

共変量を含む多因子実験データでの 交互作用を考慮した探索的な変数選択

高橋 行雄
BioStat 研究所(株)

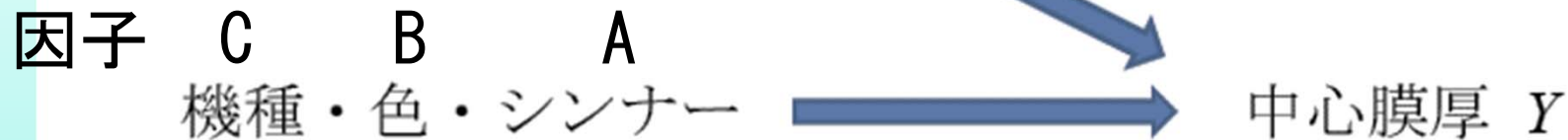
共変量がある場合の変数選択

- ◆ 繰り返し数が **10** の **3** 因子実験データ
 - 交互作用も含めて全ての要因が互いに直交
 - 分散分析表の **p** 値が大きい要因を誤差項にプーリング
- ◆ 考慮しなければならない **共変量 X** がある場合
 - 平方和の分解による分散分析表の **作成ができない**
- ◆ 宮川(2008)の「**共変量を含む3因子実験データ**」
 - 層別散布図に回帰直線を重ね書きした図を主体
 - ダミー変数を用いた回帰分析
 - 散布図における層別に勝るものはないことを強調

冷蔵庫の塗装：実験目的

- ◆ 管理限界を超える不良の発生原因の探求
- ◆ 不良品の発生を防止するために方法は

ガンスプレーの吐出量 X



		機種			
		C1 国内向け		C2 国外向け	
シンナーの種類		塗料の色		塗料の色	
		B1	B2	B1	B2
		x	y	x	y
A1	1				
	⋮				
	10				
A2	1				
	⋮				
	10				

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		色2		注
シンナー	x	y	x	y		x	y	x	y	
2	89.6	32.3	90.7	31.2		91.0	33.3	89.7	29.7	

	機種：国内					機種：国外				
	色1		色2		注	色1		色2		注
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				36.7	△	
	88.9	31.6	89.7	28.8				34.2		
	90.4	32.6	90.4	29.7						
	88.5	30.3	91.2	28.8						
	92.6	33.0	89.3	28.0						
	91.3	31.4	91.7	31.0						
	91.5	32.4	92.3	29.8						
	87.7	31.2	89.7	30.3						
	89.7	31.5	91.2	30.1						
平均	90.12	32.02	90.25	29.33						
SD	1.55	0.98	1.32	1.00						

シンナーの種類		機種							
		C1 国内向け				C2 国外向け			
		塗料の色				塗料の色			
		B1		B2		B1		B2	
		x	y	x	y	x	y	x	y
A1	1								
	:								
	10								
A2	1								
	:								
	10								

注) △:流れ不良(yが36以上), ▼

Excel による層別散布図・回帰分析

◆ Excel の散布図の活用

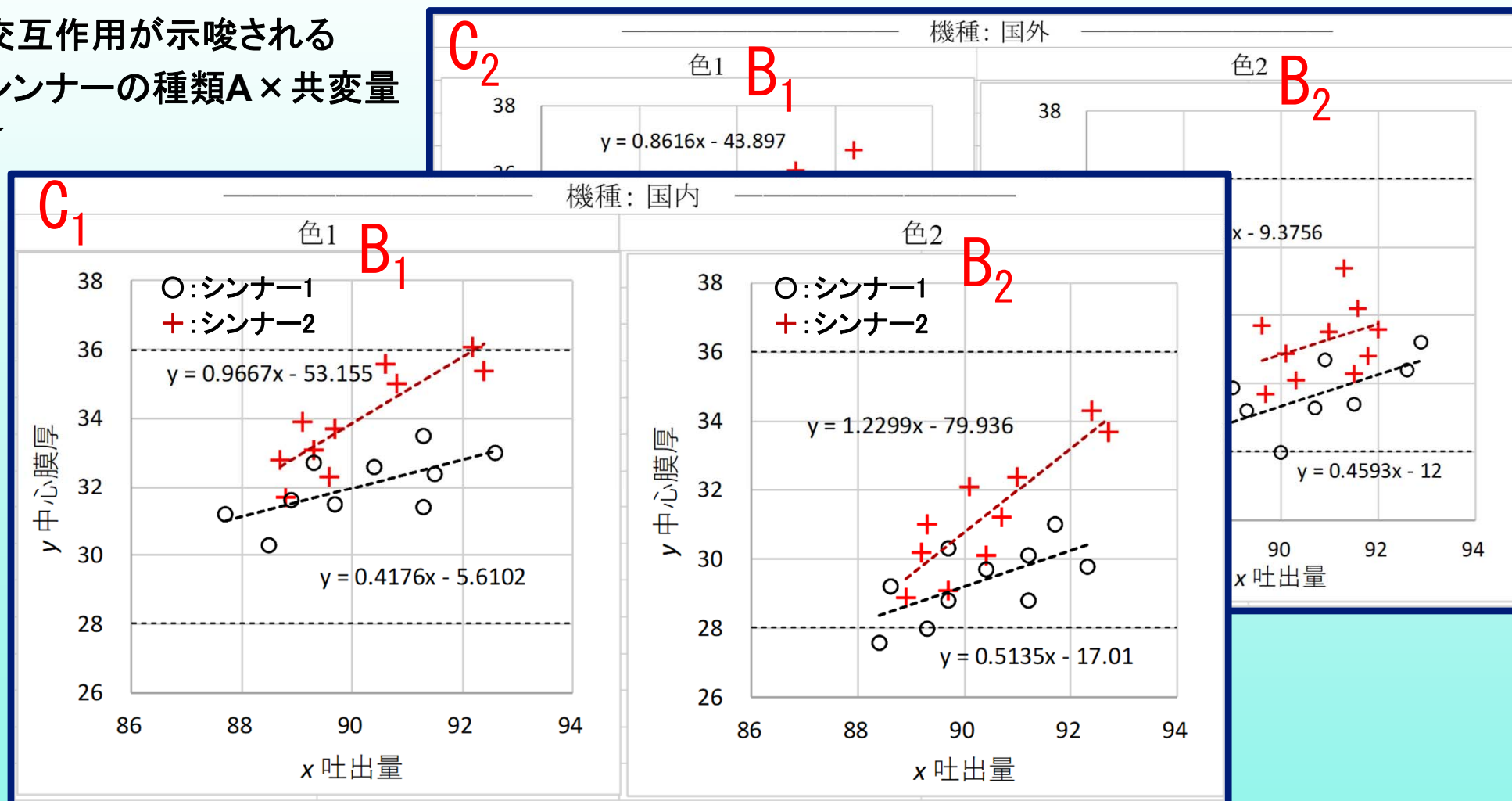
- 質的変数に対する格子状の層別(A)散布図
 - 複数の層別変数(因子)がある場合
 - 行方向に層別C, 列方向に層別B, 格子状の層別散布図行列
- Excel は, きめ細かな散布図の作成ができる

