が常に $p=1.00$ となっているのは、変数選択の対象外としているためである。

自動的な変数選択

回帰分析における変数の自動選択は、“最適なモデル”の探索に際し、恣意性を排除するための方法として多くの統計ソフトに組み込まれている。ただし、変数増加法、変数減少法、変数増減法、変数減増法、総当たり法などがあり、さらに、交互作用を含めた場合についての選択法など、恣意性を排除したいとの努力が続けられている。

表 7.19 に示した 3 因子モデルでは、吐出量 X との交互作用を 2 因子交互作用に限定していた。例えば、 $A \times B \times X$ といった 3 因子交互作用を含めていない。 $A \times B \times C \times X$ のような 4 因子交互作用をすべて含め、JMP のステップワイズによる変数の自動選択を試みる。

表 7.20 に示すのは、選択する変数として、主効果 (a, b, c, x) とし、交互作用を JMP の内部で自動的に設定するために「マクロ」の機能で「すべての組み合わせ」を選択すると、2

因子交互作用「 $a*b$ 」から4因子交互作用「 $a*b*c*x$ 」までが自動的に変数としてセットされた結果である。示されている自由度は、高次の交互作用をすべて積み上げた結果であり、変数 a の自由度8は、それぞれ自由度1の (a , $a*b$, $a*c$, $a*x$, $a*b*c$, $a*b*x$, $a*c*x$, $a*b*c*x$) を加えたものである。「閾値 p 値を」(0.1, 0.1) と設定し、すべての変数を「追加」し、「変数増減」, 「組み合わせ」のルールでの変数選択を実行する。

表 7.20 4次元交互作用まで含むフルモデル

ステップワイズ回帰の設定

停止ルール:

閾値p値

変数を追加するときのp値0.1

変数を除去するときのp値0.1

方向:

変数増減

ルール:

組み合わせ

SSE	DFE	MSE	Cp	p	AICc	BIC
57.153916	64	0.8930299	16	16	243.9988	274.6223

現在の推定値

ロック	追加	パラメータ	推定値	自由度	平方和	"F値"	p値
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	切片	-5.6102	1	0.0000	0.0000	1.0000
<input type="checkbox"/>	<input checked="" type="checkbox"/>	a	-47.545	8	70.9514	9.9313	0.0000
<input type="checkbox"/>	<input checked="" type="checkbox"/>	b	-11.400	8	167.257	23.4115	0.0000
<input type="checkbox"/>	<input checked="" type="checkbox"/>	c	4.9572	8	4.7610	0.6664	0.7189
<input type="checkbox"/>	<input checked="" type="checkbox"/>	x	0.4176	8	60.9531	8.5318	0.0000
<input type="checkbox"/>	<input checked="" type="checkbox"/>	a*b	-15.381	4	1.6514	0.4623	0.7631
<input type="checkbox"/>	<input checked="" type="checkbox"/>	a*c	4.3011	4	2.4386	0.6827	0.6065
<input type="checkbox"/>	<input checked="" type="checkbox"/>	a*x	0.5491	4	8.1424	2.2794	0.0703
<input type="checkbox"/>	<input checked="" type="checkbox"/>	b*c	0.0527	4	1.2599	0.3527	0.8412
<input type="checkbox"/>	<input checked="" type="checkbox"/>	b*x	0.0959	4	1.3638	0.3818	0.8209
<input type="checkbox"/>	<input checked="" type="checkbox"/>	c*x	-0.0550	4	3.1060	0.8695	0.4872
<input type="checkbox"/>	<input checked="" type="checkbox"/>	a*b*c	61.2491	2	0.8165	0.4572	0.6351
<input type="checkbox"/>	<input checked="" type="checkbox"/>	a*b*x	0.1674	2	0.8425	0.4717	0.6261
<input type="checkbox"/>	<input checked="" type="checkbox"/>	a*c*x	-0.0501	2	1.7276	0.9673	0.3856
<input type="checkbox"/>	<input checked="" type="checkbox"/>	b*c*x	0.0008	2	1.1493	0.6435	0.5288
<input type="checkbox"/>	<input checked="" type="checkbox"/>	a*b*c*x	-0.6791	1	0.7704	0.8627	0.3565

