

第 12 回 続高橋セミナー
層別因子を含む探索的な回帰分析入門
2024 年 2 月 6 日

第 6 章 伝統的な共分散分析からの脱却

共分散分析は、歴史的に難解な解析法として認識されてきた。質的変数に対する伝統的な平方和の分解による解析方法は、量的変数である共変量を含めた共分散分析に適用できない。そのために、偏差平方和を用いた単回帰分析を重層的に適用しつつ、交互作用を含む分散分析表を作成し、交互作用のないことを確認して交互作用項を除き、平行な直線のあてはめに帰着し、質的因子の水準間の比較を行う方法が定式化されてきた。他方、質的変数をダミー変数に置き換えて線形モデルとした共分散分析の解析法も知られてはいるのだが、統計ソフトを使うことを前提とするために、解析方法がブラック・ボックス的となり、理解の妨げになっている。そこで、先人たちの創意工夫に満ちた伝統的な解析方法を Excel の力を借りつつ忠実に再現する。さらに、Excel の回帰分析により、質的変数をダミー変数に置き換え、線形モデルによる共分散分析の解析法を導入すると共に、無償で継続的に提供されている Ondemand SAS の GLM プロシジャを用いた共分散分析の新たな活用方法についても示す。

第 6 章 目 次

6.	伝統的な共分散分析からの脱却-----	195
6.1.	伝統的な共分散分析の解析手順-----	195
	季節ごとの回帰直線, 重層的な回帰分析の適用, 原点を通る回帰直線,	
	交互作用を含む分散分析表の作成, 平行な回帰直線のあてはめ,	
	平行な直線間の差の 95%信頼区間, 共分散分析は難解なのか?	
	次ページ続く	

6.2. Excel の回帰分析を活用した共分散分析 -----	204
(0, 1)型ダミー変数を用いたデザイン行列 X , 逐次平方和による分散分析表, 共通の傾きを持つ場合の各種の推定, 水準間の差の予測プロファイル, 洗浄水の温度 x に関する予測プロファイル	
6.3. SAS の GLM プログラムによる共分散分析 -----	211
SAS でのテキストデータの取り込み, GLM プロシジャによる交互作用 $x \times A$ の 評価, GLM プロシジャが内部で生成するデザイン行列 X , 交互作用を 含まない解析, Lsmeans ステートメントによる最小 2 乗平均の推定, Estimate ステートメントによる最小 2 乗平均の差の推定	
文献索引, 索引, 解析ファイル一覧 -----	(219)

第 12 回 続・高橋セミナー

層別因子を含む探索的な回帰分析入門

目 次 (全章)

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

6. 伝統的な共分散分析からの脱却

共分散分析は、歴史的に難解な解析法として認識されてきた。質的変数に対する伝統的な平方和の分解による解析方法は、量的変数である共変量を含めた共分散分析に適用できない。そのために、偏差平方和を用いた単回帰分析を重層的に適用しつつ、交互作用を含む分散分析表を作成し、交互作用のないことを確認して交互作用項を除き、平行な直線のあてはめに帰着し、質的因子の水準間の比較を行う方法が定式化されてきた。他方、質的変数をダミー変数に置き換えて線形モデルとした共分散分析の解析法も知られてはいるのだが、統計ソフトを使うことを前提とするために、解析方法がブラック・ボックス的となり、理解の妨げになっている。そこで、先人たちの創意工夫に満ちた伝統的な解析方法を Excel の力を借りつつ忠実に再現する。さらに、Excel の回帰分析により、質的変数をダミー変数に置き換え、線形モデルによる共分散分析の解析法を導入すると共に、無償で継続的に提供されている Ondemand SAS の GLM プロシジャを用いた共分散分析の新たな活用方法についても示す。

6.1. 伝統的な共分散分析の解析手順

第 1.2 節の表 1.12 で取り上げた「季節ごとの洗浄用水の温度 x と回収液の濃度 y 」では、JMP の「モデルのあてはめ」を用いて交互作用を考慮した共分散分析の解析法を示した。さらに、Excel による交互作用を含むデザイン行列 \mathbf{X} を活用した解析法を合わせて示した。取り上げるデータは、奥野ら(1981)「多変量解析法《改訂版》、第 II 部 重回帰分析、第 7 章 層別因子を含む回帰分析」の事例 7.1 である。原著に示されている解析法は、伝統的な共分散分析の解析手順であり、共分散分析を取り上げている成書で共通の解析方法として定着している。ここでは、原著で示されている解析方法を Excel の力を借りつつ忠実に再現することから始める。

季節ごとの回帰直線

表 6.1 に表 1.12 のデータを再掲する。このデータは、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 との関係を調べたが、測定回数が少ないので、長期間にわたって収集されたデータを使わなければならない。その間に季節の変化ばかりでなく、操業条件も変わっているため、データを 4 つ季節に分けた結果が示されている。

表 6.1 季節ごとの洗浄水の温度 x と回収液の濃度 y [表 1.12 再掲]

No.	A ₁ (7月・8月)		A ₂ (9月・10月)		A ₃ (11月・12月)		A ₄ (1月・2月)	
	x	y	x	y	x	y	x	y
1	30	8.0	34	9.1	27	15.0	16	23.4
2	28	10.5	19	19.4	20	21.0	16	28.2
3	30	8.2	22	20.5	14	24.2	20	29.5
4	29	13.0	25	14.2	18	15.3	16	22.2
5	28	10.1	22	11.0	12	17.3	6	40.2
6	17	16.2	25	19.1	18	24.7	16	36.6
7	24	16.1	22	16.0	13	23.9	12	35.5
8	22	13.4	23	11.1	21	20.2	11	42.1
9	30	13.0	28	12.0	24	13.6	9	42.5
10	30	7.3	24	17.1			22	25.8
11	25	14.0	30	13.3				
12			24	8.9				
平均	26.64	11.80	24.83	14.31	18.56	19.47	14.40	32.60
			$n=11+12+9+10=42$		総平均		21.4762	19.1119

図 6.1 に示すのは、得られたデータの関連を概観するために、洗浄水の温度 x と回収液の濃度 y について Excel で作成した層別散布図上に季節別の回帰直線を上書きした結果である。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 x が低い。伝統的な共分散分析の観点からは、季節 A のそれぞれの共変量 x の分布 (平均値) が明らかに異なり、共分散分析の適用の前提を満たしていない。全ての季節で共通となる洗浄水の温度 x が 20 度で比べた場合でも、A₄ (1 月・2 月) の回収液の濃度 y が高めている。そのために、季節に関連する操業条件などの他の因子の影響があると示唆される。

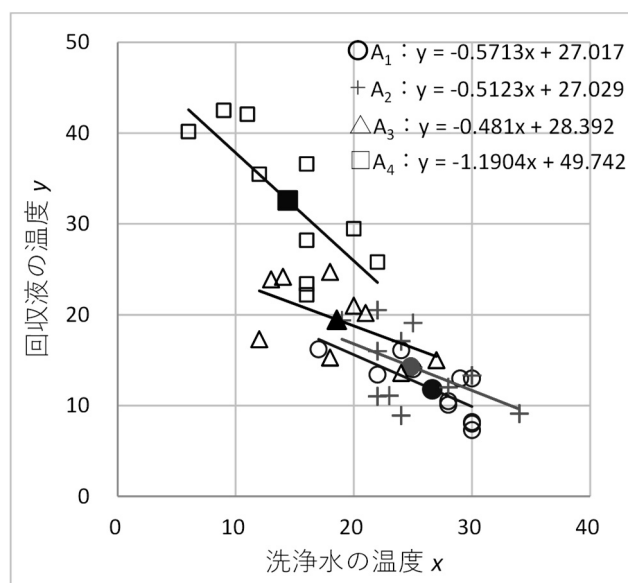


図 6.1 Excel の散布図による季節ごとの 4 本の回帰直線のあてはめ

Excel の散布図で最初に A₁ の (x , y) をプロット・○印に変更・近似直線の追加・平均をプロットする。さらに、A₂, A₃, A₄ を順次「データの選択」で重ね書きし、書式を整え、軸目盛の調整などを行っている。回帰直線は、「近似曲線の追加」の「線形近似」で重ね書きしている。

伝統的な共分散分析は、1 因子実験モデルでの解析に際し、反応 y に影響を与えることがはっきりしている量的変数があるが、実験に際し一定に保つことがどうしてもできない。この量的変数を「共変量」とし、統計的に共変量の影響を除去し、季節 A の水準間の比較をしたい。そのために、共変量の分布が季節 A に対し均一でなければ共変量として扱えないと認識されている。季節 A により洗浄水の温度 x の分布が異なるので共変量としての前提が満たされないことを無視したとしても、更なる共分散分析の適用条件とし、季節 A と洗浄水の温度 x の間に交互作用がないことも確認しなければならない。

重層的な回帰分析の適用

伝統的な共分散分析は、洗浄水の温度 x と季節 A との交互作用の平方和 S_{xxA} を含む分散分析表を作成するために、単回帰分析を重層的に適用する解析方法となっている。

(手順1) 全データに対する回帰分析

表 6.1 に示した季節 A の水準ごとのデータを全て行方向に並び替え、偏差平方和を用いた回帰分析を行なう。まず、 x と y に関する偏差平方和 ($S_{xx}^{(\text{全体})}$, $S_{xy}^{(\text{全体})}$, $S_{yy}^{(\text{全体})}$) を計算する。

$$\left. \begin{aligned} S_{xx}^{(\text{全体})} &= \sum_{i=1}^4 \sum_j^{n_i} (x_{ij} - \bar{x}_{..})^2 = 1788.48 \\ S_{xy}^{(\text{全体})} &= \sum_{i=1}^4 \sum_j^{n_i} (x_{ij} - \bar{x}_{..})(y_{ij} - \bar{y}_{..}) = -2124.24 \\ S_T^{(\text{全体})} &= S_{yy}^{(\text{全体})} = \sum_{i=1}^4 \sum_j^{n_i} (y_{ij} - \bar{y}_{..})^2 = 3654.82 \end{aligned} \right\} \quad (6.1)$$

$S_{xy}^{(\text{全体})}$ の計算は Excel で $\bar{x}_{..} = \text{Average}(x \text{ の範囲})$, $\bar{y}_{..} = \text{Average}(y \text{ の範囲})$ を前もって計算し、SumProduct (x の範囲 $-\bar{x}_{..}$, y の範囲 $-\bar{y}_{..}$) 関数により配列として計算する。 $S_{xx}^{(\text{全体})}$, $S_{yy}^{(\text{全体})}$ は、SumSq() 関数で同様に配列として計算する。これらの結果からパラメータ ($\hat{\beta}_1^{(\text{全体})}$, $\hat{\beta}_0^{(\text{全体})}$) を

$$\left. \begin{aligned} \text{傾き: } \hat{\beta}_1^{(\text{全体})} &= \frac{S_{xy}^{(\text{全体})}}{S_{xx}^{(\text{全体})}} = \frac{-2124.24}{1788.48} = -1.1877 \\ \text{切片: } \hat{\beta}_0^{(\text{全体})} &= \bar{y}_{..} - \hat{\beta}_1^{(\text{全体})} \bar{x}_{..} = 19.1119 - (-1.1877) \times 21.4762 = 44.6200 \end{aligned} \right\} \quad (6.2)$$