◆ Excelの回帰分析を用いて, 共変量を含む交互作用を考慮した変数選択

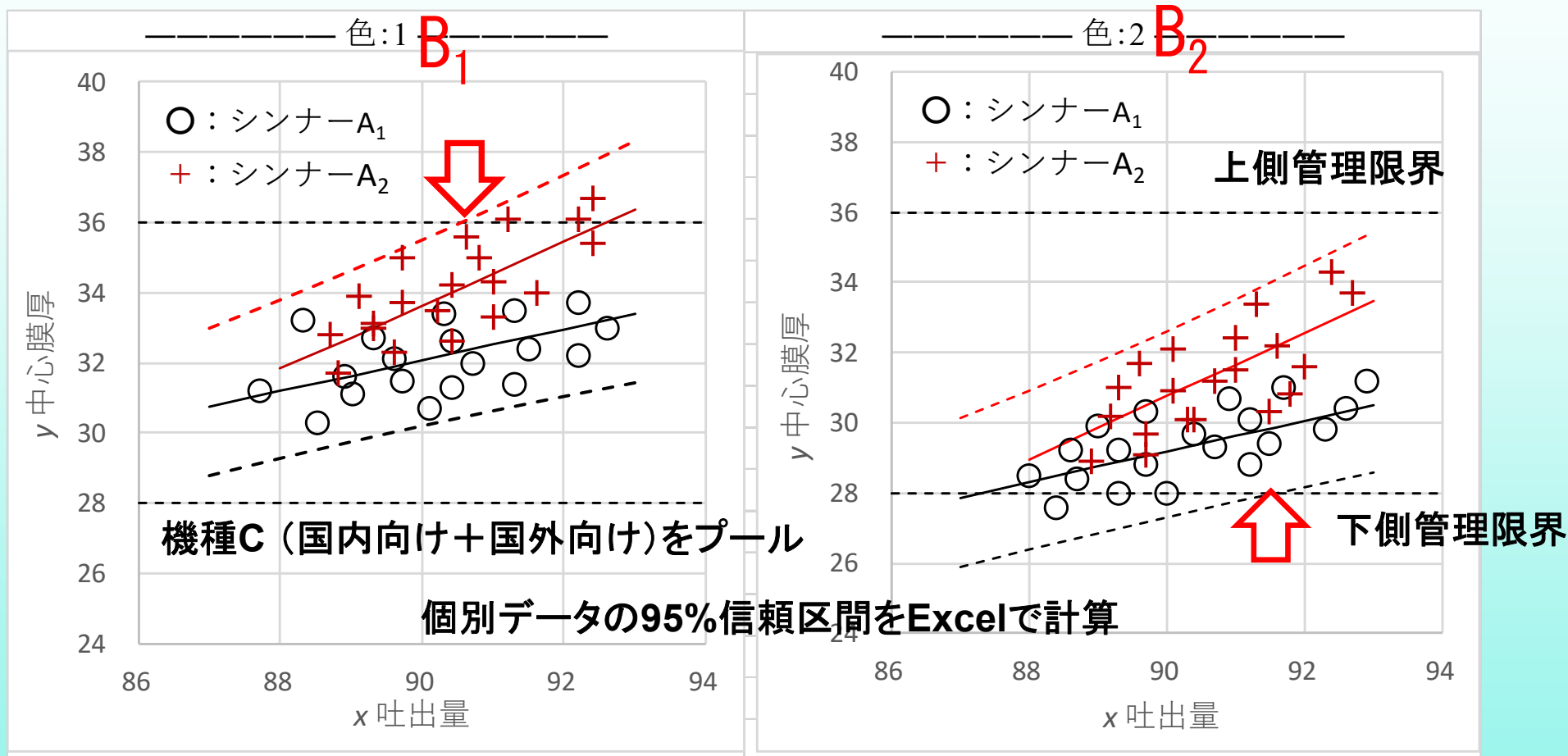
- 管理限界がある特性値
 - 回帰直線と個別データの95%信頼区間の重ね書き
 - 管理限界と個別データの95%信頼区間の交点が最重要