自動的に変数選択された結果を表 7.21 に示す。結果は、表 7.19 示した手動での結果に一致する。示されている自由度は、選択されている変数の場合は、高次の交互作用の変数の自由度を合わせた結果であり、選択されていない場合は、低次の交互作用を含む結果となっている。

表 7.21 変数増減法により選択されたモデル

現在の推定値							
ロック	追加	パラメータ	推定値	自由度	平方和	"F値"	p値
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	切片	-7.5835	1	0.0000	0.0000	1.0000
<input type="checkbox"/>	<input checked="" type="checkbox"/>	a	-40.331	2	67.6239	40.3508	0.0000
<input type="checkbox"/>	<input checked="" type="checkbox"/>	b	-2.8842	1	164.831	196.708	0.0000
<input type="checkbox"/>	<input type="checkbox"/>	c	0.0000	1	0.4481	0.5314	0.4683
<input type="checkbox"/>	<input checked="" type="checkbox"/>	x	0.4406	2	56.2733	33.5780	0.0000
<input type="checkbox"/>	<input type="checkbox"/>	a*b	0.0000	1	0.7507	0.8946	0.3473
<input type="checkbox"/>	<input type="checkbox"/>	a*c	0.0000	2	1.4216	0.8447	0.4338
<input type="checkbox"/>	<input checked="" type="checkbox"/>	a*x	0.4656	1	6.7247	8.0252	0.0059
<input type="checkbox"/>	<input type="checkbox"/>	b*c	0.0000	1	0.0245	0.0288	0.8657
<input type="checkbox"/>	<input type="checkbox"/>	b*x	0.0000	1	0.0684	0.0806	0.7772
<input type="checkbox"/>	<input type="checkbox"/>	c*x	0.0000	2	1.6661	0.9940	0.3751
<input type="checkbox"/>	<input type="checkbox"/>	a*b*c	0.0000	5	2.3958	0.5548	0.7341
<input type="checkbox"/>	<input type="checkbox"/>	a*b*x	0.0000	1	0.0338	0.0398	0.8425
<input type="checkbox"/>	<input type="checkbox"/>	a*c*x	0.0000	4	3.4510	1.0313	0.3972
<input type="checkbox"/>	<input type="checkbox"/>	b*c*x	0.0000	5	2.3434	0.5422	0.7436
<input type="checkbox"/>	<input type="checkbox"/>	a*b*c*x	0.0000	11	5.6924	0.5795	0.8385

質的変数と量的変数の交互作用を含む変数選択

観察研究で、説明変数が多数ある場合に、目的とする反応 y に対し、最もあてはまりの良い線形モデルを自動的に選択したいとの要望に沿って様々な方法が研究され、多くの統計ソフトで実用化されてきた。恣意性を排除した「自動的」な変数選択が行われたことを強調する人達に対し、様々な選択方法があり一意に決めることはできない、変数選択ではなくすべての変数の組み合わせの中から選ぶ方が望ましいのではないかと、などの反論もあり悩ましい問題でもある。

JMP では、「自動的」な変数選択だけでなく、「手動的」な変数選択が GUI 操作によって実現できる特徴がある。さらに、交互作用を自動的に変数リストに加える機能を有することも現実の課題解決に欠かせない。また、選択された主効果および交互作用を含んだ変数に対し「モデルのあてはめ」による「標準最小 2 乗法」に繋げる配慮もされており、きめ細かな対応が可能である。

これまで示してきた解析法は、質的変数をダミー変数により量的変数とした解析を行ってきたが、JMP の「ステップワイズ法」では、質的変数のままでも交互作用を含めて変数選択ができる。内部では、もちろん全て (1, -1) 対比型ダミー変数に置き換えられていて、「モデルのあてはめ」での結果の出力に際しては、質的変数として変換された表示となる。

表 7.22 に (0, 1) 型ダミー変数 (a , b , c) を元の 2 水準 (1, 2) の因子 (A, B, C) としたデータリストを示す. もちろん, Excel の回帰分析のデータとしては使うことができない.