として推定する。次いで分散分析に必要な回帰の平方和 $S_R^{(\text{全体})}$ および残差平方和 $S_e^{(\text{全体})}$ を推定する。なお、 $S_e^{(\text{全体})}$ には、季節 A の平方和 S_A および交互作用の平方和 $S_{R \times A}$ が含まれている。

$$\left. \begin{aligned} S_R^{(\text{全体})} &= \frac{(S_{xy}^{(\text{全体})})^2}{S_{xx}^{(\text{全体})}} = \frac{(-2124.24)^2}{1788.48} = 2523.03 \\ S_e^{(\text{全体})} &= S_T^{(\text{全体})} - S_R^{(\text{全体})} = 3654.82 - 2523.03 = 1131.79 \end{aligned} \right\} \quad (6.3)$$

(手順2) 季節 A の水準ごとの回帰分析の併合

表 6.2 に示す計算シート上で、季節 A の水準ごとの偏差平方和 ($S_{xy}^{(i)}$, $S_{xx}^{(i)}$, $S_{yy}^{(i)}$) を計算し、 i に関する合計を

表 6.2 季節ごとの回帰分析の併合

A ₁		A ₂		A ₃		A ₄							
x	y	x	y	x	y	x	y		A ₁	A ₂	A ₃	A ₄	和
30	8.0	34	9.1	27	15.0	16	23.4	n_i	11	12	9	10	42
28	10.5	19	19.4	20	21.0	16	28.2	$\bar{x}_{i\cdot}$	26.64	24.83	18.56	14.40	—
30	8.2	22	20.5	14	24.2	20	29.5	$\bar{y}_{i\cdot}$	11.80	14.31	19.47	32.60	—
29	13.0	25	14.2	18	15.3	16	22.2	$S_{xx}^{(i)}$	178.55	183.67	204.22	216.40	782.83
28	10.1	22	11.0	12	17.3	6	40.2	$S_{xy}^{(i)}$	-102.00	-94.08	-98.23	-257.60	-551.92
17	16.2	25	19.1	18	24.7	16	36.6	$\hat{\beta}_1^{(i)}$	-0.57	-0.51	-0.48	-1.19	—
24	16.1	22	16.0	13	23.9	12	35.5	$\hat{\beta}_0^{(i)}$	27.02	27.03	28.39	49.74	—
22	13.4	23	11.1	21	20.2	11	42.1	$S_{yy}^{(i)}$	100.36	181.85	148.76	538.44	969.41
30	13.0	28	12.0	24	13.6	9	42.5	$S_R^{(i)}$	58.27	48.19	47.25	306.64	460.36
30	7.3	24	17.1			22	25.8	$S_e^{(i)}$	42.09	133.65	101.51	231.80	509.05
25	14.0	30	13.3										
		24	8.9										

$$\left. \begin{aligned} S_{xx}^{(\text{和})} &= \sum_{i=1}^4 S_{xx}^{(i)} = 178.55 + 183.67 + 204.22 + 216.40 = 782.83 \\ S_{xy}^{(\text{和})} &= \sum_{i=1}^4 S_{xy}^{(i)} = (-102.00) + (-94.08) + (-98.23) + (-257.60) = -551.92 \\ S_{yy}^{(\text{和})} &= \sum_{i=1}^4 S_{yy}^{(i)} = 100.36 + 181.85 + 148.76 + 538.44 = 969.41 \end{aligned} \right\} \quad (6.4)$$

として求め ($S_{xy}^{(\text{和})}$, $S_{xx}^{(\text{和})}$, $S_{yy}^{(\text{和})}$) とする. また, 季節 A の水準ごとの回帰係数を

$$\hat{\beta}_1^{(i)} = \frac{S_{xy}^{(i)}}{S_{xx}^{(i)}}, \quad \hat{\beta}_0^{(i)} = \bar{y}_{i\cdot} - \hat{\beta}_1^{(i)} \bar{x}_{i\cdot}. \quad (6.5)$$

で求め, 図 6.1 に示された回帰係数に一致することを確認する. 水準ごとの回帰の平方和 $S_R^{(i)}$ を求め, それらの和を計算する. $S_R^{(i)}$ の和を $S_R^{(\text{和})}$, $S_{yy}^{(i)}$ の和を $S_{yy}^{(\text{和})} = S_T^{(\text{和})}$ とし, それらの差から残差平方和 $S_e^{(\text{和})}$ を求める.

$$\left. \begin{aligned} S_R^{(\text{和})} &= \sum_{i=1}^4 \frac{(S_{xy}^{(i)})^2}{S_{xx}^{(i)}} = 58.27 + 48.19 + 47.25 + 306.64 = 460.36 \\ S_e^{(\text{和})} &= S_T^{(\text{和})} - S_R^{(\text{和})} = 969.41 - 460.36 = 509.05 \end{aligned} \right\} \quad (6.6)$$

(手順 3) 季節 A について共通の傾き $\hat{\beta}_1^{(\text{共通})}$ の推定

手順 2 で求めた $S_{xx}^{(\text{和})} = 782.83$, $S_{xy}^{(\text{和})} = -551.92$ から共通の傾き $\hat{\beta}_1^{(\text{共通})} = -0.7050$ を算出する. 共通の傾きでの回帰の平方和 $S_R^{(\text{共通})} = 389.11$ を推定し, $S_T^{(\text{和})} = S_{yy}^{(\text{和})} = 969.41$ との差から交互作用 $S_{xxA}^{(\text{共通})}$ を含む $S_e^{(\text{共通})} = 580.29$ を算出する.

$$\hat{\beta}_1^{(\text{共通})} = \frac{S_{xy}^{(\text{和})}}{S_{xx}^{(\text{和})}} = \frac{-551.92}{782.83} = -0.7050 \quad (6.7)$$

$$\left. \begin{aligned} S_R^{(\text{共通})} &= \frac{(S_{xy}^{(\text{和})})^2}{S_{xx}^{(\text{和})}} = \frac{(-551.92)^2}{782.83} = 389.11 \\ S_e^{(\text{共通})} &= S_T^{(\text{和})} - S_R^{(\text{共通})} = 969.41 - 389.11 = 580.29 \end{aligned} \right\} \quad (6.8)$$

原点を通る回帰直線

(手順 2 および 3 の別解) 原点を通る回帰分析の適用

手順 2 および手順 3 は、季節 A の水準ごとに共通の傾きを推定するために単回帰分析を重層的に適用する方法であるが、見通しが悪い。そこで、別の方法として、表 6.3 に示すように季節 A の水準ごとの偏差 $x'_{ij} = x_{ij} - \bar{x}_{i\cdot}$, $y'_{ij} = y_{ij} - \bar{y}_{i\cdot}$ を計算し、全データ $n = 42$ に対する原点を通る Excel の回帰分析を適用することにより、共通の傾きを求めることができる。表 6.4 に分散分析表およびパラメータ（係数）の推定結果を示す。

表 6.3 季節 A の水準ごとの偏差の計算

No.	A	x	y	水準ごとの偏差	
				x'	y'
1	A ₁	30	8.0	3.4	-3.8
2	A ₁	28	10.5	1.4	-1.3
:					
11	A ₁	25	14.0	-1.6	2.2
12	A ₂	34	9.1	9.2	-5.2
:					
23	A ₂	24	8.9	-0.8	-5.4
24	A ₃	27	15.0	8.4	-4.5
:					
32	A ₃	24	13.6	5.4	-5.9
33	A ₄	16	23.4	1.6	-9.2
:					
41	A ₄	9	42.5	-5.4	9.9
42	A ₄	22	25.8	7.6	-6.8
				782.83	969.41
				$S_{x'x'}^{(和)}$	$S_{y'y'}^{(和)}$
				-551.92	$S_{x'y'}^{(和)}$

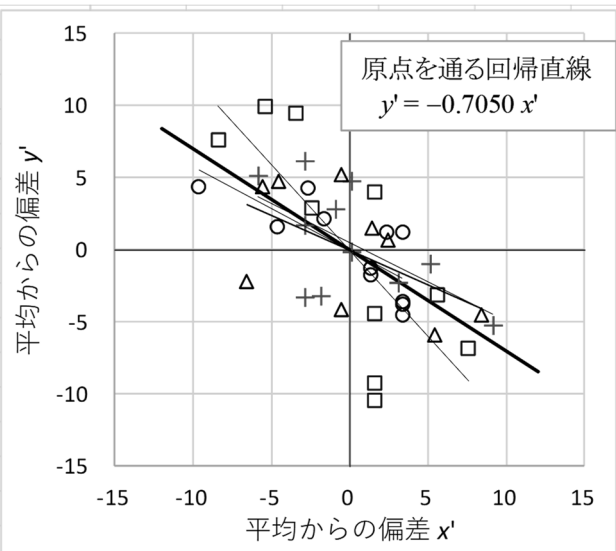


図 6.2 原点を通る回帰直線

表 6.4 Excel による原点を通る回帰分析（定数に 0 を使用） on

	自由度	変動	分散	分散比	係数	標準誤差	t	P-値
回帰	1	389.11	389.11	27.49	切片	0.0000	#N/A	#N/A
残差	41	580.29	14.15		x'	-0.7050	0.1345	-5.2433
合計	42	969.41						0.0000

傾きが $\hat{\beta}_1^{(原点)} = -0.7050$ と推定され、式 (6.7) の $\hat{\beta}_1^{(共通)}$ に一致する。また、分散分析表の平方和も

$$\left. \begin{aligned} \hat{\beta}_1^{(原点)} &= \hat{\beta}_1^{(共通)} = -0.7050 \\ S_T^{(原点)} &= S_{yy}^{(和)} = 969.41 \\ S_R^{(原点)} &= S_R^{(共通)} = 389.11 \\ S_e^{(原点)} &= S_e^{(共通)} = 580.29 \end{aligned} \right\} \quad (6.9)$$

のように一致する。

交互作用を含む分散分析表の作成

(手順4) 手順1~3で求めた平方和から分散分析表の作成

手順1の $S_{yy}^{(全体)}$ が総平方和 $S_T = S_{yy}^{(全体)} = 3654.82$ 、 $S_R^{(全体)}$ が共変量 x の平方和 $S_x = S_R^{(全体)} = 2523.03$ となる。手順1の $S_e^{(全体)} = 1131.79$ から手順2で示した $S_e^{(共通)} = 580.29$ との差が、季節Aの平方和 $S_A = 551.49$ となる。手順3の $S_e^{(共通)} = 580.29$ から手順2で示した $S_e^{(和)} = 509.05$ との差が、交互作用 $S_{x \times A} = 71.25$ となる。

$$\left. \begin{aligned} S_T &= S_{yy}^{(全体)} = 3654.82 \\ S_x &= S_R^{(全体)} = 2523.03 \\ S_A &= S_e^{(全体)} - S_e^{(共通)} = 1131.79 - 580.29 = 551.49 \\ S_{x \times A} &= S_e^{(共通)} - S_e^{(和)} = 580.29 - 509.05 = 71.25 \\ S_e &= S_e^{(和)} = 509.05 \end{aligned} \right\} \quad (6.10)$$

これらの平方和を用いて表6.5に示す分散分析表を作成し、交互作用の有無について統計的判断を行なう。表6.5に示すように交互作用の p 値は、 $p = 0.2107$ と有意ではないので、平行な直線があてはめられるとみなす。なお、第1.2節では、交互作用を無視しないとした。

表6.5 交互作用を含む分散分析表

要因	df	平方和	平均平方	F	p
x	1	2523.03	2523.03	168.52	0.0000
A	3	551.49	183.83	12.28	0.0000
$x \times A$	3	71.25	23.75	1.59	0.2107
e	34	509.05	14.97	1.00	
全体	41	3654.82			

