

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 16 日

第 10 章 層別因子を含むロジスティック曲線のあてはめ

ロジスティック曲線のあてはめは、反応が 2 値の場合の用量反応関係に対して定式化され、複数の説明変数に対応したロジスティック回帰分析が、2 値反応に対する標準的な解析方法として広く使われている。本章では、2 値反応ではなく、反応が量的変数の場合のロジスティック曲線のあてはめを扱う。ロジスティック分布の (0 から 1) の範囲を (最小値から最大値) に拡張する。層別因子を含む回帰分析を行なった際に、交互作用があり平行な直線のあてはめがためられるような場合が少なからずある。このような場合に、ロジスティック曲線をあてはめることにより、より現実の問題に即した対応が可能となる場合がある。ロジスティック曲線のあてはめは、1 つの変数に対し 1 つのパラメータが付随するような線形モデルに該当しないので、デザイン行列 X を用いたパラメータの推定が行えない。デザイン行列 X の代わりに、ロジスティック曲線式をパラメータで偏微分して得られる微係数行列 Z をデザイン行列 X と同様に扱う解析方法を示す。

第 10 章 目 次

10. 層別因子を含むロジスティック曲線のあてはめ ----- 327

10.1. カドミウムガス曝露後の肺活量の減少 ----- 327

デザイン行列を用いた 3 本の回帰直線の同時あてはめ、 3 本の回帰直線の傾きの差に対する t 検定、 残差平方和の差から交互作用の平方和の算出、
回帰直線の傾きの差を用いた検討、 回帰直線の差の 95%信頼区間、
スプライン曲線による直線のあてはめの妥当性チェック

次ページに続く

非売品，無断複製を禁ずる
連絡先:高橋 行雄, 〒105-0014 東京都 港区 芝 1-12-3 の 1005
takahashi.stat@nifty.com , FAX : 03-342-8035

10.2. 複数のロジスティック曲線の同時あてはめ -----	335
ロジスティック曲線のあてはめの基礎, ソルバーによる残差平方和の 最小化, パラメータの初期値の探索的な推定, パラメータの初期値の 設定の正攻法, Excel のソルバーの活用法	
10.3. ロジスティック曲線の 95%信頼区間 -----	342
ロジスティック曲線式のパラメータに関する偏微分, パラメータに関する 95%信頼区間, ロジスティック曲線の 95%信頼区間, ロジスティック 曲線の差の 95%信頼区間	
10.4. JMP による 3 本のロジスティック曲線の同時あてはめ -----	349
JMP によるロジスティック曲線のあてはめ, JMP による非線形回帰, ロジスティック曲線の 95%信頼区間, JMP によるロジスティック曲線式 のパラメータに関する偏微分	
10.5. SAS の非線形 NLIN プロシジャによる解析 -----	358
SAS/NLIN プロシジャによる解析, NLIN プロシジャの出力結果を Excel の折れ線グラフで	
10.6. 残差線形化法によるロジスティック曲線のパラメータ推定 -----	363
残差を反応とした回帰分析の反復, Wikipedia のガウス・ニュートン法 文献索引, 索引, 解析ファイル一覧 (367)	

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
文献索引, 索引, 解析用ファイル一覧 -----	489

10. 層別因子を含むロジスティック曲線のあてはめ

ロジスティック曲線のあてはめは、反応が 2 値の場合の用量反応関係に対して定式化され、複数の説明変数に対応したロジスティック回帰分析が、2 値反応に対する標準的な解析方法として広く使われている。本章では、2 値反応ではなく、反応が量的変数の場合のロジスティック曲線のあてはめを扱う。ロジスティック分布の（0 から 1）の範囲を（最小値から最大値）に拡張する。層別因子を含む回帰分析を行なった際に、交互作用があり平行な直線のあてはめがためられるような場合が少なからずある。このような場合に、ロジスティック曲線をあてはめることにより、より現実の問題に即した対応が可能となる場合がある。ロジスティック曲線のあてはめは、1 つの変数に対し 1 つのパラメータが付随するような線形モデルに該当しないので、デザイン行列 \mathbf{X} を用いたパラメータの推定が行えない。デザイン行列 \mathbf{X} の代わりに、ロジスティック曲線式をパラメータで偏微分して得られる微係数行列 \mathbf{Z} をデザイン行列 \mathbf{X} と同様に扱う解析方法を示す。

