

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
2023 年 12 月 17 日

第 1 章 層別因子を含む各種の回帰分析の実際

ほとんどの統計の教科書で回帰分析が取り上げられている。ただし、単回帰分析に特化した解析法が示されているだけで、層別因子を含む複数の回帰直線についての解析法を取り上げている成書はごく限られている。さらに多くは、手計算時代の偏差平方和を主体にした解析法が示されており、技巧的かつ難解であり推奨できない。多くの統計ソフトは、デザイン行列を主体にした行列計算による解析を行なっているが、その解析法はブラックボックス的である。そこで、特徴的な事例が示されている成書の事例について、デザイン行列を用いた回帰分析を主体にし、各種のパラメータの 95%信頼区間の推定に、Excel の行列関数を活用することにより、見通しの良い汎用的な解析方法を例示する。

第 1 章 目 次

1.	層別因子を含む各種の回帰分析の実際	7
1.1.	年収の比較調査データ	7
	別々の回帰直線のあてはめ、デザイン行列 X を用いた単回帰分析、 デザイン行列 X を用いた 3 本の回帰直線の同時あてはめ、共通の傾きを 持つ 3 本の回帰直線、交互作用の検討、回帰直線上の推定値にする 95% 信頼区間、会社間の推定値の差の 95%信頼区間、デザイン行列とは何か	
1.2.	季節による洗浄用水の温度と回収液の濃度の関連	19
	季節による層別、JMP の「グラフ・ビルダー」による層別散布図、JMP の 「モデルのあてはめ」による解析、JMP の「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 (1, -1) 対比型ダミー変数を用いた解析、 (1, -1) 対比型ダミー変数における回帰直線の推定式、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	
	次ページに続く	

1.3. 前値を共変量とする交互作用がある抗うつ剤の評価-----	30
非平行な 2 本の直線, 推定されたパラメータによる交互作用の検討, 分散分析表による交互作用の検討, 非平行な 2 本の回帰直線の差, 2 本の直線間の差に対する 95%信頼区間および t 検定, 2 本の回帰 直線が有意な差となる前値 x_0 , JMP による検証	
1.4. 発電プラント周辺でのカキの成長-----	38
前値を共変量とする対照群を含む 2 元配置, 共変量を含む 1 元配置として モデル化, 1 元配置の水準の組み合わせ平均, 1 元配置を 2 元配置として 再構築, SAS による検証	
1.5. 除草剤の散布による雑草の成長抑制-----	47
MCP 剤に対する共通の切片を持つ回帰直線, PCP 剤に対する原点を 共通とする回帰直線, R による検証(原点共通の場合)	
文献索引, 索引, 解析用ファイル一覧-----	(55)

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-----	53
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-----	122
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 目次, 解析用ファイル一覧-----	489

1. 層別因子を含む各種の回帰分析の実際

ほとんどの統計の教科書で回帰分析が取り上げられている。ただし、単回帰分析に特化した解析法が示されているだけで、層別因子を含む複数の回帰直線についての解析法を取り上げている成書はごく限られている。さらに多くは、手計算時代の偏差平方和を主体にした解析法が示されており、技巧的かつ難解であり推奨できない。多くの統計ソフトは、デザイン行列を主体にした行列計算による解析を行なっているが、その解析法はブラックボックス的である。そこで、特徴的な事例が示されている成書の事例について、デザイン行列を用いた回帰分析を主体にし、各種のパラメータの95%信頼区間の推定に、Excelの行列関数を活用することにより、見通しの良い汎用的な解析方法を例示する。

1.1. 年収の比較調査データ

表 1.1 に示すのは、芳賀（2014）, 「医薬品開発のための統計解析 第2部 実験計画法 改訂版, 第4章 共分散分析」に示されているデータで、就職先を検討するための会社 A (A_1 , A_2 , A_3) の従業員をランダムに 10 人を選び、年齢 X および年収 Y を調査した結果である。

表 1.1 年収の比較調査の結果 [芳賀（2014）, 表示 4.1.1]

	会社A ₁		会社A ₂		会社A ₃	
	年齢	年収	年齢	年収	年齢	年収
No	x	y	x	y	x	y
1	34	684	34	692	34	723
2	33	788	34	712	39	762
3	34	764	37	700	37	883
4	37	836	43	843	29	678
5	29	606	37	748	33	699
6	26	696	32	580	36	830
7	37	766	46	837	41	900
8	38	862	36	667	37	835
9	32	606	41	805	45	905
10	31	708	35	689	39	793
平均	33.10	731.60	37.50	727.30	37.00	800.80
				総平均	35.87	753.23

別々の回帰直線のあてはめ

年収の平均を比較すると、高い順に (A_3 : 800.8 万円, A_1 : 731.6 万円, A_2 : 727.3 万円) となる。平均年齢が 3 社で微妙に異なるので、年齢を加味して年収を比較したい。図 1.1 に示すのは、会社 A (A_1 , A_2 , A_3) 別の Excel で作成した層別散布図に、近似曲線の線形近似を使って回帰直線を重ね書きした結果である。

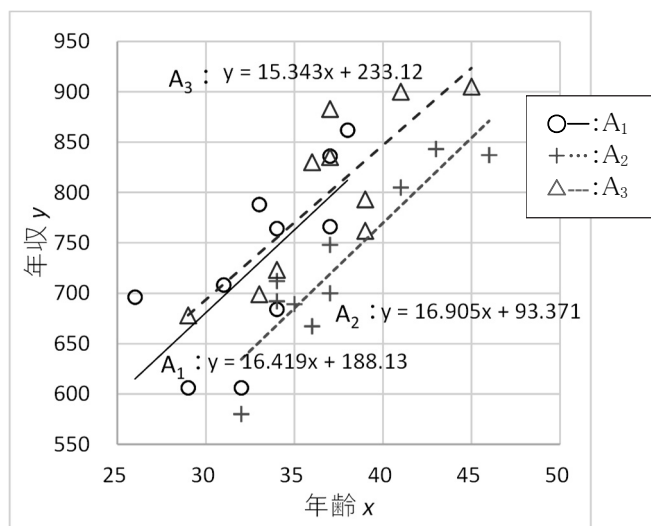


図 1.1 層別散布図による会社別の年齢と年収の関連

図 1.1 から A_3 社の平均年収が 800.8 万円と高いのは、平均年齢が 37.0 歳と A_1 社に比べ高いためと推測される。 A_1 社の平均年収 731.6 万円は、平均年齢が 33.1 歳と低いためであり、同一年齢で比較をすれば A_3 社と同程度の年収であることが、回帰直線が重なっていることから読み取れる。 A_2 社の年収が 727.3 万円と A_1 社と同程度であったが、平均年齢が 37.5 歳と高いためであり、同一年齢に固定して年収を比較すれば、3 社の中で最も低いことが読み取れる。

一般的に同業の会社であっても年齢構成は異なるし、企業規模により同一年齢であっても年収に違いがあることも知られている。図 1.2 に示すように模式的に表せば、会社の違いが年収に影響を与えていることもあれば、会社の違いが年齢構成にも影響を与えることにより、年齢が年収に影響を与えていることもあり、会社間の年収の比較をするためには、年齢の違いを統計的に補正する必要がある。

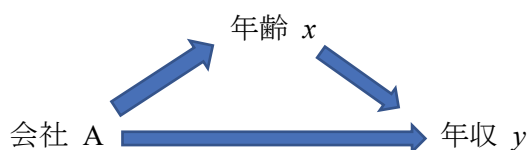


図 1.2 年収に影響をおよぼす要因

デザイン行列 X を用いた単回帰分析

A₁ 社についてデザイン行列 X を用いた単回帰分析を行なってみよう．表 1.2 に示すようにデザイン行列 X は、(切片 x_0 ，説明変数：年齢 x_1) のように、切片に対応する変数 $x_0 = 1$ を含んでいる．回帰分析を行なうための統計ソフトでは、説明変数として年齢 x_1 のみを指定するのが一般的であり、切片 $x_0 = 1$ を内部で自動的に付加している．原点を通る回帰直線を推定したい場合には、切片を含まない解析が必要となるので、統計ソフトでは、「切片の自動付与を抑制」オプションが備わっている．

表 1.2 左に示す (10 行×2 列) のデザイン行列 X をそのまま用いて回帰分析をしたい．表 1.2 右に示す Excel の回帰分析では、「定数に 0 を使用」をオン ☒ として切片を自動的に付与することを抑制し、「入力 X の範囲」で変数 x_0 と x_1 の (10 行×2 列) のデータの範囲を選択している．表 1.2 右に示す「分散分析表」の「回帰」の自由度が 2 となっているのは、変数 x_0 と x_1 の 2 変数に対応し、「合計」の自由度 10 は、データ数に、「残差」の自由度は、それらの差 $df = 10 - 2 = 8$ に対応している．回帰パラメータの推定で、切片の係数が「0」となっているのは、「定数に 0 を使用」をオンとしたためで、まったく意味のない出力である．

表 1.2 A₁ 社についてデザイン行列 X を用いた単回帰分析

デザイン行列 X								
	切片	年齢	年収					
No	x_0	x_1	y	分散分析表, 「定数に 0 を使用」を on				
1	1	34	684		自由度	変動	分散	分散比
2	1	33	788	回帰	2	5387135	2693567	629.9015
3	1	34	764	残差	8	34209	4276.17	
4	1	37	836	合計	10	5421344		
5	1	29	606					
6	1	26	696		係数	標準誤差	t	P -値
7	1	37	766	切片	0	#N/A	#N/A	#N/A
8	1	38	862	β_0	x_0	188.1334	191.7648	0.9811
9	1	32	606	β_1	x_1	16.4189	5.7597	2.8506
10	1	31	708	回帰の自由度が 2なのは、 x_0 と x_1 の 2 変数に対応				

推定されたパラメータ (係数) から回帰式

$$\left. \begin{aligned} \hat{y} &= \hat{\beta}_0 x_0 + \hat{\beta}_1 x_1 \\ &= 188.1334 x_0 + 16.4189 x_1 \end{aligned} \right\} \quad (1.1)$$

が得られ、図 1.1 に示した A₁ 社に対する近似曲線の重ね書きした回帰式に一致する．パラメータ (係数) の標準誤差 SE は、

$$\left. \begin{aligned} SE(\hat{\beta}_0) &= 191.7648 \\ SE(\hat{\beta}_1) &= 5.7597 \end{aligned} \right\} \quad (1.2)$$

で、 t 値は、

$$\left. \begin{aligned} t(\hat{\beta}_0) &= \frac{\hat{\beta}_0}{SE(\hat{\beta}_0)} = \frac{188.1334}{191.7648} = 0.9811 \\ t(\hat{\beta}_1) &= \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} = \frac{16.4189}{5.7597} = 2.8506 \end{aligned} \right\} \quad (1.3)$$

として計算されている．これらは，残差の自由度 8 の t 分布に従うことから p 値は，

$$\left. \begin{aligned} p(\hat{\beta}_0) &= 0.3553 \\ p(\hat{\beta}_1) &= 0.0215 \end{aligned} \right\} \quad (1.4)$$

となり，切片 $\hat{\beta}_0 = 188.1334$ は，統計的に有意ではなく，傾き $\hat{\beta}_1 = 16.4189$ は， $p < 0.05$ と統計的に有意な差となっている．

デザイン行列 X を用いた 3 本の回帰直線の同時あてはめ

単回帰分析を拡張して A_2 社， A_3 社についても一括して回帰分析によりパラメータを推定する方法はあるのだろうか．図 1.1 に示した 3 本の回帰直線の回帰式を同時推定するために，ダミー変数を用いたデザイン行列 X を生成する．ダミー変数は，表 1.3 に示すように質的変数を何らかの量的変数に置き換えた変数の総称であり，どのような量的変数に置き換えるかは自由であり，解析の目的に合ったダミー変数を生成する必要がある．図 1.1 に示した 3 本の回帰直線の切片と傾きを同時に推定するために表 1.3 に示した各種のダミー変数の型の中から (1, 1) 標示型ダミー変数を選択する．

表 1.3 各種のダミー変数

(1, 1) 標示型				(1, -1) 対比型				(0, 1) 型				(1, 0) 型			
会社A	a_1	a_2	a_3	会社A	a_1	a_2	a_3	会社A	a_1	a_2	a_3	会社A	a_1	a_2	a_3
A_1	1	0	0	A_1	1	0	.	A_1	.	0	0	A_1	1	0	.
A_2	0	1	0	A_2	0	1	.	A_2	.	1	0	A_2	0	1	.
A_3	0	0	1	A_3	-1	-1	.	A_3	.	0	1	A_3	0	0	.

(1, 1) 標示型ダミー変数は， A_1 社に対し $a_1 = 1$ ， A_1 社以外は 0，同様に A_2 社に対し $a_2 = 1$ ， A_2 社以外は 0， A_3 社に対し $a_3 = 1$ ， A_3 社以外は 0 とする．表 1.2 では，切片に対する変数として x_0 ，年齢を x_1 としたのであるが，会社 A ごとに異なる変数とするために，(1, 1) 標示型ダミー変数 (a_1 ， a_2 ， a_3) を会社ごとの切片，会社 A ごとの年齢を (x_1 ， x_2 ， x_3) とする．表 1.4 左に示すように，これらの 6 変数を (30 行×6 列) のデザイン行列 X とし，Excel の回帰分析で，「定数に 0 を使用」をオンとすることにより，3 本の回帰直線の同時のあてはめを行なう．

会社	切片	年齢
A_1 社	a_1	x_1
A_2 社	a_2	x_2
A_3 社	a_3	x_3

表 1.4 会社 A (A_1, A_2, A_3) 別の回帰直線の同時あてはめ

—— デザイン行列 X ——								分散分析表, 「定数に 0 を使用」を on					
No	a_1	x_1	a_2	x_2	a_3	x_3	y		自由度	変動	分散	分散比	
1	1	34	0	0	0	0	684						
2	1	33	0	0	0	0	788	回帰	6	17182504.95	2863750.82	1061.47	
:								残差	24	64750.05	2697.92		
9	1	32	0	0	0	0	606	合計	30	17247255.00			
10	1	31	0	0	0	0	708						
11	0	0	1	34	0	0	692		係数	標準誤差	t	P -値	
12	0	0	1	34	0	0	712	切片	0	#N/A	#N/A	#N/A	
:								$\beta_0^{A_1}$	a_1	188.1334	152.3196	1.2351	0.2287
19	0	0	1	41	0	0	805	$\beta_1^{A_1}$	x_1	16.4189	4.5750	3.5889	0.0015
20	0	0	1	35	0	0	689	$\beta_0^{A_2}$	a_2	93.3714	146.7120	0.6364	0.5305
21	0	0	0	0	1	34	723	$\beta_1^{A_2}$	x_2	16.9048	3.8877	4.3482	0.0002
22	0	0	0	0	1	39	762	$\beta_0^{A_3}$	a_3	233.1202	144.9811	1.6079	0.1209
:								$\beta_1^{A_3}$	x_3	15.3427	3.8932	3.9409	0.0006
29	0	0	0	0	1	45	905	Excelシート上では, 有効数字15桁である					
30	0	0	0	0	1	39	793						

A_1 社に対する変数 (a_1, x_1) は, 他の A_2, A_3 社の場合の変数を全て 0 としている. 同様に, A_2 社に対する変数 (a_2, x_2) は, 他の A_1, A_3 社の場合の変数を全て 0 とし, A_3 社に対する変数 (a_3, x_3) は, 他の A_1, A_2 社の変数を全て 0 とし, これらをまとめて (30 行×6 列) のデザイン行列 X を設定している. このデザイン行列 X には, すべてのデータに共通の切片がない. (1, 1) 標示型ダミー変数を設定したデザイン行列に対して回帰分析を行なうためには, 「定数に 0 を使用」をオンとする必要がある. 表 1.4 右下段から, 次に示すように 3 本の回帰直線のパラメータが別々に同時推定されている.

$$\left. \begin{aligned} A_1 \text{ 社: } \hat{y}^{(A_1)} &= \hat{\beta}_0^{(A_1)} a_1 + \hat{\beta}_1^{(A_1)} x_1 = 188.1334 + 16.4189 x_1 \\ A_2 \text{ 社: } \hat{y}^{(A_2)} &= \hat{\beta}_0^{(A_2)} a_2 + \hat{\beta}_1^{(A_2)} x_2 = 93.3714 + 16.9048 x_2 \\ A_3 \text{ 社: } \hat{y}^{(A_3)} &= \hat{\beta}_0^{(A_3)} a_3 + \hat{\beta}_1^{(A_3)} x_3 = 233.1202 + 15.3427 x_3 \end{aligned} \right\} \quad (1.5)$$

表 1.4 右の「分散分析表」の「回帰」の自由度が 6 となっているのは, デザイン行列 X の 6 変数を「回帰モデル」に含めたことに対応している. また, 「合計」の自由度 30 は, データ数に対応している. A_1 社に対する傾き $\hat{\beta}_1^{(A_1)}$ の p 値は, $p=0.0015$ と A_1 社の対する単回帰分析の表 1.2 に示した結果 $p=0.0215$ に比べて小さくなっている. これは, 分散分析表で示されている誤差分散の推定値 (残差の分散) が $\hat{\sigma}_{(A_1 \text{ 単独})}^2 = 4276.17$ から $\hat{\sigma}_{(共通)}^2 = 2697.92$ と半分強と小さくなったことに起因している.

共通の傾きを持つ 3 本回帰直線

図 1.1 から, 3 本の回帰直線の傾きの差は, ほとんどなく共通の傾きを持つと推測される. そこで, 共通の傾きを持つ 3 本の回帰直線をあてはめ, 35 歳における 3 社の年収を推定し統計的に比較したい.

表 1.5 左に示すように、会社 A (A_1, A_2, A_3) ごとの切片を求めるために (1, 1) 標示型ダミー変数を (a_1, a_2, a_3) とし、年齢は共通の x とするような (30 行×4 列) のデザイン行列 X を設定する。Excel の回帰分析で、「定数に 0 を使用」をオンとすることにより、切片は異なるが、傾きが共通の回帰直線をあてはめることができる。

表 1.5 右に示すように、分散分析表の「回帰」に対する自由度は 4 となり、デザイン行列 X で設定した変数の数に一致する。「合計」の自由度は、データ数 30 に一致している。

表 1.5 会社 A (A_1, A_2, A_3) に共通の傾きを持つパラメータの推定

— デザイン行列 X —										
No	a_1	a_2	a_3	x	y	分散分析表, 「定数に 0 を使用」を on				
1	1	0	0	34	684		自由度	変動	分散	分散比
2	1	0	0	33	788	回帰	4	17182279.29	4295569.82	1718.87
:						残差	26	64975.71	2499.07	
10	1	0	0	31	708	合計	30	17247255.00		
11	0	1	0	34	692					
12	0	1	0	34	712		係数	標準誤差	t	P -値
:						切片	0	#N/A	#N/A	#N/A
20	0	1	0	35	689	$\hat{\beta}_0^{A_1}$	a_1	195.2832	76.7504	2.5444
21	0	0	1	34	723	$\hat{\beta}_0^{A_2}$	a_2	119.6903	86.5444	1.3830
22	0	0	1	39	762	$\hat{\beta}_0^{A_3}$	a_3	201.2918	85.4293	2.3562
:						$\hat{\beta}_1^{共通}$	x	16.2029	2.2690	7.1409
30	0	0	1	39	793					0.0000

推定されたパラメータから、次に示すように傾きが $\hat{\beta}_1^{共通} = 16.2029$ と共通で、

$$\left. \begin{aligned} A_1 \text{社: } \hat{y}^{(A_1)} &= \hat{\beta}_0^{(A_1)} a_1 + \hat{\beta}_1^{共通} x = 195.2832 + 16.2029x \\ A_2 \text{社: } \hat{y}^{(A_2)} &= \hat{\beta}_0^{(A_2)} a_2 + \hat{\beta}_1^{共通} x = 119.6903 + 16.2029x \\ A_3 \text{社: } \hat{y}^{(A_3)} &= \hat{\beta}_0^{(A_3)} a_3 + \hat{\beta}_1^{共通} x = 201.2918 + 16.2029x \end{aligned} \right\} \quad (1.6)$$

のように、切片が異なる 3 本の回帰直線が推定されている。

表 1.5 右で得られたパラメータ (係数) から、回帰直線の (始点, 終点) を会社ごとの年齢の (最小値, 最大値) に対する年収の推定値 \hat{y} を式 (1.6) を用いて推定し、図 1.3 に、年齢と年収の散布図上に回帰直線を重ね書きした結果を示す。この図からも明らかなように、同一年齢で比較すれば、 A_1 社と A_2 社の年収は同程度であり、 A_2 社が他の 2 社よりも低いことが一目で判断される。

交互作用の検討

共通の傾きを持つ 3 本の直線のあてはめの統計的妥当性は、得られた 2 つの分散分析表の回帰の平方和 $S_{回帰}$ の差分により評価することができる。3 本の異なる直線をあてはめた場合、

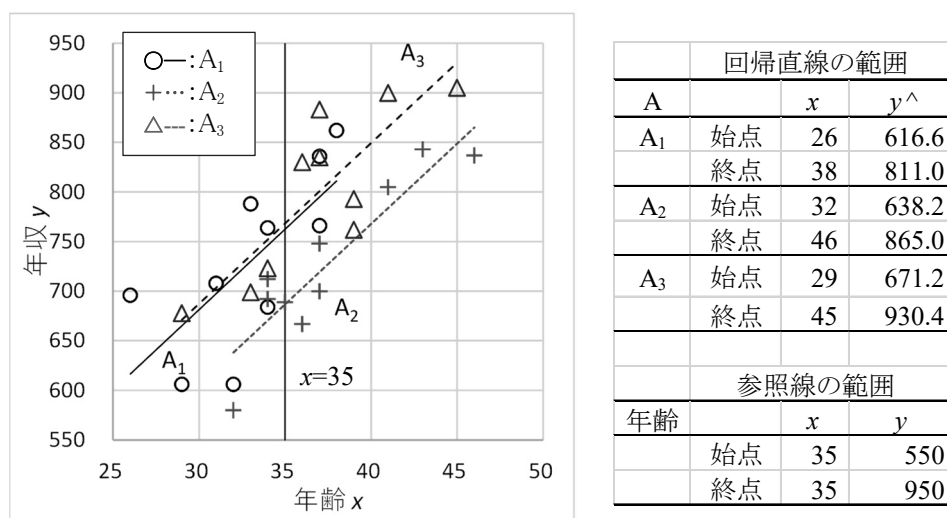


図 1.3 Excel の散布図による共通の傾きを持つ 3 本の直線

表 1.4 右の「分散分析表」の回帰の平方和（変動）は、 $S_{\text{回帰}, df=6} = 17,182,504.95$ で、共通の傾きに対する表 1.5 右の「分散分析表」の回帰の平方和は、 $S_{\text{回帰}, df=4} = 17,182,279.29$ であり、それらの差分を計算する。この回帰の平方和の差が、交互作用の平方和 $S_{A \times x} = 225.66$ となり、自由度も差分を取り $df = 6 - 4 = 2$ となる。