表 7.22 質的変数のままのデータリスト

	No.	A	B	C	x	y
1	1	A1	B1	C1	91.3	33.5
2	2	A1	B1	C1	89.3	32.7
:						
9	9	A1	B1	C1	87.7	31.2
10	10	A1	B1	C1	89.7	31.5
11	11	A1	B1	C2	89.0	31.1
:						
79	79	A2	B2	C2	90.1	30.9
80	80	A2	B2	C2	91.0	31.5

表 7.23 に示したのは, JMP の「ステップワイズ法」を起動し, 表 7.20 に示したと同様に質的変数 (A, B, C) と共変量 x を選択し, 「マクロ」の「すべての組み合わせ」で 4 因子交互作用までを自動作成し, 「変数増減」による変数選択をした結果である. (平方和, F 値, p 値) は, 表 7.20 と同じ結果が得られているが, パラメータの推定値はまったく異なる. これは, JMP の内部で生成されるダミー変数が (1, -1) 対比型ダミー変数となっていることに起因する. 表示されている変数名 “A{A1-A2}” は, A1 に対し 1 が, A2 に対し -1 となる対比型ダミー変数と理解される. “B{B2-B1}” は, B1 に対し -1, B2 に対し 1 とする対比型ダミー変数が自動設定されている. また, 変数 x に関する交互作用において “A{A1-A2} * ($x - 90.3662$)” のように変数 x の総平均 90.3662 が差し引く中心化が行われている. これは, 多重共線性による致命的な計算誤差をなくするためである.

変数選択の結果をさらに深堀する

表 7.23 右上に, モデルの作成, モデルの実行 の選択ができるようになっていて, 選択されている交互作用を含む変数について「モデルのあてはめ」の「標準最小 2 乗法」を起動できるようになっている. 実際に実行すると, 表 7.24 に示すように分散分析表が出力される. 表 7.23 の $SSE=62.8463$ と表 7.24 の誤差の平方和 $S_e=62.8463$ と一致することが確認される. ただし, 効果の検定の平方和は, 自由度が同じならば一致するが, 交互作用を含む主効果 A および x は一致しない.

表 7.23 質的変数と量的変数を込みにした変数選択の結果

yのステップワイズ									
ステップワイズ回帰の設定									
停止ルール:	閾値p値								
変数を追加するときのp値	0.1	→	すべて追加	モデルの作成					
変数を除去するときのp値	0.1	←	すべて削除	モデルの実行					
方向:	変数増減								
ルール:	組み合わせ								
SSE	DFE	RMSE	R2乗	自由度調整R2乗	Cp	p	AICc	BIC	
62.846289	75	0.9153964	0.8181	0.8084	0.374224	5	220.8741	234.0155	
現在の推定値									
ロック	追加	パラメータ	推定値	自由度	平方和	"F値"	"p値"		
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	切片	-29.1911	1	0.0000	0.000	1.0000		
<input type="checkbox"/>	<input checked="" type="checkbox"/>	A{A1-A2}	-0.8701	2	67.6239	40.351	0.0000		
<input type="checkbox"/>	<input type="checkbox"/>	C{C1-C2}	0.0000	1	0.4481	0.531	0.4683		
<input type="checkbox"/>	<input checked="" type="checkbox"/>	B{B2-B1}	-1.4421	1	164.8314	196.708	0.0000		
<input type="checkbox"/>	<input checked="" type="checkbox"/>	x	0.6734	2	56.2733	33.578	0.0000		
<input type="checkbox"/>	<input type="checkbox"/>	A{A1-A2}*C{C1-C2}	0.0000	2	1.4216	0.845	0.4338		
<input type="checkbox"/>	<input type="checkbox"/>	A{A1-A2}*B{B2-B1}	0.0000	1	0.7507	0.895	0.3473		
<input type="checkbox"/>	<input checked="" type="checkbox"/>	A{A1-A2}*(x-90.3662)	-0.2328	1	6.7247	8.025	0.0059		
<input type="checkbox"/>	<input type="checkbox"/>	C{C1-C2}*B{B2-B1}	0.0000	1	0.0245	0.029	0.8657		
<input type="checkbox"/>	<input type="checkbox"/>	C{C1-C2}*(x-90.3662)	0.0000	2	1.6661	0.994	0.3751		
<input type="checkbox"/>	<input type="checkbox"/>	B{B2-B1}*(x-90.3662)	0.0000	1	0.0684	0.081	0.7772		
<input type="checkbox"/>	<input type="checkbox"/>	A{A1-A2}*C{C1-C2}*B{B2-B1}	0.0000	5	2.3958	0.555	0.7341		
<input type="checkbox"/>	<input type="checkbox"/>	A{A1-A2}*C{C1-C2}*(x-90.3662)	0.0000	4	3.4510	1.031	0.3972		
<input type="checkbox"/>	<input type="checkbox"/>	A{A1-A2}*B{B2-B1}*(x-90.3662)	0.0000	1	0.0338	0.040	0.8425		
<input type="checkbox"/>	<input type="checkbox"/>	C{C1-C2}*B{B2-B1}*(x-90.3662)	0.0000	5	2.3434	0.542	0.7436		
<input type="checkbox"/>	<input type="checkbox"/>	A{A1-A2}*C{C1-C2}*B{B2-B1}*(x-90.3662)	0.0000	11	5.6924	0.579	0.8385		