10.1. カドミウムガス曝露後の肺活量の減少

アーミテージら著、椿・椿訳（2001）、「医学研究のための統計的方法、第 9.4 節 群内の回帰」には、カドミウム産業で働いている 84 人の男性を、（ A_1 ：曝露されていない群、 A_2 ：少なくとも 10 年はカドミウムガスに曝露されている群、 A_3 ：10 年はカドミウムガスに曝露されている群）の 3 群に分け、年齢および肺活量（リットル）のデータが示されている。ただし、 A_3 群の男性は、 A_1 群や A_2 群に属する男性よりも年齢が高くなっている。肺活量は、加齢と共に衰えることが知られており、群間の平均年齢の違いを考慮した比較が必要となる。

アーミテージら（2001）には、「第 9.4 節 群内の回帰」に引き続き「第 9.5 節 共分散分析」でこの事例が取り上げられている。「共分散分析」は、共変量と層別因子に交互作用がない場合に適用でき、交互作用がある場合については、回帰直線の傾きの差の検討が必要となると「第 9.4 節 群内の回帰」で強調されている。原著での計算方法は、第 6 章で示した単回帰分析を重層的に使う伝統的な解析方法が示されているので、Excel の回帰分析を用いた線形モデルによる計算法により、回帰直線の傾きの差の検討を行い、ロジスティック曲線のあてはめの必要性について導入する。

表 10.1 に示すのは、群ごとの年齢 x と肺活量 y についての A_1 群（非曝露）44 人、 A_2 群（曝露 < 10 年）28 人、 A_3 群（曝露 > 10 年）12 人のデータである。 A_3 群の平均年齢は、49.75 歳と他の群より 10 歳ほど高くなっている。

表 10.1 カドミウムガス曝露による肺活量の減少（ x ：年齢， y ：肺活量リットル）

A ₁ 群 非曝露						A ₂ 群 曝露 < 10年				A ₃ 群 曝露 > 10年	
<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>
27	5.29	35	4.09	48	3.86	29	5.21	43	4.61	39	4.62
25	3.67	35	4.24	47	4.68	29	5.17	39	4.73	40	5.29
24	5.82	41	3.88	53	4.74	33	4.88	38	4.58	41	5.52
32	4.77	38	4.85	49	3.76	32	4.50	42	5.12	41	3.71
23	5.71	41	4.79	54	3.98	31	4.47	43	3.89	45	4.02
25	4.47	36	4.36	48	5.00	29	5.12	43	4.62	49	5.09
32	4.55	36	4.02	49	3.31	29	4.51	37	4.30	52	2.70
18	4.61	41	3.77	47	3.11	30	4.85	50	2.70	47	4.31
19	5.86	41	4.22	52	4.76	21	5.22	50	3.50	61	2.70
26	5.20	37	4.94	58	3.95	28	4.62	45	5.06	65	3.03
33	4.44	42	4.04	62	4.60	23	5.07	48	4.06	58	2.73
27	5.52	39	4.51	65	4.83	35	3.64	51	4.51	59	3.67
33	4.97	41	4.06	62	3.18	38	3.64	46	4.66		
25	4.99	43	4.02	59	3.03	38	5.09	58	2.88		
42	4.89	41	4.99								
				<i>n</i> =	44			<i>n</i> =	28	<i>n</i> =	12
			年齢 <i>x</i> 平均 =	39.80				<i>x</i> 平均 =	37.79	<i>x</i> 平均 =	49.75
			肺活量 <i>y</i> 平均 =	4.46				<i>y</i> 平均 =	4.47	<i>y</i> 平均 =	3.95
アーミテージら(2001), 表9.4 より引用, ただし, 群のコードは付け替えている.											