ここに示した平方和は、逐次平方和（タイプI）であり、表1.13のタイプIIIの平方和とは異なる。

平行な回帰直線のあてはめ

(手順5) 交互作用の平方和 $S_{x \times A}$ を残差平方和へプーリング

交互作用（ $x \times A$ ）が有意ではないので、交互作用の平方和 $S_{x \times A} = 71.25$ を残差平方和 $S_e = 509.05$ にプーリングして表6.6示すように新たな残差平方和 S'_e

$$S'_e = S_e^{(共通)} = 509.05 + 71.25 = 580.29$$

とし、分散分析表を作り直す。なお、 S'_e は、手順3別解で求めた $S_e^{(原点)} = 580.29$ に一致する。

表6.6 交互作用をプーリングした分散分析表

要因	df	平方和	平均平方	F	p
x	1	2523.03	2523.03	160.87	0.0000
A	3	551.49	183.83	11.72	0.0000
e	37	580.29	15.68	1.00	
全体	41	3654.82			

これまでの手順は、先人たちの努力の賜物であるが、化石的な手順なのでまったく推奨できない。次節のダミー変数を用いたExcelの回帰分析を参照のこと。

(手順 6) 平行線をあてはめた場合の季節 A の水準ごとの切片の推定

手順 3 で求めた季節 A についての共通の傾き $\hat{\beta}_1^{(\text{共通})} = -0.7050$ とする切片 $\hat{\beta}_0^{(i)}$ を

$$\hat{\beta}_0^{(i)} = \bar{y}_{i\cdot} - \hat{\beta}_1^{(\text{共通})} \bar{x}_{i\cdot} \quad (6.11)$$

で推定すると表 6.9 に示すように $\hat{\beta}_0^{(1)} = 30.58$, $\hat{\beta}_0^{(2)} = 31.82$, $\hat{\beta}_0^{(3)} = 32.55$, $\hat{\beta}_0^{(4)} = 42.75$ が得られる.

表 6.7 傾きを共通とする場合の切片の推定

A	$\bar{x}_{i\cdot}$	$\bar{y}_{i\cdot}$	切片 $\hat{\beta}_0^{(i)}$	$\hat{\beta}_1^{(\text{共通})}$
A ₁	26.64	11.80	30.58	-0.705
A ₂	24.83	14.31	31.82	-0.705
A ₃	18.56	19.47	32.55	-0.705
A ₄	14.40	32.60	42.75	-0.705

これらを用いて図 6.3 に示すように共通の傾きを持つ平行な回帰直線を引く.

表 6.8 回帰直線の範囲の設定

A		x	共通 y [^]
A ₁	min	17	18.59
	平均	26.64	11.80
	max	30	9.43
A ₂	min	19	18.42
	平均	24.83	14.31
	max	34	7.85
A ₃	min	12	24.09
	平均	18.56	19.47
	max	27	13.51
A ₄	min	6	38.52
	平均	14.40	32.60
	max	22	27.24

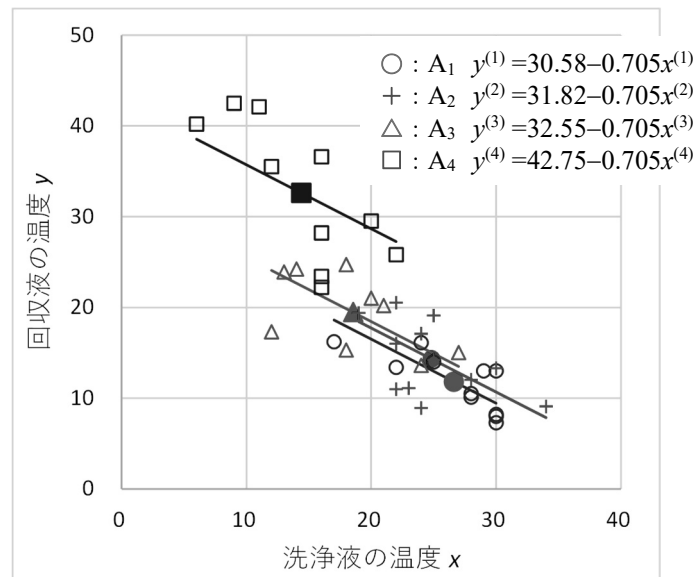


図 6.3 平行な 4 本の回帰直線のあてはめ

平行な直線のあてはめは, 図の左の直線の始点と終点をあらかじめ設定し, 「データの選択」で順次「追加」して描く. 各水準の重心については, マーカ ● などを設定している.

平行な直線間の差の 95%信頼区間.

(手順 7) 季節 A の第 4 水準を基準とする水準間の差

図 6.3 に示したように, 季節 A ごとの平行線に着目したときに, 季節 A の A₄ 水準を基準とし, それぞれの A_i 水準との平行線間の差と 95%信頼区間を推定し, 図 6.4 に示すようにひげ付き線グラフを作成したい.

平行な回帰直線の任意の x に対し、基準となる水準を i' とすると、Y 軸方向の差 $d^{(i-i')}$ は、

$$\left. \begin{aligned} d^{(i-i')} &= (\hat{\beta}_0^{(i)} + \hat{\beta}_1^{(\text{共通})}x) - (\hat{\beta}_0^{(i')} + \hat{\beta}_1^{(\text{共通})}x) \\ &= \hat{\beta}_0^{(i)} - \hat{\beta}_0^{(i')} \end{aligned} \right\} \quad (6.12)$$

のように、切片の差に帰着する。切片の差の分散を求めるために、 $\bar{y}^{(i)} = \hat{\beta}_0^{(i)} + \hat{\beta}_1^{(\text{共通})}\bar{x}^{(i)}$ を $\hat{\beta}_0^{(i)}$ について解いた式、および、 $\bar{y}^{(i')} = \hat{\beta}_0^{(i')} + \hat{\beta}_1^{(\text{共通})}\bar{x}^{(i')}$ を $\hat{\beta}_0^{(i')}$ について解いた式を合成し、

$$\left. \begin{aligned} \hat{\beta}_0^{(i)} - \hat{\beta}_0^{(i')} &= (\bar{y}^{(i)} - \hat{\beta}_1^{(\text{共通})}\bar{x}^{(i)}) - (\bar{y}^{(i')} - \hat{\beta}_1^{(\text{共通})}\bar{x}^{(i')}) \\ &= \bar{y}^{(i)} - \bar{y}^{(i')} - (\bar{x}^{(i)} - \bar{x}^{(i')})\hat{\beta}_1^{(\text{共通})} \end{aligned} \right\} \quad (6.13)$$

に置き換える。

(手順 8) 水準間の差の 95%信頼区間

回帰直線間の差 $d^{(i-i')}$ の分散 $Var(d^{(i-i')})$ は、平均値 $\bar{y}^{(i)}$, $\bar{y}^{(i')}$, $(\bar{x}^{(i)} - \bar{x}^{(i')})\hat{\beta}_1^{(\text{共通})}$ で表すことができ、これらは互いに独立で、 $(\bar{x}^{(i)} - \bar{x}^{(i')})$ は誤差を持たない定数なので、 $\hat{\beta}_1^{(\text{共通})}$ の分散は、式 (5.22) に示した $Var(\beta_1) = \sigma^2 / \sum_i (X_i - \bar{X})^2$ を用いて

$$\left. \begin{aligned} Var(d^{(i-i')}) &= Var[\bar{y}^{(i)} - \bar{y}^{(i')} - (\bar{x}^{(i)} - \bar{x}^{(i')})\hat{\beta}_1^{(\text{共通})}] \\ &= Var(\bar{y}^{(i)}) + Var(\bar{y}^{(i')}) + Var[(\bar{x}^{(i)} - \bar{x}^{(i')})\hat{\beta}_1^{(\text{共通})}] \\ &= \frac{\hat{\sigma}^2}{n^{(i)}} + \frac{\hat{\sigma}^2}{n^{(i')}} + (\bar{x}^{(i)} - \bar{x}^{(i')})^2 \frac{\hat{\sigma}^2}{S_{xx}^{(\text{共通})}} \\ &= \left(\frac{1}{n^{(i)}} + \frac{1}{n^{(i')}} + \frac{(\bar{x}^{(i)} - \bar{x}^{(i')})^2}{S_{xx}^{(\text{共通})}} \right) \hat{\sigma}^2 \end{aligned} \right\} \quad (6.14)$$

となる。ここで、誤差分散の推定値 $\hat{\sigma}^2$ は、表 6.6 に示した交互作用をプーリングした残差の平均平方の 15.6836 を用いる。表 6.9 に示すように、季節 A₄ の切片 $\hat{\beta}_0^{(4)}$ 基準とした季節 A_i の切片 $\hat{\beta}_0^{(i)}$ の差について式 (6.14) を用いて差の分散を計算し、次いで、95%信頼区間を求める。

表 6.9 季節 A₄ を基準とする水準間の差の 95%信頼区間

	季節		切片	平均	差 ($\beta_0^{(i)}$ $-\beta_0^{(i')}$)	差 ($\beta_0^{(i)}$ $-\beta_0^{(i')}$)	幅	差の95%CL	
	A _i	n _i	$\beta_0^{(i)}$	$\bar{x}_{i\cdot}$			$t_{0.05} \times SE$	L 95%	U 95%
規 準	A ₄	10	42.75	14.40	—	—	—	—	—
	A ₁	11	30.58	26.64	-12.17	5.9939	4.9606	-17.13	-7.21
	A ₂	12	31.82	24.83	-10.94	5.0562	4.5561	-15.49	-6.38
	A ₃	9	32.55	18.56	-10.20	3.6570	3.8747	-14.08	-6.33
	A ₄	10	42.75	14.40	0.00				
	$\sum_i S_{yy}^{(i)} = S_{yy}^{(\text{共通})} =$			782.83			$\hat{\sigma}^2 =$	15.6836	
			df =	37			$t(0.05, 37) =$	2.0262	

次節で示すパラメータの共分散行列 $\Sigma(\hat{\beta})$ を用いた表 6.19 の結果に一致する。

Excel の線グラフを用い図 6.4 季節 A₄ を基準とする水準間の差の 95%信頼区間を重ね書きする。図から明らかなように、他の季節との間に、有意な差があることが示されている。

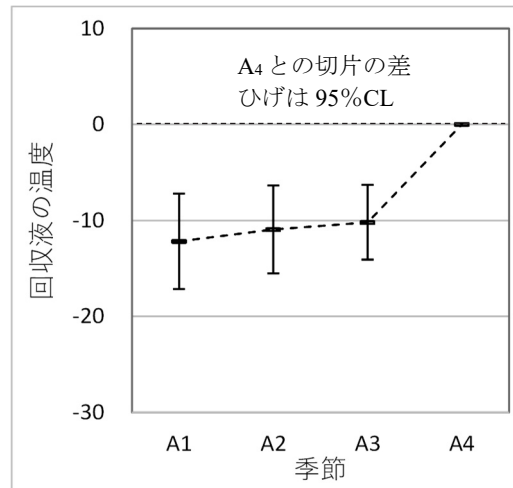


図 6.4 季節 A4 を基準とする水準間の差の 95%信頼区間

伝統的な共分散分析で定式化されているのは、ここまでである。更なる結果の吟味をするために、推定された回帰直線の 95%信頼区間を示し、季節 A ごとの洗浄水の温度 x が共通となる 20 度における回収液の推定温度 \hat{y}_i での 95%信頼区間付きの折れ線グラフも欲しいが、どうしたら良いのであろうか。これについては、次節の図 6.5 を参照のこと。