表 7.24 選択された変数による「モデルのあてはめ」での分散分析表

分散分析					
要因	自由度	平方和	平均平方	F値	
モデル	4	282.7236	70.6809	84.3497	
誤差	75	62.8463	0.8380	p値(Prob>F)	
全体(修正済み)	79	345.5699		<.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	1	59.3629	59.3629	70.8430	<.0001*
B	1	164.8314	164.8314	196.7078	<.0001*
x	1	55.7923	55.7923	66.5819	<.0001*
A*x	1	6.7247	6.7247	8.0252	0.0059*

表 7.25 に示すのは、「モデルのあてはめ」で出力されたパラメータの推定値である。この推定値は、表 7.15 右に示した (0, 1) 型ダミー変数での推定値と全く異なるが、切片を除いて p 値は同じである。このように、用いるダミー変数が異なれば、推定値も全く異なるので、パラメータの推定値を用いた考察は、理解の妨げになる。

表 7.25 選択された変数による「モデルのあてはめ」でのパラメータ推定値

パラメータ推定値				
項	推定値	標準誤差	t値	p値(Prob> t)
切片	-29.1911	7.4616	-3.9122	0.0002*
A[A1]	-0.8701	0.1034	-8.4168	<.0001*
B[B1]	1.4421	0.1028	14.0253	<.0001*
x	0.6734	0.0825	8.1598	<.0001*
A[A1]*(x-90.3662)	-0.2328	0.0822	-2.8329	0.0059*

JMP の優れている点は、表 7.25 の推定値を使った各種の予測プロファイルの出力にある。特に交互作用がある場合に、交互作用プロットは、結果の解釈に役立つ。図 7.7 右上段に示したのは、 B_1 を選択した場合の A_1 および A_2 の回帰直線であり、図 7.7 右下段に示したのは、 B_2 を選択した場合の A_1 および A_2 の回帰直線である。