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

図 10.1 に示すように、Excel で作成した層別散布図に対し、「近似曲線の追加」で群ごとに「線形近似」を用いて回帰直線を重ね書きしている。回帰直線は、それぞれの群の年齢の範囲（ A_1 ：18～65 歳， A_2 ：21～58 歳， A_3 ：39～65 歳）を反映している。 A_3 群の傾きが A_1 群および A_2 群に比べて大きいようなので、有意な差があるのか検討したい。そのために、 A_1 群を基準とする（0，1）型ダミー変数（ a_2 ， a_3 ）を用いた線形モデル

$$y_i = \theta_0 x_{0,i} + \theta_1 a_{2,i} + \theta_2 a_{3,i} + \theta_3 x_i + \theta_4 a_{2,i} x_i + \theta_5 a_{3,i} x_i + \varepsilon_i, \quad \varepsilon_i \sim \text{正規分布} \quad (10.1)$$

を適用する。このモデルから得られるパラメータの推定値 $\hat{\theta}_4$ が A_1 群と A_2 群の傾きの差の推定値となり、 $\hat{\theta}_5$ が A_1 群と A_3 群の傾きの差の推定値となる。なお、 $\hat{\theta}_0$ は A_1 群の切片、 $\hat{\theta}_1$ は A_1 群と A_2 群の切片の差、 $\hat{\theta}_2$ は A_1 群と A_3 群の切片の差、 $\hat{\theta}_3$ は A_1 群の傾きの推定値となる。

表 10.2 に示すのは、線形モデル式(10.1)に対応するデザイン行列 X を用いた回帰分析の結果である。 A_1 群の傾きは、 $\hat{\beta}_1^{(A_1)} = \hat{\theta}_3 = -0.0306$ ， A_1 群と A_2 群の傾きの差は $\hat{\beta}_1^{(A_2-A_1)} = \hat{\theta}_4$

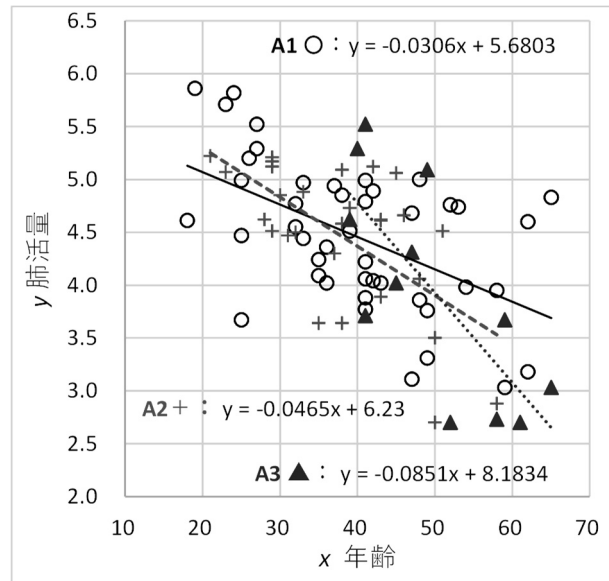


図 10.1 Excel の層別散布図に 3 本の回帰直線の重ね書き

= -0.0159 なので、 A_2 群の傾きは、

$$\left. \begin{aligned} \hat{\beta}_1^{(A_2)} &= \hat{\theta}_3 + \hat{\theta}_4 \\ &= -0.0306 + (-0.0159) = -0.0465 \end{aligned} \right\} \quad (10.2)$$

であり、 A_3 群の傾きは、

$$\left. \begin{aligned} \hat{\beta}_1^{(A_3)} &= \hat{\theta}_3 + \hat{\theta}_5 \\ &= -0.0306 + (-0.0545) = -0.0851 \end{aligned} \right\} \quad (10.3)$$