交互作用の平方和（変動） : $S_{A \times x} = 17,182,504.95 - 17,182,279.29 = 225.66$

交互作用の自由度 : $df_{A \times x} = 6 - 4 = 2$

交互作用の平均平方（分散） : $MS_{A \times x} = 225.66 / 2 = 112.83$

これらを、表 1.6 に示す分散分析表にまとめる。交互作用 $A \times x$ の平均平方（分散） $MS_{A \times x} = 112.83$ を自由度 24 の誤差分散の推定値（残差の平均平方） $\hat{\sigma}^2 = 2697.92$ で割り、検定統計量としての分散比（ F 値）を求める。 F 分布の右側確率のための $F.\text{dist.RT}()$ 関数で求めた p 値

$$\left. \begin{aligned} F_{A \times x} &= \frac{112.83}{2697.92} = 0.0418 \\ p_{A \times x} &= F.\text{dist.RT}(0.0418, 2, 24) = 0.9591 \end{aligned} \right\} \quad (1.7)$$

表 1.6 回帰の変動（平方和）の差分による交互作用の評価

	自由度	平方和(変動)	平均平方(分散)	分散比 F 値	p 値
回帰(別々の直線)	6	17182504.95	2863750.82	1061.47	
回帰(傾き共通)	4	17182279.29	4295569.82	1592.18	
交互作用 $A \times x$	2	225.66	112.83	0.0418	0.9591
残差	24	64750.05	2697.92		
合計	30	17247255.00			

から、交互作用 $A \times x$ は全くないと判定される。ただし、このような統計的方法によらずとも、図 1.1 から視覚的には交互作用がないことは明らかである。

表 1.6 に示した交互作用の評価は、残差の平方和に着目しても同じ結果が得られる。表 1.5 の自由度 26 の残差の平方和は、 $S_{\text{残差}, df=26} = 64,975.71$ であり、表 1.4 の自由度 24 の回帰の平方和は、 $S_{\text{残差}, df=24} = 64,750.05$ でありその差から、交互作用 $S_{A \times x, df=2} = 225.66$ として求めることもできる。

回帰直線上の推定値に対する 95%信頼区間

傾きが共通な平行な直線のあてはめが、“統計的”にも支持され、3 社の年齢の総平均は、 $\bar{x} = 35.9$ 歳であり、切のよい 35 歳とした場合の、3 社の年収の推定値と 95%信頼区間を推定し比較検討を行いたい。それぞれの会社の年収の推定値は、

$$\left. \begin{array}{l} A_1 \text{社: } \hat{y}_{x=35}^{(A_1)} = 195.2832 + 16.2029 \times 35 = 762.4 \\ A_2 \text{社: } \hat{y}_{x=35}^{(A_2)} = 119.6903 + 16.2029 \times 35 = 686.8 \\ A_3 \text{社: } \hat{y}_{x=35}^{(A_3)} = 201.2918 + 16.2029 \times 35 = 768.4 \end{array} \right\} \quad (1.8)$$

のように推定されたパラメータを用い、年齢 $x=35$ を式 (1.6) に代入することにより簡単に求まる。さて、これらの 95%信頼区間を求めるために、推定値の分散が必要となる。それぞれの推定されたパラメータの標準誤差 SE は、表 1.5 に示されているのであるが、年齢が $x=35$ における各社の年収 y の推定値 $\hat{y}_{x=35}^{(A_i)}$ の分散は、合成分散の一般式

$$\left. \begin{aligned} Var(\hat{y}_{x=35}^{(A_1)}) &= Var(\hat{\beta}_0^{(A_1)} + \hat{\beta}_1^{\text{共通}} x) \\ &= Var(\hat{\beta}_0^{(A_1)}) + 2Cov(\hat{\beta}_0^{(A_1)}, \hat{\beta}_1^{\text{共通}})x + Var(\hat{\beta}_1^{\text{共通}})x^2 \end{aligned} \right\} \quad (1.9)$$

によって求めることができる。

合成分散の一般式で用いられている分散 $Var(\hat{y}_{x=35}^{(A_1)})$ および $Var(\hat{\beta}_1^{\text{共通}})$ は、それぞれのパラメータの SE を 2 乗して求められるが、共分散 $Cov(\hat{\beta}_0^{(A_1)}, \hat{\beta}_1^{\text{共通}})$ は Excel の回帰分析からは得られない。そこで、線形モデルで定式化されている方法を用い、新たに計算することが必要となる。パラメータの共分散行列 $\Sigma(\hat{\beta})$ は、デザイン行列を X 、誤差分散の推定値 $\hat{\sigma}^2$ を表 1.5 の分散分析表の「残差の行の分散」 $\hat{\sigma}^2 = 2499.07$ としたときに、

$$\Sigma(\hat{\beta}) = (X^T X)^{-1} \hat{\sigma}^2 \quad (1.10)$$

によって求めることができる。

実際に、平行な 3 本の回帰直線のパラメータの共分散行列 $\Sigma(\hat{\beta})$ を求めてみよう。表 1.8 に示すように Excel の行列関数を用いて計算する。計算式中の X^T は、(30 行 \times 4 列) のデザ

イン行列 \mathbf{X} を転置したサイズが (4 行 \times 30 列) の行列である。行列の転置は、表 1.7 に示すように Excel のシート上でデザイン行列 \mathbf{X} の範囲を選択・コピーし、プルダウン・メニューから「形式を選択して貼り付け」を選択し、「行/列の入れ替え」でペーストして得られる。また、転置のための行列関数 Transpose (\mathbf{X} の範囲) によっても得られる。なお、行列計算に不慣れな場合は、第 5 章に Excel による行列計算の手順について丁寧に解説してあるので先に読んでもらいたい。

表 1.7 転置行列 \mathbf{X}^T および積和行列 $(\mathbf{X}^T \mathbf{X})$

転置行列 X^T						デザイン行列 X				積和行列 ($X^T X$)					
No	1	2	3		29	30	a_1	a_2	a_3	x					
a_1	1	1	1	...	0	0	1	0	0	34	=	10	0	0	331
a_2	0	0	0		0	0	1	0	0	33		0	10	0	375
a_3	0	0	0		1	1	1	0	0	34		0	0	10	370
x	34	33	34		45	39	:					331	375	370	39194
							0	0	1	45					
							0	0	1	39					
(4×30)						(30×4)				(4×4)					

表 1.8 に示すのは,

表 1.5 での Excel の回帰分析による回帰パラメータの推定に代え、デザイン行列 \mathbf{X} を用いた Excel の行列計算による回帰パラメータの推定結果である．積和行列 $(\mathbf{X}^T \mathbf{X})$ は、 \mathbf{X}^T の行方向の 30 個のデータに \mathbf{X} の列方向の 30 個のデータを順番に掛けて加えたもので、 (4×30) の転置行列 \mathbf{X}^T 、 (30×4) の \mathbf{X} の積の計算で $(\mathbf{X}^T \mathbf{X})$ は、 (4×4) の行列となり、Excel の行列関数 `Mmult (\mathbf{X}^T の範囲, \mathbf{X} の範囲)` により求められる．

積和行列 $(\mathbf{X}^T \mathbf{X})$ の逆行列 $(\mathbf{X}^T \mathbf{X})^{-1}$ は、スカラーの演算での逆数に対応し、Excel の行列関数 Minverse $(\mathbf{X}^T \mathbf{X}$ の範囲) で求める。パラメータの共分散行列 $\Sigma(\hat{\boldsymbol{\beta}})$ は、逆行列 $(\mathbf{X}^T \mathbf{X})^{-1}$ のそれぞれの要素に残差の行の分散（誤差分散の推定値） $\hat{\sigma}^2 = 2499.07$ を掛けて求める。

表 1.8 右下段の $(X^T Y)$ は、デザイン行列 X の転置行列に年収 Y の列ベクトルを掛けた結果が示されている。この $(X^T Y)$ に逆行列 $(X^T X)^{-1}$ を前から掛けることにより、表 1.5 に示した Excel の回帰分析の結果と同様に、回帰パラメータの推定値 $\hat{\beta}$ が求められている。また、推定されたパラメータの SE は、パラメータの共分散行列 $\Sigma(\hat{\beta})$ の対角要素の平方根で求められている。このように、Excel の行列関数を使うことにより、Excel の回帰分析と同様の結果が得られる。詳しくは、第 5 章を参照のこと。

表 1.9 に示すように、パラメータの共分散行列 $\Sigma(\hat{\beta})$ の対角要素は、各パラメータの分散となっており、これらの平方根は表 1.5 に示した Excel の回帰分析で計算したパラメータの標

表 1.8 平行な3本の回帰直線のパラメータの共分散行列 $\Sigma(\hat{\beta})$ の推定

— デザイン行列 X —						分散分析表「定数に0を使用」を on				
No	a_1	a_2	a_3	x	y		自由度	変動	分散	分散比
1	1	0	0	34	684	回帰	4	17182279	4295570	1718.87
2	1	0	0	33	788	残差	26	64976	2499.07	$=\sigma^2$
3	1	0	0	34	764	合計	30	17247255		
4	1	0	0	37	836					
5	1	0	0	29	606					
6	1	0	0	26	696	積和行列 ($X^T X$)				
7	1	0	0	37	766	a_1	10	0	0	331
8	1	0	0	38	862	a_2	0	10	0	375
9	1	0	0	32	606	a_3	0	0	10	370
10	1	0	0	31	708	x	331	375	370	39194
11	0	1	0	34	692	=Mmult (Transpose (X の範囲), X の範囲)				
12	0	1	0	34	712					
13	0	1	0	37	700	逆行列 ($X^T X$) ⁻¹				
14	0	1	0	43	843	a_1	2.3571	2.5572	2.5231	-0.0682
15	0	1	0	37	748	a_2	2.5572	2.9971	2.8585	-0.0773
16	0	1	0	32	580	a_3	2.5231	2.8585	2.9204	-0.0762
17	0	1	0	46	837	x	-0.0682	-0.0773	-0.0762	0.0021
18	0	1	0	36	667	=Minverse ($X^T X$ の範囲)				
19	0	1	0	41	805					
20	0	1	0	35	689	パラメータの共分散行列 $\Sigma(\hat{\beta}^{\wedge}) = (X^T X)^{-1} \sigma^2$				
21	0	0	1	34	723	β_0^{A1}	5890.62	6390.53	6305.33	-170.41
22	0	0	1	39	762	β_0^{A2}	6390.53	7489.94	7143.50	-193.07
23	0	0	1	37	883	β_0^{A3}	6305.33	7143.50	7298.16	-190.49
24	0	0	1	29	678	$\beta_1^{\text{共通}}$	-170.41	-193.07	-190.49	5.1485
25	0	0	1	33	699	$(X^T X)^{-1}$ の範囲 $\times \sigma^2$				
26	0	0	1	36	830	回帰パラメータ				
27	0	0	1	41	900	$(X^T Y)$		$\hat{\beta}^{\wedge} = (X^T X)^{-1} X^T Y$		SE
28	0	0	1	37	835		7316	$\beta_0^{A1} =$	195.2832	76.7504
29	0	0	1	45	905		7273	$\beta_0^{A2} =$	119.6903	84.5192
30	0	0	1	39	793		8008	$\beta_0^{A3} =$	201.2918	85.4293
							819058	$\beta_1^{\text{共通}} =$	16.2029	2.2690

準誤差 SE に一致する。ちなみに、共通の傾き $SE(\hat{\beta}_1) = 2.2690$ は、

$$SE(\hat{\beta}_1) = \sqrt{Var(\hat{\beta}_1)} = \sqrt{5.1485} = 2.2690$$

となり、表 1.5 の標準誤差に一致することが確認できる。

表 1.9 パラメータの共分散行列 $\Sigma(\hat{\beta})$

$Var(\beta_0^{A1})$				=	5890.62	6390.53	6305.33	-170.41
$Cov(\beta_0^{A1}, \beta_0^{A2})$	$Var(\beta_0^{A2})$				6390.53	7489.94	7143.50	-193.07
$Cov(\beta_0^{A1}, \beta_0^{A3})$	$Cov(\beta_0^{A2}, \beta_0^{A3})$	$Var(\beta_0^{A3})$			6305.33	7143.50	7298.16	-190.49
$Cov(\beta_0^{A1}, \beta_1^{\text{共通}})$	$Cov(\beta_0^{A2}, \beta_1^{\text{共通}})$	$Cov(\beta_0^{A3}, \beta_1^{\text{共通}})$	$Var(\beta_1^{\text{共通}})$		-170.41	-193.07	-190.49	5.1485
					$\Sigma(\hat{\beta}) = (X^T X)^{-1} \sigma^2$			

表 1.10 に年齢が 35 歳に場合の回帰の推定値 ($\hat{y}_{x=35}^{(A_1)}$, $\hat{y}_{x=35}^{(A_2)}$, $\hat{y}_{x=35}^{(A_3)}$), それらの分散 [$Var(\hat{y}_{x=35}^{(A_1)})$, $Var(\hat{y}_{x=35}^{(A_2)})$, $Var(\hat{y}_{x=35}^{(A_3)})$], および, 95%信頼区間の計算結果を示す。

表 1.10 回帰直線の 35 歳における推定値, 分散, SE, および 95%信頼区間

会社	推定値 $\hat{y}_{x=35}^{(A_i)}$	分散 $Var(\hat{y}_{x=35}^{(A_i)})$	標準誤差 SE	幅 $t_{0.05} \times SE$	95%信頼区間	
					$L_{95\%}$	$U_{95\%}$
A ₁	762.39	268.49	16.3857	33.6814	728.70	796.07
A ₂	686.79	282.08	16.7954	34.5234	652.27	721.32
A ₃	768.39	270.50	16.4469	33.8071	734.59	802.20
		T.Inv.2T (0.05,30-4)=		2.0555		

A₁ 社の 35 歳における分散は, 推定値 $\hat{y}_{x=35}^{(A_1)} = 762.29$ に対する合成分散の一般式 (1.9) を用い,

$$\begin{aligned}
 Var(\hat{y}_{x=35}^{(A_1)}) &= Var(\hat{\beta}_0^{(A_1)} + \hat{\beta}_1^{共通} x) \\
 &= Var(\hat{\beta}_0^{(A_1)}) + 2Cov(\hat{\beta}_0^{(A_1)}, \hat{\beta}_1^{共通})x + Var(\hat{\beta}_1^{共通})x^2 \\
 &= 5890.62 + 2 \times (-170.41) \times 35 + 5.1485 \times 35^2 \\
 &= 268.49
 \end{aligned} \quad (1.11)$$

として求められ, 95%信頼区間は, 一般的な方法により

$$\begin{aligned}
 95\%CL &= \hat{y}_{x=35}^{(A_1)} \pm t_{0.05}(30-4) \sqrt{Var(\hat{y}_{x=35}^{(A_1)})} \\
 &= 762.39 \pm 2.0555 \times 16.3857 \\
 &= (728.70, 796.07)
 \end{aligned} \quad (1.12)$$

として推定される。

会社間の推定値の差の 95%信頼区間

図 1.3 から, 年齢を揃えた場合には, A₂ 社の給料が他社に比べて低いことは, 散布図上の点が A₁ 社および A₃ 社よりも明らかに下方にあることから明らかである. 表 1.10 から, 35 歳での A₂ 社の給料の推定値は, 686.8 万円であり, 95%信頼区間の上限が 721.3 万円と A₁ 社の下限 728.70 万円および A₃ 社の下限 734.59 万円よりも低いので統計的にも明らかな差があると判断される。

統計的に厳密に 35 歳での A₂ 社の給与の推定値と A₁ 社の推定値を比較する方法を示す. A₁ 社と A₂ 社の推定値は, 次式で

$$\begin{aligned}
 A_1 \text{社}: \quad \hat{y}_{x=35}^{(A_1)} &= \hat{\beta}_0^{(A_1)} + \hat{\beta}_1^{共通} x \\
 A_2 \text{社}: \quad \hat{y}_{x=35}^{(A_2)} &= \hat{\beta}_0^{(A_2)} + \hat{\beta}_1^{共通} x
 \end{aligned} \quad (1.13)$$

で与えられている. この式で, A₁ 社から A₂ 社を引くと $\hat{\beta}_1 x_1$ 項は相殺され,

$$(A_1 \text{社}) - (A_2 \text{社}) : \left\{ \begin{aligned} \hat{y}_{x=35}^{(A_1)} - \hat{y}_{x=35}^{(A_2)} &= \hat{\beta}_0^{(A_1)} - \hat{\beta}_0^{(A_2)} \\ &= 762.39 - 686.79 = 75.59 \end{aligned} \right\} \quad (1.14)$$

となる. その分散 $Var(\hat{y}_{x=35}^{(A_1)} - \hat{y}_{x=35}^{(A_2)})$ は,

$$\left. \begin{aligned} Var(\hat{y}_{x=35}^{(A_1)} - \hat{y}_{x=35}^{(A_2)}) &= Var(\hat{\beta}_0^{(A_1)} - \hat{\beta}_0^{(A_2)}) \\ &= Var(\hat{\beta}_0^{(A_1)}) - 2Cov(\hat{\beta}_0^{(A_1)}, \hat{\beta}_0^{(A_2)}) + Var(\hat{\beta}_0^{(A_2)}) \\ &= 5890.62 - 2 \times 6390.53 + 7489.94 = 599.49 \end{aligned} \right\} \quad (1.15)$$

となり、表 1.11 に示すように、95%信頼区間を計算することができる。

表 1.11 回帰直線の差の推定値、分散、SE、および 95%信頼区間

	差の推定値	分散	標準誤差	幅	95%信頼区間	
会社	$\hat{y}_{35}^{\text{差}}$	$Var(\hat{y}_{35}^{\text{差}})$	SE	$t_{0.05} \times SE$	L 95%	U 95%
A ₁ -A ₂	75.59	599.49	24.4844	50.3285	25.26	125.92
A ₂ -A ₂	0.00	0.00	0.0000	0.0000	0.00	0.00
A ₃ -A ₂	81.60	501.10	22.3853	46.0136	35.59	127.62
		T.Inv.2T (0.05,30-4)=		2.0555		

表 1.10 および表 1.11 で計算された推定値と 95%信頼区間を Excel の折れ線グラフで作図した結果を図 1.4 に示す。A₁社は、A₂社および A₃社に比べて平均年収が低かったが、同一年齢で比べれば、A₂社よりも年収は高く、A₃社と同水準の年収であることが明らかになった。

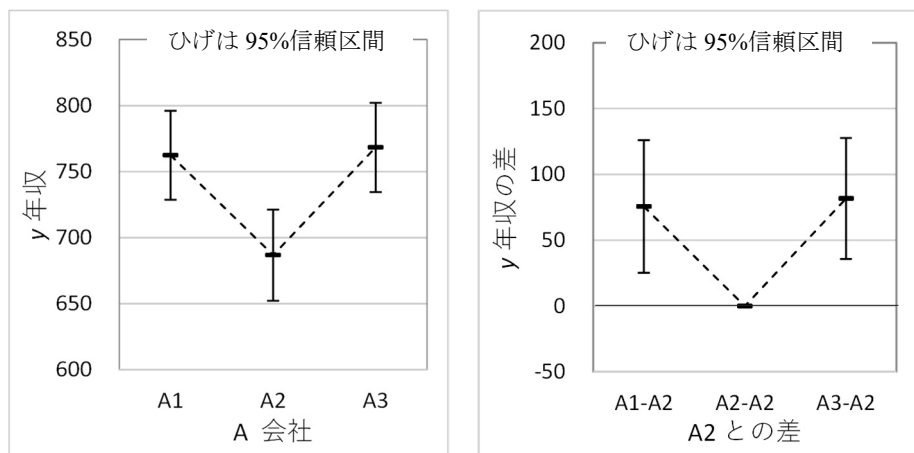


図 1.4 Excel の折れ線グラフによる年齢が 35 歳における年収の比較

デザイン行列とは何か

デザイン行列 (design matrix) は、計画行列とも称されるが、本書では、デザイン行列で統一している。デザイン行列は、変幻自在で説明に窮するが、最小 2 乗法などによるパラメータの推定のための用いられる行列で、通常 \mathbf{X} と表記する。単回帰分析であれば、表 1.2 に即して言えば切片 β_0 を推定するための変数 x_0 、傾き β_1 を推定するための変数 x_1 の 10 行分のデータの (10 行×2 列) の矩形データがデザイン行列 \mathbf{X} となる。また、表 1.4 では、(30 行×6 列)、表 1.5 では、(30 行×4 列) のデザイン行列 \mathbf{X} となり、推定の目的に合わせて行列のサイズが変わる。なお、第 5 章の 5.2 節の末尾に、design matrix について英語版の Wikipedia を引用した解説を示してあるので参考にされたい。

1.2. 季節による洗浄用水の温度と回収液の濃度の関連

奥野・久米・芳賀・吉澤(1981),「多変量解析法《改訂版》」は,多くの人達に読み継がれてきた名著である. その中の「第 II 部 重回帰分析 第 7 章 層別因子を含む回帰分析」の[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. 原著では, 第 6 章で示すように伝統的な層ごとの単回帰分析を積み上げる解析方法が示されている. この方法は, PC による回帰分析が普及する以前の手計算時代の一般的な解法であり, 技巧的で見通しが悪い. そこで, ダミー変数を用いた回帰分析による解析方法を用いる. なお, この事例については, 高橋 (2021),「最尤法によるポアソン回帰分析入門, 第 13.2 節 交互作用を考慮した共分散分析」でも取り上げている.

季節による層別

表 1.12 に示すように, 亜硫酸ガス回収塔の洗浄用水の温度 x と, 回収液の濃度 y との関係を調べたい. 長期間にわたって取られたデータなので, 季節の変化だけでなく, 操業条件も変わっていると思われる. そこで, 表 1.12 に示すようにデータを 4 つ季節に分けている.