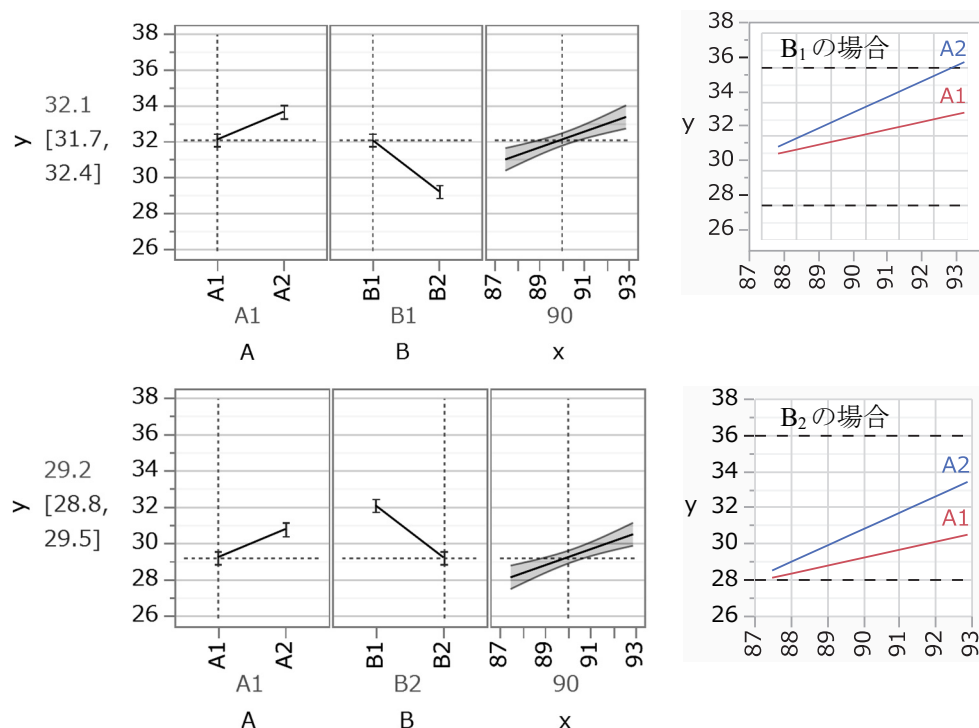


図 7.7 JMP の予測プロファイルと交互作用プロファイルの連動

第7章 文献 索引

宮川(2008) - 問題発見と解決の科学—SQCの基本—, 第6.3節 計画的に採取した3元配置データの解析例







219

第7章 索引

あ	1 因子実験 - 共分散分析	230	か	交互作用を念頭に - 分野	219
	(1, -1) 対比型 - (0, 1) 型	230		格子状 - 層別散布図行列	221
	if() 関数 - Excel	224		誤差分散の推定値 - 残差の分散	226
	Minverse() 関数 - 逆行列	226		個別データ - 95%信頼区間	224, 227
	ウス不良 - 流れ不良	220		- 95%信頼区間の表示	229
	SQCの基本 - 宮川(2008)	219		個別データの95%信頼区間 - 回帰直線	224
	Excel - if() 関数	224		- 管理限界	235
	Excel の回帰分析 - 変数は隣り合う	231		込みにした変数選択 - 質的変数と量的変数	241
	- 変数減少法	230	さ	最終モデル - 主効果+交互作用モデル	234
	大きく変化 - 回帰パラメータ	233		3 因子モデル - 2 因子モデル	232
か	回帰パラメータ - 大きく変化	233		3 因子実験データ - 共変量を含む	219, 230
	回帰直線 - 個別データの95%信頼区間	224		残差の分散 - 誤差分散の推定値	226
	- 推定されたパラメータ	234		層別散布図行列 - 格子状	221
	- 層別散布図	219		恣意性を排除 - 変数の自動選択	237
	回帰直線の推定値 - 分散	225		$\Sigma(\theta^{\wedge})$ - パラメータの共分散行列	227
	回帰分析 - 探索的な解析	230		質的変数と量的変数 - 交互作用	239
	- デザイン行列 X	225		- 込みにした変数選択	241
	重ね書き - 平均 \pm SD	222		質的変数のまま - データリスト	240
	ガンスプレー - 吐出力	219		自動的 - 変数選択	237
	管理限界 - 個別データの95%信頼区間	235		JMP - 交互作用プロファイル	242
	- 上限および下限	221		- ステップワイズ法	256
	- ソルバー	228		- 変数選択	256
	- 中心膜厚	219		- 予測プロファイル	242
	- 統計的な判断	224		主効果+交互作用モデル - 最終モデル	234
	管理限界を超える - 吐出力 x	235		- 2 因子モデル	233
	基準 - 要因の取舍選択	232		主効果に重き - 分野	219
	逆行列 - Minverse() 関数	226		上限および下限 - 管理限界	221
	95%信頼区間 - 個別データ	224, 227		推定されたパラメータ - 回帰直線	234
	95%信頼区間の表示 - 個別データ	229		数値の基づく対策 - 統計的分析	229
	共分散行列 - パラメータ	226		ステップワイズ法 - JMP	256
	共分散分析 - 1 因子実験	230		- (1, -1) 対比型	239
	共変量 x - 均一性	222		- 手作業による変数減少法	237
	共変量を含む - 3 因子実験データ	219, 230		- 名義尺度, 順序尺度	256
	行列の積 - Mmult() 関数	226		- モデルのあてはめ	256
	均一性 - 共変量 x	222		すべての組み合わせ - マクロ	240
	区別 - 要因	232		積和 10^{\wedge} - 線形和 L	227
	現業実習 - 冷蔵庫の塗装不良	219		切片の差の推定値 - 交互作用	225
	交互作用 - 質的変数と量的変数	239		(0, 1) 型 - (1, -1) 対比型	230
	- 切片の差の推定値	225		- 線形モデル	224
	- 線形モデル	224		線形モデル - 交互作用	224
	- ダミー変数	225		- 交互作用を含む	231
	- 中心化	240		- (0, 1) 型	224
	- 変数減少法	230		線形和 L - 積和 10^{\wedge}	227
	交互作用プロット - 予測プロファイル	242		- 分散	227
	交互作用プロファイル - JMP	242		選択されたモデル - 変数増減法	239
	交互作用を含む - 線形モデル	231		選択された変数 - モデルのあてはめ	241
	- 変数選択	239		総当たり法 - 変数選択	237