表 10.2 Excel の回帰分析による 3 本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) off						
No	A	x_0	a_2	a_3	x	a_2x	a_3x	y		自由度	変動	分散	分散比		
1	A_1	1	0	0	27	0	0	5.29	回帰	5	20.1057	4.0211	11.3909		
2	A_1	1	0	0	25	0	0	3.67	残差	78	27.5352	0.3530	$=\sigma^2$		
3	A_1	1	0	0	24	0	0	5.82	合計	83	47.6410				
:															
43	A_1	1	0	0	62	0	0	3.18		係数	標準誤差	t	P -値		
44	A_1	1	0	0	59	0	0	3.03	θ^{\wedge}_0 切片 x_0	5.6803	0.3134	18.1232	0.0000		
45	A_2	1	1	0	29	29	0	5.21	θ^{\wedge}_1 a_2	0.5497	0.5759	0.9546	0.3427		
46	A_2	1	1	0	29	29	0	5.17	θ^{\wedge}_2 a_3	2.5031	1.0418	2.4026	0.0187		
47	A_2	1	1	0	33	33	0	4.88	θ^{\wedge}_3 x	-0.0306	0.0075	-4.0560	0.0001		
:									θ^{\wedge}_4 a_2x	-0.0159	0.0145	-1.0943	0.2772		
71	A_2	1	1	0	46	46	0	4.66	θ^{\wedge}_5 a_3x	-0.0545	0.0211	-2.5866	0.0116		
72	A_2	1	1	0	58	58	0	2.88							
73	A_3	1	0	1	39	0	39	4.62	パラメータの共分散行列 $\Sigma(\theta^{\wedge})=(X^T X)^{-1} \sigma^2$						
74	A_3	1	0	1	40	0	40	5.29	θ^{\wedge}_0	0.0982	-0.0982	-0.0023	0.0023	0.0023	
75	A_3	1	0	1	41	0	41	5.52	θ^{\wedge}_1	-0.0982	0.3316	0.0982	0.0023	-0.0081	-0.0023
:									θ^{\wedge}_2	-0.0982	0.0982	1.0854	0.0023	-0.0023	-0.0215
83	A_3	1	0	1	58	0	58	2.73	θ^{\wedge}_3	-0.0023	0.0023	0.0023	0.0001	-0.0001	-0.0001
84	A_3	1	0	1	59	0	59	3.67	θ^{\wedge}_4	0.0023	-0.0081	-0.0023	-0.0001	0.0002	0.0001
									θ^{\wedge}_5	0.0023	-0.0023	-0.0215	-0.0001	0.0001	0.0004
									切片 x_0	a_2	a_3	x	a_2x	a_3x	

と推定される．これらは，図 10.1 に示した Excel の近似直線で求めた回帰直線の傾きに一致する．

3 本の回帰直線の傾きの差に対する t 検定

表 10.2 に示した Excel の回帰分析の結果から， A_1 群に対する A_2 群の傾きの差は， $\hat{\theta}_4 = \hat{\beta}_1^{(A_2)} - \hat{\beta}_1^{(A_1)} = -0.0159$ であり， t 検定の結果は $p_4 = 0.2772$ と有意な差ではない． A_1 群に対する A_3 群の傾きの差は， $\hat{\theta}_5 = \hat{\beta}_1^{(A_3)} - \hat{\beta}_1^{(A_1)} = -0.0545$ であり， t 検定の結果は $p_5 = 0.0116^*$ と有意な差であることが示されている． A_3 群と A_2 群の傾きの差は，

$$\begin{aligned}\hat{\beta}_1^{(A_3)} - \hat{\beta}_1^{(A_2)} &= (\hat{\theta}_3 + \hat{\theta}_5) - (\hat{\theta}_3 + \hat{\theta}_4) \\ &= \hat{\theta}_5 - \hat{\theta}_4 \\ &= -0.0545 - (-0.0159) = -0.0386\end{aligned}\quad (10.4)$$

として求められる．その分散 $Var(\hat{\theta}_5 - \hat{\theta}_4)$ は，表 10.2 のパラメータの共分散行列 $\Sigma(\hat{\theta})$ を抜粋し有効数字を増やした結果を用いて合成分散の一般式により