共分散分析は難解なのか？

奥野ら（1981）に示されている共分散分析の手順に忠実に Excel の力も借りて再現したのであるが、なんとも技巧的であり気が遠くなりそうである。私も実務で共分散分析を使ってきたのであるが、ダミー変数を用いた線形モデルによる共分散分析を主体にしてきた。伝統的な解析手順については、面倒なことをしていると流し読みしていた。あらためて、共分散分析について記載されている成書を紐解いてみると、伝統的な解析法のオンパレードである。他方、Web 上での解説記事では、ダミー変数を用いた回帰分析を用いた解析方法がメジャーであるが、表層的な説明で終わっている。

この事例は、「層別因子を含む回帰分析」であって、いわゆる共変量を含む「共分散分析」の事例ではない。だが、質的因子と量変数を含む解析は、短絡的に「共分散分析」の事例とみなされがちで、「交互作用のないことを確認し、平行な直線をあてはめる」ような紋きり型の解析に絡めとられるが常である。交互作用があると、式 (6.14) に示したような「水準間の差の 95%信頼区間」が適用できなくなってしまうために、交互作用がないことが「共分散分析」の適用の絶対条件であるかの風評が蔓延している。次の第 7 章では、本物の“共変量”と質的変数との間で交互作用が存在する事例を示す。

6.2. Excel の回帰分析を活用した共分散分析

第 6.1 節で示した回帰分析を重層的に用いる共分散分析の解析法は、手計算時代の先人たちの創意工夫の賜物である。Excel を用いて原著に示されている計算方法を忠実に再現したのは、デザイン行列 X を用いた共分散分析と対比することにより、統計解析の多面性を浮き彫りにするためである。

(0, 1)型ダミー変数を用いたデザイン行列 X

季節 A は 4 水準なので、表 6.10 に示すように (0, 1) 型ダミー変数 (a_2, a_3, a_4) を与える。洗浄水の温度 x との交互作用のための変数は、(xa_2, xa_3, xa_4) のように積で与える。

表 6.10 (0, 1) 型ダミー変数

A	a_2	a_3	a_4
A ₁	0	0	0
A ₂	1	0	0
A ₃	0	1	0
A ₄	0	0	1

A	xa_2	xa_3	xa_4
A ₁	0	0	0
A ₂	x	0	0
A ₃	0	x	0
A ₄	0	0	x

表 6.11 左に (0, 1) 型ダミー変数を用いた交互作用を含むデザイン行列 X を示す。表 6.11 右には、Excel の回帰分析で得られた (回帰モデル, 主効果モデル, 交互作用モデル) での分散分析表を示す。

表 6.11 交互作用モデルに対する段階的な Excel の回帰分析の適用

3) 交互作用モデル														
2) 主効果モデル										1) 回帰モデル (定数に 0 を使用) off				
1) 回帰モデル										自由度	変動	分散	分散比	
No.	A	x_0	x	a_2	a_3	a_4	xa_2	xa_3	xa_4	y	回帰			
1	A ₁	1	30	0	0	0	0	0	0	8.0	残差	40	1131.79	28.29
2	A ₁	1	28	0	0	0	0	0	0	10.5	合計	41	3654.82	
:		:								:				
11	A ₁	1	25	0	0	0	0	0	0	14.0	2) 主効果モデル (定数に 0 を使用) off			
12	A ₂	1	34	1	0	0	34	0	0	9.1	自由度	変動	分散	分散比
:		:								:	回帰	4	3074.53	768.63
23	A ₂	1	24	1	0	0	24	0	0	8.9	残差	37	580.29	15.68
24	A ₃	1	27	0	1	0	0	27	0	15.0	合計	41	3654.82	
:		:								:				
32	A ₃	1	24	0	1	0	0	24	0	13.6	3) 交互作用モデル (定数に 0 を使用) off			
33	A ₄	1	16	0	0	1	0	0	16	23.4	自由度	変動	分散	分散比
:		:								:	回帰	7	3145.78	449.40
41	A ₄	1	9	0	0	1	0	0	9	42.5	残差	34	509.05	14.97
42	A ₄	1	22	0	0	1	0	0	22	25.8	合計	41	3654.82	

逐次平方和による分散分析表

表 6.11 に示した 3 通りの分散分析表は、説明変数を逐次増加させた Excel の回帰分析の分散分析表のみを示した結果である。表 6.12 に示すのは、3 通りの分散分析表の回帰の平方和（変動）、残差の平方和（変動）に含まれている各要因の平方和（ S_x , S_A , $S_{(x \times A)}$, S_e ）の合成方法を示している。回帰分析に含める説明変数を増加すれば、「回帰」の平方和も逐次増加し、その差分から S_A および $S_{(x \times A)}$ を得ることができる。同様に残差の平方和の減少分からも S_A および $S_{(x \times A)}$ が計算できる。

表 6.12 逐次的な回帰分析による平方和の増加

1) 回帰モデル				2) 主効果モデル				3) 交互作用モデル			
要因	df	平方和		要因	df	平方和		要因	df	平方和	
x	1	2523.03	S_x	x	1		S_{x+A}	x	1		$S_{x+A+(x \times A)}$
A	(3)		$S_{A+(x \times A)+e}$	A	3	3074.53		A	3	3145.78	
$x \times A$	(3)	1131.79		$x \times A$	(3)		$S_{(x \times A)+e}$	$x \times A$	3		S_e
e	40			e	37	580.29		e	34	509.05	
T	41	3654.82		T	41	3654.82		T	41	3654.82	
	$S_x =$	2523.03			$S_{x+A} - S_x =$	551.49	S_A		$S_{x+A+(x \times A)} - S_{x+A} =$	71.25	$S_{x \times A}$

表 6.13 に示すのは、表 6.12 のそれぞれの「回帰」の平方和から

$$\left. \begin{aligned} S_A &= S_{x+A} - S_x = 3074.53 - 2523.03 = 551.49 \\ S_{x \times A} &= S_{x+A+(x \times A)} - S_{x+A} = 3145.78 - 3074.53 = 71.25 \end{aligned} \right\} \quad (6.15)$$

S_A および $S_{(x \times A)}$ を計算し、逐次（タイプ I）平方和による分散分析表にまとめた結果で、表 6.5 に示した分散分析表に一致する。このように変数を逐次増加した Excel の回帰分析により、先人たちの苦勞を全くせずに、共分散分析における交互作用が、あっさりと求められることは、喜ばしいことである。なお、交互作用を残差平方和 S_e にプールした結果は、すでに表 6.6 に示してあるので割愛する。

表 6.13 逐次（タイプ I）平方和による分散分析表（表 6.5 に一致）

要因	df	逐次平方	平均平方	F	p
x	1	2523.03	2523.03	168.52	0.0000
A	3	551.49	183.83	12.28	0.0000
$x \times A$	3	71.25	23.75	1.59	0.2107
e	34	509.05	14.97	1.00	
全体	41	3654.82			

共通の傾きを持つ場合の各種の推定

前節の手順 7 で、季節 A の A_4 水準を基準として、他の水準との平行な直線間の差については、式 (6.12) に示した切片の差に帰着し、式 (6.14) に示した分散を用いて 95%信頼区間を推定し、統計的な考察を行った。ここでは、パラメータの共分散行列 $\Sigma(\hat{\theta})$ を活用し、各種

の95%信頼区間の算出法を示す．表6.14に主効果モデルにおけるExcelの回帰分析の結果を示し，パラメータの共分散行列 $\Sigma(\hat{\theta})$ の計算結果も合わせて示す．

これまでも繰り返し述べてきたが，パラメータの共分散行列 $\Sigma(\hat{\theta})$ の対角要素が，推定された回帰パラメータ $\hat{\theta}_i$ の分散 $Var(\hat{\theta}_i)$ となり，この平方根が標準誤差 $SE(\hat{\theta}_i)$ となる．傾きの推定値は， $\hat{\theta}_1 = -0.7050$ の $SE(\hat{\theta}_1) = 0.1415$ は， $\Sigma(\hat{\theta})$ の2行2列目の平方根 $\sqrt{0.0200} = 0.1415$ となっている．パラメータの共分散行列 $\Sigma(\hat{\theta})$ の1行2列目は， $\hat{\theta}_0$ と $\hat{\theta}_1$ の共分散 $Cov(\hat{\theta}_0, \hat{\theta}_1) = -0.5336$ であり，各種の線形和 L の分散の推定に用いられる．

表6.14 主効果モデルに対するパラメータの共分散行列

デザイン行列 X													
No.	A	x_0	x	a_2	a_3	a_4	y	2) 主効果モデル (定数に 0 を使用) off					
1	A_1	1	30	0	0	0	8.0		自由度	変動	分散	分散比	
2	A_1	1	28	0	0	0	10.5	回帰	4	3074.53	768.63	49.01	
:		:	:	:	:	:	:	残差	37	580.29	15.68	$=\sigma^{\wedge^2}$	
11	A_1	1	25	0	0	0	14.0	合計	41	3654.82			
12	A_2	1	34	1	0	0	9.1						
:		:	:	:	:	:	:		係数	標準誤差	t	P -値	
23	A_2	1	24	1	0	0	8.9	θ^{\wedge}_0 切片	30.5793	3.9548	7.7323	0.0000	
24	A_3	1	27	0	1	0	15.0	θ^{\wedge}_1 x	-0.7050	0.1415	-4.9810	0.0000	
:		:	:	:	:	:	:	θ^{\wedge}_2 a_2	1.2372	1.6727	0.7396	0.4642	
32	A_3	1	24	0	1	0	13.6	θ^{\wedge}_3 a_3	1.9695	2.1158	0.9309	0.3580	
33	A_4	1	16	0	0	1	23.4	θ^{\wedge}_4 a_4	12.1731	2.4482	4.9722	0.0000	
:		:	:	:	:	:	:						
41	A_4	1	9	0	0	1	42.5	パラメータの共分散行列 $\Sigma(\theta^{\wedge})=(X^T X)^{-1} \sigma^{\wedge^2}$					
42	A_4	1	22	0	0	1	25.8	x_0	x	a_2	a_3	a_4	
								θ^{\wedge}_0	15.6401	-0.5336	-2.3880	-5.7381	-7.9557
								θ^{\wedge}_1	-0.5336	0.0200	0.0361	0.1619	0.2451
								θ^{\wedge}_2	-2.3880	0.0361	2.7979	1.7177	1.8678
								θ^{\wedge}_3	-5.7381	0.1619	1.7177	4.4767	3.4068
								θ^{\wedge}_4	-7.9557	0.2451	1.8678	3.4068	5.9939
=Minverse (Mmult (Transpose (X の範囲), X の範囲)) * σ^{\wedge^2}													

デザイン行列 X の変数と回帰分析で推定されたパラメータ（係数）を組合わせ，各種の線形和 $L^{(i)}$ による推定値を定義する．デザイン行列の変数 (x_0, x_1, a_2, a_3, a_4) に対し別名として，行ベクトル $\mathbf{l}^{(i)} = (l_0^{(i)}, l_1^{(i)}, l_2^{(i)}, l_3^{(i)}, l_4^{(i)})$ を与える．推定されたパラメータを列ベクトル $\hat{\theta} = (\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4)^T$ としたときに，線形和 $L^{(i)}$ を

$$L^{(i)} = \sum_{j=0}^4 l_j^{(i)} \hat{\theta}_j = \mathbf{l}^{(i)} \hat{\theta} \quad (6.16)$$