表 1.12 季節ごとの洗浄用水の温度 x と回収液の濃度 y [奥野ら (1981), 表 7.2]

	A ₁ (7月・8月)		A ₂ (9月・10月)		A ₃ (11月・12月)		A ₄ (1月・2月)	
No.	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	26.64	11.80	24.83	14.31	18.56	19.47	14.40	32.60
		データ数	11+12+9+10=42			総平均	21.48	19.11

JMP の「グラフ・ビルダー」による層別散布図

図 1.5 に示すように, 洗浄用水の温度 x は, 層別因子として取り上げた季節 A の影響を受ける. 表 1.12 に示したように, A_1 (7 月・8 月) で $\bar{x}_1 = 26.64$ 度であったのが, A_2 (9 月・10 月) に $\bar{x}_2 = 24.83$ 度と減少し, さらに, A_4 (1 月・2 月) に $\bar{x}_4 = 14.40$ 度と減少している. 回

収液の濃度 y の平均値は、逆に (11.80, 14.31, 19.47, 32.60) と上昇している。回収液の濃度 y の上昇は、季節 A 以外の他の因子の影響があるかも知れない。

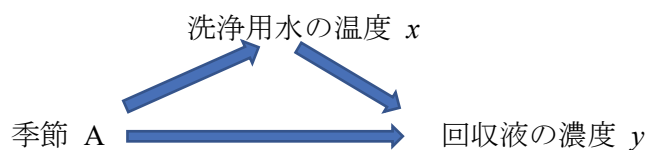


図 1.5 回収液の濃度に影響をおよぼす要因

得られたデータの関連を概観するために、JMP のグラフ・ビルダーを用いて洗浄水の温度 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 1.6 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。そのために、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ではないかと推測される。

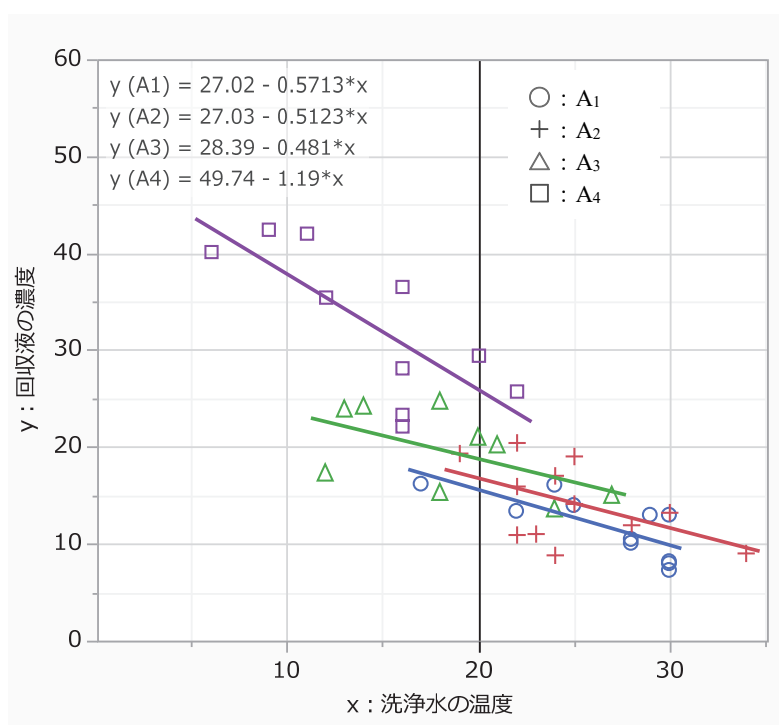


図 1.6 JMP のグラフ・ビルダーによる 4 本の回帰直線のあてはめ

JMP の「モデルのあてはめ」による解析

JMP の「モデルのあてはめ」を用いた回帰分析を用い、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、表 1.13 に示すような分散分析表および効果の検定が得られる。

表 1.13 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	246.2924	82.0975	5.4834	0.0035*
x	1	369.0901	369.0901	24.6520	<.0001*
A*x	3	71.2462	23.7487	1.5862	0.2107

「▼ モデルの設定」の「多項式の中心化」をオフとした結果である。

交互作用 $A \times x$ が $p = 0.2107$ と有意ではないので原著では、解析モデルから除いて季節 A に共通の傾きを持つ回帰直線をあてはめる根拠としている。だが、季節 A₄ (1 月, 2 月) の洗浄水の温度の範囲は、6~22 度と他の季節に比べて著しく低く、低くなるほど回収液の濃度は高くなり、他の季節とは異なる別の要因が影響していると推測される。そこで、図 1.6 に示した結果から各季節に共通している洗浄水の温度が 20 度の場合に、各季節の回帰直線の推定値とその 95%信頼区間を求めて比較検討を行う。

JMP の「予測プロファイル」

JMP の「予測プロファイル」は、解析結果のグラフ表示に優れている。図 1.7 に示すように、質的変数に対しては、他の量的変数の設定値に連動し、水準ごとの推定値および 95%信頼区間が表示され、量的変数の場合には、他の質的変数で選択された水準に連動した回帰直線および 95%信頼区間が動的に表示される。さらに、それらを組み合わせた場合の推定値および 95%信頼区間の推定結果も示される。質的変数の水準の選択、量的変数の設定も GUI (Graphical User Interface) により自在に変更することができる。JMP の「予測プロファイル」は、他の統計ソフトにはない動的なグラフ表示が特徴である。

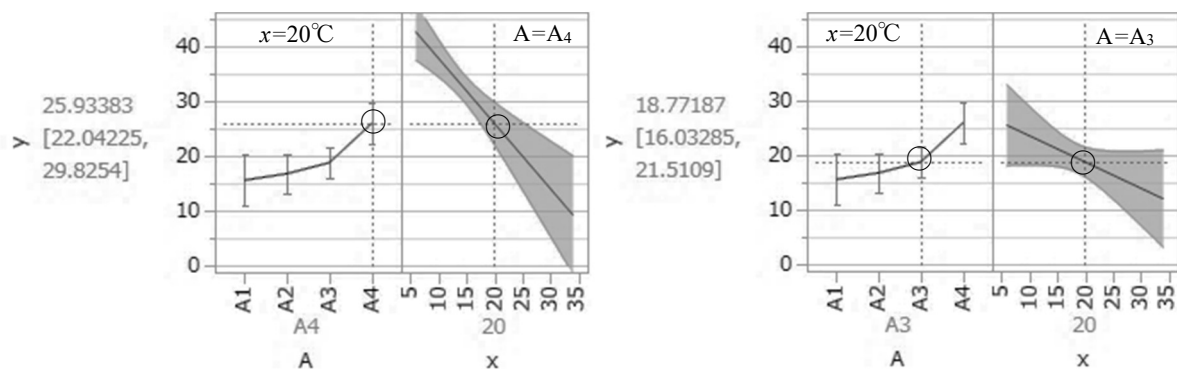
図 1.7 A₄および A₃を選択した場合の予測プロファイル

図 1.7 左は、予測プロファイル上で因子 A の A₄ 水準を選択し、洗浄水の温度 x を 20 度と設定した結果であり、それらを組み合わせた推定値として 25.93, および、その 95%信頼区間 [22.04, 29.83] が図 1.7 左端に表示され、他の (A₁, A₂, A₃) 水準に対しても、洗浄水の温度を $x=20$ とした場合の各水準の推定値とその 95%信頼区間が図示されている。この予測プロファイルから A₄ 水準と A₃ 水準の 95%信頼区間のひげが重なりあっていないので、統計的に有意な差があると判断される。

図 1.7 右は、GUI により A₃ 水準を選択し、洗浄水の温度 x を 20 度はそのままにした結果であり、A₃ 水準の推定値 18.77 と 95%信頼区間 [16.03, 21.51] が表示され、A₃ 水準に対応する洗浄水の温度 x に関する回帰直線と 95%信頼区間が示されている。回帰直線の傾きは、図 1.6 から示唆されるように A₄ 水準に比べて平坦化し、95%信頼区間の上限が洗浄水の温度 x の上昇に伴い上方に発散気味であり傾きが有意な差とは言えないことが読み取れる。

洗浄水の温度 x を 20 度と設定しているが、デフォルトでは、洗浄水の温度 x の総平均の $\bar{x} = 21.48$ 度が設定される。ただし、その温度は自由に変更することができ、図 1.8 に示すように温度を 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 A 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。洗浄水の温度が 10 度の場合には、A₄ 水準以外は外挿となり 95%信頼区間の髭の差が広がってはいるが、有意な差となっている。図 1.8 右に示すように洗浄水の温度が 30 度の場合には、A₄ 水準は外挿となることが図 1.6 から明らかであり、回収液の濃度 y の推定値に関し季節 A 間に有意な差は認められない。

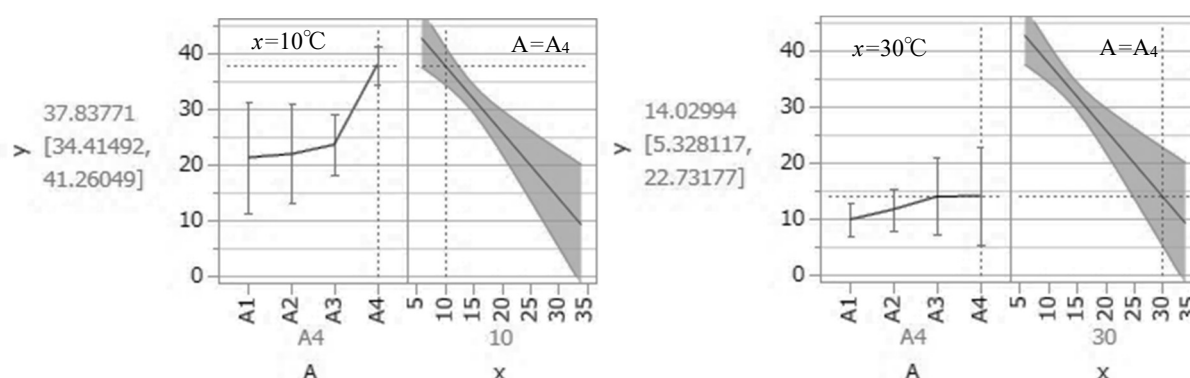


図 1.8 洗浄水の温度を変えた場合の季節 A の各水準の推定値と 95%信頼区間

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 1.14 に示すように、洗浄水の温度を 20 度に固定し、季節 A₃ を +1, 季節 A₄ を -1 とする対比の指定により、図 1.7 に示した A₃ の推定値の 18.7719 と A₄ の推定値 25.9338 との

表 1.14 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		対比		パラメータ関数	
対比の指定		検定の詳細		パラメータ	
A A1 0 A2 0 A3 1 A4 -1 +または-をクリックして対比值を作成。		A1 0 A2 0 A3 1 A4 -1 推定値 -7.16195 標準誤差 2.34167 t値 -3.05848 p値(Prob> t) 0.004317 平方和 140.0529 下側95% -11.9208 上側95% -2.40311		切片 0 A[A1] 1 A[A2] 1 A[A3] 2 x 0 x*A[A1] 20 x*A[A2] 20 x*A[A3] 40	
温度を 20°C に設定		平方和 140.1 分子自由度 1 分母自由度 34 F値 9.3543 p値(Prob>F) 0.0043*			
x 20					

差が -7.1620, 標準誤差が 2.3417 と推定されている. 95%信頼区間の出力が, (-11.9208, -2.4031) と 0 を含まないので有意な差であることが分かる.

JMP での対比の設定は, 直観的にわかりやすい ($+A_3 - A_4$) となっているが, JMP の内部での設定は, 表 1.14 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 (1, -1) 対比型のダミー変数をデフォルトにしていることに関係している. 表 1.15 に示す対比型ダミー変数から明らかのように, ($+A_3 - A_4$) は, パラメータ関数で示されているように, 内部では ($+a_1 + a_2 + 2a_3$) と置き換えられている.

表 1.15 (1, -1) 対比型ダミー変数における A₃ と A₄ の差

A	x_0	a_1	a_2	a_3	x	a_1x	a_2x	a_3x
A ₁	1	1	0	0	20	20	0	0
A ₂	1	0	1	0	20	0	20	0
A ₃	1	0	0	1	20	0	0	20
A ₄	1	-1	-1	-1	20	-20	-20	-20
A ₃ -A ₄	0	1	1	2	0	20	20	40

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

x	0
A[A1]*x	20
A[A2]*x	20
A[A3]*x	40

Excel による (1, -1) 対比型ダミー変数を用いた解析

JMP の「モデルのあてはめ」により, 質的変数としての季節 A, 量的変数としての洗浄水の温度 x , それらの交互作用 $A \times x$ を含めた解析モデル

$$\text{解析モデル: } y = A + x + (A \times x) \quad (1.16)$$

を適用した. JMP の内部では, 質的変数について, (1, -1) 対比型ダミー変数に変換し, 線

形モデルとした解析が行われている。さらに、「予測プロファイル」による推定値の出力のグラフ表示（予測プロファイル）は、結果の解釈に有用であり、Excel で再現することにより理解を深めたい。

質的変数を線形モデルで扱うためには、量的変数のダミー変数に置き換える必要がある。表 1.3 に示したようにダミー変数は、幾つかの型に分類でき、第 1.1 節では (1, 1) 標示型ダミー変数を用い、3 本の回帰直線の切片と傾きを直接推定した。JMP の内部では、表 1.16 に示すように (1, -1) 対比型ダミー変数が使われている。季節 A は 4 水準なので、ダミー変数名として (a_1, a_2, a_3) を与え、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 1.16 (1, -1) 対比型ダミー変数における交互作用

A	a_1	a_2	a_3	A	a_1x	a_2x	a_3x
A ₁	1	0	0	A ₁	x	0	0
A ₂	0	1	0	A ₂	0	x	0
A ₃	0	0	1	A ₃	0	0	x
A ₄	-1	-1	-1	A ₄	$-x$	$-x$	$-x$

式 (1.16) に示した交互作用を含む解析モデルに対しダミー変数を用いた式は、

$$y_i = \theta_0 x_{0,i} + \theta_1 a_{1,i} + \theta_2 a_{2,i} + \theta_3 a_{3,i} + \theta_4 x_i + \theta_5 a_{1,i} x_i + \theta_6 a_{2,i} x_i + \theta_7 a_{3,i} x_i + \varepsilon_i \quad (1.17)$$

となる。データ数は、 $n=11+12+9+10=42$ で、切片 x_0 を含め変数の数は 8 であり、Excel の回帰分析のためのデザイン行列 \mathbf{X} は、表 1.17 に示すように (42 行×8 列) と大きくなるが、回帰分析の結果は瞬時に得られる。

表 1.17 交互作用モデルに対する回帰分析

		——— デザイン行列 X ———														
No.	A	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」 on とする				
1	A ₁	1	1	0	0	30	30	0	0	8.0			自由度	変動	分散	分散比
2		1	1	0	0	28	28	0	0	10.5		回帰	8	18486.90	2310.86	154.35
:												残差	34	509.05	14.9720	= σ^2
11		1	1	0	0	25	25	0	0	14.0		合計	42	18995.95		
12	A ₂	1	0	1	0	34	0	34	0	9.1						
13		1	0	1	0	19	0	19	0	19.4			係数	SE	t	P -値
:												切片	0	#N/A	#N/A	#N/A
23		1	0	1	0	24	0	24	0	8.9	θ^{\wedge}_0	x_0	33.0450	3.1137	10.6127	0.0000
24	A ₃	1	0	0	1	27	0	0	27	15.0	θ^{\wedge}_1	a_1	-6.0281	6.3343	-0.9517	0.3480
25		1	0	0	1	20	0	0	20	21.0	θ^{\wedge}_2	a_2	-6.0157	5.9544	-1.0103	0.3195
:											θ^{\wedge}_3	a_3	-4.6528	4.8112	-0.9671	0.3403
32		1	0	0	1	24	0	0	24	13.6	θ^{\wedge}_4	x	-0.6887	0.1387	-4.9651	0.0000
33	A ₄	1	-1	-1	-1	16	-16	-16	-16	23.4	θ^{\wedge}_5	a_1x	0.1175	0.2473	0.4749	0.6379
34		1	-1	-1	-1	16	-16	-16	-16	28.2	θ^{\wedge}_6	a_2x	0.1765	0.2450	0.7205	0.4762
:											θ^{\wedge}_7	a_3x	0.2077	0.2364	0.8786	0.3858
42		1	-1	-1	-1	22	-22	-22	-22	25.8						
総平均						21.4762				19.1119						

Excel の回帰分析により，分散分析表およびパラメータの推定結果が得られる．ただし，第 1.1 節でも示したように，パラメータ間の共分散が得られないために，Excel の行列関数を用いてパラメータの共分散行列 $\Sigma(\hat{\theta})$ を式 (1.10) と同様に

$$\Sigma(\hat{\theta}) = (X^T X)^{-1} \hat{\sigma}^2 \quad (1.18)$$

により計算する必要がある．行列のサイズは異なるが，第 1.1 節の表 1.8 で示したと全く同じ計算手順である．

パラメータの共分散行列 $\Sigma(\hat{\theta})$ は，式 (1.18) に示したようにデザイン行列と分散分析表の残差の分散（誤差分散の推定値） $\hat{\sigma}^2 = 14.972$ を用いて

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

= Minverse (Mmult (Transpose (Xの範囲), Xの範囲)) * 14.972

	x_0	a_1	a_2	a_3	x	$a_1 x$	$a_2 x$	$a_3 x$
θ^{\wedge}_0	9.6953	5.5188	3.1845	-2.9689	-0.4134	-0.1450	-0.0927	0.0733
θ^{\wedge}_1	5.5188	40.1234	-18.3985	-12.2452	-0.1450	-1.5302	0.6511	0.4851
θ^{\wedge}_2	3.1845	-18.3985	35.4548	-9.9109	-0.0927	0.6511	-1.4256	0.4328
θ^{\wedge}_3	-2.9689	-12.2452	-9.9109	23.1481	0.0733	0.4851	0.4328	-1.0936
θ^{\wedge}_4	-0.4134	-0.1450	-0.0927	0.0733	0.0192	0.0017	0.0011	-0.0009
θ^{\wedge}_5	-0.1450	-1.5302	0.6511	0.4851	0.0017	0.0612	-0.0221	-0.0200
θ^{\wedge}_6	-0.0927	0.6511	-1.4256	0.4328	0.0011	-0.0221	0.0600	-0.0195
θ^{\wedge}_7	0.0733	0.4851	0.4328	-1.0936	-0.0009	-0.0200	-0.0195	0.0559

のように計算することができる．表 1.17 は，表示スペースの制約で少数点以下 2 桁あるいは 4 桁で表示しているが，Excel のセル内では，倍精度実数（有効数字で 15 桁）となっている．なお，これら行列計算の詳細は，第 5 章を参照のこと．

(1, -1)対比型ダミー変数における回帰直線の推定式

前節では，質的変数について (1, 1) 標示型ダミー変数を用いたので，推定されたパラメータが切片および傾きの推定値となっていたが，本節では (1, -1) 対比型ダミー変数を用いているので，季節 A_i の回帰直線のパラメータ推定には，次に示すような計算が必要となる．

$$\left. \begin{aligned} y^{(A_1)} &= (\hat{\theta}_0 + \hat{\theta}_1) + (\hat{\theta}_4 + \hat{\theta}_5)x \\ &= 33.04 - 6.03 + (-0.69 + 0.12)x \\ &= 27.02 - 0.57x \end{aligned} \right\} \quad (1.19)$$

$$\left. \begin{aligned} y^{(A_2)} &= (\hat{\theta}_0 + \hat{\theta}_2) + (\hat{\theta}_4 + \hat{\theta}_6)x \\ &= 27.03 - 0.51x \end{aligned} \right\} \quad (1.20)$$

$$\left. \begin{aligned} y^{(A_3)} &= (\hat{\theta}_0 + \hat{\theta}_3) + (\hat{\theta}_4 + \hat{\theta}_7)x \\ &= 28.39 - 0.48x \end{aligned} \right\} \quad (1.21)$$

$$\left. \begin{aligned} y^{(A_4)} &= [\hat{\theta}_0 - (\hat{\theta}_1 + \hat{\theta}_2 + \hat{\theta}_3)] + [\hat{\theta}_4 - (\hat{\theta}_5 + \hat{\theta}_6 + \hat{\theta}_7)]x \\ &= 33.04 - (-16.70) + (-0.69 - 0.50)x \\ &= 49.74 - 1.19x \end{aligned} \right\} \quad (1.22)$$

これらの推定値は、図 1.6 に示されている回帰式に一致し、Excel による推定方法が適切であることが確認できる。

Excel による予測プロファイル

図 1.7 に示した JMP の予測プロファイルで推定した洗浄水の温度 $x=20$ とした場合の A_3 の推定を Excel の行列計算で行う。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 $\hat{y}_{x=20}^{(A_3)}$ と分散 $Var(\hat{y}_{x=20}^{(A_3)})$ を推定したい。式 (1.21) で $x=20$ とすると、

$$\hat{y}_{x=20}^{(A_3)} = (\hat{\theta}_0 + \hat{\theta}_3) \times 1 + (\hat{\theta}_4 + \hat{\theta}_7)x = 28.39 - 0.48 \times 20 = 18.77 \quad (1.23)$$

が得られ、図 1.7 右に示されている推定値に一致する。

式 (1.23) を拡張できるように一般化すると、表 1.18 に示すように、 A_3 の推定値 $\hat{y}_{x=20}^{(A_3)}$ を算出するための行ベクトルを \mathbf{l} としたときに、推定された回帰パラメータ $\hat{\boldsymbol{\theta}}$ との積についての線形和 $L_{x=20}^{(A_3)} = \mathbf{l}\hat{\boldsymbol{\theta}}$ として求めることができる。

表 1.18 季節 A_3 の推定値に対する線形和 L による計算

		$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	$\hat{\theta}_5$	$\hat{\theta}_6$	$\hat{\theta}_7$	
	$\boldsymbol{\theta}^T$	33.0450	-6.0281	-6.0157	-4.6528	-0.6887	0.1175	0.1765	0.2077	
		l_0	l_1	l_2	l_3	l_4	l_5	l_6	l_7	
A_3	\mathbf{l}	1	0	0	1	20	0	0	20	線形和
										L
積和	$L = \mathbf{l}\hat{\boldsymbol{\theta}}$	33.0450	0.0000	0.0000	-4.6528	-13.7747	0.0000	0.0000	4.1544	= 18.7719

線形和 $L_{x=20}^{(A_3)}$ の分散は、式 (1.9) で示した 2 変数の合成分散の一般式を 8 変数に拡張した 2 次形式、

$$Var(L_{x=20}^{(A_3)}) = \mathbf{l} \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) \mathbf{l}^T \quad (1.24)$$

					\mathbf{l}					パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$														\mathbf{l}^T		
=	1	0	0	1	20		0	0	20		9.70	5.52	3.18	-2.97	-0.41	-0.14	-0.09	0.07	1	=	1.8165					
											5.52	40.12	-18.40	-12.25	-0.14	-1.53	0.65	0.49	0							
											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	1							
											-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	0							
											-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	20							

によって $Var(\hat{y}_{x=20}^{(A_3)})=1.8165$ として求めることができる. その 95%信頼区間は,

$$\left. \begin{aligned} (L95\%, 95\%) &= \hat{y}_{x=20}^{(A_3)} \pm t_{0.05}(42-8)\sqrt{Var(\hat{y}_{x=20}^{(A_3)})} \\ &= 18.7719 \pm 2.0322\sqrt{1.8165} \\ &= (16.0328, 21.5109) \end{aligned} \right\} \quad (1.25)$$

が得られ, 図 1.7 右に示されている推定値に一致する. 同様に他の季節 A の水準に対して計算した結果を表 1.19 に示す. 季節 A₄ の推定値は, $L_{x=20}^{(A_4)} = 25.9338$, 95%信頼区間は, (22.0423, 29.8254) と図 1.7 左の推定値に一致する.