	$\Sigma(\hat{\theta})$ の抜粋	
$\hat{\theta}_4$	0.0002116	0.0000570
$\hat{\theta}_5$	0.0000570	0.0004439
	a_2x	a_3x

$$\begin{aligned}Var(\hat{\theta}_5 - \hat{\theta}_4) &= Var(\hat{\theta}_5) - 2Cov(\hat{\theta}_5, \hat{\theta}_4) + Var(\hat{\theta}_4) \\ &= 0.0004439 - 2 \times 0.0000570 + 0.0002116 \\ &= 0.00054162\end{aligned}\quad (10.5)$$

が得られる． A_3 群と A_2 群の傾きの差の t 値は，

$$\begin{aligned}t_{(A_3-A_2)} &= \frac{\hat{\theta}_5 - \hat{\theta}_4}{\sqrt{Var(\hat{\theta}_5 - \hat{\theta}_4)}} \\ &= \frac{-0.0386}{\sqrt{0.00054162}} = -1.6577\end{aligned}\quad (10.6)$$

となり，有意な差ではないことは自明である．なお，このような計算を行わなくとも， A_2 群を基準とした (0, 1) 型ダミー変数 (a_1 , a_3) を用いれば，Excel の回帰分析で A_3 群と A_2 群の傾きの差を直接推定することができる．

残差平方和の差から交互作用の平方和の算出

表 10.2 に示した分散分析表には，3 本の直線の傾きの差の全体に関する交互作用項が含まれていないので，主効果モデルと交互作用モデルの残差平方和の差から交互作用の平方和を算出する．表 10.3 左に示した主効果モデルは，交互作用を含まない分散分析表であり，表 10.3 中に示した交互作用モデルの分散分析表である．この 2 つの分散分析表の残差平方和の差分を用いて，表 10.3 右に示すように交互作用 $A \times x$ を分離する．主効果モデルの残差の平方

和（変動） $S_e^{(主)} = 30.0347$ と交互作用モデルの残差の平方和 $S_e^{(交)} = 27.5352$ の差から交互作用 $A \times x$ の平方和 $S_{A \times x}$

$$\left. \begin{aligned} S_{A \times x} &= S_e^{(主)} - S_e^{(交)} \\ &= 30.0347 - 27.5352 = 2.4995 \end{aligned} \right\} \quad (10.7)$$

を求めることができる。

表 10.3 残差平方和(変動)の差による交互作用 $A \times x$ の分離

主効果モデル			交互作用モデル			交互作用項 分離					
	自由度	変動		自由度	変動		自由度	変動	分散	F 値	p 値
回帰	3	17.6063	回帰	5	20.1057	(A+x)	3	17.6063	5.8688	16.6247	0.0000
残差	80	30.0347	残差	78	27.5352	A×x	2	2.4995	1.2497	3.5402	0.0338
合計	83	47.6410	合計	83	47.6410	残差	78	27.5352	0.3530	1	
						合計	83	47.6410			
S _{A×x} = 主効果モデルの残差変動－交互作用モデルの残差変動＝30.0347 - 27.5352 = 2.4995											

また、モデル全体の平方和 $S_{回帰}^{(交)} = 20.1057$ ， $S_{回帰}^{(主)} = 17.6063$ の差からも交互作用の平方和を

$$\left. \begin{aligned} S_{A \times x} &= S_{回帰}^{(交)} - S_{回帰}^{(主)} \\ &= 20.1057 - 17.6063 = 2.4995 \end{aligned} \right\} \quad (10.8)$$