Excel による格子状の層別散布図

交互作用が示唆される
シナーの種類A × 共変量
x



層別散布図による試行錯誤的な解決



統計的方法により結果の妥当性を示したい

JMPの「ステップワイズ法」

- ◆ JMP「ステップワイズ法」の特徴
 - 交互作用を含む階層的な変数選択に対応
 - 3水準以上の質的変数にも対応
 - 量的変数との交互作用では中心化が自動設定
 - 量的な反応だけでなく、2値反応、順序反応にも対応
- ◆ 変数選択に引き続き「標準最小2乗法」が起動できる
 - 「予測プロファイル」と「交互作用プロファイル」の活用
- ◆ 「グラフビルダー」による格子状の層別散布図
 - 多面的・総合的な問題解決法のサポート役
 - 個別データの95%信頼区間がサポートされている

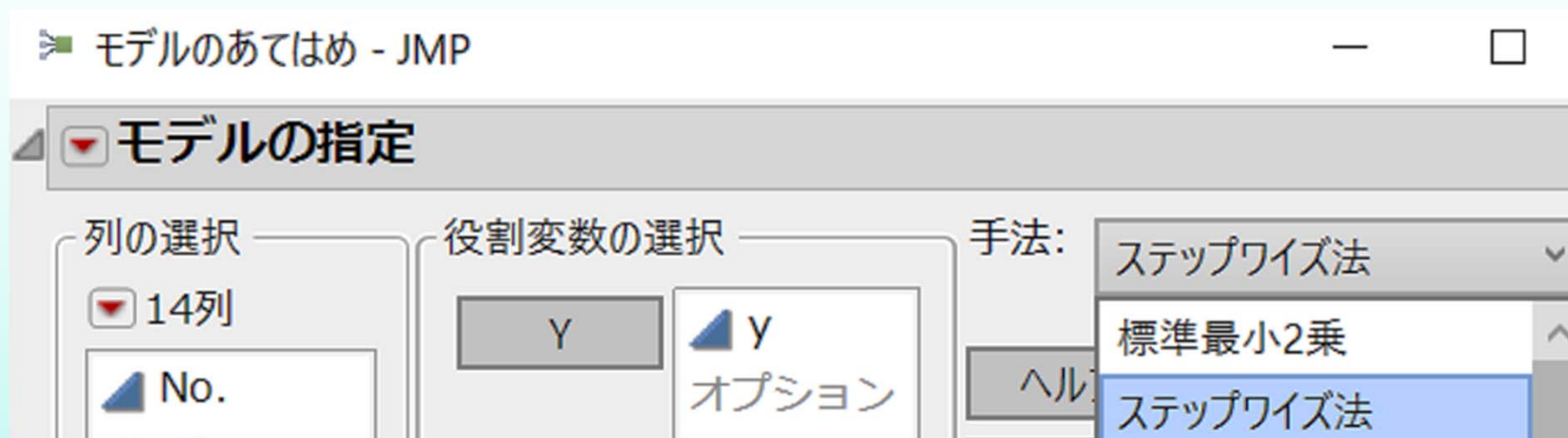
変数選択のためのJMPデータファイル

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

主効果： A シンナー
 B 色
 C 機種
 吐出量 X

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

「モデルのあてはめ」から「ステップワイズ法」



ステップワイズ法による変数選択を行う。この手法では、標準最小2乗法、順序ロジスティックモデル、2値応答の名義ロジスティックモデルにおいて、ステップワイズ法による変数選択を行える。

「ステップワイズ法」は、
「モデルのあてはめ」の
「手法」から選択

「ステップワイズ法」の実行

▼ 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

変数を除去するときのp値 0.1

方向: 変数増減

実行 学習行

すべて追加 モデルの作成

すべて削除 モデルの実行

- ◆ 変数を「すべて追加」する
- ◆ 停止ルール:「閾値p値」を選択
- ◆ 変数増減のp値: 0.1 と設定
- ◆ 方向:「変数増減」を選択
- ◆ 手作業による変数選択を

ステップワイズ法：手作業による変数増減法

(3 因子モデル)

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

(2 因子モデル)

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

(最終モデル)