表 1.19 交互作用を考慮した季節別の推定値と 95%信頼区間

		l_0	l_1	l_2	l_3	l_4	l_5	l_6	l_7	$l\theta^{\wedge}$	分散	幅	95%信頼区間	
L	A	x_0	a_1	a_2	a_3	x	a_1x	a_2x	a_3x	y^{\wedge}	$Var(y^{\wedge})$	$t_{0.05} \times SE$	L95%	U95%
L_1	A ₁	1	1	0	0	20	20	0	0	15.5912	5.0542	4.5688	11.0224	20.1600
L_2	A ₂	1	0	1	0	20	0	20	0	16.7842	3.1520	3.6080	13.1762	20.3922
L_3	A ₃	1	0	0	1	20	0	0	20	18.7719	1.8165	2.7390	16.0328	21.5109
L_4	A ₄	1	-1	-1	-1	20	-20	-20	-20	25.9338	3.6669	3.8916	22.0423	29.8254
										$t_{0.05}(0.05, 42-8)=$	2.0322			

この結果を Excel の「折れ線」グラフで作成した結果を図 1.9 左に示す. 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 $t_{0.05}(df) \times SE$ により, ひげの長さを計算する必要がある.

表 1.20 に示すのは, 洗浄水の温度 x を変化させた場合の回帰直線推定値および 95%信頼区間の計算結果である. 図 1.7 では別々に示されている回帰直線の推定値と 95%信頼区間の図であるが, 図 1.9 右に示すように, Excel の散布図により A₃ と A₄ についてこれらをまとめて重ね書きした結果を示す.

表 1.20 季節 A₃ と A₄ に対する回帰直線の 95%信頼区間の計算

		l_0	l_1	l_2	l_3	l_4	l_5	l_6	l_7	$l\theta^{\wedge}$	分散	幅	95%信頼区間	
L	A	x_0	a_1	a_2	a_3	x	a_1x	a_2x	a_3x	y^{\wedge}	$Var(y^{\wedge})$	$t_{0.05} \times SE$	L95%	U95%
L_5	A ₃	1	0	0	1	10	0	0	10	23.5820	7.0298	5.3883	18.1937	28.9703
L_6	A ₃	1	0	0	1	15	0	0	15	21.1769	2.5904	3.2708	17.9061	24.4478
L_7	A ₃	1	0	0	1	20	0	0	20	18.7719	1.8165	2.7390	16.0328	21.5109
L_8	A ₃	1	0	0	1	25	0	0	25	16.3668	4.7083	4.4097	11.9571	20.7765
L_9	A ₃	1	0	0	1	30	0	0	30	13.9618	11.2657	6.8211	7.1407	20.7829
L_{10}	A ₄	1	-1	-1	-1	5	-5	-5	-5	43.7896	7.6105	5.6064	38.1833	49.3960
L_{11}	A ₄	1	-1	-1	-1	10	-10	-10	-10	37.8377	2.8367	3.4228	34.4149	41.2605
L_{12}	A ₄	1	-1	-1	-1	15	-15	-15	-15	31.8858	1.5221	2.5073	29.3785	34.3930
L_{13}	A ₄	1	-1	-1	-1	20	-20	-20	-20	25.9338	3.6669	3.8916	22.0423	29.8254
L_{14}	A ₄	1	-1	-1	-1	25	-25	-25	-25	19.9819	9.2710	6.1879	13.7940	26.1697
										$t_{0.05}(42-8)=$	2.0322			

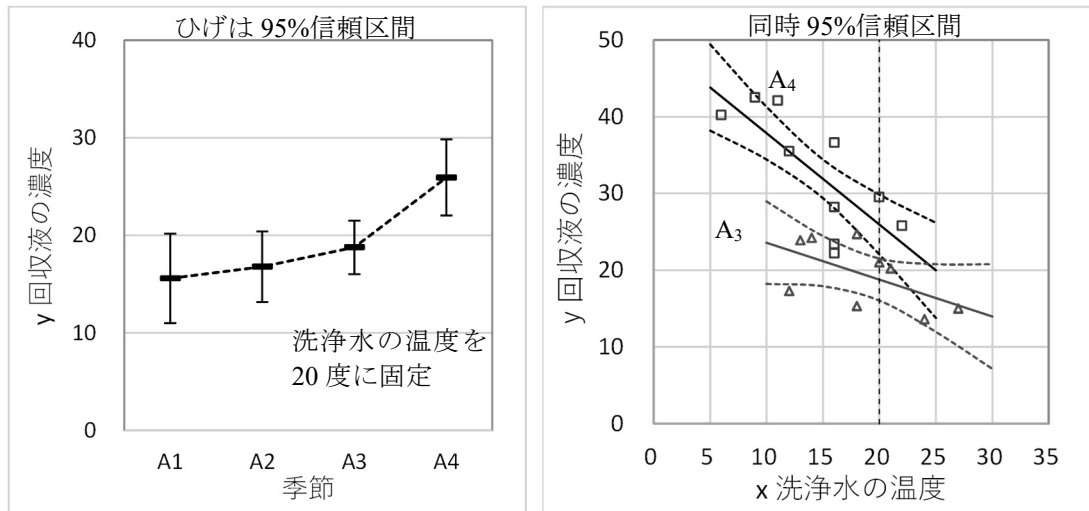


図 1.9 Excel の「折れ線」による予測プロファイルおよび散布図上の回帰直線

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）

のように慎重に設定する。気を抜くと尤もらしいが、い物の誤差範囲となってしまうので細心の注意が必要である。

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 1.21 に示すように、線形和のための行ベクトル \mathbf{l} を用いて差を計算する。推定値は、

$$\begin{aligned} L_{x=20}^{(A_3-A_4)} &= L_{x=20}^{(A_3)} - L_{x=20}^{(A_4)} \\ &= 18.7719 - 25.9338 = -7.1620 \end{aligned}$$

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 $\mathbf{l}^{(A_3)}$ と季節 A_4 のベクトル $\mathbf{l}^{(A_4)}$ の差

$$\begin{array}{rcl} \mathbf{l}^{(A_3)} & = & [1 \quad 0 \quad 0 \quad 1 \quad 20 \quad 0 \quad 0 \quad 20] \\ -) \quad \mathbf{l}^{(A_4)} & = & [1 \quad -1 \quad -1 \quad -1 \quad 20 \quad -20 \quad -20 \quad -20] \\ \hline \mathbf{l}^{(A_3-A_4)} = \mathbf{l}^{(A_3)} - \mathbf{l}^{(A_4)} & = & [0 \quad 1 \quad 1 \quad 2 \quad 0 \quad 20 \quad 20 \quad 40] \end{array}$$

であり、 $L_{x=20}^{(A_3-A_4)}$ の分散 $\text{Var}[L_{x=20}^{(A_3-A_4)}]$ は、

$$\begin{aligned} \text{Var}[L_{x=20}^{(A_3-A_4)}] &= \mathbf{l}^{(A_3-A_4)} \boldsymbol{\Sigma}(\hat{\theta}) (\mathbf{l}^{(A_3-A_4)})^T \\ &= 5.4834 \end{aligned}$$

として求められ. この結果は, 表 1.14 に示した「パラメータ関数」に示された結果に一致する. 差 $\bar{A}_3 - \bar{A}_4 = -7.1620$ 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 1.21 2 水準間の差の線形和 L と推定値

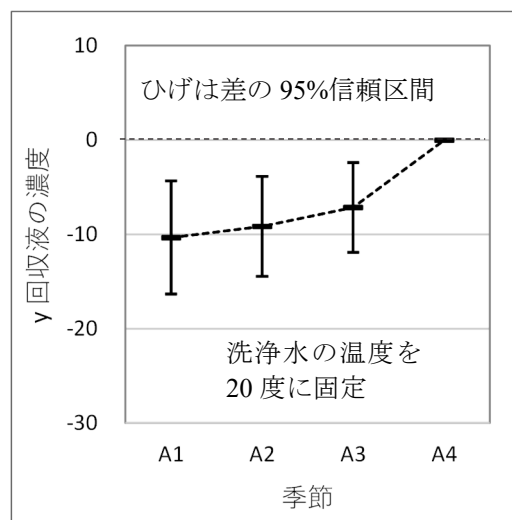
		l_0	l_1	l_2	l_3	l_4	l_5	l_6	l_7	$l\hat{\theta}$	分散	幅	95%信頼区間	
L	A	x_0	a_1	a_2	a_3	x	a_1x	a_2x	a_3x	y^\wedge	$Var(y^\wedge)$	$t_{0.05} \times SE$	$L_{95\%}$	$U_{95\%}$
L_3	A_3	1	0	0	1	20	0	0	20	18.7719				
L_4	A_4	1	-1	-1	-1	20	-20	-20	-20	25.9338				
差	$A_3 - A_4$	0	1	1	2	0	20	20	40	-7.1620	5.4834	4.7588	-11.9208	-2.4031

同様に季節 A_4 を基準とし, 他の季節の差を表 1.22 に示す. 季節感の差の計算のための行ベクトル l 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行ベクトル l の場合と同様に推定し, 95%信頼区間を計算する. JMP にも表 1.14 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 1.22 季節 A_4 を基準とした 2 水準間の差の推定

		l_0	l_1	l_2	l_3	l_4	l_5	l_6	l_7	$l\hat{\theta}$	分散	幅	95%信頼区間	
L	A	x_0	a_1	a_2	a_3	x	a_1x	a_2x	a_3x	y^\wedge	$Var(y^\wedge)$	$t_{0.05} \times SE$	$L_{95\%}$	$U_{95\%}$
L_{15}	$A_1 - A_4$	0	2	1	1	0	40	20	20	-10.3426	8.7211	6.0015	-16.3441	-4.3411
L_{16}	$A_2 - A_4$	0	1	2	1	0	20	40	20	-9.1496	6.8189	5.3068	-14.4564	-3.8428
L_{17}	$A_3 - A_4$	0	1	1	2	0	20	20	40	-7.1620	5.4834	4.7588	-11.9208	-2.4031
L_{18}	$A_4 - A_4$	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0000

図 1.10 に季節 A_4 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ァイルを示す. この図は, 洗浄水の温度を 20 度に固定した場合であり, 図 1.8 に示したように温度が低くなれば季節 A_4 との差は広がり, 温度が高くなれば差は縮まる.

図 1.10 Excel の「折れ線」による季節 A_4 を基準とした差の予測プロファイル

1.3. 前値を共変量とする交互作用がある抗うつ剤の評価

フライス著, KR(関西臨床データ解析)研究会訳 (2004), 「臨床試験のデザインと解析, 第7章 共分散分析と変化量の検討」にランダム化ブロック試験の30組の患者のうち, 表1.23に示すように, 初めの10組について治療前のうつ症状のHamiltonの評価尺度によるスコア X , 治療後のスコア Y が示されている. この試験データは, 薬剤Aと前値 X に交互作用があるような場合について“共分散分析”による解析の考え方が丁寧に示されている. なお, Hamiltonの評価尺度スコアは, 大きいほどうつ症状は重くなる. ここでは, 簡単のためペアを考慮した解析ではなく, 前値 X を共変量, 薬剤をA, 前後差(変化量) D を反応変数とする解析を行う.

表1.23 抗うつ薬投与前後のHamilton評価尺度データ [フライス (2004), 表7.6]

No	A ₁ : プラセボ				A ₂ : イミプラミン		
	前値 x	後値 y	前後差 d		前値 x	後値 y	前後差 d
1	18	4	-14		27	6	-21
2	21	7	-14		18	4	-14
3	24	10	-14		25	6	-19
4	24	10	-14		23	7	-16
5	18	2	-16		22	5	-17
6	27	11	-16		24	6	-18
7	21	9	-12		25	8	-17
8	21	5	-16		22	7	-15
9	26	11	-15		26	8	-18
10	20	8	-12		19	3	-16
平均	22.00	7.70	-14.30		23.10	6.00	-17.10
SE	0.99	0.99	0.47		0.92	0.52	0.64
L 95%	19.76	5.46	-15.37		21.01	4.83	-18.55
U 95%	24.24	9.94	-13.23		25.19	7.17	-15.65

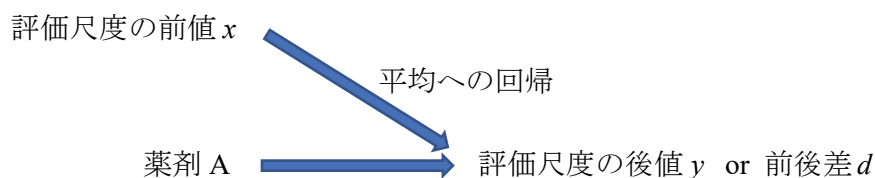


図1.11 前値が後値に与える影響

前後差 D に対して薬剤間の比較をする際に, 前値 X を共変量とするのは, 「平均への回帰」と言われている現象の影響を取り除いて薬剤間の比較を行うためである. 後値 Y を反応とする場合でも, 同様に前値 X を共変量として含める必要がある. 前後差 D のみによる対応のない t 検定を適用したくなるのであるが, 慎まなければならない.

「平均への回帰」について詳しくは、第 9 章を参照のこと。なお、「平均への回帰」とは、測定される反応に上限と下限がある場合に、前値 X が上限に近ければ、後値 Y は、さらに増加するよりも平均値方向に減少しやすく、逆に前値 X が下限に近ければ、後値 Y は、平均方向に増加しやすくなる現象のことである。

非平行な 2 本の回帰直線

図 1.12 に示すのは、前値 x と前後差 d の散布図に Excel の「近似曲線の追加」による線形近似（回帰直線）を重ね書きした結果である。前値が小さい場合に薬剤間の前後差 d は小さく、前値が大きい場合に薬剤間の前後差 d が広がっている。このような現象は、評価スコアが 0 以上なので、スコアの減少量に下限ができてしまうためである。

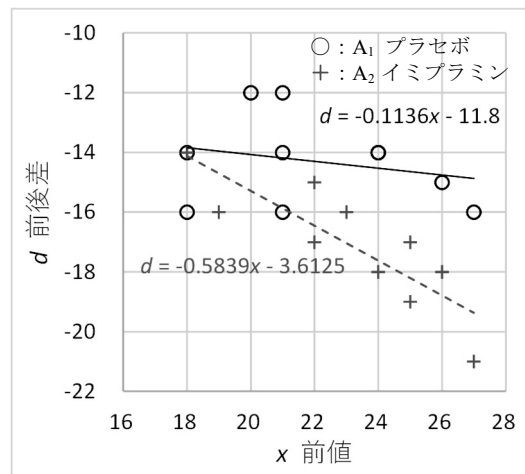


図 1.12 薬剤ごとの前後差 d に対する Excel の線形近似（回帰直線）

この図から、 A_1 プラセボ群は、回帰直線の傾きが小さいので、前値 x の大きさに関わらず前後差 d にはっきりとした違いが見いだせない。他方、 A_2 イミプラミン群は、回帰直線の傾きが大きく、前値 x が大きければ前後差 d も大きく、前値 x が小さくなると前後差 d は、 A_1 プラセボ群に近い前後差 d となることが、読み取れる。

前値 x を共変量とする共分散分析を適用する際に、薬剤 A と前値 x に交互作用 $A \times x$ がないことが必須であるとの思い込みが蔓延している。交互作用 $A \times x$ を統計的に評価するため、薬剤 A についての (1, -1) 対比型ダミー変数を a_1 として、交互作用モデルの回帰式、主効果モデルの回帰式

$$\left. \begin{array}{ll} \text{交互作用モデル: } d_i = \theta_0 + \theta_1 a_{1,i} + \theta_2 x_i + \theta_3 (a_{1,i} x_i) + \varepsilon_i \\ \text{主効果モデル: } d_i = \theta_0 + \theta_1 a_{1,i} + \theta_2 x_i + \varepsilon_i \end{array} \right\} \quad (1.26)$$

を用いる。薬剤 A が 2 水準なので、交互作用の検討は、分散分析表によらず $a_{1,i} x_i$ のパラメータ θ_3 に関する p 値で行なうことができる。

表 1.24 左に示すのは、交互作用モデルに対応するデザイン行列 \mathbf{X} である。デザイン行列 \mathbf{X} は、(20 行×4 列) であるが、「定数に 0 を使用」をオフ (デフォルト) とした場合に、選択する変数は、切片 x_0 を除いた (a_1, x, a_1x) である。表 1.24 右上段に Excel の回帰分析を適用した交互作用モデルの結果を示す。表 1.24 右下段に示す主効果モデルの場合は、交互作用のための変数 a_1x を除いた結果である。

表 1.24 前後差 d に対する交互作用モデルおよび主効果モデル

		デザイン行列 X											
		交互作用モデル											
		主効果モデル							交互作用モデル				
No	A	x_0	a_1	x	a_1x	y	d		分散分析表 「定数に 0 を使用」 off				
1	A_1	1	1	18	18	4	-14			自由度	変動	分散	分散比
2		1	1	21	21	7	-14		回帰	3	66.5524	22.1841	11.9721
3		1	1	24	24	10	-14		残差	16	29.6476	1.8530	$=\sigma^2$
4		1	1	24	24	10	-14		合計	19	96.2000		
5		1	1	18	18	2	-16						
6		1	1	27	27	11	-16			係数	標準誤差	t	P -値
7		1	1	21	21	9	-12	θ^{\wedge}_0	切片 x_0	-7.7062	2.4197	-3.1848	0.0058
8		1	1	21	21	5	-16	θ^{\wedge}_1	a_1	-4.0938	2.4197	-1.6918	0.1101
9		1	1	26	26	11	-15	θ^{\wedge}_2	x	-0.3488	0.1062	-3.2825	0.0047
10		1	1	20	20	8	-12	θ^{\wedge}_3	a_1x	0.2351	0.1062	2.2130	0.0418
11	A_2	1	-1	27	-27	6	-21						
12		1	-1	18	-18	4	-14		主効果モデル分散分析表「定数に 0 を使用」 off				
13		1	-1	25	-25	6	-19			自由度	変動	分散	分散比
14		1	-1	23	-23	7	-16		回帰	2	57.4778	28.7389	12.6171
15		1	-1	22	-22	5	-17		残差	17	38.7222	2.2778	$=\sigma^2$
16		1	-1	24	-24	6	-18		合計	19	96.2000		
17		1	-1	25	-25	8	-17						
18		1	-1	22	-22	7	-15			係数	標準誤差	t	P -値
19		1	-1	26	-26	8	-18	$\theta^{\wedge'}_0$	切片 x_0	-8.1924	2.6717	-3.0664	0.0070
20		1	-1	19	-19	3	-16	$\theta^{\wedge'}_1$	a_1	1.2169	0.3436	3.5415	0.0025
平均				22.550	6.850		-15.70	$\theta^{\wedge'}_2$	x	-0.3329	0.1175	-2.8327	0.0115

推定されたパラメータによる交互作用の検討

表 1.24 右上段のパラメータの推定値から交互作用を含む回帰式は、

$$\left. \begin{aligned} \hat{d}^{(\text{交})} &= \hat{\theta}_0 + \hat{\theta}_1 a_1 + \hat{\theta}_2 x + \hat{\theta}_3 a_1 x \\ &= -7.7062 - 4.0938 a_1 - 0.3488 x + 0.2351 a_1 x \end{aligned} \right\} \quad (1.27)$$

となる。(A₁, A₂) それぞれの回帰式は、

$$\left. \begin{aligned} \hat{d}^{(\text{交}, A_1)} &= -7.7062 - 4.0938 \times 1 - 0.3488 x + 0.2351 \times 1 \times x \\ &= -11.8000 - 0.1136 x \\ \hat{d}^{(\text{交}, A_2)} &= -7.7062 - 4.0938 \times (-1) - 0.3488 x + 0.2351 \times (-1) \times x \\ &= -3.6125 - 0.5839 x \end{aligned} \right\} \quad (1.28)$$

として求められ、これらは、図 1.12 で示されている回帰式に一致する。

交互作用の検討は、変数 a_1x のパラメータ $\hat{\theta}_3 = 0.2351$ 、標準誤差 $SE(\hat{\theta}_3) = 0.1062$ を用い、 t 値と p 値は、

$$\left. \begin{aligned} t &= \frac{\hat{\theta}_3}{SE(\hat{\theta}_3)} = \frac{0.2351}{0.1062} = 2.2130 \\ p &= \text{T.dist.2T}(2.2130, 16) = 0.0418 \end{aligned} \right\} \quad (1.29)$$

となり、 $p < 0.05$ なので、有意な交互作用があると判断される。

分散分析表による交互作用の検討

交互作用の検討は、交互作用モデルの分散分析表と主効果モデルの分散分析表を合成することで行うことができる。表 1.24 右下段に示されているパラメータから主効果モデルの回帰式は、

$$\left. \begin{aligned} \hat{d}^{(\pm)} &= \hat{\theta}_0' + \hat{\theta}_1' a_1 + \hat{\theta}_2' x \\ &= -8.1924 + 1.2169 a_1 - 0.3329 x \end{aligned} \right\} \quad (1.30)$$

となる。共通の傾きは、 $\hat{\theta}_2' = -0.3329$ であり、それぞれの切片は、 $\hat{\theta}_0' \pm \hat{\theta}_1' = (-6.9756, -9.4093)$ となる。

分散分析表の残差の平方和（変動） $S_{\text{残差}}$ は、前後差 d_i とそれぞれの推定値 $\hat{d}_i^{(\text{交})}$ および $\hat{d}_i^{(\pm)}$ からの差の偏差平方和

$$\left. \begin{aligned} S_{\text{残差}}^{(\text{交})} &= \sum_i (d_i - \hat{d}_i^{(\text{交})})^2 = 29.6476 \\ S_{\text{残差}}^{(\pm)} &= \sum_i (d_i - \hat{d}_i^{(\pm)})^2 = 38.7222 \end{aligned} \right\} \quad (1.31)$$

であり、分散分析表の残差の平方和（変動）として求められている。表 1.25 に示すように、これらの差が交互作用の平方和

$$\left. \begin{aligned} S_{\text{Axx}} &= S_{\text{残差}}^{(\pm)} - S_{\text{残差}}^{(\text{交})} \\ &= 38.7222 - 29.6476 = 9.0746 \end{aligned} \right\} \quad (1.32)$$

となる。これは、交互作用モデルの残差の平方和 $S_{\text{残差}}^{(\text{交})}$ には、交互作用の平方和 S_{Axx} が含まれていないために、交互作用が含まれている主効果モデルの残差平方和 $S_{\text{残差}}^{(\pm)}$ からの差を取ることによって交互作用の平方和 S_{Axx} を取り出すことができる。

残差の平方和ではなく回帰の平方和の差からも交互作用の平方和を求めることができる。表 1.24 から交互作用モデルの回帰の平方和は、 $S_{\text{回帰}}^{(\text{交})} = 66.5524$ であり、主効果モデルの回帰の平方和は、 $S_{\text{回帰}}^{(\pm)} = 57.4778$ である。この差からも交互作用の平方和 S_{Axx}

$$\left. \begin{aligned} S_{\text{Axx}} &= S_{\text{回帰}}^{(\text{交})} - S_{\text{回帰}}^{(\pm)} \\ &= 66.5524 - 57.4778 = 9.0746 \end{aligned} \right\} \quad (1.33)$$

を取り出すことができる。表 1.25 に示すように、交互作用の F 値から p 値を求めると 0.0418 となり、式 (1.29) で示した回帰パラメータに対する p 値に一致する。

表 1.25 分散分析表 (交互作用モデル, 平方和の分解)

	自由度 df	平方和(変動)	平均平方(分散)	F	p
回帰(a_1+x+a_1x)	3	66.5524	22.1841	11.9721	0.0002
(a_1+x) 主効果	2	57.4778	28.7389	15.5096	0.0002
(a_1x) 交互作用	1	9.0746	9.0746	4.8973	0.0418
残差	16	29.6476	1.8530	1.0000	
合計	19	96.2000			
$p=F.\text{dist.RT}$ (F 値, 分子の df , 分母の df)					

非平行な2本の回帰直線の差

交互作用が有意な差ではなく, 平行な2本の直線があてはめられれば, A_2 イミプラミン群の A_1 プラセボ群に対する薬効は, 式 (1.14) で示したように2本の直線間の差が, 切片の差と同じになる. さて, 交互作用が有意となった場合には, 前値 x の大きさにより2本の直線間の差は異なり, どのような要約統計量によって薬効を論ずるかが課題となる. 恣意性が入り込まないように, 交互作用が有意となった場合は, 表 1.24 左下段に示した前値 x の総平均 $\bar{x}_.. = 22.550$ における回帰直線の推定値の差で評価するなど, 前もって定めておく必要がある.