として定義すると，線形和 $L^{(i)}$ の分散 $Var(L^{(i)})$ は，パラメータの共分散行列 $\Sigma(\hat{\theta})$ の j 行 j' 列の要素を $c_{j,j'}$ として

$$Var(L^{(i)}) = \sum_{j=0}^4 \sum_{j'=0}^4 l_j^{(i)} c_{j,j'} l_{j'}^{(i)} = \mathbf{l}^{(i)} \Sigma(\hat{\theta}) \mathbf{l}^{(i)T} \quad (6.17)$$

で与えられ，パラメータの共分散行列 $\Sigma(\hat{\theta})$ に関する $\mathbf{l}^{(i)}$ の2次形式となる．

表 6.14 に示した傾きを共通とする回帰パラメータの切片は、表 6.15 に示すように、それぞれの線形和 $L^{(i)}$ に対応する行ベクトル $\mathbf{l}^{(i)}$ とパラメータの推定値の列ベクトル $\hat{\boldsymbol{\theta}}$ の積 $L^{(i)} = \mathbf{l}^{(i)} \hat{\boldsymbol{\theta}}$ で計算されている。

表 6.15 線形和 $L^{(i)}$ を用いた傾きを共通とする回帰パラメータの切片の推定

		l_0	l_1	l_2	l_3	l_4			切片	
A	$L^{(i)}$	x_0	x	a_2	a_3	a_4	$\hat{\boldsymbol{\theta}}$	=	$L^{(i)}$	
A ₁	$L^{(1)}$	1	0	0	0	0	30.5793	=	30.58	$\hat{\beta}_0^{(1)}$
A ₂	$L^{(2)}$	1	0	1	0	0	-0.7050		31.82	$\hat{\beta}_0^{(2)}$
A ₃	$L^{(3)}$	1	0	0	1	0	1.2372		32.55	$\hat{\beta}_0^{(3)}$
A ₄	$L^{(4)}$	1	0	0	0	1	1.9695		42.75	$\hat{\beta}_0^{(4)}$
							12.1731			

前節の表 6.9 では、A₄ 水準と他水準の切片の差の分散を求め、95%信頼区間の推定を行ったが、推定された切片の 95%信頼区間の推定は行わなかった。これは、推定された回帰直線の外挿であり、結果の解釈に不必要なためである。

さて、季節 A の水準ごとの意味のある推定値として、洗浄水の温度 x が 20 度場合の回収液の温度 $\hat{y}_{x=20}^{(i)}$ を推定し、その 95%信頼区間を求め、報告書に記載したいとしよう。表 6.15 では、 $x=l_1=0$ としたので、 $x=l_1=20$ とすれば、20 度における線形和の推定値が計算される。A₁ 水準の場合であれば、推定値は $\hat{y}_{x=20}^{(1)}=16.4788$ であり、

表 6.16 A₁ 水準の洗浄水の温度が 20 度場合の回収液の温度の推定

	l_0	l_1	l_2	l_3	l_4				
	x_0	x	a_2	a_3	a_4	$\hat{\boldsymbol{\theta}}$			
$y^{(1)}_{x=20}=$	1	20	0	0	0	30.5793	=	16.4788	
						-0.7050			
						1.2372			
						1.9695			
						12.1731			

その線形和の分散は、式 (6.17) に示したパラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ に関する $\mathbf{l}^{(i)}$ の 2 次形式を用い、表 6.17 に示すように $Var(\hat{y}_{x=20}^{(1)})=2.3081$ と推定される。

表 6.17 A₁ 水準の洗浄水の温度が 20 度場合の回収液の温度の分散の推定

l_0	l_1	l_2	l_3	l_4						
x_0	x	a_2	a_3	a_4	$\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$		\mathbf{l}^T	$Var(\hat{y}_{x=20}^{(1)})$		
1	20	0	0	0	15.64	-0.53	-2.39	-5.74	-7.96	1 = 2.3081
					-0.53	0.02	0.04	0.16	0.25	$L_{95\%}=13.4005$
					-2.39	0.04	2.80	1.72	1.87	$U_{95\%}=19.5571$
					-5.74	0.16	1.72	4.48	3.41	
					-7.96	0.25	1.87	3.41	5.99	$t_{0.05}(42-5)=2.0262$

95%信頼区間は,

$$\begin{aligned}
 (L95\%, U95\%) &= \hat{y}_{x=20}^{(1)} \pm t(0.05, 37) \sqrt{\text{Var}(\hat{y}_{x=20}^{(1)})} \\
 &= 16.4788 \pm 2.0262 \times \sqrt{2.3081} \\
 &= (13.4005, 19.5571)
 \end{aligned} \tag{6.18}$$

として求められる. 表 6.16 に示した計算式により, 他の季節における推定結果を表 6.18 に示す. なお, 計算式の設定を A₁ 水準で行い, セルのフィルハンドルを用いて他の季節の計算式をコピーして求めている.

表 6.18 季節別の洗浄水の温度が 20 度の場合の回収液の濃度の推定

		l_0	l_1	l_2	l_3	l_4	推定値	分散	幅	95%信頼区間	
A	$L^{(i)}$	x_0	x	a_2	a_3	a_4	$L^{(i)}$	$\text{Var}(L^{(i)})$	$t_{0.05} \times SE$	L 95%	U 95%
A ₁	$L^{(1)}$	1	20	0	0	0	16.48	2.3081	3.08	13.40	19.56
A ₂	$L^{(2)}$	1	20	1	0	0	17.72	1.7750	2.70	15.02	20.42
A ₃	$L^{(3)}$	1	20	0	1	0	18.45	1.7844	2.71	15.74	21.15
A ₄	$L^{(4)}$	1	20	0	0	1	28.65	2.1966	3.00	25.65	31.65
							$t(0.05, 42-5)=$		2.0262		

この結果を Excel の「折れ線」グラフで作成した結果を図 6.5 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別途計算する必要がある.

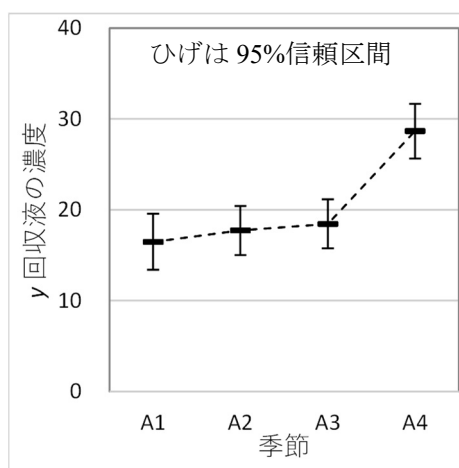


図 6.5 洗浄水の温度が 20 度の場合の予測プロファイル

Excel での「折れ線グラフ」のひげの長さの設定は, かなり深いところにあり,

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

のように設定する. 少しでも気を抜くとまがい物のひげが付く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、デザイン行列の変数 (x_0, x_1, a_2, a_3) に対し別名として与えた行ベクトル $\mathbf{l}^{(i)} = (l_0^{(i)}, l_1^{(i)}, l_2^{(i)}, l_3^{(i)}, l_4^{(i)})$ 間の差を計算する。推定値は、

$$\left. \begin{aligned} \hat{y}_{x=20}^{(A_3-A_4)} &= \hat{y}_{x=20}^{A_3} - \hat{y}_{x=20}^{A_4} \\ &= 18.45 - 28.65 = -10.20 \end{aligned} \right\} \quad (6.19)$$

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

$$\left. \begin{aligned} \mathbf{l}^{(A_3)} &= [1 \quad 20 \quad 0 \quad 1 \quad 0] \\ -) \quad \mathbf{l}^{(A_4)} &= [1 \quad 20 \quad 0 \quad 0 \quad 1] \\ \hline \mathbf{d}^{(A_3-A_4)} &= [0 \quad 0 \quad 0 \quad 1 \quad -1] \end{aligned} \right\} \quad (6.20)$$

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

$$\begin{aligned} \text{Var}[\hat{y}^{(A_3-A_4)}] &= \mathbf{d}^{(A_3-A_4)} \boldsymbol{\Sigma}(\hat{\theta}) [\mathbf{d}^{(A_3-A_4)}]^T \\ &= 3.6570 \end{aligned} \quad (6.21)$$

として求められ、差 -10.20 の 95%信頼区間が $(-18.10, -2.31)$ 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。

同様に季節 A_4 を基準とし、他の季節の差を表 6.19 に示す。差のデザイン行列については、季節 A_3 と季節 A_4 の場合と同様にする。差の分散についてもそれぞれの差のデザイン行列の場合と同様に推定し、95%信頼区間を計算する。この結果は、前節の表 6.9 に示した結果に一致し、また、図 6.4 に示した水準間の差の 95%信頼区間の折れ線グラフが作成することができる。

表 6.19 季節 A_4 を基準とした 2 水準間の差の 95%信頼区間

		l_0	l_1	l_2	l_3	l_4	推定値	分散	幅	95%信頼区間	
A	$L^{(i)}$	x_0	x	a_2	a_3	a_4	$L^{(i)}$	$\text{Var}(L^{(i)})$	$t_{0.05 \times SE}$	L 95%	U 95%
A_1-A_4	$L^{(5)}$	0	0	0	0	-1	-12.17	5.9939	4.96	-17.13	-7.21
A_2-A_4	$L^{(6)}$	0	0	1	0	-1	-10.94	5.0562	4.56	-15.49	-6.38
A_3-A_4	$L^{(7)}$	0	0	0	1	-1	-10.20	3.6570	3.87	-14.08	-6.33
A_4-A_4	$L^{(8)}$	0	0	0	0	0	0.00	0.0000	0.00	0.00	0.00
							$t(0.05, 42-5)=$	2.0262			

線形モデルにおける各種の推定は、線形和 $L^{(i)}$ に対する式 (6.16) , その分散 $\text{Var}(L^{(i)})$ は、式 (6.17) として定式化されていることが、汎用的であり伝統的な方法に対する優位性でもある。

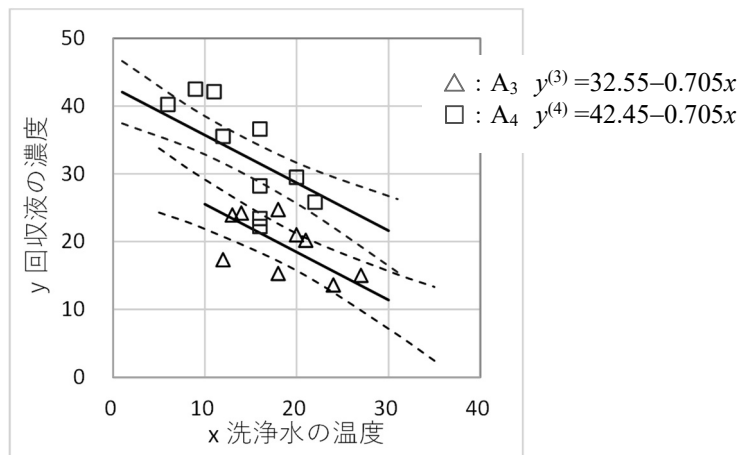
洗浄水の温度 x に関する予測プロファイル

季節 A_3 と季節 A_4 について洗浄水の温度を変化させた場合の予測プロファイルを作成する. 表 6.20 に示すように, 季節 A_3 の洗浄水の温度 x を 1 度から 35 度まで変化させる. 季節 A_3 も同様に洗浄水の温度 x を 5 度から 31 度まで変化させた行ベクトル $L^{(i)}$ を生成する. これらのベクトルに対し, パラメータの推定値 $\hat{\theta}$ およびパラメータの共分散行列 $\Sigma(\hat{\theta})$ を用いて予測値と分散を推定し, 95%信頼区間を求める.