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。

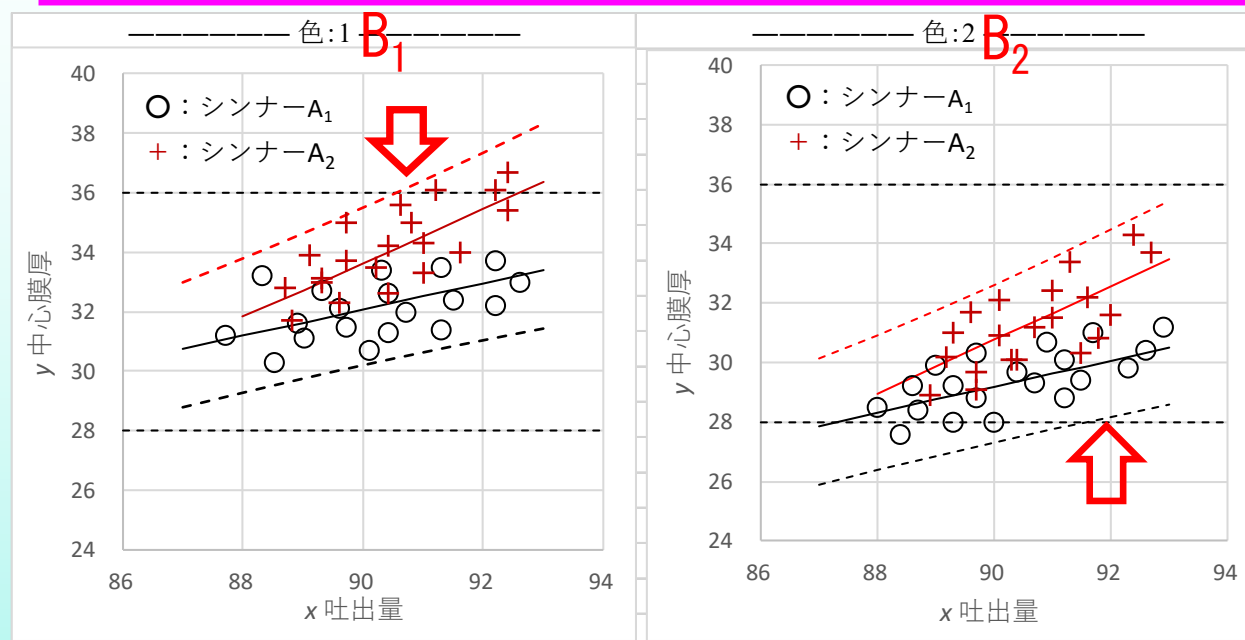
交互作用を含む変数増減法における留意点

- ◆ 主効果, 2 因子交互作用, 3因子交互作用を**区別**
- ◆ 高次の交互作用から段階的に
 - 主効果も含め低次の交互作用の p 値を完全に無視
- ◆ 同じレベルの交互作用の p 値を考慮
 - **3 因子交互作用 abc が有意ではない**
- ◆ 交互作用 abc を除いた 2 因子モデル
 - 2 因子交互作用の p 値が大きいものから適宜減少
 - 交互作用 ax は, 小さいまま残り続ける.
- ◆ 主効果 c は, 交互作用項にも含まれていないので除く
- ◆ 主効果 b の p 値の変化 $p=0.475 \rightarrow 0.504 \rightarrow 0.000$
なぜこのような変化が起きるのか

交互作用が含まれている場合の主効果

- ◆ 量的変数 x と 質的変数をダミー変数 b とし, それらの積 bx が含まれている場合
 - 積 bx がモデルに含まれていなければ, ダミー変数 b の推定されたパラメータは, 平行な直線をあてはめた場合の $b = 0$ の切片からの差である.
 - 積 bx をモデルに含めると, 平行でない 2 つの直線の切片の差が, b で推定されたパラメータとなる.
- ◆ その結果, 2 つの直線の傾きの差により b で推定されたパラメータは, 量的変数 x の範囲に依存し揺れ動くので, 使い物にならない.

層別散布図での結果の統計的妥当性



(最終モデル)

	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

- ◆ シンナー A とガンスプレーの吐出量 X に明らかな交互作用があることが確認された。
- ◆ 色 B により, 中心膜厚 Y に明らかな影響があることが確認された。
- ◆ 管理限界を超える不良品を軽減させる方法については, 変数選択の結果からでは, **何もヒントが得られない**。

「ステップワイズ法」のさらなる活用

- ◆ 質的変数 A, B, C の水準 (1, 2) のまま使いたい.
- ◆ 交互作用についても全自動で設定したい.
- ◆ 変数 (A, B, C, X) の 4 次の交互作用も含め, 自動的な変数選択を行い, 手作業で得られた変数選択の結果と一致することを確認したい.
- ◆ (変数増加法, 変数減少法, 変数増減法) のすべての結果が一致することを確認したい.