として求めることもできる。これらを表 10.3 右の分散分析表にまとめると交互作用の F 値が 3.5402 ($p=0.0338^*$) と有意であることが確認される。なお、交互作用を分離した分散分析表では、因子 A および年齢 x については分離せずに、合算したままにしている。これらの主効果を分離することは、交互作用が有意なので、無意味であるだけでなく、判断を錯誤させる危険性がある。交互作用を含む場合の因子 A の主効果は、年齢が $x=0$ 歳，すなわち図 10.1 における A_1 群と A_2 群の切片の差 $\hat{\theta}_1$ ，および， A_1 群と A_3 群の切片の差 $\hat{\theta}_2$ を合わせた平方和から計算されるので，結果の解釈にまったく役立たない。なお，因子 A および年齢 x の平方和の分離法については，第 3 章を参照のこと。

回帰直線の傾きの差を用いた検討

A_1 群(非曝露)と A_3 群(曝露 > 10 年)の傾きに統計的に差があることが判明したので，その差異をどのように説明したらよいのであろうか。図 10.1 の 2 本の回帰直線を見ながら考察してみよう。 A_1 群は，20 歳で肺活量が約 5 リットルあり，加齢と共に肺活量が減少し，65 歳で約 3.8 リットルと 1 リットル強の減少となる。 A_3 群は，カドミウムガスの曝露が 10 年以上の人達なので，最少年齢が 39 歳と A_1 群の 18 歳に比べ高年齢となっている。 A_3 群の 40 歳の肺活量は 5 リットル弱であり， A_1 群の 20 歳の肺活量と同程度であるが，加齢と共に肺活量の減少は大きく，65 歳で約 2.6 リットルに減少し， A_1 群に比べ肺活量が約 1 リットル少ない。その後，さらに加齢を重ねた場合の推定は，外挿となるので，さらに肺活量に差が広がるのかは，データがないので更なる観察が必要である。

交互作用がある場合の考察は、このような説明をするのが精いっぱいであり、傾きの差を用いた説明とはなっていない。あえて用いれば、 A_1 群と A_3 群の回帰直線がクロスする 45 歳あたりから 65 歳における 20 歳分の傾きの差は、

$$20 \times \hat{\theta}_5 = 20 \times (-0.0545) = -1.0900 \text{ リットル}$$

と推定される。グラフから読みとり考察した同様の結果であり、代わり映えない。なお、第 1.3 節で例示したように、 A_1 群の回帰直線と A_3 群の回帰直線について、同一年齢で差の推定および 95%信頼区間を計算することにより、統計量を用いた説明ができるようになる。

回帰直線の差の 95%信頼区間

曝露 10 年以上の A_3 群と非曝露の A_1 群の回帰直線間の肺活量の差の推定値について 95%信頼区間を描き、何歳から肺活量が有意な差になるかを検討する。交互作用モデルの式 (10.1) において、 A_3 群と A_1 群の場合の回帰式を特定し、 A_3 群と A_1 群の回帰式の差を求める。式 (10.1) のダミー変数 a_2 および a_3 に (0, 1) を反映させると、その差が、

$$\begin{array}{rcl} A_3 : \hat{y}^{(A_3)} & = & \hat{\theta}_0 + \hat{\theta}_2 + \hat{\theta}_3 x + \hat{\theta}_5 x \\ -) \quad A_1 : \hat{y}^{(A_1)} & = & \hat{\theta}_0 + \hat{\theta}_3 x \\ \hline (A_3 - A_1) : \hat{d} & = & \hat{\theta}_2 + \hat{\theta}_5 x \end{array} \quad (10.9)$$

のように求められる。差 \hat{d} の分散 $Var(\hat{d})$ は、合成分散の一般式から

$$Var(\hat{\theta}_2 + \hat{\theta}_5 x) = Var(\hat{\theta}_2) + 2Cov(\hat{\theta}_2, \hat{\theta}_5)x + Var(\hat{\theta}_5)x^2 \quad (10.10)$$

で求められる。これらのパラメータの分散および共分散は、表 10.4 に示すように表 10.2 のパラメータの共分散行列 $\Sigma(\hat{\theta})$ から該当するセルを抜き出した 2×2 の行列となる。