表 6.20 洗浄水の温度 x を変えた場合の回収液の濃度の 95%信頼区間

		l_0	l_1	l_2	l_3	l_4	推定値	分散	幅	95%信頼区間	
A	$L^{(i)}$	x_0	x	a_2	a_3	a_4	$L^{(i)}$	$Var(L^{(i)})$	$t_{0.05} \times SE$	$L_{95\%}$	$U_{95\%}$
A_3	$L^{(9)}$	1	5	0	1	0	29.02	5.42	4.72	24.30	33.74
	$L^{(10)}$	1	10	0	1	0	25.50	3.21	3.63	21.87	29.13
	$L^{(11)}$	1	20	0	1	0	18.45	1.78	2.71	15.74	21.15
	$L^{(12)}$	1	30	0	1	0	11.40	4.37	4.23	7.16	15.63
	$L^{(13)}$	1	35	0	1	0	7.87	7.16	5.42	2.45	13.29
A_4	$L^{(14)}$	1	1	0	0	1	42.05	5.17	4.61	37.44	46.65
	$L^{(15)}$	1	10	0	0	1	35.70	1.96	2.83	32.87	38.54
	$L^{(16)}$	1	20	0	0	1	28.65	2.20	3.00	25.65	31.65
	$L^{(17)}$	1	30	0	0	1	21.60	6.44	5.14	16.46	26.75
	$L^{(18)}$	1	31	0	0	1	20.90	7.09	5.39	15.50	26.29
							$t(0.05, 42-5)=$		2.0262		

図 6.6 に示す予測プロファイルは, 表 6.20 で計算された季節 A_4 と季節 A_3 の予測値と 95% 信頼区間を散布図に重ね書きした結果である. 季節 A_4 の洗浄水の温度 x が 10 度の場合, 回収液の濃度 \hat{y} は, 37.8377 と季節 A_3 の 23.5820 と 14.2557 の差がある. 洗浄水の温度 x が 10 度の場合, 差が 7.1620 と縮小するが, 95%信頼区間が互いに重なっていないので統計的な差があると判断される. このように, 推定された平行な回帰直線に対し, 95%信頼区間を計算し重ね書きして示せることは, 一連の定式化された線形モデルの解析手順の賜物であるし, Excel の行列計算の柔軟性のおかげでもある.

図 6.6 洗浄水の温度 x に対する季節 A_4 および季節 A_3 の予測プロファイ

6.3. SAS の GLM プロシジャによる共分散分析

SAS でのテキストデータの取り込み

無償で継続的に提供されている OnDemand SAS の GLM プロシジャを用いて、これまでに示した解析法と対応付けた解析方法を示す。GLM プロシジャは、共変量 x 、質的変数 A 、それらの交互作用を扱うことができる。まず、GLM プロシジャを使うための SAS ファイルの作成から始める。SAS は、様々な形式の外部ファイルを読み込めるが、特徴的なのは、テキストデータを読み込むための種々のプログラミング機能を持っているので少量のデータの入力に対し、柔軟な対応が可能である。

表 6.1 に示した Excel で作成した表形式のデータを、テキストエディタ（メモ帳）にコピー & ペーストし、tab 記号をスペースに変換した結果を SAS プログラムの中に取り込む。SAS の DATA ステップで、do ステートメントなどを使い、各行に 4 セット分の季節 A に対応する共変量 x 、反応 y を読み込み、行方向に 42 行を SAS データファイル「d01」に出力する。元のデータの並びのテキストデータがどのような形であれ、解析用の SAS データファイルを作成することができる。

SAS プログラム 6.1 /* SAS データファイルの作成 */

```

Title1 'ANCOVA_奥野.sas 2022-6-13 Y.Takahashi ' ;
data d01 ;
    input No @ ;
    do A = "A1", "A2", "A3", "A4" ;
        input x y @;
        if (x ne .) then output ;
    end ;
/*      A1          A2          A3          A4
No      x      y      x      y      x      y      x      y  */ ;
datalines ;
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   .   .   .   .
;
proc print data=d01 ; run ;

```

GLM プロシジャによる交互作用 $x \times A$ の評価

GLM プロシジャを用いて交互作用 $x \times A$ の評価を行なう。季節 A は、文字データであるが、GLM プロシジャでは、class ステートメントで質的変数であることを明示し、model ステートメントで `model y = x A x*A` のようにモデル式を与え、交互作用は「*」で設定する。「/」以後がオプションで、solution オプションは、パラメータの推定値の出力要求であり、SS1 SS2 SS3 は、タイプ I, II, III の平方和の出力要求である。

SAS プログラム 6.2 /* GLM プロシジャによる交互作用の解析 */

```
proc glm data=d01 ;
  class A ;
  model y = x A x*A / solution SS1 SS2 SS3 ;
run ;
```

OnDemand SAS のデフォルトの出力は、HTML 形式なので、出力された分散分析表コピーし、Excel にペーストし、若干の整形後にワードにペーストした結果を表 6.21 に示す。この結果は、表 6.11 左下段の Excel の回帰分析の分散分析表に一致する。

表 6.21 GLM プロシジャによる包括的な分散分析表

要因	自由度	平方和	平均平方	F 値	Pr > F
Model	7	3145.7754	449.3965	30.02	<.0001
Error	34	509.0487	14.9720		
Corrected Total	41	3654.8240			

表 6.22 に示すのは、SAS の GLM プロシジャで出力される 3 種の平方和を Excel 上で編集した結果である。交互作用 $x \times A$ の平方和は、3 種の平方和が共に $S_{x \times A} = 71.2462$ であり、交互作用が統計的に無いとみなせる。主効果 x および季節 A のタイプ I の平方和は、Excel での回帰分析から導出した表 6.13 の逐次（タイプ I）の平方和に一致する。なお、3 種の平方和についての詳細は、第 3.3 節を参照のこと。

表 6.22 交互作用モデルにおけるタイプ I, II, III の平方和の比較

要因	自由度	Type I 平方和		Type II 平方和		Type III 平方和	F 値	Pr > F
x	1	2523.0347	≠	389.1143	≠	369.0901	-	-
A	3	551.4944	≠	246.2924	=	246.2924	-	-
x*A	3	71.2462	=	71.2462	=	71.2462	1.59	0.2107

GLM プロシジャが内部で生成するデザイン行列 X

分散分析表の出力結果だけで満足するならば、SAS が内部で生成しているデザイン行列の知識は必要ない。表 6.23 に示すように solution オプションで出力したパラメータの推定値から 4 本の回帰直線のパラメータ推定を行いたい、どのようにしたら良いのだろうか。

表 6.23 GLM プロシジャによるパラメータの推定値

パラメータ	推定値		標準誤差	t 値	Pr > t
Intercept	49.7416	B	3.9804	12.5	<.0001
x	-1.1904	B	0.2630	-4.53	<.0001
A A1	-22.7247	B	8.7579	-2.59	0.0139
A A2	-22.7124	B	8.2075	-2.77	0.0091
A A3	-21.3495	B	6.5383	-3.27	0.0025
A A4	0	B	.	.	.
x*A A1	0.6191	B	0.3912	1.58	0.1228
x*A A2	0.6781	B	0.3882	1.75	0.0897
x*A A3	0.7094	B	0.3775	1.88	0.0688
x*A A4	0	B	.	.	.

注) 「B」は、一意的な推定値ではないとの警告である。

(1, 0) 型ダミー変数を用いて Excel の回帰分析で再現できる。

表 6.24 に示すように GLM プロシジャの内部のデザイン行列 X は、表 6.23 に示すパラメータの出力から、(1, 1) 標示型ダミー変数としていることがうかがわれる。ダミー変数間が一次従属となり、このままでは解析できない。そのために GLM プロシジャでは、(A A4) および (x*A A4) を解析に際し無視してパラメータの推定を行っている。そのために、パラメータの推定は、最後の水準を基準とした (1, 0) 型ダミー変数と同じ結果となっている。

表 6.24 SAS が内部で生成しているデザイン行列

SAS元データ														
No.	A	x	y	切片	x	A1	A2	A3	A4	x*A1	x*A2	x*A3	x*A4	y
1	A1	30	8.0	1	30	1	0	0	0	30	0	0	0	8.0
2	A1	28	10.5	1	28	1	0	0	0	28	0	0	0	10.5
:			:	:	:									:
11	A1	25	14.0	1	25	1	0	0	0	25	0	0	0	14.0
12	A2	34	9.1	1	34	0	1	0	0	0	34	0	0	9.1
:			:	:	:									:
23	A2	24	8.9	1	24	0	1	0	0	0	24	0	0	8.9
24	A3	27	15.0	1	27	0	0	1	0	0	0	27	0	15.0
:			:	:	:									:
32	A3	24	13.6	1	24	0	0	1	0	0	0	24	0	13.6
33	A4	16	23.4	1	16	0	0	0	1	0	0	0	16	23.4
:			:	:	:									:
41	A4	9	42.5	1	9	0	0	0	1	0	0	0	9	42.5
42	A4	22	25.8	1	22	0	0	0	1	0	0	0	22	25.8

SAS/GLM プロシジャには、JMP のように内部で生成されたデザイン行列を出力するすべがないので、別途 Excel で作成した結果を示している。

したがって、表 6.25 に示すように切片 $\hat{\theta}_0 = 49.7416$ が A₄ 水準の切片となり、 $\hat{\theta}_1 = -1.1904$ が傾きとなる。A₁ 水準の切片 $\hat{\beta}_0^{(1)}$ は、 $\hat{\theta}_0 = 49.7416$ に $\hat{\theta}_2 = -22.7247$ を加えて

$$\hat{\beta}_0^{(1)} = \hat{\theta}_0 + \hat{\theta}_2 = 49.7416 - 22.7247 = 27.0169 \quad (6.22)$$

表 6.25 GLM プロシジャの推定値から4本の回帰直線の切片と傾きの推定

タ	推定値				切片	
Intercept	49.7416	$=\theta^{\wedge}_0$		A ₁	$\beta^{\wedge}_0^{(1)}=$	27.0169 $=\theta^{\wedge}_0+\theta^{\wedge}_2$
x	-1.1904	$=\theta^{\wedge}_1$		A ₂	$\beta^{\wedge}_0^{(2)}=$	27.0292 $=\theta^{\wedge}_0+\theta^{\wedge}_3$
A A1	-22.7247	$=\theta^{\wedge}_2$		A ₃	$\beta^{\wedge}_0^{(3)}=$	28.3921 $=\theta^{\wedge}_0+\theta^{\wedge}_4$
A A2	-22.7124	$=\theta^{\wedge}_3$		A ₄	$\beta^{\wedge}_0^{(4)}=$	49.7416 $=\theta^{\wedge}_0$
A A3	-21.3495	$=\theta^{\wedge}_4$				
A A4	0					傾き
x*A A1	0.6191	$=\theta^{\wedge}_5$		A ₁	$\beta^{\wedge}_1^{(1)}=$	-0.5713 $=\theta^{\wedge}_1+\theta^{\wedge}_5$
x*A A2	0.6781	$=\theta^{\wedge}_6$		A ₂	$\beta^{\wedge}_1^{(2)}=$	-0.5123 $=\theta^{\wedge}_1+\theta^{\wedge}_6$
x*A A3	0.7094	$=\theta^{\wedge}_7$		A ₃	$\beta^{\wedge}_1^{(3)}=$	-0.4810 $=\theta^{\wedge}_1+\theta^{\wedge}_7$
x*A A4	0			A ₄	$\beta^{\wedge}_1^{(4)}=$	-1.1904 $=\theta^{\wedge}_1$