用いるJMPデータファイル

	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

マクロを用いてすべての組み合わせ

モデルの あてはめ

完全実施要因

設定された次数まで

すべての組み合わせ

応答曲面

配合応答曲面

多項式の次数

Scheffeの3次多項式

部分 3次

モデル効果の構成

追加 交差 枝分かれ マクロ

次数 2

属性 変換

切片なし

A
B
C
x
A*B
A*C
A*x
B*C
B*x
C*x
A*B*C
A*B*x
A*C*x
B*C*x
A*B*C*x

ロック	追加	パラメータ
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	切片
<input type="checkbox"/>	<input type="checkbox"/>	A{A1-A2} ?
<input type="checkbox"/>	<input type="checkbox"/>	B{B2-B1} ?
<input type="checkbox"/>	<input type="checkbox"/>	⋮
<input type="checkbox"/>	<input type="checkbox"/>	(x-90.3662) ?
<input type="checkbox"/>	<input type="checkbox"/>	A{A1-A2}*B{B2-B1}
<input type="checkbox"/>	<input type="checkbox"/>	A{A1-A2}*C{C1-C2}
<input type="checkbox"/>	<input type="checkbox"/>	A{A1-A2}*(x-90.3662)
<input type="checkbox"/>	<input type="checkbox"/>	B{B2-B1}*C{C1-C2}
<input type="checkbox"/>	<input type="checkbox"/>	B{B2-B1}*(x-90.3662)
<input type="checkbox"/>	<input type="checkbox"/>	C{C1-C2}*(x-90.3662)
<input type="checkbox"/>	<input type="checkbox"/>	A{A1-A2}*B{B2-B1}*C{C1-C2}
<input type="checkbox"/>	<input type="checkbox"/>	A{A1-A2}*B{B2-B1}*(x-90.3662)
<input type="checkbox"/>	<input type="checkbox"/>	A{A1-A2}*C{C1-C2}*(x-90.3662)
<input type="checkbox"/>	<input type="checkbox"/>	B{B2-B1}*C{C1-C2}*(x-90.3662)
<input type="checkbox"/>	<input type="checkbox"/>	A{A1-A2}*B{B2-B1}*C{C1-C2}*(x-90.3662)

ステップ
ワイズ

列とすべての交互作用を、主効果、2次の交互作用...というように次数の順に追加する。

ステップワイズの実行

ステップワイズ回帰の設定

停止ルール:

閾値p値



すべて追加

モデルの作成

変数を追加するときのp値 0.25



すべて削除

モデルの実行

変数を除去するときのp値 0.1

方向:

変数増加

ルール:

組み合わせ

組み合わせ

制限

ルールなし

効果全体

より低次の項をもつ項に関して、該当の項とその低次の項を組み合わせたときのp値、および、低次の項がすでにモデルに追加されている状態で該当の項を追加したときのp値を計算する。これら2つのp値のうち大きい方のp値を、モデル選択におけるp値とする。

変数選択の結果(方法によらず一致)

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

以下の3因・4因子子交互作用は表示に含めず

(モデルの作成) → (実行), 推定値?

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	3		

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*

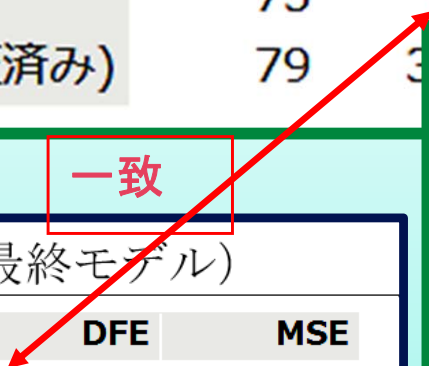
(最終モデル)		
SSE	DFE	MSE
62.846	75	0.838

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

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*

一致

一致しない



JMPの内部(1, -1)対比型ダミー変数

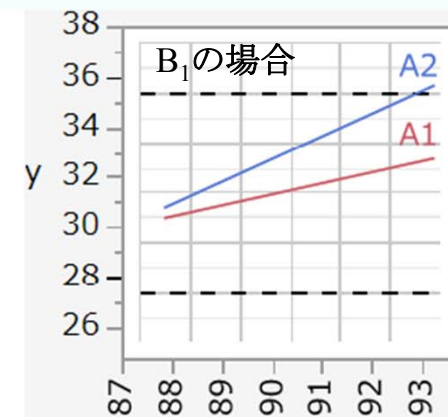
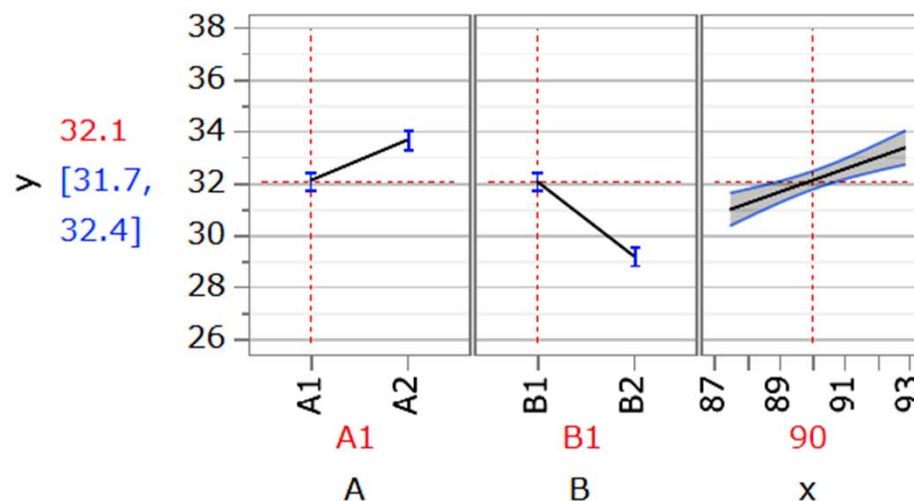
シンナー	a	色	b	機種	c	シンナー	色	a	b	ab
1	0	1	0	国内	0	1	1	0	0	0
2	1	2	1	国外	1		2	0	1	0
						2	1	1	0	0
							2	1	1	1
シンナー	a	色	b	機種	c	シンナー	色	a	b	ab
1	1	1	1	国内	1	1	1	1	1	1
2	-1	2	-1	国外	-1		2	1	-1	-1
						2	1	-1	1	-1
							2	-1	-1	1