さ	層別散布図 - 回帰直線	219	は	不良の発生 - 中心膜厚	220
	- 2×2の格子状	220		フルモデル - 4次元交互作用	238
	- 併合	223		分散 - 回帰直線の推定値	225
	ソルバー - 管理限界	228		- 線形和 L	227
た	対角要素の平方根 - 標準誤差	226		分野 - 交互作用を念頭に	219
	(1, -1) 対比型 - ステップワイズ法	239		- 主効果に重き	219
	多重共線性 - 中心化	240		平均±SD - 重ね書き	222
	ダミー変数 - 交互作用	225		併合 - 層別散布図	223
	探索的な解析 - 回帰分析	230		別名 - 変数名	227
	中心化 - 交互作用	240		変数名 - 別名	227
	- 多重共線性	240		変数の自動選択 - 恣意性を排除	237
	中心膜厚 - 管理限界	219		変数は隣り合う - Excel の回帰分析	231
	- 不良の発生	220		変数減少法 - Excel の回帰分析	230
	繋げる配慮 - 標準最小2乗法	239		- 交互作用	230
	データリスト - 質的変数のまま	240		変数選択 - 交互作用を含む	239
	デザイン行列 X - 回帰分析	225		- 自動的	237
	手作業 - 変数選択	256		- JMP	256
	手作業による変数減少法 - ステップワイズ法	237		- 総当たり法	237
	手順 - 変数選択	232		- 手作業	256
	転置行列 - Transpose () 関数	226		- 手順	232
	統計的な判断 - 管理限界	224		変数増減法 - 選択されたモデル	239
	統計的分析 - 数値の基づく対策	229	ま	マクロ - すべての組み合わせ	240
	吐出量 x - ガンスプレー	219		Mmult () 関数 - 行列の積	226
	- 管理限界を超える	235		宮川(2008) - SQCの基本	219
	塗装不良の改善 - 宮川(2008)	219		- 塗装不良の改善	219
	Transpose () 関数 - 転置行列	226		名義尺度, 順序尺度 - ステップワイズ法	256
な	流れ不良 - ウス不良	220		モデルのあてはめ - ステップワイズ法	256
	2因子モデル - 3因子モデル	232		- 選択された変数	241
	- 主効果+交互作用モデル	233		- パラメータ推定値	242
	2×2の格子状 - 層別散布図	220	や	要因 - 区別	232
は	パラメータ - 共分散行列	226		要因の取捨選択 - 基準	232
	パラメータの共分散行列 - $\Sigma(\theta^*)$	227		4次元交互作用 - フルモデル	238
	パラメータ推定値 - モデルのあてはめ	242		予測プロファイル - 交互作用プロット	242
	標準誤差 - 対角要素の平方根	226		- JMP	242
	標準最小2乗法 - 繋げる配慮	239	ら	冷蔵庫の塗装不良 - 現業実習	219

第7章 解析用ファイル一覧

サイズ	名前		種類
 73 KB	第07章_01_層別散布図		Microsoft Excel ワークシート
 79 KB	第07章_02_95%信頼区間		Microsoft Excel ワークシート
 58 KB	第07章_03_変数選択		Microsoft Excel ワークシート
 13 KB	第07章_04_1_JMP_ステップ		JMP Data Table
 9 KB	第07章_04_2_JMP_ステップ名義		JMP Data Table
 13 KB	第07章_04_3_JMP_グラフビルダー		JMP Data Table

非売品, 無断複製を禁ずる

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 7 章 共変量を含む 3 因子実験データの探索的解析

BioStat 研究所(株)
〒105-0014 東京都 港区 芝 1-12-3 の1005
2024 年 2 月 7 日 高橋 行雄
takahashi.stat@nifty.com , FAX : 03-342-8035