GLM プロシジャの出力を Excel に取り込み、回帰直線のパラメータを計算した結果である。

として求められる。A₁ 水準の傾き $\hat{\beta}_1^{(1)}$ は、 $\hat{\theta}_1 = -1.1904$ に $\hat{\theta}_5 = 0.6191$ を加えて

$$\hat{\beta}_1^{(1)} = \hat{\theta}_1 + \hat{\theta}_5 = -1.1904 + 0.6191 = -0.5713 \quad (6.23)$$

として求められる。

なお、表 6.25 に示した回帰パラメータの推定値は、表 6.2 に示した別々に推定した4本の回帰直線の推定値に一致する。

表 6.26 季節ごとの回帰パラメータ推定 (表 6.2 抜粋)

	A ₁	A ₂	A ₃	A ₄	和
n_i	11	12	9	10	
$\bar{x}_{i\cdot}$	26.64	24.83	18.56	14.40	—
$\bar{y}_{i\cdot}$	11.80	14.31	19.47	32.60	—
$S_{xx}^{(i)}=$	178.55	183.67	204.22	216.40	782.83
$S_{xy}^{(i)}=$	-102.00	-94.08	-98.23	-257.60	-551.92
$\beta^{\wedge}_1^{(i)}=$	-0.57	-0.51	-0.48	-1.19	—
$\beta^{\wedge}_0^{(i)}=$	27.02	27.03	28.39	49.74	—
$\beta^{\wedge}_1^{(i)} = S_{xy}^{(i)} / S_{xx}^{(i)} \quad \beta^{\wedge}_0^{(i)} = \bar{y}_{i\cdot} - \beta^{\wedge}_1^{(i)} \bar{x}_{i\cdot}$					

散布図上に4本の回帰直線を重ね書きした結果が、図 6.7 に示すようにデフォルトで出力される。この結果に、何らかの方法で、手を加えてブラッシュアップしようとしたが、まったく手が加えられなかった。結果の信頼性の確保の観点からは、望ましいのであるが、残念なので、表 6.25 の回帰パラメータの推定値をワードで重ね書きしてある。

交互作用を含まない解析

交互作用を含まない解析は、model y = x A のように x*A を落とせばよい。表 6.27 に示すように、包括的な分散分析表が得られる。Excel の回帰分析による表 6.14 の分散分析表に一致する。

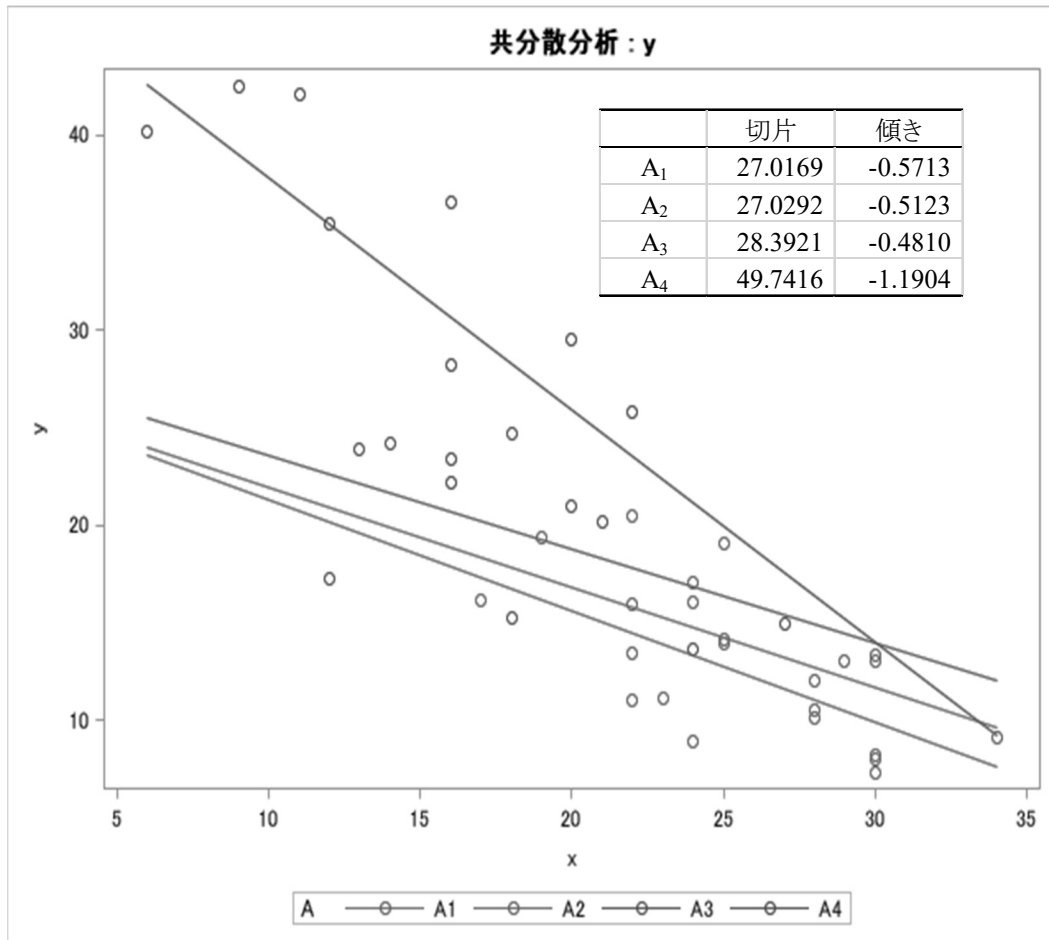


図 6.7 GLM プロシジャでデフォルト出力される散布図上の 4 本の回帰直線

SAS プログラム 6.3 /* GLM プロシジャによる主効果モデル */

```

proc glm data=d01 ;
  class A ;
  model y = x A / solution SS1 SS2 SS3 ;
run ;

```

表 6.27 GLM プロシジャによる包括的な分散分析表

要因	自由度	平方和	平均平方	F 値	Pr > F
Model	4	3074.5291	768.63229	49.01	<.0001
Error	37	580.2949	15.6836		
Corrected Total	41	3654.8240			

表 6.28 に示すように、交互作用を含まない主効果モデルでは、タイプ II, III の平方和が一致することが確認される。もちろんタイプ I の平方和は、季節 A については、他と同じ $S_A = 551.4944$ であるが、共変量 x については $S_x^{(I)} = 2523.0347$ とタイプ I の平方和のみが異

なる．主効果 x および A のタイプ I の平方和は，Excel での回帰分析から導出した表 6.12 の逐次（タイプ I）の平方和に一致する．季節 A は，統計的に明らかな差があるので，これまでと同様に各種の推定値による比較検討を行なう．

表 6.28 主効果モデルにおけるタイプ I, II, III の平方和の比較

要因	自由度	Type I 平方和	Type II 平方和	Type III 平方和	F 値	Pr > F
x	1	2523.0347	389.1143	389.1143	-	-
A	3	551.4944	551.4944	551.4944	11.72	<.0001

表 6.29 に，solution オプションで出力したパラメータ推定値から，平行な 4 本の回帰直線のパラメータを推定する． A_1 の切片は，

$$\hat{\beta}_0^{(1)} = \hat{\theta}_0 + \hat{\theta}_2 = 42.7523 - 12.1731 = 30.5793 \quad (6.24)$$

であり，傾きは， $\hat{\beta}_1^{(1)} = \hat{\theta}_1 = -0.7050$ である．この結果は，Excel の回帰分析の結果から推定した表 6.15 に一致する．

表 6.29 GLM プロシジャのパラメータ推定値から 4 本の回帰直線のパラメータ推定

パラメータ	推定値								
Intercept	42.7523	$=\theta^0_0$							
x	-0.7050	$=\theta^1_1$			切片			傾き	
A A1	-12.1731	$=\theta^2_2$	A_1	$\beta^{(1)}_0 =$	30.5793	$=\theta^0_0 + \theta^2_2$	$\beta^{(1)}_1 =$	-0.7050	$=\theta^1_1$
A A2	-10.9359	$=\theta^3_3$	A_2	$\beta^{(2)}_0 =$	31.8164	$=\theta^0_0 + \theta^3_3$	$\beta^{(2)}_1 =$	-0.7050	$=\theta^1_1$
A A3	-10.2036	$=\theta^4_4$	A_3	$\beta^{(3)}_0 =$	32.5488	$=\theta^0_0 + \theta^4_4$	$\beta^{(3)}_1 =$	-0.7050	$=\theta^1_1$
A A4	0		A_4	$\beta^{(4)}_0 =$	42.7523	$=\theta^0_0$	$\beta^{(4)}_1 =$	-0.7050	$=\theta^1_1$

散布図上に 4 本の回帰直線を重ね書きした結果が，図 6.8 に示すようにデフォルトで出力される．

Lsmeans ステートメントによる最小 2 乗平均の推定

SAS の GLM プロシジャは，推定値としての最小 2 乗平均を最初に提供した統計ソフトである．最小 2 乗平均は，GLM プロシジャの lsmeans ステートメントを用いて推定することができる．最小 2 乗平均は，デフォルトでは， x の総平均 $\bar{x}.. = 21.4762$ に対する推定値が用いられるが，任意の x の値を設定して求めることができる．そこで，at オプションで切りの良い $x=20$ と設定する．

SAS プログラム 6.4 /* GLM プロシジャによる最小 2 乗平均の推定 */

```
proc glm data=d01 ;
  class A ;
  model y = x A / ;
  lsmeans A / at x=20 stderr cl ;
run ;
```

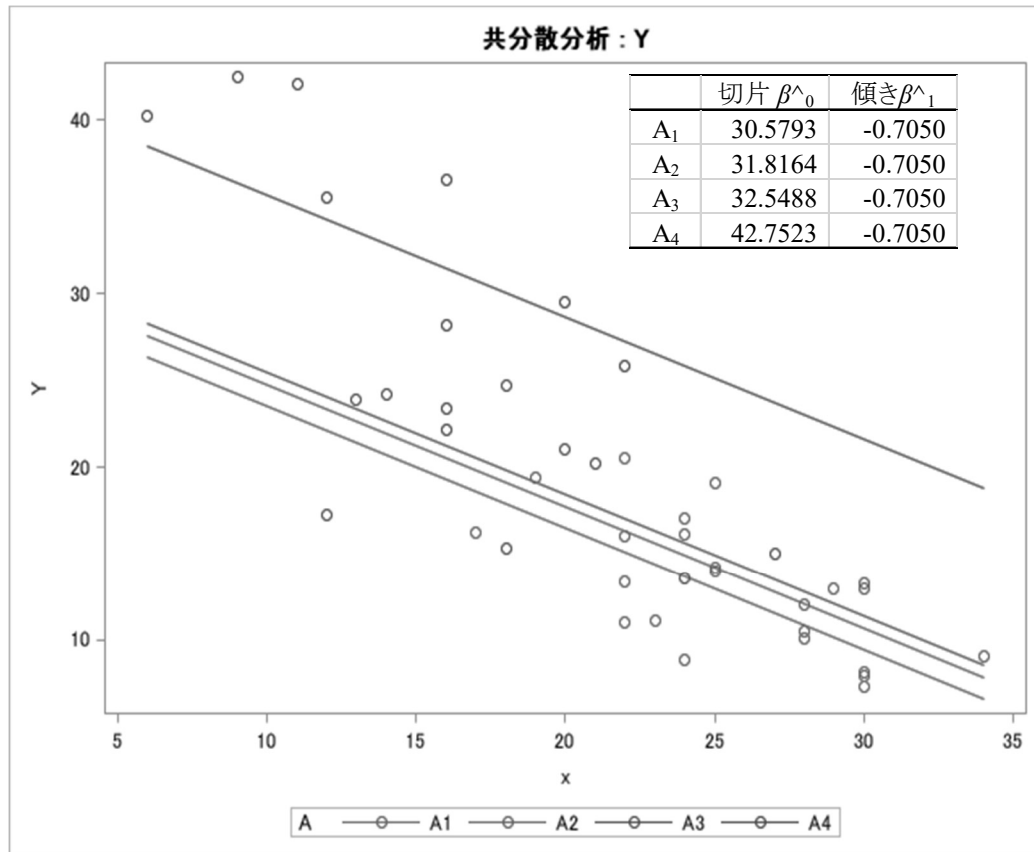



図 6.8 GLM プロシジャによるデフォルトの散布図上の平行な 4 本の回帰直線

表 6.30 に GLM プロシジャによって計算されたにおける $x=20$ 最小 2 乗平均と **stderr** および **ci** オプションによって出力された標準誤差と 95%信頼区間が示されている。この結果は、Excel で作成した表 6.18 に一致する。SAS でこの結果を線グラフで作成するのは難儀であるので、Excel で作成した図 6.5 を図 6.9 に再掲する。