表 10.4 パラメータの共分散行列 $\Sigma(\hat{\theta})$ の抜粋

	$\Sigma(\hat{\theta})$ の抜粋	
$\hat{\theta}_2$	1.0854347	-0.0215188
$\hat{\theta}_5$	-0.0215188	0.0004439
	a_3	$a_3 x$

小数点以下の桁数を増やして表示している

表 10.5 に傾きが異なる A_3 群と A_1 群の 2 本の回帰直線の推定値、それらの差、差の分散と SE 、さらに 95%信頼区間および t 検定の結果を示す。65 歳における A_1 群と A_3 群の回帰直線の差の 95%信頼区間を示す。 A_1 群の推定値は $\hat{y}_{x=65}^{(A_1)} = 3.6905$ 、 A_3 群の推定値は $\hat{y}_{x=65}^{(A_3)} = 2.6512$ 、その差は $\hat{d}_{x=65}^{(A_3-A_1)} = -1.0392$ として求められている。

$$\left. \begin{aligned} \hat{y}_{x=65}^{(A_1)} &= \hat{\theta}_0 + \hat{\theta}_3 x \\ &= 5.6803 + (-0.0306) \times 65 = 3.6905 \\ \hat{y}_{x=65}^{(A_3)} &= \hat{\theta}_0 + \hat{\theta}_2 + \hat{\theta}_3 x + \hat{\theta}_5 x \\ &= 5.6803 + 2.5031 + (-0.0306) \times 65 + (-0.0545) \times 65 = 2.6512 \\ \hat{d}_{x=65}^{(A_3-A_1)} &= \hat{y}_{x=65}^{(A_3)} - \hat{y}_{x=65}^{(A_1)} \\ &= 2.6512 - 3.6905 = -1.0392 \end{aligned} \right\} \quad (10.11)$$

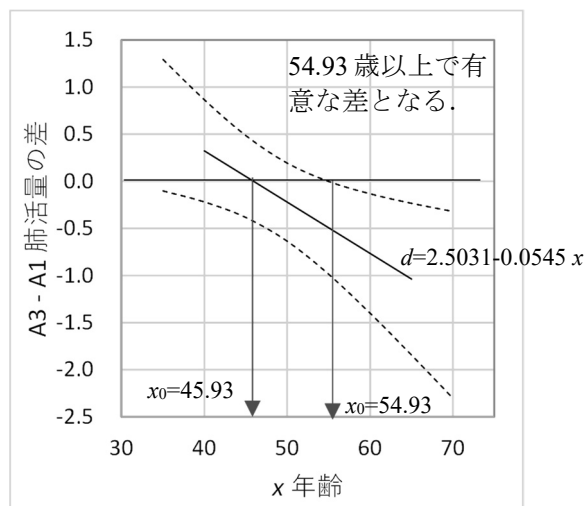
表 10.5 傾きが異なる 2 本の回帰直線の差に関する 95%信頼区間と t 検定

年齢	A_3	A_1	A_3-A_1						
x	\hat{y}_{A_3}	\hat{y}_{A_1}	\hat{d}	$Var(\hat{d})$	$SE(\hat{d})$	$L95\%$	$U95\%$	t	p
35	5.2046	4.6088	0.5957	0.1229	0.3506	-0.1023	1.2938	1.6990	0.0933
40	4.7790	4.4558	0.3232	0.0742	0.2725	-0.2192	0.8656	1.1863	0.2391
45	4.3534	4.3027	0.0507	0.0477	0.2184	-0.3842	0.4856	0.2322	0.8170
50	3.9279	4.1497	-0.2218	0.0434	0.2083	-0.6365	0.1930	-1.0646	0.2904
55	3.5023	3.9966	-0.4943	0.0613	0.2475	-0.9871	-0.0014	-1.9967	0.0493 *
60	3.0768	3.8435	-0.7668	0.1014	0.3184	-1.4006	-0.1330	-2.4085	0.0184 *
65	2.6512	3.6905	-1.0392	0.1636	0.4045	-1.8445	-0.2339	-2.5692	0.0121 *
70	2.2257	3.5374	-1.3117	0.2481	0.4981	-2.3034	-0.3201	-2.6335	0.0102 *
75	1.8001	3.3843	-1.5842	0.3548	0.5956	-2.7700	-0.3984	-2.6598	0.0095 *
# 45.93	4.2742	4.2742	0.0000	0.0452	0.2127	-0.4234	0.4234	0.0000	1.0000
# 54.95	3.5063	3.9980	-0.4917	0.0610	0.2470	-0.9835	0.0000	-1.9908	0.0500 *
# ソルバーを用いた推定値				$t_{0.05}(84-6)=$		1.9908			* $P<0.05$