交互作用がない場合の平行な2本の直線間の差は, 前値 x が変化しても同じであり, そのために, 前値 $x=0$ の切片のパラメータの差が群間差となり, その95%信頼区間から群間差について論ずることができる. ただし, 交互作用がある場合は, 前値 x の変化によって差が変化するので, A_2 イミプラミンと A_1 プラセボの2本の回帰直線の差 ($\hat{d}^{(\text{交}, A_2)} - \hat{d}^{(\text{交}, A_1)}$) について, 前値 x を変化させたときの分散が必要となる. 2本の回帰直線の差は,

$$\left. \begin{array}{l} A_2 \text{の回帰式: } \hat{d}^{(\text{交}, A_2)} = (\hat{\theta}_0 - \hat{\theta}_1) + (\hat{\theta}_2 - \hat{\theta}_3)x = -3.6125 - 0.5839x \\ A_1 \text{の回帰式: } \hat{d}^{(\text{交}, A_1)} = (\hat{\theta}_0 + \hat{\theta}_1) + (\hat{\theta}_2 + \hat{\theta}_3)x = -11.8000 - 0.1136x \end{array} \right\} \quad (1.34)$$

$$\text{差} \quad \left\{ \begin{array}{l} (\hat{d}^{(\text{交}, A_2)} - \hat{d}^{(\text{交}, A_1)}) = -2\hat{\theta}_1 - 2\hat{\theta}_3x \\ \quad \quad \quad = -2 \times (-4.0938) - 2 \times 0.2351x \\ \quad \quad \quad = 8.1875 - 0.4702x \end{array} \right\} \quad (1.35)$$

である. この回帰式の差の推定値の分散は, 合成分散の一般公式から

$$\left. \begin{array}{l} \text{Var}(\hat{d}^{(\text{交}, A_2)} - \hat{d}^{(\text{交}, A_1)}) = \text{Var}(-2\hat{\theta}_1 - 2\hat{\theta}_3x) \\ \quad \quad \quad = 4\text{Var}(\hat{\theta}_1) + 8\text{Cov}(\hat{\theta}_1, \hat{\theta}_3)x + 4\text{Var}(\hat{\theta}_3)x^2 \end{array} \right\} \quad (1.36)$$

が得られる. これらのパラメータの分散および共分散は, 表 1.24 に示したデザイン行列 \mathbf{X} と交互作用モデルの分散分析表の誤差分散の推定値 $\hat{\sigma}_{(\text{交互})}^2 = 1.8539$ を用いて, 式 (1.10) および式 (1.18) と同様にパラメータの共分散行列 $\Sigma(\hat{\theta})$ が,

$$\Sigma(\hat{\theta}) = (\mathbf{X}^T \mathbf{X})^{-1} \hat{\sigma}_{(\text{交互})}^2 \quad (1.37)$$

	$\Sigma(\theta^\wedge) = (X^T X)^{-1} \sigma^{\wedge 2}_{交互}$				
$\Sigma(\theta^\wedge) =$	5.8550	-0.6666	-0.2550	0.0233	θ^\wedge_0
	-0.6666	5.8550	0.0233	-0.2550	θ^\wedge_1
	-0.2550	0.0233	0.0113	-0.0008	θ^\wedge_2
	0.0233	-0.2550	-0.0008	0.0113	θ^\wedge_3
	x_0	a_1	x	$a_1 x$	

として得られ, $Var(\hat{\theta}_1) = 5.8550$, $Cov(\hat{\theta}_1, \hat{\theta}_3) = -0.2550$, $Var(\hat{\theta}_3) = 0.0113$ となるので, 式 (1.35) の x を変化させ, 分散を計算することができる.

2本の直線間の差に対する95%信頼区間および t 検定

表 1.26 に示すのは, 図 1.12 に示したように, 前値 x を 16 から 2 刻みで 30 まで増加させた場合の各種の推定値である. ちなみに, $x = 22$ における $(\hat{d}_{A_2, x=22}^{(交)} - \hat{d}_{A_1, x=22}^{(交)})$ の p 値が, $p = 0.0036$ と有意な差となっている.

表 1.26 非平行な2本の回帰直線の差の推定値に対する t 検定

x	$d_{A1}^{(交)}$	$d_{A2}^{(交)}$	差	$Var(\text{差})$	$SE(\text{差})$	$L95\%$	$U95\%$	t 値	p 値	
16	-13.62	-12.95	0.66	2.3433	1.5308	-2.5814	3.9088	0.4336	0.6704	
18	-13.85	-14.12	-0.28	1.3342	1.1551	-2.7255	2.1719	-0.2396	0.8137	
20	-14.07	-15.29	-1.22	0.6864	0.8285	-2.9736	0.5390	-1.4693	0.1611	
22	-14.30	-16.46	-2.16	0.3998	0.6323	-3.4981	-0.8174	-3.4127	0.0036	*
22.550	-14.36	-16.78	-2.42	0.3843	0.6199	-3.7305	-1.1023	-3.8981	0.0013	*
24	-14.53	-17.63	-3.10	0.4743	0.6887	-4.5582	-1.6382	-4.4985	0.0004	*
26	-14.75	-18.79	-4.04	0.9101	0.9540	-6.0611	-2.0163	-4.2334	0.0006	*
28	-14.98	-19.96	-4.98	1.7072	1.3066	-7.7490	-2.2093	-3.8108	0.0015	*
30	-15.21	-21.13	-5.92	2.8654	1.6928	-9.5081	-2.3312	-3.4970	0.0030	*
			$t_{0.05}(20-4) =$	2.1199						

図 1.13 に示すのは, 表 1.26 で求められた2本の回帰直線の差の推定値および差の95%信頼区間である.

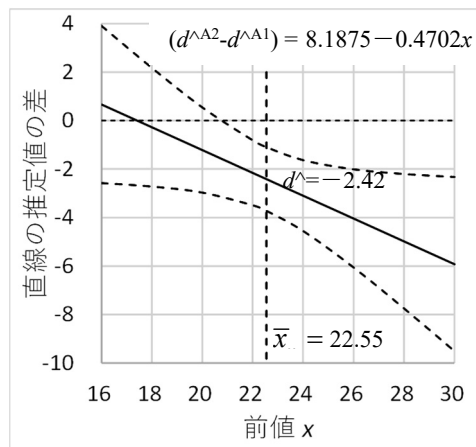


図 1.13 2本の回帰直線の差の推定値に対する95%信頼区間

前値が総平均 $\bar{x}_{..} = 22.550$ の場合についての予測値は、式 (1.34) および式 (1.35) より

$$\left. \begin{aligned} \hat{d}_{x=22.55}^{(交, A_1)} &= -11.8000 - 0.1136 \times 22.550 = -14.36 \\ \hat{d}_{x=22.55}^{(交, A_2)} &= -3.6125 - 0.5839 \times 22.550 = -16.78 \\ \hat{d}_{x=22.55}^{(交, A_2)} - \hat{d}_{x=22.55}^{(交, A_1)} &= -2.42 \end{aligned} \right\} \quad (1.38)$$

となる。分散は、式 (1.36) より、

$$\begin{aligned} Var(\hat{d}_{x=22.55}^{(交, A_2)} - \hat{d}_{x=22.55}^{(交, A_1)}) &= 4Var(\hat{\theta}_1) + 8Cov(\hat{\theta}_1, \hat{\theta}_3)x + 4Var(\hat{\theta}_3)x^2 \\ &= 4 \times 5.8550 + 8 \times (-0.2550) \times 22.550 + 4 \times 0.0113 \times 22.550^2 \\ &= 0.3843 \end{aligned} \quad (1.39)$$

となり、 t 値および p 値は、

$$\left. \begin{aligned} t_{x=20} &= \frac{\hat{d}_{x=22.55}^{(交, A_2)} - \hat{d}_{x=22.55}^{(交, A_1)}}{\sqrt{Var(\hat{d}_{x=22.55}^{(交, A_2)} - \hat{d}_{x=22.55}^{(交, A_1)})}} \\ &= \frac{-2.42}{\sqrt{0.3843}} = -3.8981 \end{aligned} \right\} \quad (1.40)$$

$$\left. \begin{aligned} p &= T.dist.2T(|t|, 20 - 4) \\ &= T.dist.2T(|-3.8981|, 16) = 0.0013 \end{aligned} \right\} \quad (1.41)$$

として計算されている。したがって、有意な差があると判定される。前値が $x = 22$ の場合も同様に $p = 0.0036$ が得られ、有意な差と判定される。 $x = 22$ 以上の場合も同様に有意な差となっている。これらの結果から、 A_2 イミブタミンは、 A_1 プラセボに対し、前値が $x = 22$ 以上の場合に有意な差があると判定される。

2 本の回帰直線の差が有意となる前値 x_0

前値の総平均 $\bar{x} = 22.550$ における p 値による判定だけでなく、2 本の回帰直線の差の 95% 信頼区間の上限が 0 となる前値 x_0 が推定できれば、「 x_0 以上であれば、統計的に有意な差があると言える」ので、歯切れがよい。95% 信頼区間の上限の式は、

$$\hat{y}_0 = (2\hat{\theta}_1 + 2\hat{\theta}_3 x_0) + t_{0.05}(20 - 4) \sqrt{4Var(\hat{\theta}_1) + 8Cov(\hat{\theta}_1, \hat{\theta}_3)x_0 + 4Var(\hat{\theta}_3)x_0^2} \quad (1.42)$$

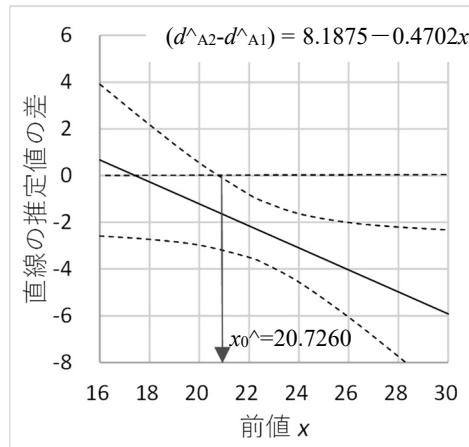
で与えられるので、 $\hat{y}_0 = 0$ となる x_0 を求めるためには、

$$(-2\hat{\theta}_1 - 2\hat{\theta}_3 x_0)^2 = t_{0.05}^2(16) \cdot [4Var(\hat{\theta}_1) + 8Cov(\hat{\theta}_1, \hat{\theta}_3)x_0 + 4Var(\hat{\theta}_3)x_0^2] \quad (1.43)$$

を整理すると x_0 に関して 2 次式となるので、2 次式の解の公式を用いて解を求めればよいが、煩雑な式となる。救いは、Excel のソルバーであり、式 (1.42) で $\hat{y}_0 = 0$ となるような x_0 を探索的に求めてくれる。表 1.27 に示すのは、 $x = 21.0000$ の場合の上限が -0.1916 であり、Excel のソルバーで上限が 0.0000 となるように x を探索的に変化させた結果とし、 $x = 20.7260$ が得られた結果である。この結果として、 $p = 0.0500$ と有意な差となる x の推定値が得られたことになる。

表 1.27 Excel ソルバーによる 95%信頼区間上限が 0 となる x_0 の推定

	x	d^{\wedge}_{A1}	d^{\wedge}_{A2}	$d^{\wedge}_{A2}-d^{\wedge}_{A1}$	$Var(差)$	$SE(差)$	$L95\%$	$U95\%$	t 値	p 値
	20	-14.07	-15.29	-1.22	0.6864	0.8285	-2.9736	0.5390	-1.4693	0.1611
	21.0000	-14.19	-15.87	-1.69	0.4979	0.7056	-1.6875	-0.1916	-2.3915	0.0294
	22	-14.30	-16.46	-2.16	0.3998	0.6323	-2.1577	-0.8174	-3.4127	0.0036 *
	20	-14.07	-15.29	-1.22	0.6864	0.8285	-1.2173	0.5390	-1.4693	0.1611
#	20.7260	-14.16	-15.71	-1.56	0.5406	0.7352	-1.5586	0.0000	-2.1199	0.0500 *
	22	-14.30	-16.46	-2.16	0.3998	0.6323	-2.1577	-0.8174	-3.4127	0.0036 *
#	ソルバー解				$t_{0.05}(20-4)=$		2.1199			*: $p < 0.05$

図 1.14 有意な差となる前値 x_0 の推定

JMP による検証

JMP による交互作用を含む分散分析表を表 1.28 に示す。分散分析表は、逐次（タイプ 1）検定（平方和）を選択する。薬剤 A の平均平方 $S_A = 39.2000$ に前値 x の平方和 $S_x = 18.2778$ を加えた $S_{A \times x} = 57.4778$ は、表 1.25 の主効果の平方和に一致する。さらに交互作用 $S_{A \times x} = 9.0746$ は、表 1.25 の交互作用に一致する。

表 1.28 JMP による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	3	66.5524	22.1841	11.9721
誤差	16	29.6476	1.8530	p値(Prob>F)
全体(修正済み)	19	96.2000		0.0002*

逐次(タイプ1)検定				
要因	自由度	逐次平方和	F値	p値
A	1	39.2000	21.1551	0.0003*
x	1	18.2778	9.8640	0.0063*
A*x	1	9.0746	4.8973	0.0418*

1.4. 発電プラント周辺でのカキの成長

Littell ら (2002), 「SAS for Linear Models 4th ed., Chapter 7 Analysis of Covariance」に興味深い事例がある. 第 7.2 節の「1 元配置型の共分散分析」は, カキの成長に関する実験データの解析事例である. ただし, 2×2 元配置型の 4 群と対照群の 1 群を加えた実験であり, 全体で 5 群の 1 元配置型としている.

前値を共変量とする対照群を含む 2 元配置

実験は, 発電プラントの (取水路, 放水路), 近くの湾の浅い部分, 3 箇所でのカキの成長の比較である. 発電プラントの取水路と放水路では, それぞれの水路の (上部, 下部) に 4 籠を設置し, 湾内では中ほどの水深に 4 籠が設置された. 表 1.29 に示すように, それぞれの籠の中には, 10 個のカキが入れられ, 実験の最初に洗浄されカキ重量が測定され, 1 か月後に再び洗浄されカキ重量が測定された結果である.

実験の目的は, A_1 取水路に比較して, 発電プラントで人工的に加熱された A_2 放水路でのカキの成長に与える影響を知りたい. さらに, B : 水深 (上部, 下部) での成長与える影響も合わせて検討したい. それらを C : 湾内での成長とも比較したいなどである. 図 1.15 に示すように, initial 重量を共変量とするが, 重量が減少することはなく増加のみなので, 「平均への回帰」現象は起きない.

表 1.29 カキの成長実験データ [Littell ら (2002), Output 7.1]

B: 上下	A: 水路						C: 湾内		
	A_1 : 取水路			A_2 : 放水路			B : 位置: 中ほど		
	籠	x : initial	y : final	x : initial	y : final		籠	x : initial	y : final
B_1 : 上部	1	27.2	32.6	28.6	35.2		1	20.4	24.6
	2	32.0	36.6	22.4	29.1		2	19.6	23.4
	3	33.0	37.7	23.2	28.9		3	25.1	30.3
	4	26.8	31.0	24.4	30.2		4	18.1	21.8
B_2 : 下部	1	28.6	33.8	29.3	35.0				
	2	26.8	31.7	21.8	27.0				
	3	26.5	30.7	30.3	36.4				
	4	26.8	30.4	24.3	30.5				

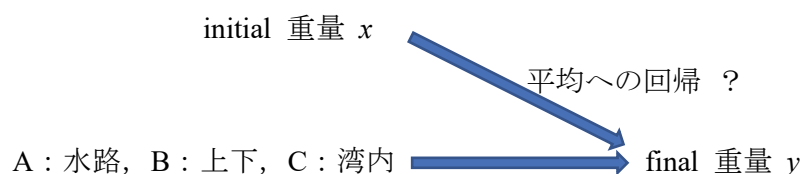


図 1.15 設置場所がカキの成長に与える影響

図 1.16 に示すのは、データの全体像を把握するために、Excel の散布図で 5 群の前値 x 、後値 y を重ね書きし、それぞれの群に回帰直線を引いた結果である。C：湾内設置の前値 x が他に比べて明らかに小さい。カキの洗浄後の設置場所による 4 籠の前値の平均値が揃うように“群分け”が行われなかったようであり、設置が同時ではなく何日かけて逐次的に行なわれているとすれば、その間にカキがいくばくか成長した後に設置したように推測されるが、詳細は不明である。

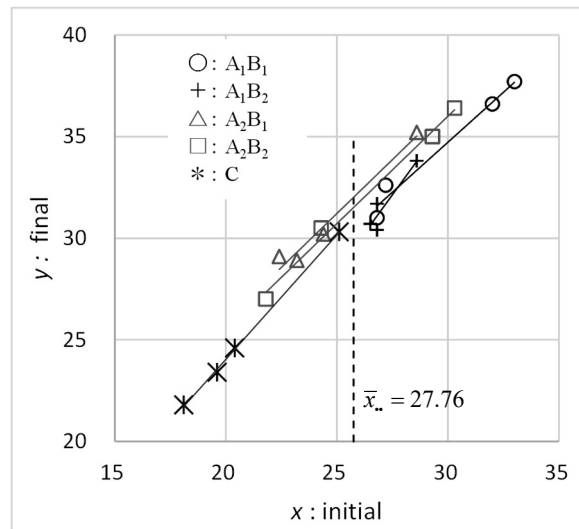


図 1.16 設置場所によるカキの成長

共変量を含む 1 元配置としてモデル化

1 か月後のカキの成長は、設置場所により 4～6 単位程度であり、C：湾内と A₁：取水路は、前値 x が全く異なるものの回帰直線を引き延ばせば互いに重なり合い、カキの成長は同程度と思われる。それらに対し、A₂：放水路は、2 単位ほど成長が早いようである。

込み入った群構成になっているので、表 1.30 に示すように全体で 5 群 (G₁～G₅) としてダミー変数の設定を行う。SAS の結果と合わせるため、最後の水準を 0 とする (1, 0) 型のダミー変数 (g_1 , g_2 , g_3 , g_4) を使うことにする。推定される回帰パラメータは、G₅ 群 (C：湾内) との差となる。切片の推定値は、G₅ 群の切片となるので、平行線をあてはめた場合には、前値 x の大きさに関わらず G₅ 群 (C：湾内) との差の推定値となる。

表 1.30 (1, 0) 型ダミー変数による群構成

A:水路	B:上下	群G	(1, 0) 型ダミー変数			
			g_1	g_2	g_3	g_4
A ₁ :取水路	B ₁ :上部	G ₁	1	0	0	0
"	B ₂ :下部	G ₂	0	1	0	0
A ₂ :放水路	B ₁ :上部	G ₃	0	0	1	0
"	B ₂ :下部	G ₄	0	0	0	1
C:湾内	中ほど	G ₅	0	0	0	0

表 1.31 に示すように、切片を x_0 、群に対し (g_1, g_2, g_3, g_4)、共変量を x とした (20 行×6 列) のデザイン行列 X に対し、Excel の回帰分析を適用する。回帰のモデル式は、図 1.16 から交互作用はないと判断されるので、主効果モデル

$$y_i = \theta_0 x_{0,i} + \theta_1 g_{1,i} + \theta_2 g_{2,i} + \theta_3 g_{3,i} + \theta_4 g_{4,i} + \theta_5 x_i + \varepsilon_i, \quad \varepsilon_i \sim \text{正規分布} \quad (1.44)$$

を適用する。

1 元配置の水準の組み合わせ平均

この実験で知りたのは、水路 A (A_1, A_2) による違いはあるのか、水深 B (B_1, B_2) による違いはあるのか、さらには、A と B の交互作用はあるのか、などである。このためには、湾内 C を除いて 2 元配置型での共分散分析とすれば可能と思われるが、比較の基準となる湾内 C を外すことはできない。式 (1.44) で推定されるのは湾内 C との差のみであるが、パラメータの共分散分行列を活用すれば、事後的にどのような推定でも可能となることを示す。