◆ ガンスプレーの吐出量 x

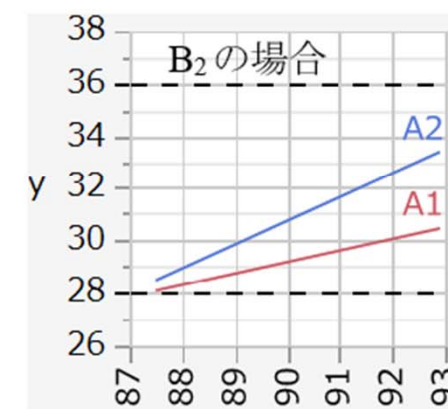
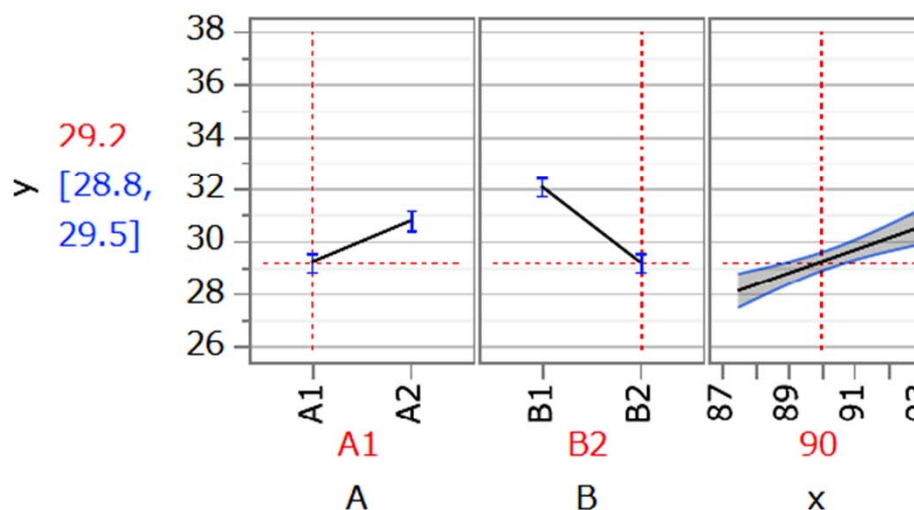
- 交互作用 $A^*(x - 90.3662)$: 総平均からの偏差(中心化)

(予測プロファイル) + (交互作用プロファイル)