表 6.30 GLM プロシジャによる最小 2 乗平均

最小 2 乗平均 : $x=20$				
A	最小 2 乗平均	標準誤差	95% 信頼限界	
A1	16.4788	1.5193	13.4005	19.5571
A2	17.7159	1.3323	15.0165	20.4154
A3	18.4483	1.3358	15.7417	21.1549
A4	28.6519	1.4821	25.6488	31.6549

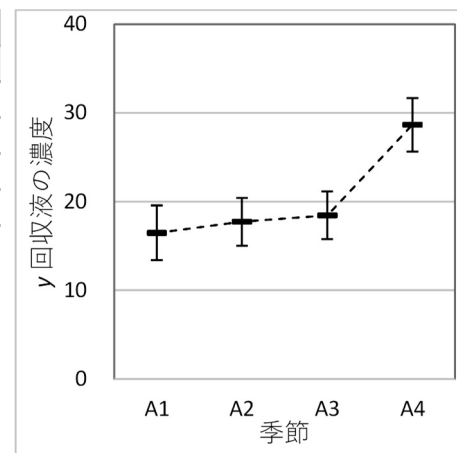


図 6.9 最小 2 乗平均プロット (図 6.5 再掲)

SAS ユーザにとって GLM プロシジャによって計算された最小 2 乗平均は、極めて有用な統計量であり、結果の提示に際し何の断りもなく LSMEANS と称して使われているのが現状である。ただし、SAS・JMP ユーザ以外にとっては、未知の統計量であることを認識し、どのような計算によって推定されたかを Excel などによって再現できることが、他者からの質疑に対応する基礎知識を与える。たとえば、「線形モデルで推定されたパラメータを用いた線形和」と一般論を示し、「共分散分析では、推定された平行な直線上の推定値」などと続け、表 6.30 のような推定値を示し、表 6.18 に示した線形和について解説をする。

Estimate ステートメントによる最小 2 乗平均の差の推定

Lsmeans ステートメントを用いた最小 2 乗平均について示したきたのであるが、ある基準となる水準、たとえば A₄ 水準からの他の水準の差および標準誤差、95%信頼区間を推定したい。Lsmeans ステートメントでは、pdiff オプションを使って総当り式に水準間の差の推定値が得られるが、出力が冗長なので、estimate ステートメントを用いて、水準間の差の係数ベクトルを与える。

SAS プログラム 6.5 /* estimate ステートメントを用いた最小 2 乗平均の差の推定 */

```
proc glm data=d01 ;
  class A ;
  model y = x A ;
  estimate 'A1-A4 full' intercept 0 x 0.0 A 1 0 0 -1 ;
  estimate 'A1-A4' A 1 0 0 -1 ;
  estimate 'A2-A4' A 0 1 0 -1 ;
  estimate 'A3-A4' A 0 0 1 -1 ;
  estimate 'A4-A4' A 0 0 0 0 ;
run ;
```

Estimate ステートメントで、水準間の差の推定に際し、0 となる係数 intercept 0 あるいは x 0.0 などは、省略することができる。表 6.31 に示すように、Excel による追加の計算によって、表 6.19 に完全に一致することが確かめられる。

表 6.31 Estimate ステートメントによる最小 2 乗平均の計算

パラメータ	推定値	標準誤差	t 値	Pr > t	L95%	U95%
A1-A4 full	-12.1731	2.4482	-4.97	<.0001	-17.1337	-7.2125
A1-A4	-12.1731	2.4482	-4.97	<.0001	-17.1337	-7.2125
A2-A4	-10.9359	2.2486	-4.86	<.0001	-15.4920	-6.3798
A3-A4	-10.2036	1.9123	-5.34	<.0001	-14.0783	-6.3288
A4-A4	0	0	.	.	0.0000	0.0000
				$t_{0.05}(42-5)=$	2.0262	

パラメータの推定値、95%信頼区間が得られたならば、その結果をグラフ化する必要がある。統計ソフトで対応できないならば、何らかの手段で作成する努力を怠ってはならない。

第6章 文献索引





奥野ら (1981) - 多変量解析法《改訂版》, 第II部 重回帰分析, 第7章 層別因子を含む回帰分析 195

第6章 索引

あ (1, 0)型 - GLMプロシジャ	213	か 交互作用モデル - 回帰分析	204
Excel - 層別散布図	196	- 主効果モデル	204
Excel の回帰分析 - 共分散分析	204	交互作用を含む - 分散分析表	200
SS1 SS2 SS3 - タイプ I, II, III	212	交互作用を考慮 - 共分散分析	195
estimateステートメン - 最小2乗平均の差	218	- デザイン行列 X	195
LSMEANS - 最小2乗平均	218	さ 最小2乗平均 - LSMEANS	218
lsmeansステートメント - 最小2乗平均	216	- lsmeansステートメント	216
奥野ら(1981) - 共分散分析	203	- GLMプロシジャ	217
- 伝統的な共分散分析	195	- 推定値	217
折れ線グラフ - ひげの長さの設定	208	最小2乗平均の差 - estimateステートメン	218
on - 定数に0を使用	199	SAS - DATAステップ	211
OnDemand SAS - 無償で継続的	195, 211	- テキストデータ	211
か 回帰の平方和 - 水準ごと	198	差の95%信頼区間 - 平行な回帰直線	201
回帰モデル - 主効果モデル	204	残差平方和ヘプーリング - 交互作用	200
回帰直線 - 季節ごと	195	GLMプロシジャ - 共分散分析	195, 211
- 原点を通る	199	- 交互作用 $x \times A$	212
回帰分析 - 交互作用モデル	204	- (1, 0)型	213
- 全データ	197	- 最小2乗平均	217
回帰分析の適用 - 重層的	197	- デザイン行列 X	212
回帰分析の併合 - 水準ごと	197	- パラメータの推定値	213
回帰の平方和 - 共通の傾き	198	$\Sigma(\beta^{\wedge})$ - パラメータの共分散行列	206
各種の推定 - 共通の傾き	205	質的変数 - classステートメント	212
各種の線形和 L - 推定値	206	重層的 - 回帰分析の適用	197
季節ごと - 回帰直線	195	重層的に適用 - 単回帰分析	195
95%信頼区間 - 2水準間の差	209	主効果モデル - 回帰モデル	204
- ひげの長さ	208	- 交互作用モデル	204
- 平行な回帰直線	210	水準ごと - 回帰の平方和	198
- 予測プロファイル	210	- 回帰分析の併合	197
共通の傾き - 回帰の平方和	198	水準ごとの切片 - 平行線	201
- 各種の推定	205	水準間の差 - 予測プロファイル	209
共分散 - パラメータ	206	推定値 - 各種の線形和 L	206
共分散分析 - Excel の回帰分析	204	- 最小2乗平均	216, 217
- 奥野ら(1981)	203	推定値の出力要求 - solutionオプション	212
- 交互作用を考慮	195	切片の推定 - 線形和 L	207
- GLMプロシジャ	195, 211	(0, 1)型ダミー変数 - デザイン行列 X	204
- 線形モデル	195	線グラフ - ひげ付き	201
- ダミー変数	203	線形モデル - 共分散分析	195
- 難解なのか?	203	線形和 L - 切片の推定	207
- 難解な解析法	195	線形和 L の分散 - 2次形式	206
共分散分析の適用 - 難点	196	洗浄水の温度 x - 予測プロファイル	210
共変量の分布 - 均一	197	全データ - 回帰分析	197
均一 - 共変量の分布	197	層別散布図 - Excel	196
classステートメント - 質的変数	212	solutionオプション - 推定値の出力要求	212
原点を通る - 回帰直線	199	た タイプ I, II, III - SS1 SS2 SS3	212
交互作用 - 残差平方和ヘプーリング	200	- 平方和の比較	216
- ダミー変数	204	ダミー変数 - 共分散分析	203
交互作用 $x \times A$ - GLMプロシジャ	212	- 交互作用	204

た	単回帰分析 - 重層的に適用	195	は	パラメータの共分散行列 - $\Sigma(\beta^{\wedge})$	206
	逐次的な回帰分析 - 平方和の増加	205		パラメータの推定値 - GLMプロシジャ	213
	逐次平方和 - 分散分析表	205		ひげ付き - 線グラフ	201
	定数に0を使用 - on	199		ひげの長さ - 95%信頼区間	208
	DATAステップ - SAS	211		ひげの長さの設定 - 折れ線グラフ	208
	- doステートメント	211		分散分析表 - 交互作用を含む	200
	テキストデータ - SAS	211		- 逐次平方和	205
	デザイン行列 X - 交互作用を考慮	195		平行な回帰直線 - 差の95%信頼区間	201
	- GLMプロシジャ	212		- 95%信頼区間	210
	- (0, 1)型ダミー変数	204		平行線 - 水準ごとの切片	201
	デザイン行列の変数 - 別名	209		平方和の増加 - 逐次的な回帰分析	205
	伝統的な共分散分析 - 奥野ら(1981)	195		平方和の比較 - タイプ I, II, III	216
	doステートメント - DATAステップ	211		別名 - デザイン行列の変数	209
	統計ソフト - 理解の妨げ	195	ま	無償で継続的 - OnDemand SAS	195, 211
な	難解なのか? - 共分散分析	203		modelステートメント - モデル式	212
	難解な解析法 - 共分散分析	195		モデル式 - modelステートメント	212
	難点 - 共分散分析の適用	196	や	予測プロファイル - 95%信頼区間	210
	2次形式 - 線形和 L の分散	206		- 水準間の差	209
	2水準間の差 - 95%信頼区間	209		- 洗浄水の温度 x	210
は	パラメータ - 共分散	206	ら	理解の妨げ - 統計ソフト	195

第 6 章 解析用ファイル一覧

サイズ	名前 ^		種類
 280 KB	第06章_01_伝統的共分散分析		Microsoft Excel ワークシート
 194 KB	第06章_02_層別共分散		Microsoft Excel ワークシート
 61 KB	第06章_03_SASの出力編集		Microsoft Excel ワークシート
 2 KB	第06章_03_SASプログラムSAS		テキストドキュメント

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 6 章 伝統的な共分散分析からの脱却

BioStat 研究所(株)
〒105-0014 東京都 港区 芝 1-12-3 の1005
2024 年 2 月 6 日 高橋 行雄
takahashi.stat@nifty.com , FAX : 03-342-8035