表 1.31 (1, 0) 型のダミー変数を用いた回帰分析の結果

				デザイン行列 X													
A	B	G	籠	x_0	g_1	g_2	g_3	g_4	x	y	分散分析表「定数に 0 を使用」off						
A_1	B_1	G_1	1	1	1	0	0	0	27.2	32.6		自由度	変動	分散	分散比		
			2	1	1	0	0	0	32.0	36.6	回帰	5	354.45	70.8894	235.05		
			3	1	1	0	0	0	33.0	37.7	残差	14	4.2223	0.3016	$=\sigma^{\wedge 2}$		
			4	1	1	0	0	0	26.8	31.0	合計	19	358.67				
	B_2	G_2	1	1	0	1	0	0	28.6	33.8							
			2	1	0	1	0	0	26.8	31.7		係数	標準誤差	t	P -値		
			3	1	0	1	0	0	26.5	30.7	θ^{\wedge}_0 切片 x_0	2.4949	1.0279	2.4272	0.0293		
			4	1	0	1	0	0	26.8	30.4	θ^{\wedge}_1 g_1	-0.2445	0.5766	-0.4240	0.6780		
A_2	B_1	G_3	1	1	0	0	1	0	28.6	35.2	θ^{\wedge}_2 g_2	-0.2803	0.4929	-0.5686	0.5786		
			2	1	0	0	1	0	22.4	29.1	θ^{\wedge}_3 g_3	1.6548	0.4294	3.8534	0.0018		
			3	1	0	0	1	0	23.2	28.9	θ^{\wedge}_4 g_4	1.1071	0.4718	2.3468	0.0342		
			4	1	0	0	1	0	24.4	30.2	θ^{\wedge}_5 x	1.0832	0.0476	22.7461	0.0000		
	B_2	G_4	1	1	0	0	0	1	29.3	35.0							
			2	1	0	0	0	1	21.8	27.0	パラメータの共分散行列 $\Sigma(\theta^{\wedge})=(X^T X)^{-1} \sigma^{\wedge 2}$						
			3	1	0	0	0	1	30.3	36.4	θ^{\wedge}_0	1.0565	0.3468	0.2253	0.1062	0.1899	-0.0472
			4	1	0	0	0	1	24.3	30.5	θ^{\wedge}_1	0.3468	0.3324	0.2048	0.1535	0.1896	-0.0203
C		G_5	1	1	0	0	0	0	20.4	24.6	θ^{\wedge}_2	0.2253	0.2048	0.2430	0.1311	0.1567	-0.0145
			2	1	0	0	0	0	19.6	23.4	θ^{\wedge}_3	0.1062	0.1535	0.1311	0.1844	0.1245	-0.0087
			3	1	0	0	0	0	25.1	30.3	θ^{\wedge}_4	0.1899	0.1896	0.1567	0.1245	0.2225	-0.0128
			4	1	0	0	0	0	18.1	21.8	θ^{\wedge}_5	-0.0472	-0.0203	-0.0145	-0.0087	-0.0128	0.0023
総平均				$\bar{x}_{\cdot}=25.76$						30.85		x_0	g_1	g_2	g_3	g_4	x

表 1.31 右下段に示すのは、パラメータの共分散行列 $\Sigma(\hat{\theta})$ であり、Excel の行列関数を使いデザイン行列 X と分散分析表の誤差分散の推定値 $\hat{\sigma}^2 = 0.3016$ を用いて

$$\begin{aligned} \Sigma(\hat{\theta}) &= (X^T X)^{-1} \hat{\sigma}^2 \\ &= \text{Minverse}(\text{Mmult}(\text{Transpose}(X \text{ の範囲}), X \text{ の範囲})) * \hat{\sigma}^2 \end{aligned}$$

のように計算した結果である。

表 1.31 右中段は, Excel の回帰分析で求めた係数(パラメータ)で, 切片の推定値 $\hat{\theta}_0 = 2.4949$ は, G₅ 群 (C: 湾内) の切片の推定値となる. G₁ 群 (A₁: 取水路, B₁: 上部) 推定値は, $\hat{\theta}_1 = -0.2445$ であるが, これは G₅ 群 (C: 湾内) の切片との差であり, G₁ 群の切片の推定値は,

$$G_1 \text{ 群の切片} = \hat{\theta}_0 + \hat{\theta}_1 = 2.4949 + (-0.2445) = 2.2504 \quad (1.45)$$

となる. 交互作用を含まない主効果モデルなので, G₅ 群 (C: 湾内) は, 対照群として設定され, 対照群と G₁ 群 (A₁: 取水路, B₁: 上部) の切片の差が, 前値 x に関係なく 2 本の回帰直線の差の推定値になっている. したがって, $\hat{\theta}_1 = -0.2445$ に対する t 検定の p 値は $p_1 = 0.6780$ であり, 有意な差ではないことが分かる. G₃ 群および G₄ 群は, (A₂: 放水路) に設置した場合であり, (B₁: 上部, B₂: 下部) とともに G₅ 群 (C: 湾内) に比べて有意な差 ($p_3 = 0.0018$, $p_4 = 0.0342$) であることが読み取れる.

パラメータの共分散行列 $\Sigma(\hat{\theta})$ の対角要素の平方根は, それぞれの係数 (パラメータ) の標準誤差に相当する. 第 3 群 g_3 の推定値 $\hat{\beta}_3 = 1.6548$ の $SE_3 = 0.4294$ は, $\Sigma(\hat{\theta})$ の 4 行 4 列目 $Var(\hat{\beta}_3) = 0.1844$ の平方根 $\sqrt{0.1844} = 0.4294$ に一致する. パラメータの共分散行列 $\Sigma(\hat{\theta})$ は, 回帰分析で出力されていない各種の推定に対する 95%信頼区間を求めるために使われる. カキ籠の設置位置 (B₁: 上部, B₂: 下部) 間の推定値の差の 95%信頼区間を求めたい, 推定されたそれぞれの回帰直線の 95%信頼区間を図で示したい, などである.

最も基本的な推定は, 各群の共変量である前値 x の影響を除くために前値 x の総平均 $\bar{x} = 25.76$ におけるそれぞれの回帰直線の推定値とその 95%信頼区間であり, それらを図示することも結果の解釈に役立つ. そのためには, 表 1.32 に示すようにパラメータの推定値の列ベクトル $\hat{\theta}$ に対し, 推定目的に合致した行ベクトル $\mathbf{l}^{(i)}$ を設定し, それらの線形和 $\hat{L}^{(i)}$

$$\hat{L}^{(i)} = \mathbf{l}^{(i)} \hat{\theta} = \sum_{k=0}^5 l_k^{(i)} \hat{\theta}_k \quad (1.46)$$

で推定する. 線形和 $\hat{L}^{(i)}$ の分散 $Var(\hat{L}^{(i)})$ は,

$$Var(\hat{L}^{(i)}) = \mathbf{l}^{(i)} \Sigma(\hat{\theta}) (\mathbf{l}^{(i)})^T \quad (1.47)$$

表 1.32 前値の総平均 $\bar{x} = 25.76$ に対する回帰直線上の推定値 (最小 2 乗平均)

A, B,		線形和	l_0	l_1	l_2	l_3	l_4	l_5			
C	G	L	x_0	g_1	g_2	g_3	g_4	x	$\hat{\theta}$	$\mathbf{l}^{(i)} \hat{\theta}$	
A ₁ B ₁	G ₁	$L^{(1)}$	1	1	0	0	0	25.76	2.4949	=	30.1531
A ₁ B ₂	G ₂	$L^{(2)}$	1	0	1	0	0	25.76	-0.2445		30.1173
A ₂ B ₁	G ₃	$L^{(3)}$	1	0	0	1	0	25.76	-0.2803		32.0523
A ₂ B ₂	G ₄	$L^{(4)}$	1	0	0	0	1	25.76	1.6548		31.5047
C	G ₅	$L^{(5)}$	1	0	0	0	0	25.76	1.1071		30.3976
									1.0832		
$L^{(i)} = \text{Mmult}(\mathbf{l}^{(i)} \text{ の範囲}, \hat{\theta} \text{ の範囲})$											

によって推定することができる．各群の $\bar{x}_{..} = 25.76$ における推定に必要な線形和 $L^{(i)}$ に対する行ベクトル $\mathbf{l}^{(i)}$ を与え、 $\hat{L}^{(i)} = \mathbf{l}^{(i)} \hat{\boldsymbol{\theta}}$ によって求めている． $L^{(1)}$ の場合であれば、

$$L^{(1)} = 1 \times 2.49491 + 1 \times (-0.2445) + 0 + 0 + 0 + 25.76 \times 1.0832 = 30.1531 \quad (1.48)$$

が得られる．推定値を得るためには、四則演算で対応できるのであるが、その推定値の分散を手計算レベルで計算するのは難儀であり、表 1.33 に示すように Excel の行列関数によるパラメータの共分散行列に関する 2 次形式により簡単に計算できる．

表 1.33 各群の回帰直線上の推定値および第 5 群との差の 95%信頼区間

	線形和	A B		l_0	l_1	l_2	l_3	l_4	l_5	推定値	分散	幅	95%信頼区間		
	L	C	G	x_0	g_1	g_2	g_3	g_4	x	L^\wedge	$Var(L^\wedge)$	$t_{0.05} \times SE$	$L_{95\%}$	$U_{95\%}$	
前値 x	$L^{(1)}$	A ₁ B ₁	G ₁	1	1	0	0	0	25.76	30.1531	0.1115	0.72	29.44	30.87	
総平均	$L^{(2)}$	A ₁ B ₂	G ₂	1	0	1	0	0	25.76	30.1173	0.0799	0.61	29.51	30.72	
25.76	$L^{(3)}$	A ₂ B ₁	G ₃	1	0	0	1	0	25.76	32.0523	0.0782	0.60	31.45	32.65	
	$L^{(4)}$	A ₂ B ₂	G ₄	1	0	0	0	1	25.76	31.5047	0.0764	0.59	30.91	32.10	
	$L^{(5)}$	C	G ₅	1	0	0	0	0	25.76	30.3976	0.1312	0.78	29.62	31.17	
G ₅ との差 (C群)	$L^{(6)}$	A ₁ B ₁	G ₁	0	1	0	0	0	0.00	-0.2445	0.3324	1.24	-1.48	0.99	
	$L^{(7)}$	A ₁ B ₂	G ₂	0	0	1	0	0	0.00	-0.2803	0.2430	1.06	-1.34	0.78	
	$L^{(8)}$	A ₂ B ₁	G ₃	0	0	0	1	0	0.00	1.6548	0.1844	0.92	0.73	2.58	
	$L^{(9)}$	A ₂ B ₂	G ₄	0	0	0	0	1	0.00	1.1071	0.2225	1.01	0.10	2.12	
	$L^{(10)}$	C	G ₅	0	0	0	0	0	0.00	0.0000	0.0000	0.00	0.00	0.00	
										$t_{0.05}(20-4)=$			2.1448		
$\mathbf{l}^{(1)}=$	l_0	l_1	l_2	l_3	l_4	l_5	パラメータの共分散行列 $\boldsymbol{\Sigma}(\boldsymbol{\theta}^\wedge)=(\mathbf{X}^T\mathbf{X})^{-1}\sigma^{\wedge^2}$				$\mathbf{l}^{(1)T}$				
$Var(L^\wedge^{(1)})=$	1	1	0	0	0	25.76	1.0565	0.3468	0.2253	0.1062	0.1899	-0.0472	1		
	x_0	g_1	g_2	g_3	g_4	x	0.3468	0.3324	0.2048	0.1535	0.1896	-0.0203	1		
							0.2253	0.2048	0.2430	0.1311	0.1567	-0.0145	0		
							0.1062	0.1535	0.1311	0.1844	0.1245	-0.0087	0		
							0.1899	0.1896	0.1567	0.1245	0.2225	-0.0128	0		
							-0.0472	-0.0203	-0.0145	-0.0087	-0.0128	0.0023	25.76		
=	0.1115						θ^\wedge_0	θ^\wedge_1	θ^\wedge_2	θ^\wedge_3	θ^\wedge_4	θ^\wedge_5			

分散が計算できれば、推定値 \hat{y} の 95%信頼区間は、一般的な計算方式により

$$\begin{aligned}
 (L_{95\%}, U_{95\%}) &= \hat{L}^{(1)} \pm t(0.05, 20-6) \sqrt{Var(\hat{L}^{(1)})} \\
 &= 30.1531 \pm 2.1448 \sqrt{0.1115} \\
 &= (29.44, 30.87)
 \end{aligned}$$

として求められる．

さらに、第 5 群との差については、表 1.31 に示した Excel による回帰分析の推定値が使えるのであるが、表 1.33 の下段に示すように、差のベクトル $\mathbf{l}^{(6)} \sim \mathbf{l}^{(10)}$ を与えて計算式のフィルハンドルを用いて簡単に計算することもできる．これらの推定結果を、図 1.17 に示すように Excel の折れ線グラフで表す．

この線グラフから A_1 : 取水路 では, C : 湾内 に比べ有意な差ではなく, A_2 : 排水路 では, 位置 (B_1 : 上部, B_2 : 下部) 共に有意な差があることが読み取れる. 各水路内の位置 (B_1 : 上部, B_2 : 下部) の違いは, 95%信頼区間がそれぞれの推定値を含んでいるので, 統計的な差はないと結論付けられる.

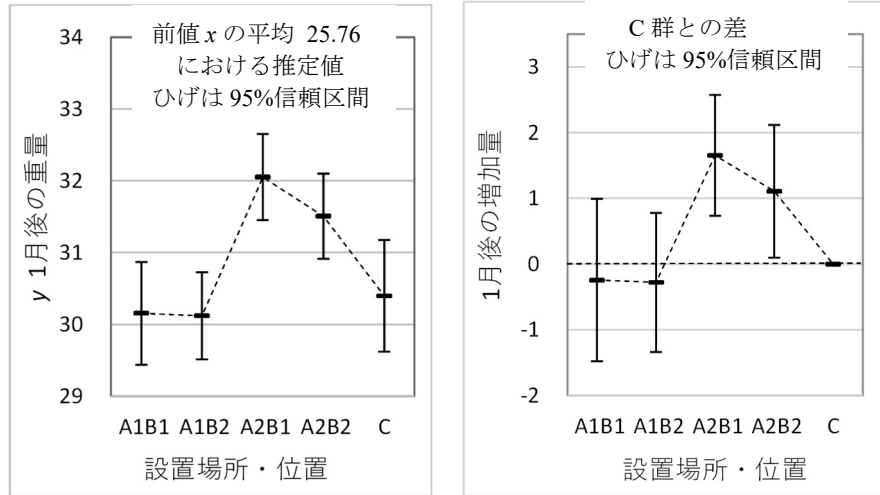


図 1.17 設置場所および位置の組み合わせによる推定平均 (最小 2 乗平均)

1 元配置を 2 元配置として再構築

線グラフから有意な差がないといっても報告書には, きちっとした解析結果が求められる. 設置場所 A_1 における位置の差 ($B_1 - B_2$) の推定値および 95%信頼区間を求めたい. そのためには, 単純に A_1B_1 の推定値 $\hat{y}_{A_1B_1}$ から A_1B_2 の推定値 $\hat{y}_{A_1B_2}$ を引けばいいのだが, 差の分散を求めるためには, パラメータに関する線形式,

$$\begin{array}{rcl}
 A_1B_1 : & \hat{y}_{A_1B_1} = \hat{\theta}_0 + \hat{\theta}_1 & + \hat{\theta}_5 x_i \\
 -) A_1B_2 : & \hat{y}_{A_1B_2} = \hat{\theta}_0 & + \hat{\theta}_2 + \hat{\theta}_5 x_i \\
 \hline
 A_1B_1 - A_1B_2 : & \hat{y}_{A_1B_1} - \hat{y}_{A_1B_2} = \hat{\theta}_1 - \hat{\theta}_2 &
 \end{array} \quad (1.49)$$

が必要となる. この分散は, パラメータが 2 つなので, 行列計算の代わりに, 合成分散の一般式でも推定できる.

$$\begin{aligned}
 Var(\hat{\theta}_1 - \hat{\theta}_2) &= Var(\hat{\theta}_1) - 2Cov(\hat{\theta}_1, \hat{\theta}_2) + Var(\hat{\theta}_2) \\
 &= 0.3324 - 2 \times 0.2048 + 0.2430 \\
 &= 0.1658
 \end{aligned} \quad (1.50)$$

設置場所 A_1 について位置を (B_1, B_2) をプールした場合の平均値を推定し, 95%信頼区間を求めたい. そのためには, A_1B_1 の推定値 $\hat{y}_{A_1B_1}$ と A_1B_2 の推定値 $\hat{y}_{A_1B_2}$ を足して 2 で割ればよい.

表 1.34 A 内 B の差, A および B の回帰直線上の推定値と 95%信頼区間

		A B	l_0	l_1	l_2	l_3	l_4	l_5	推定値	分散	標準誤差	95%信頼区間	
	L	C	x_0	g_1	g_2	g_3	g_4	x	L^\wedge	$Var(L^\wedge)$	SE	$L_{95\%}$	$U_{95\%}$
A ₁ 内B ₁	$L^{(11)}$	A ₁ :(B ₁ -B ₂)	0	1	-1	0	0	0.00	0.036	0.1658	0.8734	-0.838	0.909
B ₂ との差	$L^{(12)}$	A ₁ :(B ₂ -B ₂)	0	0	0	0	0	0.00	0.000	0.0000	0.0000	0.000	0.000
A ₂ 内B ₁	$L^{(13)}$	A ₂ :(B ₁ -B ₂)	0	0	0	1	-1	0.00	0.548	0.1579	0.8524	-0.305	1.400
B ₂ との差	$L^{(14)}$	A ₂ :(B ₂ -B ₂)	0	0	0	0	0	0.00	0.000	0.0000	0.0000	0.000	0.000
A ₁ 全体	$L^{(15)}$	A ₁	1	0.5	0.5	0	0	25.76	30.135	0.0543	0.4996	29.636	30.635
A ₂ 全体	$L^{(16)}$	A ₂	1	0	0	0.5	0.5	25.76	31.779	0.0378	0.4171	31.361	32.196
B ₁ 全体	$L^{(17)}$	B ₁	1	0.5	0	0.5	0	25.76	31.103	0.0424	0.4416	30.661	31.544
B ₂ 全体	$L^{(18)}$	B ₂	1	0	0.5	0	0.5	25.76	30.811	0.0402	0.4298	30.381	31.241
C	$L^{(19)}$	C	1	0	0	0	0	25.76	30.398	0.1312	0.7768	29.621	31.174
									$t_{0.05}(20-6)=$	2.1448			

$$\begin{aligned}
 A_1 B_1 : \quad & \hat{y}_{A_1 B_1} = \hat{\theta}_0 + \hat{\theta}_1 + \hat{\theta}_5 x_i \\
 +) \quad A_1 B_2 : \quad & \hat{y}_{A_1 B_2} = \hat{\theta}_0 + \hat{\theta}_2 + \hat{\theta}_5 x_i \\
 \hline
 \frac{A_1 B_1 + A_1 B_2}{2} : \quad & \frac{\hat{y}_{A_1 B_1} + \hat{y}_{A_1 B_2}}{2} = \hat{\theta}_0 + \frac{\hat{\theta}_1}{2} + \frac{\hat{\theta}_2}{2} + \hat{\theta}_5 x_i
 \end{aligned} \tag{1.51}$$

これを、推定に用いる線形和のベクトル \mathbf{l} として

$$\mathbf{l}^{(15)} = [1 \quad 0.5 \quad 0.5 \quad 0 \quad 0 \quad 25.76]$$

を設定し、これまでと同様に推定値を $L^{(15)} = \mathbf{l}^{(15)} \hat{\boldsymbol{\theta}}$ で計算し、分散は

$$Var(L^{(15)}) = \mathbf{l}^{(15)} \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) (\mathbf{l}^{(15)})^T \tag{1.52}$$

で計算する。

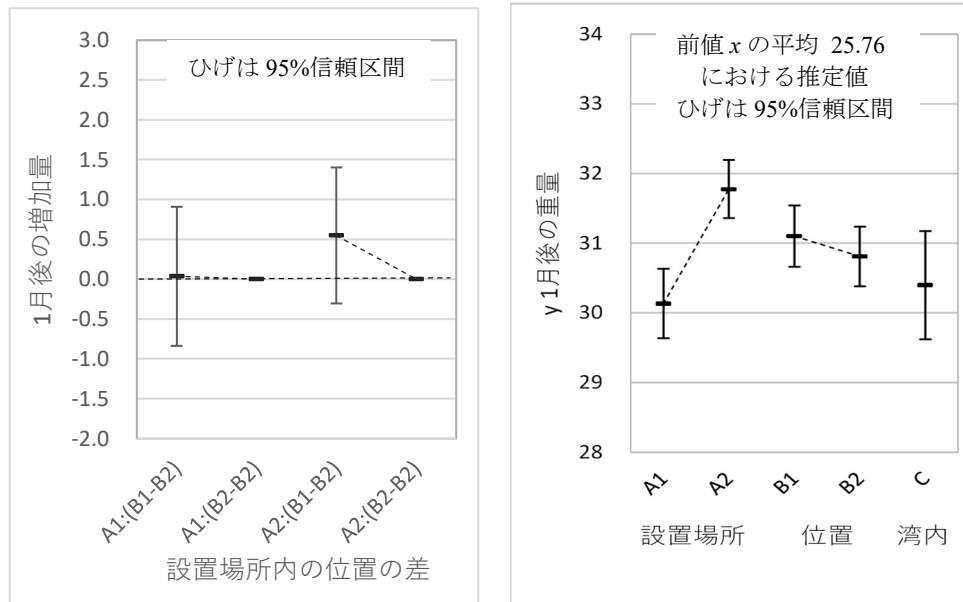


図 1.18 A 内の B 間の差, および, A の水準平均と B の水準平均

これらを総合すると, A₂の放水路においてカキの成長が他に比べて早いことが明らかとなり, その位置 (B₁上部, B₂下部) 間には有意な差が認められないと結論される。

SAS による検証

無償で継続的に提供されている OnDemand SAS の一般線形モデル GLM プロシジャにより、Excel での結果を検証する。GLM プロシジャには、最小 2 乗平均を算出するための `lsmeans` (`least squares means`) ステートメントがあり、表 1.33 に示したと同じ結果が得られる。また、`estimate` ステートメントで、表 1.34 に示したと同様に、各種の推定値を求めることができる。ただし、群 G の最後の水準に対してダミー変数として“0”を追加する必要がある。このことが、交互作用がある場合にダミー変数のインフレーションを引き起こし、`estimate` ステートメントの設定が煩雑となる原因である。

SAS プログラム 1.1 /* SAS データファイルの作成, GLM プロシジャ */

```

TITLE1 ' Oysters.SAS 2020/08/23 Y.Takahashi ' ;
data d01 ;
  input G$ @ ;
  do rep=1 to 4 ;
    input x y @ ;
    output ;
  end ;
  datalines ;
G1_A1B1 27.2 32.6 32.0 36.6 33.0 37.7 26.8 31.0
G2_A1B2 28.6 33.8 26.8 31.7 26.5 30.7 26.8 30.4
G3_A2B1 28.6 35.2 22.4 29.1 23.2 28.9 24.4 30.2
G4_A2B2 29.3 35.0 21.8 27.0 30.3 36.4 24.3 30.5
G5_C 20.4 24.6 19.6 23.4 25.1 30.3 18.1 21.8
;
proc glm data=d01 ;
  class G ;
  model Y = G x / solution ;
  lsmeans G / stderr CL ;
  estimate 'A1:(B1-B2)' G 1 -1 0 0 0 0 ;
  estimate 'A2:(B1-B2)' G 0 0 1 -1 0 0 ;
  estimate 'A1' intercept 1 G 0.5 0.5 0 0 0 x 25.76;
  estimate 'A2' intercept 1 G 0 0 0.5 0.5 0 x 25.76;
  estimate 'B1' intercept 1 G 0.5 0 0.5 0 0 x 25.76;
  estimate 'B2' intercept 1 G 0 0.5 0 0.5 0 x 25.76;
run ;