$A_1 B_1$
を選択



$A_1 B_2$
を選択



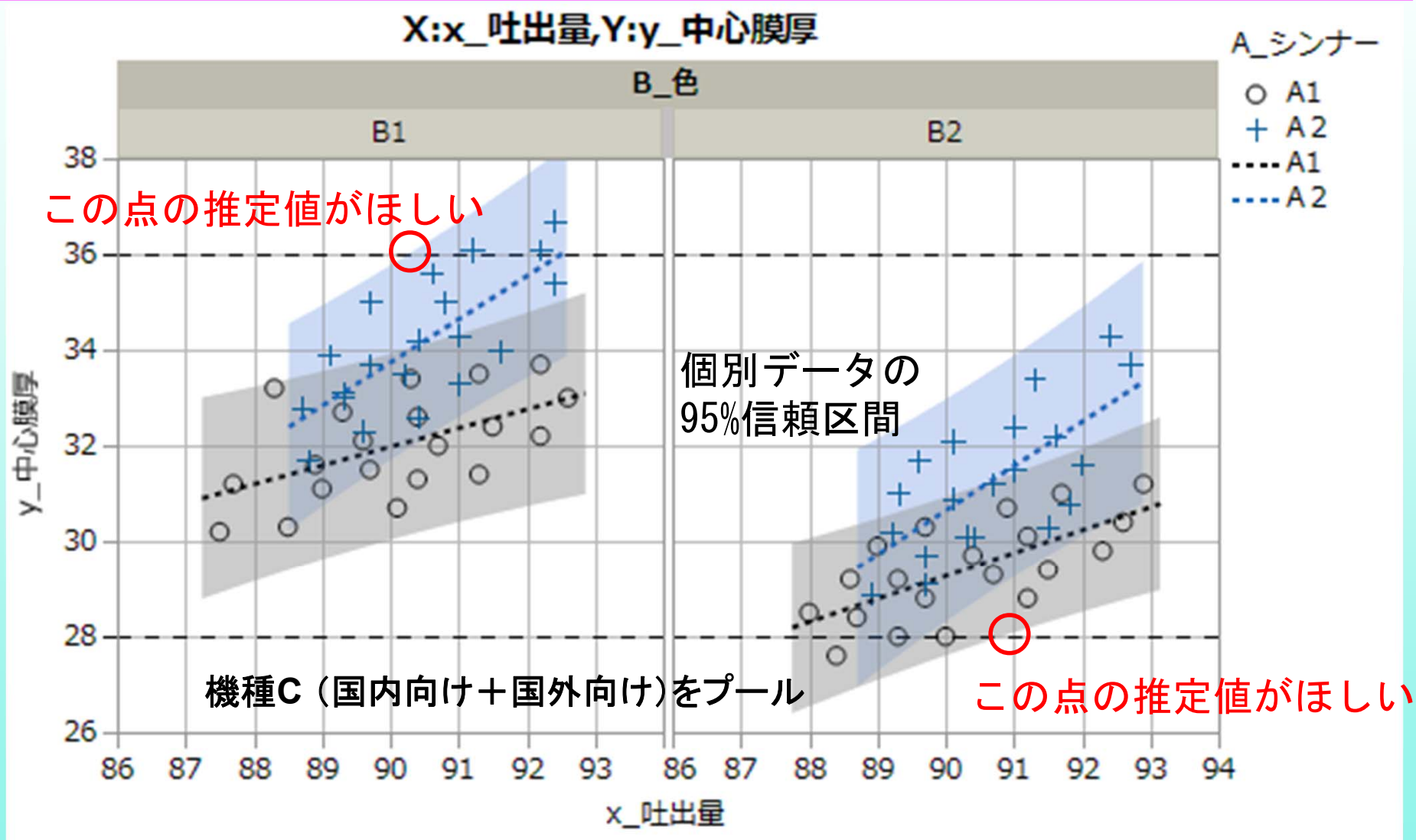
JMPの予測プロファイル

- ◆ JMPの予測プロファイルは、理解が困難なパラメータの意味のある推定値(Lsmeans:最小2乗平均)および95%信頼区間を可視化する優れた機能で、他のソフトに類を見ない優れたものである。
- ◆ Rでは、lsmeansパッケージで対応できる。
- ◆ 予測プロファイルに、個別データの95%信頼区間の表示はサポートされていない。
- ◆ 交互作用プロファイルに、95%信頼区間の表示はサポートされていない。
- ◆ JMPのさらなる発展に期待したい。

JMPの「グラフビルダー」

- ◆ 「グラフビルダー」は、多機能なので試行錯誤的な対応は生産的でないので、作成ストーリーを明確にする。
 - 変数選択の結果を参考にする.
- ◆ 層別散布図に回帰直線 および 個別データの 95%信頼区間を重ね書きする. 管理限界線も入れる。
 - ガンスプレーの吐出量 X と交互作用がある A:シンナーの種類を層別因子とし、マーカで区別する.
 - A:シンナーの種類を「重ね合わせ」とし、回帰直線と個別データの95%信頼区間を重ね書きする.
 - 色 B は、主効果で有意な差なので、「グループ変数」とし、2つの層別散布図を横に並べる.