```

SAS の結果:HTML 出力を Excel に取り込み、有効数字などを揃えている。

Model ステートメントによる分散分析表

要因	自由度	平方和	平均平方	F 値	Pr > F
Model	5	354.4472	70.8894	235.05	<.0001
Error	14	4.2223	0.3016		
Corrected Total	19	358.6695			

Excel の表 1.31 右上段の結果に一致する。

群 G および前値 x に対する F 検定

要因	自由度	Type III 平方和	平均平方	F 値	Pr > F
G	4	12.0894	3.0223	10.02	0.0005
x	1	156.0402	156.0402	517.38	<.0001

Excel による各因子の平方和は、回帰の平方和の差分から別途計算する必要がある

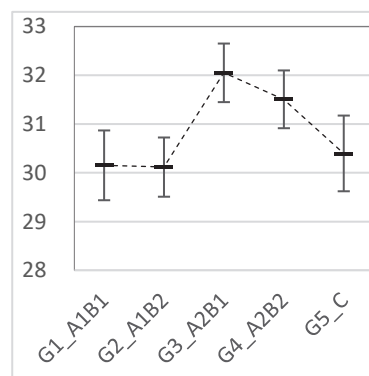
Model ステートメントによる回帰パラメータの推定 (solution オプション)

パラメータ	推定値		標準誤差	t 値	Pr > t		係数	標準誤差
Intercept	2.4949	B	1.0279	2.43	0.0293	切片	2.4949	1.0279
G G1_A1B1	-0.2445	B	0.5766	-0.42	0.6780	g_1	-0.2445	0.5766
G G2_A1B2	-0.2803	B	0.4929	-0.57	0.5786	g_2	-0.2803	0.4929
G G3_A2B1	1.6548	B	0.4294	3.85	0.0018	g_3	1.6548	0.4294
G G4_A2B2	1.1071	B	0.4718	2.35	0.0342	g_4	1.1071	0.4718
G G5_C	0	B	.	.	.	x	1.0832	0.0476
x	1.0832		0.0476	22.75	<.0001			
Note: X'X は特異行列です。w 正規方程式には一般化逆行列が使用されています。								
w 文字 'B' が付けられた推定値は一意的な推定値ではありません。								

Excel の表 1.31 右中段の結果に一致する。

lsmeans ステートメントによる最小 2 乗平均 (表 1.33 上段に対応)

	y の最小					
G	2 乗平均	標準誤差	Pr > t	95% 信頼限界		信頼幅
G1_A1B1	30.1531	0.3339	<.0001	29.44	30.87	0.72
G2_A1B2	30.1173	0.2827	<.0001	29.51	30.72	0.61
G3_A2B1	32.0523	0.2796	<.0001	31.45	32.65	0.60
G4_A2B2	31.5047	0.2764	<.0001	30.91	32.10	0.59
G5_C	30.3976	0.3622	<.0001	29.62	31.17	0.78
表1.35の上段に一致			信頼幅は、Excelで計算			
	Excelの線グラフによる作図。図1.17左と同じ⇒					



estimate ステートメントによる推定値 (表 1.34 に対応)

パラメータ	推定値	標準誤差	t 値	Pr > t	95% 信頼区間	
A1:(B1-B2)	0.0358	0.4072	0.09	0.9312	L95%	U95%
A2:(B1-B2)	0.5476	0.3974	1.38	0.1898	-0.84	0.91
A1	30.1352	0.2329	129.37	<.0001	-0.30	1.40
A2	31.7785	0.1945	163.43	<.0001	29.64	30.63
B1	31.1027	0.2059	151.05	<.0001	31.36	32.20
B2	30.8110	0.2004	153.76	<.0001	30.66	31.54
				$t_{0.05}(20-6)=$	2.1448	30.38
						31.24

estimate ステートメントでの各種の推定は、SAS/GLM プロシジャでの内部で生成される「デザイン行列」について理解した上で設定する必要がある。

1.5. 除草剤の散布による雑草の成長抑制

奥野ら（1978）、「応用統計ハンドブック，3.7.3 節 乱塊法における共分散分析」に除草剤の効果と比較するための実験データが示されている．表 1.35 に示すように，無散布群と 4 種類の除草剤に対し，2 または 3 濃度を設定した実験であり，延べ 11 群の処理について 3 ブロックで除草剤の散布実験を行ない，所定期日の雑草の重量 y を測定した．雑草重量 y は，初期の雑草の生え方 x の影響を受けるので，被覆度を目測で 10～50%の間で採点して前値 x とした．

表 1.35 除草剤の散布試験データ [奥野ら（1978），表 7.4]

除草剤	散布量 g/3m ²	群	ブロック I		ブロック II		ブロック III	
			x	y	x	y	x	y
無 散 布		1	10	41	30	157	30	168
M.C.P.	100	2	15	56	10	28	15	68
	150	3	15	31	35	129	30	92
C.M.u.	25	4	10	17	20	54	40	123
	50	5	10	12	10	13	50	141
	75	6	10	10	15	20	40	61
P.C.P.	300	7	20	49	15	45	40	115
	600	8	20	49	15	33	20	25
	900	9	10	9	20	43	30	60
24-D	100	10	15	27	20	90	40	123
	150	11	10	17	15	51	20	98
平均			13.2		18.6		32.3	

除草剤の種類は 4 剤で，それぞれ 2～3 用量となっており，対照群として「無散布」が設定されている．ブロック別の被覆度 x の平均は，（13.2%，18.6%，32.3%）となっており，雑草の生育状況でブロック化されて，各ブロック内で全 11 群が割り付けられている．ただし，仔細にみると，ブロック内での前値 x の大きさの逆転が若干あるので，ブロックを無視して除草剤別に前値 x の小さい順位に並べ変えた結果を表 1.36 に示す．

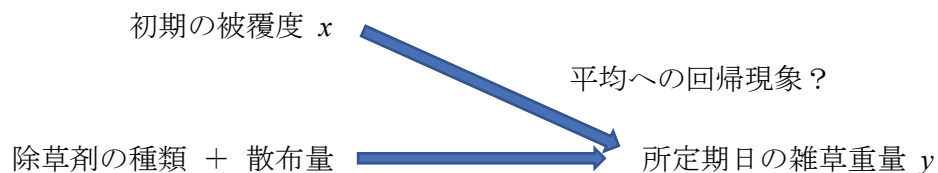


図 1.19 除草剤の雑草の成長抑制に与える影響

表 1.36 右には，除草剤の種類別に散布量を（＋：低，△：中，□：高）と区分けした散布図に，線形近似による回帰直線が重ね書きされている．ほとんどの回帰直線は，原点方向に

向かっており、平行でないことは明らかである。また、第1.3節で取り上げた交互作用を含む共分散分析ではなく、切片を共通とする回帰直線のあてはめ、あるいは、原点を共通の「切片」とする回帰直線のあてはめが妥当である。

所定期日の雑草重量 y (後値とする) は、初期の雑草の被覆度 x (前値とする) の影響が大きいので、前値 x の被覆度 (15%, 30%) とした場合に推定される回帰直線の推定値とその95%信頼区間により、各除草剤の用量反応関係の形状を概観することにする。

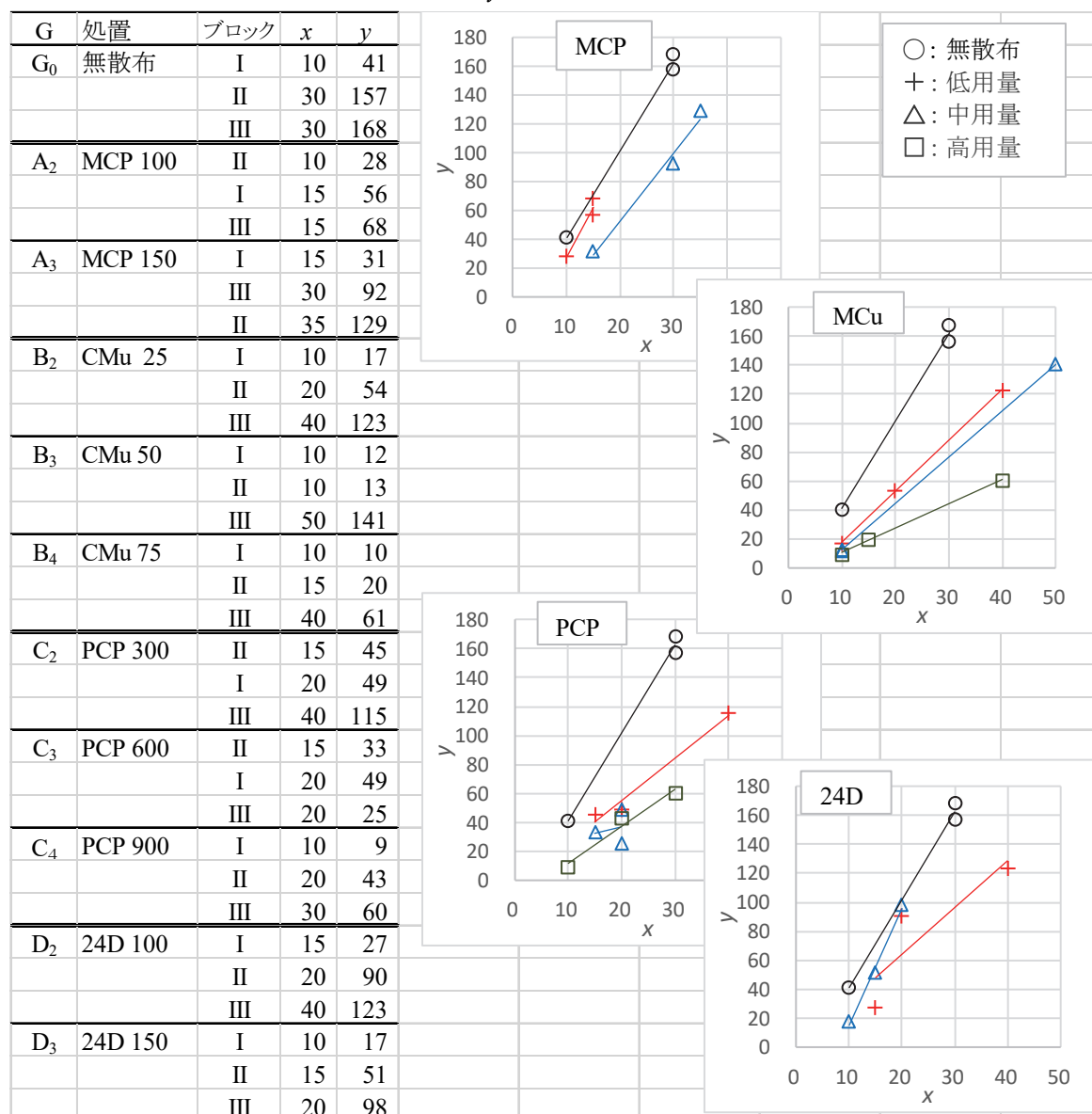
表 1.36 除草剤別の前値 x に対する後値 y の関連

図 1.20 除草剤別の反応関係

MCP 剤に対する共通の切片を持つ回帰直線

無処置と MCP 剤の 2 用量を質的変数として、前値を共変量とし、切片 β_0 を共通とする 3 本の回帰式を

$$\left. \begin{array}{ll} A_1 \text{ 無散布:} & y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \\ A_2 \text{ MCP100:} & y_i = \beta_0 + \beta_2 x_i + \varepsilon_i \\ A_3 \text{ MCP150:} & y_i = \beta_0 + \beta_3 x_i + \varepsilon_i \end{array} \right\} \quad (1.53)$$

とする。これらの式を表 1.3 で例示した (1, 1) 標示型ダミー変数を用いて統合すると、

$$y_i = \beta_0 + \beta_1 a_1 x_i + \beta_2 a_2 x_i + \beta_3 a_3 x_i + \varepsilon_i \quad (1.54)$$

のように一つの式にまとめることができる。表 1.37 に示すように、(1, 1) 標示型ダミー変数を (a_1, a_2, a_3) とし、解析に用いるデザイン行列 X は、ダミー変数 (a_1, a_2, a_3) と前値 x との積とする。

表 1.37 に示すのは、Excel の回帰分析を用いて共通の切片を持つ回帰直線をあてはめた結果である。共通の切片 $\hat{\beta}_0 = -29.7171$ は、 t 検定の結果 $p = 0.0213$ であり、原点よりも統計的に下回っている。切片がマイナスであることは現実には解釈できない推定値であり、共通の切片の推定値に対し、意味付けすることには無理がある。推定された回帰パラメータ（係数）から、前値 $x = 15\%$ における各群の推定値は、

$$A_1 : \hat{y}_{x=15}^{(A_1)} = -29.7271 + 6.4426 \times 15 = 66.9116$$

$$A_2 : \hat{y}_{x=15}^{(A_2)} = -29.7271 + 6.0529 \times 15 = 61.0661$$

$$A_3 : \hat{y}_{x=15}^{(A_3)} = -29.7271 + 4.3056 \times 15 = 34.8570$$

表 1.37 共通の切片を持つ回帰直線のあてはめ

					デザイン行列 X					分散分析表(共通切片)「定数に 0 を使用」off					
A	処置	a_1	a_2	a_3	x	x_0	a_1x	a_2x	a_3x	y		自由度	変動	分散	分散比
A ₁	無散布 (G ₀)	1	0	0	10	1	10	0	0	41	回帰	3	22967.47	7655.82	120.09
		1	0	0	30	1	30	0	0	157	残差	5	318.75	63.75	$= \sigma^2$
		1	0	0	30	1	30	0	0	168	合計	8	23286.22		
A ₂	MCP 100	0	1	0	10	1	0	10	0	28			$t_{0.05}(9-4)=$	2.5706	
		0	1	0	15	1	0	15	0	56		係数	標準誤差	t	P - 値
		0	1	0	15	1	0	15	0	68	β^{\wedge}_0 切片 x_0	-29.7271	8.9913	-3.3062	0.0213
A ₃	MCP 150	0	0	1	15	1	0	0	15	31	β^{\wedge}_1 a_1x	6.4426	0.3785	17.0200	0.0000
		0	0	1	30	1	0	0	30	92	β^{\wedge}_2 a_2x	6.0529	0.7372	8.2103	0.0004
		0	0	1	35	1	0	0	35	129	β^{\wedge}_3 a_3x	4.3056	0.3476	12.3871	0.0001
											パラメータの共分散行列 $\Sigma(\beta)=(X^T X)^{-1}\sigma^2$				
											β^{\wedge}_0	80.8432	-2.9784	-5.8795	-2.7521
											β^{\wedge}_1	-2.9784	0.1433	0.2166	0.1014
											β^{\wedge}_2	-5.8795	0.2166	0.5435	0.2002
											β^{\wedge}_3	-2.7521	0.1014	0.2002	0.1208
											x_0	a_1x	a_2x	a_3x	

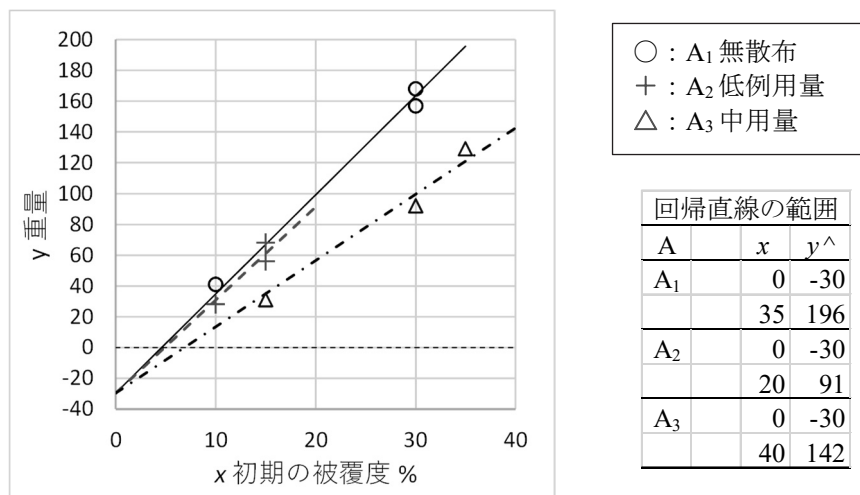


図 1.21 MCP 剤に対する切片を共通とする回帰直線のあてはめ

と推定される. それぞれの 95%信頼区間を求めるために, パラメータの共分散行列 $\Sigma(\hat{\beta})$ が, 表 1.37 右下段に計算されている. 無散布 A_1 に対する推定値の分散 $Var(\hat{y}_{x=15}^{(A_1)})$ は, 合成分散の一般式により

$$\begin{aligned}
 Var(\hat{y}_{x=15}^{(A_1)}) &= Var(\hat{\beta}_0 + \hat{\beta}_1 \times 15) \\
 &= Var(\hat{\beta}_0) + 2 \times Cov(\hat{\beta}_0, \hat{\beta}_1) \times 15 + Var(\hat{\beta}_1) \times 15^2 \\
 &= 80.8432 + 2 \times (-2.9784) \times 15 + 0.1433 \times 15^2 = 23.7291
 \end{aligned}$$

として求めることができる.

表 1.38 に示すのは, $x=15$ の場合も含め, 前値 $x=30$ における各種の推定値, 無散布 A_1 からの減少量に対する推定値, 分散, 95%信頼区間を計算した結果を示す. 推定値および分散の計算には, これまでと同様に Excel の行列関数を用いた 2 次形式で求めている.

表 1.38 共通の切片を持つ回帰直線における各種の推定

x	線形和			l_0	l_1	l_2	l_3	推定値	分散	幅	95%信頼区間	
	A	L	用量	x_0	a_1x	a_2x	a_3x	L^{\wedge}	$Var(L^{\wedge})$	$t_{0.05} \times SE$	$L_{95\%}$	$U_{95\%}$
15	A ₁	$L^{(1)}$	0	1	15	0	0	66.91	23.73	12.52	54.39	79.43
	A ₂	$L^{(2)}$	100	1	0	15	0	61.07	26.75	13.29	47.77	74.36
	A ₃	$L^{(3)}$	200	1	0	0	15	34.86	25.46	12.97	21.89	47.83
30	A ₁	$L^{(4)}$	0	1	30	0	0	163.55	31.09	14.33	149.22	177.88
	A ₂	$L^{(5)}$	100	1	0	30	0	151.86	217.23	37.89	113.97	189.75
	A ₃	$L^{(6)}$	200	1	0	0	30	99.44	24.45	12.71	86.73	112.15
15	A ₁ -A ₁	$L^{(7)}$	0	0	0	0	0	0.00	0.00	0.00	0.00	0.00
	A ₂ -A ₁	$L^{(8)}$	100	0	-15	15	0	-5.85	57.05	19.42	-25.26	13.57
	A ₃ -A ₁	$L^{(9)}$	200	0	-15	0	15	-32.05	13.80	9.55	-41.60	-22.51
30	A ₁ -A ₁	$L^{(10)}$	0	0	0	0	0	0.00	0.00	0.00	0.00	0.00
	A ₂ -A ₁	$L^{(11)}$	100	0	-30	30	0	-11.69	228.21	38.83	-50.52	27.14
	A ₃ -A ₁	$L^{(12)}$	200	0	-30	0	30	-64.11	55.18	19.10	-83.20	-45.01
								T.inv.2T(0.05,9-4)=	2.5706			

図 1.22 に表 1.38 で推定された切片を共通とする回帰直線の前値 $x=15$ および $x=30$ における回帰直線の推定値と 95%信頼区間を Excel の折れ線グラフで作図した結果を示す。無散布 A_1 群に対し、 A_2 MCP 100 g/3m² 群は、有意な差ではなく、 A_2 : MCP 200 g/3m² 群では、無散布群に対し 40%~50%程度の除草効果があると推定される。

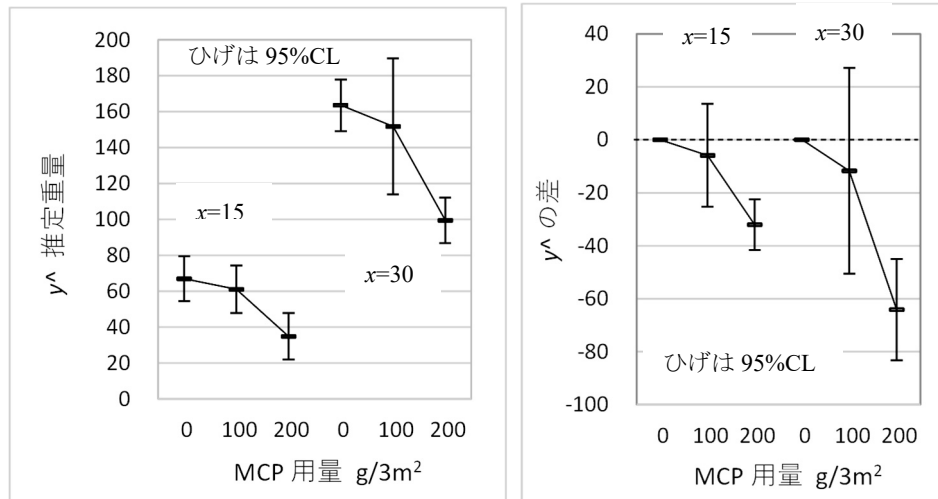


図 1.22 共通の切片を持つ回帰直線における各種の推定値と 95%信頼区間

PCP 剤に対する原点を共通とする回帰直線

PCP 剤については、原点を共通とする 4 本の回帰直線をあてはめる。原点を通る回帰式の場合は、切片がゼロなので、パラメータとして設定せず、それぞれ別々の傾きをもつ 4 本の回帰直線の同時あてはめとなる。

$$\left. \begin{array}{l} C_1 \text{ 無処置: } y_i = \beta_1 x_i + \varepsilon_i \\ C_2 \text{ PCP 300: } y_i = \beta_2 x_i + \varepsilon_i \\ C_3 \text{ PCP 600: } y_i = \beta_3 x_i + \varepsilon_i \\ C_4 \text{ PCP 900: } y_i = \beta_4 x_i + \varepsilon_i \end{array} \right\} \quad (1.55)$$

(1, 1) 標示型ダミー変数 (c_1, c_2, c_3, c_4) を用いると式 (1.55) は、

$$y_i = \beta_1 c_1 x_i + \beta_2 c_2 x_i + \beta_3 c_3 x_i + \beta_4 c_4 x_i + \varepsilon_i \quad (1.56)$$

となり、一つの式でまとめて表すことができる。表 1.39 に示すように、共通の切片がない傾きだけの ($c_1 x_i, c_2 x_i, c_3 x_i, c_4 x_i$) をデザイン行列 \mathbf{X} としている。Excel の回帰分析でも切片がないモデルの設定は、「定数に 0 を使用」設定を ☒ オンとすることに可能となる。

原点を通る回帰直線のデザイン行列 \mathbf{X} は、切片がないので積和行列 ($\mathbf{X}^T \mathbf{X}$) は、対角行列となり、その逆行列もまた対角行列となるので、誤差分散の推定値 $\hat{\sigma}^2 = 91.15$ を掛けたパラメータの共分散行列も表 1.39 右下に示すように対角行列となり、パラメータ間の共分散は、全て 0 である。

表 1.39 原点を共通とする回帰直線のあてはめ

		デザイン行列 X										分散分析表「定数に0を使用」を on					
C	処置	x	c_1	c_2	c_3	c_4	c_1x	c_2x	c_3x	c_4x	y		自由度	変動	分散	分散比	
C ₁	無散布 (G ₀)	10	1	0	0	0	10	0	0	0	41		回帰	4	81120.8	20280.2	222.50
		30	1	0	0	0	30	0	0	0	157		残差	8	729.2	91.15	$=\sigma^2$
		30	1	0	0	0	30	0	0	0	168		合計	12	81850.0		
C ₂	PCP 300	15	0	1	0	0	0	15	0	0	45			$t_{0.05}(12-4)=$ 2.3060			
		20	0	1	0	0	0	20	0	0	49		係数	標準誤差	t	P -値	
		40	0	1	0	0	0	40	0	0	115		切片	0.0000	#N/A	#N/A	#N/A
C ₃	PCP 600	15	0	0	1	0	0	0	15	0	33	β^{\wedge}_1	c_1x	5.3474	0.2190	24.4143	0.0000
		20	0	0	1	0	0	0	20	0	49	β^{\wedge}_2	c_2x	2.8112	0.2024	13.8896	0.0000
		20	0	0	1	0	0	0	20	0	25	β^{\wedge}_3	c_3x	1.9268	0.2982	6.4615	0.0002
C ₄	PCP 900	10	0	0	0	1	0	0	0	10	9	β^{\wedge}_4	c_4x	1.9643	0.2552	7.6983	0.0001
		20	0	0	0	1	0	0	0	20	43						
		30	0	0	0	1	0	0	0	30	60						
												パラメータの共分散行列 $\Sigma(\beta^{\wedge})=(X^T X)^{-1}\sigma^2$					
												β^{\wedge}_1	0.0480	0.0000	0.0000	0.0000	
												β^{\wedge}_2	0.0000	0.0410	0.0000	0.0000	
												β^{\wedge}_3	0.0000	0.0000	0.0889	0.0000	
												β^{\wedge}_4	0.0000	0.0000	0.0000	0.0651	
												c_1x	c_2x	c_3x	c_4x		

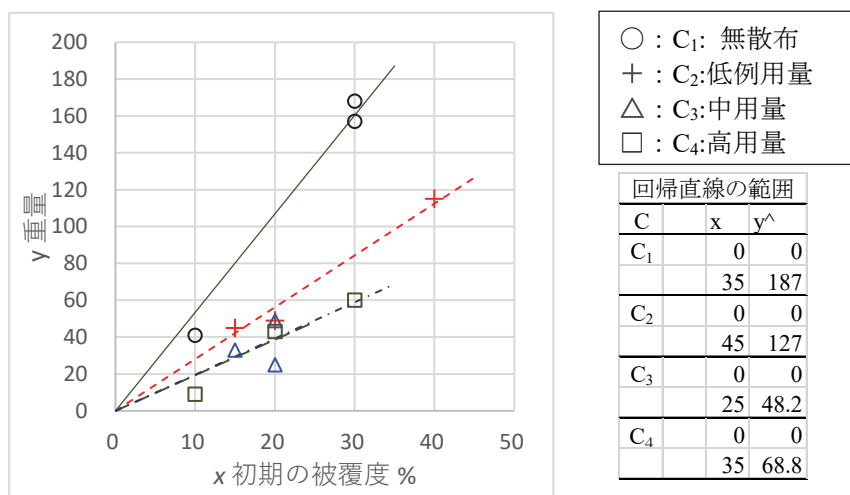


図 1.23 原点を共通とする回帰直線のあてはめ

表 1.40 に示すように、前値が $x=15$ および $x=30$ 単位の場合の各群の推定値とその 95% 信頼区間、無散布群との差についての推定値とその 95% 信頼区間の計算結果を示す。これまでと同様に行列計算を使っているが、共分散が全て 0 なので、 $\boldsymbol{l}^{(1)}=[15 \ 0 \ 0 \ 0]$ の推定値と分散は、

$$\begin{aligned}
 \hat{y}_{x=15}^{(C_1)} &= \boldsymbol{l}^{(1)} \hat{\boldsymbol{\beta}} \\
 &= 15 \hat{\beta}_1 \\
 &= 15 \times 5.3474 = 80.21
 \end{aligned} \tag{1.57}$$

$$\begin{aligned}
 \text{Var}(\hat{y}_{x=15}^{(C_1)}) &= \text{Var}(15 \hat{\beta}_1) \\
 &= 15^2 \times 0.0480 = 10.79
 \end{aligned} \tag{1.58}$$

として計算できる。

無散布からの差の推定値のためのベクトル $\mathbf{l}^{(10)} = [-15 \ 15 \ 0 \ 0]$ は、すでに計算されている

$$L^{(1)} = \hat{y}_{x=15}^{(C_1)} = 80.21, \quad L^{(2)} = \hat{y}_{x=15}^{(C_2)} = 42.17 \quad (1.59)$$

から

$$L^{(2)} - L^{(1)} = \hat{y}_{x=15}^{(C_2-C_1)} = 42.17 - 80.21 = -38.04 \quad (1.60)$$

として計算され、その分散は、分散の加法性により

$$\begin{aligned} \text{Var}(\hat{y}_{x=15}^{(C_2-C_1)}) &= \text{Var}(\hat{y}_{x=15}^{(C_2)}) + \text{Var}(\hat{y}_{x=15}^{(C_1)}) \\ &= 9.22 + 10.79 = 20.01 \end{aligned}$$

によっても計算できる．図 1.24 に表 1.40 で推定された結果を用いて，Excel の線グラフで作図した結果を示す．

表 1.40 原点を共通とする回帰直線における各種の推定

				l_1	l_2	l_3	l_4	推定値	分散	幅	95%信頼区間	
x	C	L	用量	c_1x	c_2x	c_3x	c_4x	y^\wedge	$\text{Var}(y^\wedge)$	$t_{0.05} \times \text{SE}$	$L_{95\%}$	$U_{95\%}$
15	C_1	$L^{(1)}$	0	15	0	0	0	80.21	10.79	7.58	72.63	87.79
	C_2	$L^{(2)}$	300	0	15	0	0	42.17	9.22	7.00	35.17	49.17
	C_3	$L^{(3)}$	600	0	0	15	0	28.90	20.01	10.31	18.59	39.22
	C_4	$L^{(4)}$	900	0	0	0	15	29.46	14.65	8.83	20.64	38.29
30	C_1	$L^{(5)}$	0	30	0	0	0	160.42	43.18	15.15	145.27	175.57
	C_2	$L^{(6)}$	300	0	30	0	0	84.34	36.87	14.00	70.34	98.34
	C_3	$L^{(7)}$	600	0	0	30	0	57.80	80.03	20.63	37.18	78.43
	C_4	$L^{(8)}$	900	0	0	0	30	58.93	58.60	17.65	41.28	76.58
15	C_1-C_1	$L^{(9)}$	0	0	0	0	0	0.00	0.00	0.00	0.00	0.00
	C_2-C_1	$L^{(10)}$	300	-15	15	0	0	-38.04	20.01	10.32	-48.36	-27.73
	C_3-C_1	$L^{(11)}$	600	-15	0	15	0	-51.31	30.80	12.80	-64.11	-38.51
	C_4-C_1	$L^{(12)}$	900	-15	0	0	15	-50.75	25.44	11.63	-62.38	-39.11
30	C_1-C_1	$L^{(13)}$	0	0	0	0	0	0.00	0.00	0.00	0.00	0.00
	C_2-C_1	$L^{(14)}$	300	-30	30	0	0	-76.08	80.04	20.63	-96.72	-55.45
	C_3-C_1	$L^{(15)}$	600	-30	0	30	0	-102.62	123.21	25.60	-128.21	-77.02
	C_4-C_1	$L^{(16)}$	900	-30	0	0	30	-101.49	101.77	23.26	-124.76	-78.23
T.inv.2T (0.05,12-4)=										2.3060		

Excel の回帰分析で、原点を通る（「定数に 0 を使用」をオン）回帰式をあてはめた場合の分散分析表の結果が怪しいとの指摘が、岩崎（2006）、「単回帰分析，第 4-3 節 原点を通る回帰式（Excel）」にあるので、同じデータを用いて Excel 2016 で確認したところ、SPSS および JMP の出力に一致しているので、このバグは解消されている．また、表 1.39 の Excel の回帰分析での分散分析表は、後述の R での検証で、一致することが確認されている．この指摘に限らず、Excel での計算自体に操作ミス、計算式の設定ミスなどがあっても気が付きにくいなどの脆弱性があることに注意する必要がある．

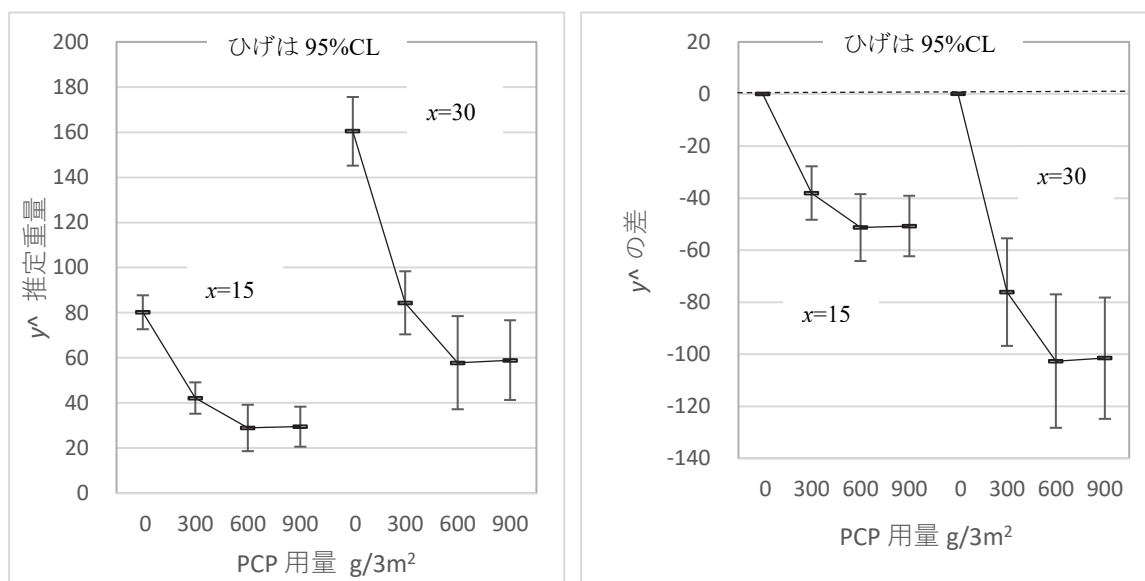


図 1.24 原点を共通とする回帰直線における各種の推定と 95%信頼区間

原点を通る回帰直線の推定は、Excel で「定数に 0 を使用」をオンとする対応であり、SAS の GLM プロシジャで Model ステートメントの「nointercept」オプションを使い、JMP で「切片なし」をオンとし、R の `lm()` 関数で `lm` の回帰式の末尾に「+0」を加えることにより実現できる。

R による検証(原点共通の場合)

R の `lm()` 関数で原点を通る回帰直線は、モデル式の最後に「+0」を加えることにより切片無のモデルになる。実際に表 1.39 で示したデザイン行列を用いて計算した結果を次に示す。表 1.39 の Excel でのパラメータの推定値は、完全に一致し、「Residual standard error: 9.547」の平方は、 $9.547^2 = 91.15$ と残差の平均平方(分散)に一致する。

Call:

```
lm(formula = y ~ c1x + c2x + c3x + c4x + 0, data = Dataset)
```

Residuals:

Min	1Q	Median	3Q	Max
-13.537	-8.079	1.811	3.810	10.463

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
c1x	5.3474	0.2190	24.414	0.00000000846 ***
c2x	2.8112	0.2024	13.890	0.00000069849 ***
c3x	1.9268	0.2982	6.461	0.000196 ***
c4x	1.9643	0.2552	7.698	0.00005752696 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.547 on 8 degrees of freedom

Multiple R-squared: 0.9911, Adjusted R-squared: 0.9866

F-statistic: 222.5 on 4 and 8 DF, p-value: 0.00000003127

第1章 文献索引

岩崎(2006) - 単回帰分析, 第4-3節 原点を通る回帰式(Excel)	53
奥野・久米・芳賀・吉澤(1981) - 多変量解析法《改訂版》, 第II部, 第7章 層別因子を含む回帰分析	19
奥野ら(1978) - 応用統計ハンドブック, 3.7.3節 乱塊法における共分散分析	47
高橋(2021) - 最尤法によるポアソン回帰分析入門, 第13.2節 交互作用を考慮した共分散分析	19
芳賀(2014) - 医薬品開発のための統計解析, 第2部 実験計画法 改訂版, 第4章 共分散分析	7
フライス(2004) - 臨床試験のデザインと解析, 第7章 共分散分析と変化量の検討	30
Littellら(2002) - SAS for Linear Models 4th ed., Chapter 7 Analysis of Covariance	38









第1章 索引

あ Rの lm () 関数 - 原点共通	54	か 回帰直線の差 - 非平行	34
1元配置型 - 共分散分析	38	回帰分析 - デザイン行列	7
(1, 0)型 - ダミー変数	10, 39	- モデルのあてはめ	20
一般線形モデル - SAS	45	回収液の濃度 - 洗浄用水の温度	19
- GLMプロシジャ	45	カキの成長 - 設置場所	39
うつ症状 - Hamilton の評価尺度	30	- 発電プラント	38
HTML出力を - Excel に	45	各種 - ダミー変数	10
Excel - 折れ線グラフ	18, 28	各種の推定値 - estimateステートメント	45
- 行列計算	7	拡張 - 合成分散の一般式	26
- 計算式の設定ミス	53	季節 - 洗浄用水の温度	19
- 線形近似	8	逆行列 - Minverse ()	15
- 操作ミス	53	- 対角行列	51
- 層別散布図	8	95%信頼区間 - 回帰直線	27, 42
- 予測プロファイル	26	- 推定値	14
Excel に - HTML出力を	45	- 推定値の差	17
Excel の回帰分析 - 定数に0を使用	9	- ソルバー	36
- 結果が怪しい	53	- 直線間の差	35
estimateステートメント - 各種の推定値	45, 46	- 2次式の解の公式	36
F.dist.RT () - F分布の右側確率	13	共通の傾き - 3本の回帰直線	11
F分布の右側確率 - F.dist.RT ()	13	共通の切片 - MCP剤	49
Minverse () - 逆行列	15	共分散行列 - 誤差分散の推定値	40
MCP剤 - 共通の切片	49	- 対角要素	15, 41
Mmult () - 関 / 行列	15	- デザイン行列 X	40
lsmeansステートメント - 最小2乗平均	45, 46	- パラメータ	14, 25, 40
off - 定数に 0 を使用	40, 49	共分散分析 - 1元配置型	38
オフ(デフォルト) - 定数に 0 を使用	32	- 交互作用	30
折れ線グラフ - Excel	18, 28	- 芳賀(2014)	7
on - 定数に 0 を使用	24	- 変化量の検討	30
オン - 定数に 0 を使用	9, 51	- 乱塊法	47
OnDemand SAS - SAS	44	共変量 - 前値	30, 38
- 無償で継続的に提供	44	行 / 列の入れ替え - 転置/行列	15
か 回帰の平方和の差 - 交互作用の平方和	33	行列計算 - Excel	7
回帰パラメータの推定 - 行列計算	15	- 回帰パラメータの推定	15
回帰式 - 交互作用	32	グラフ・ビルダー - JMP	19
回帰直線 - 95%信頼区間	27, 42	- 層別散布図	20
- 原点	9	計画行列 - デザイン行列とは何か	18
- 原点を共通	48, 52	計算式の設定ミス - Excel	53
- 散布図	28	結果が怪しい - Excelの回帰分析	53
- 切片を共通	48	原点 - 回帰直線	9
- 非平行	31	原点を共通 - 回帰直線	48, 52
- 別々	8	- PCP剤	51

か	原点共通 - R の <code>lm()</code> 関数	54	さ	関 / 行列 - <code>Mmult()</code>	15
	交互作用 - 回帰式	32		積和行列 - 対角行列	51
	- 共分散分析	30		設置場所 - カキの成長	39
	- 残差の平方和	14		切片が異なる - 3本の回帰直線	12
	- 自由度	13		切片を共通 - 回帰直線	48
	- ダミー変数	24		(0, 1)型 - ダミー変数	10
	- 分散分析表	21, 33		線形モデル - 質的変数	24
	- 平方和	13		線形近似 - Excel	8
	交互作用の平方和 - 回帰の平方和の差	33		線形式 - 合成分散の一般式	43
	- 残差平方和の差	33		線形和 L - 推定値	26, 44
	交互作用モデル - 主効果モデル	31		- 2水準間の差	29
	- デザイン行列 X	32		- 分散	26, 41
	合成分散の一般式 - 分散/差の	34		洗浄用水の温度 - 回収液の濃度	19
	- 拡張	26		- 季節	19
	- 推定値の分散	14		前値 - 共変量	30, 38
	- 線形式	43		- 2元配置	38
	誤差分散の推定値 - 共分散行列	40		前値 x - 平均への回帰	30
さ	最小2乗法などパラメータの推定 - デザイン行列	18		操作ミス - Excel	53
	最小2乗平均 - <code>lsmeans</code> ステートメント	45, 46		層別散布図 - Excel	8
	- 推定平均	43		- グラフ・ビルダー	20
	SAS - 一般線形モデル	45	た	ソルバー - 95%信頼区間	36
	- OnDemand SAS	44		対角行列 - 逆行列	51
	雑草重量 - 被覆度	47		- 積和行列	51
	残差の平方和 - 交互作用	14		対角要素 - 共分散行列	15, 41
	残差平方和の差 - 交互作用の平方和	33		- 分散	15
	散布図 - 回帰直線	28		対照群 - 2元配置	38
	3本の回帰直線 - 共通の傾き	11		(1, -1)対比型 - ダミー変数	10, 23
	- 切片が異なる	12		対比 - 水準間の差	22
	- デザイン行列 X	10		対比の設定 - JMP	23
	GLMプロシジャ - 一般線形モデル	45		- パラメータ関数	23
	質的変数 - 線形モデル	24		対比型ダミー変数 - パラメータ推定	25
	JMP - グラフ・ビルダー	19		タイプ1平方和 - 逐次平方和	37
	- 対比の設定	23		多項式の中心化 - モデルの設定	21
	- 分散分析表	37		ダミー変数 - (1, 0)型	10, 39
	- モデルのあてはめ	20		- 各種	10
	- 予測プロファイル	21		- 交互作用	24
	自由度 - 交互作用	13		- (0, 1)型	10
	主効果モデル - 交互作用モデル	31		- (1, -1)対比型	10, 23
	取水路・放水路 - 水深	38		- (1, 1)標示型	10, 49, 51
	上限と下限 - 平均への回帰	31		単回帰分析 - デザイン行列 X	9
	除草剤 - 成長抑制	47		逐次平方和 - タイプ1平方和	37
	- 用量反応関係	48		直線間の差 - 95%信頼区間	35
	水深 - 取水路・放水路	38		定数に 0 を使用 - Excel の回帰分析	9
	水準間の差 - 対比	22		- off	40, 49
	- 予測プロファイル	28		- オフ(デフォルト)	32
	推定値 - <code>estimate</code> ステートメント	46		- on	24
	- 95%信頼区間	14		- オン	9, 51
	- 線形和 L	26, 44		手計算時代 - 偏差平方和	7
	推定値の分散 - 合成分散の一般式	14		Design Matrix - デザイン行列とは何か	18
	推定値の差 - 95%信頼区間	17		デザイン行列 - 回帰分析	7
	推定平均 - 最小2乗平均	43		- 最小2乗法などパラメータの推定	18
	成長抑制 - 除草剤	47		- 統計ソフト	7

た	デザイン行列 X - 共分散行列	40	は	非平行 - 回帰直線	31
	- 交互作用モデル	32		- 回帰直線の差	34
	- 3本の回帰直線	10		(1, 1) 標示型 - ダミー変数	10, 49, 51
	- 単回帰分析	9		分散 - 線形和 L	26, 41
	デザイン行列 X - 変幻自在	18		- 対角要素	15
	デザイン行列とは何か - 計画行列	18		分散/差の - 合成分散の一般公式	34
	- Design Matrix	18		分散分析表 - 交互作用	21, 33
	転置 / 行列 - 行/列の入れ替え	15		- JMP	37
	- Transpose ()	15		平均への回帰 - 上限と下限	31
	統計ソフト - デザイン行列	7		- 前値 x	30
	Transpose () - 転置/行列	15		平方和 - 交互作用	13
な	2元配置 - 前値	38		別々 - 回帰直線	8
	- 対照群	38		変化量の検討 - 共分散分析	30
	2次式の解の公式 - 95%信頼区間	36		変幻自在 - デザイン行列 X	18
	2水準間の差 - 線形和 L	29		偏差平方和 - 手計算時代	7
	年収 - 比較調査	7	ま	無償で継続的 - OnDemand SAS	44
は	芳賀 (2014) - 共分散分析	7		モデルのあてはめ - 回帰分析	20
	発電プラント - カキの成長	38		- JMP	20
	Hamilton の評価尺度 - うつ症状	30		モデルの設定 - 多項式の中心化	21
	パラメータ - 共分散行列	14, 25, 50	や	用量反応関係 - 除草剤	48
	パラメータ関数 - 対比の設定	23		予測プロファイル - Excel	26
	パラメータ推定 - 対比型ダミー変数	25		- JMP	21
	PCP剤 - 原点を共通	51		- 水準間の差	28
	比較調査 - 年収	7	ら	乱塊法 - 共分散分析	47
	被覆度 - 雑草重量	47			

第1章 解析用ファイル一覧

サイズ	名前 ^		種類
	88 KB	第1章_01_会社と年齢と給料	Microsoft Excel ワークシート
	9 KB	第1章_02_奥野_層別回帰	JMP Data Table
	66 KB	第1章_02_奥野_層別回帰	Microsoft Excel ワークシート
	8 KB	第1章_03_フライズ_抗うつ薬_交互作用	JMP Data Table
	91 KB	第1章_03_フライズ_抗うつ薬_交互作用	Microsoft Excel ワークシート
	87 KB	第1章_04_牡蠣	Microsoft Excel ワークシート
	1 KB	第1章_04_牡蠣sas	テキストドキュメント
	101 KB	第1章_05_除草剤	Microsoft Excel ワークシート

空白ページ

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 1 章 層別因子を含む各種の回帰分析の実際

BioStat 研究所(株)
〒105-0014 東京都 港区 芝 1-12-3 の1005
2023 年 12 月 17 日 高橋 行雄
takahashi.stat@nifty.com , FAX : 03-342-8035