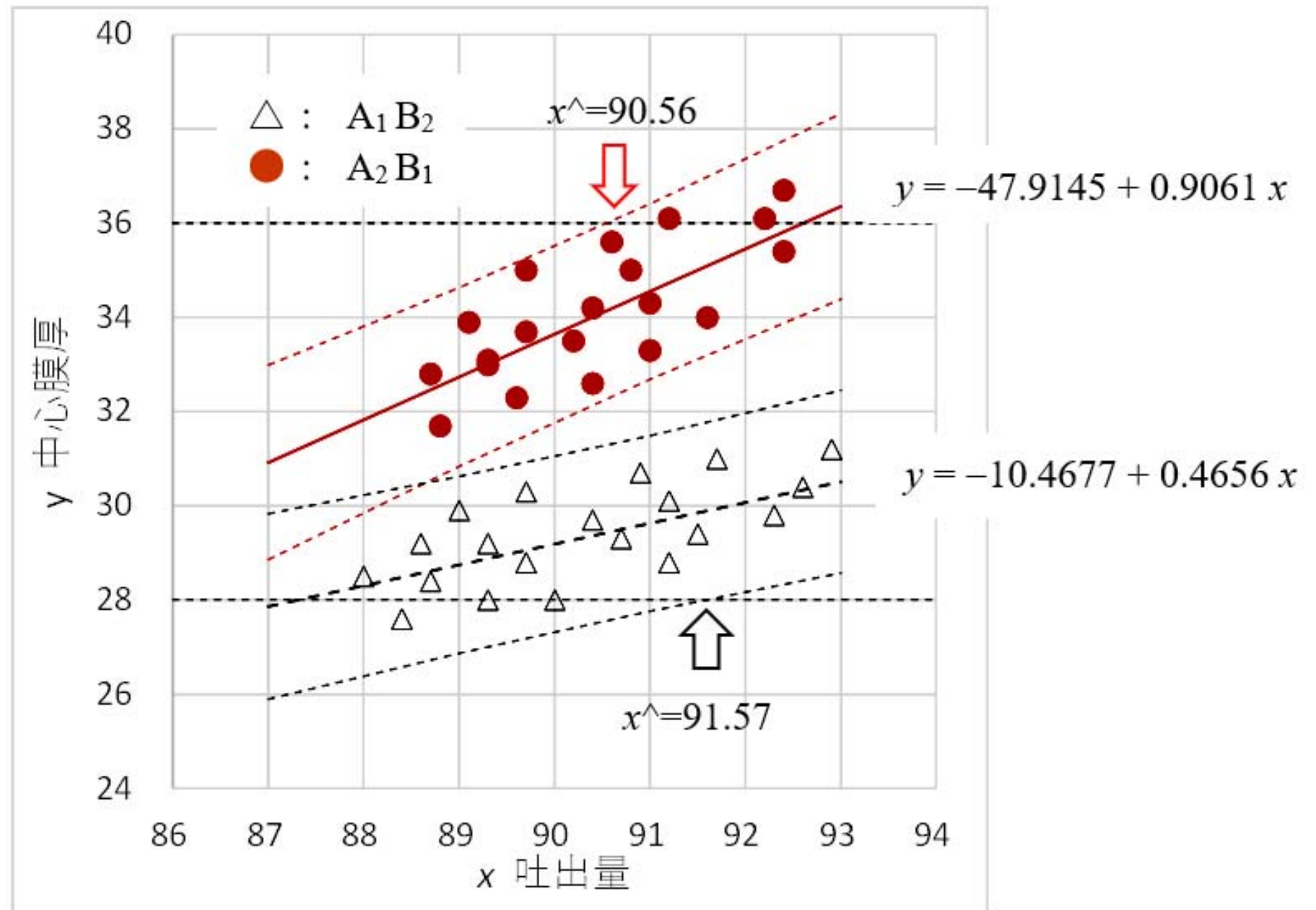
「グラフビルダー」による作図



Excel の回帰分析 + 行列計算

- ◆ Excelの回帰分析に引き続き，デザイン行列からExcelの行列計算により，パラメータの共分散行列 $\Sigma(\theta^{\wedge})$ を求め，個別データの95%信頼区間を推定する．Excelのソルバーで管理限界との交点を求めた結果を示す．

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}	$Vat(y^{\wedge})$	個別L 95%	個別U 95%
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。											
1	2	$L^{(20)}$	1	0	1	91.57	0	29.8762	0.0491	28.0000	31.7525 #



まとめ

- ◆ 伝統的な実験計画法は、「平方和の分解」による分散分析表の作成を目指している。
 - 共変量を含む場合には、「平方和の分解」が使えない。
- ◆ JMPの「ステップワイズ」は、質的変数を(1, -1)対比型ダミー変数に変換、共変量も含め、交互作用を考慮した変数選択が行える優れたものである。
- ◆ 変数選択後に、通常の「モデルのあてはめ」による解析が連携していることはうれしい。
- ◆ 結果の解釈には、グラフビルダーが大活躍する。







発表に関連する添付資料

◆ 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

pdf 暫定版

◆ Excelファイル および JMPファイル 一式 zip

サイズ	名前
 75 KB	第7章_01_層別散布図_JMP
 90 KB	第7章_02_95%信頼区間_JMP
 123 KB	第7章_03_変数選択_JMP
 13 KB	第7章_04_1_JMP_グラフ
 13 KB	第7章_04_2_JMP_ステップ
 9 KB	第7章_04_2_JMP_ステップ名義

参考文献

- ◆ 宮川雅巳(2008), SQCの基本・問題の発見と解決の科学, 日本規格協会.
6.3節 計画的に採取した3元配置データの解析例,
10.2節 変動要因解析のための回帰分析
- ◆ 守屋和幸ら(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
https://www.jstage.jst.go.jp/article/chikusan/89/1/89_1/_pdf/-char/ja
- ◆ 高橋行雄(2020), 最尤法によるポアソン回帰分析入門
＜第4章＞デザイン行列を用いた回帰分析入門
<https://www.yukms.com/biostat/takahasi2/rec/009-04.htm>
- ◆ 高橋行雄(2020), 最尤法によるポアソン回帰分析入門
＜第13章＞最小2乗平均の謎を予測プロファイルで解く
<https://www.yukms.com/biostat/takahasi2/rec/009-13.htm>
- ◆ 高橋行雄(2022), 線形モデルによる欠測値がある直交表の解析
－謎めいた最小2乗平均と95%信頼区間の活用－
<https://www.yukms.com/biostat/takahasi2/rec/010.htm>
- ◆ 高橋行雄(2023), JMPで繰り返し不揃いの2元配置データの解析ができるの？
－平方和の分解ではなくデザイン行列と最小2乗平均の活用－
<https://www.yukms.com/biostat/takahasi2/rec/011.htm>

ご清聴ありがとうございました