差の分散は、式 (10.10) の合成分散の一般式から $Var(\hat{\theta}_2 + \hat{\theta}_5 x) = 0.1636$ と計算され、95%信頼区間は、

$$\begin{aligned}
 Var(\hat{\theta}_2 + \hat{\theta}_5 x) &= Var(\hat{\theta}_2) + 2Cov(\hat{\theta}_2, \hat{\theta}_5)x + Var(\hat{\theta}_5)x^2 \\
 &= 1.0854347 + 2 \times (-0.0215186) \times 65 + 0.0004439 \times 65^2 = 0.1636 \\
 (L95\%, U95\%) &= \hat{d}_{x=65}^{(A_3-A_1)} \pm t_{0.05}(84-6) \sqrt{Var(\hat{\theta}_2 + \hat{\theta}_5 x)} \\
 &= -1.0392 \pm 1.9908 \times \sqrt{0.1636} \\
 &= (-1.8445, -0.2339)
 \end{aligned} \quad (10.12)$$

とゼロを含まないので有意な差となっている。もちろん $t = -2.5692$ に対し $p = 0.0121$ と有意な差であることが示されている。表 10.5 の計算結果を図 10.2 に示す。

図 10.2 A_3 群と A_1 群の回帰直線の差に対する 95%信頼区間

年齢が 45 歳以下の場合、 A_3 群は A_1 群よりも推定値は大きく、55 歳以上では逆に推定値が有意に小さくなり、さらに、差が拡大することが読み取れる。年齢が 65 歳の場合 A_1 群は 3.6905 リットル、 A_3 群は 2.6512 リットルと、1.0392 リットルの差となり、 $p = 0.0121$ と有意

な差がある．まとめると，カドミウムガスに 10 年は曝露されていた A_3 群の 40 歳前後では，非曝露の A_1 群と同程度の肺活量であったが，それ以後に急激に開くことが示されている．

表 10.5 の下から 2 行目は，Excel のソルバーを起動し，「目的セル」を 2 群の差「 $A_3 - A_1$ 」のセルを「目標値」として「指定値 = 0」と設定し，「変化セル」に「年齢 x 」のセルを設定して求めた結果である．この結果から，年齢が 45.93 歳で， A_3 群と A_1 群の肺活量が等しくなることが読み取れる．また，回帰直線の上側 95%信頼区間が「0」となる年齢を探索することもできる．Excel のソルバーで「U95%」のセルが「0」となるように「変化セル」に「年齢 x 」のセルを設定して求めた結果が 54.95 歳であることが示されている．その場合の肺活量の差は 0.4917 リットルで， t 検定の結果は， $p = 0.05$ と有意な差があることが確認される．

交互作用がある場合に，このような細かな考察を行なうことになるのだが，表層的な考察としか言いようがない．加齢と共に肺活量は，減少するのであるが直線的ではなく，シグモイド曲線状になると思われるが，得られたデータの範囲からでは確証が持てない．

スプライン曲線による直線のあてはめの妥当性チェック

図 10.3 に示すのは，JMP のグラフビルダーを用いてスプライン曲線を因子 A の 3 群に別々にあてはめた結果である．曲線の滑らかさを調整する λ スライダーで変更し，直線のあてはめの妥当性を視覚的に確認する．群ごとに見た場合には，はっきりとしたシグモイド曲線とは言い難いが，3 群全体から直線のあてはめではなくシグモイド曲線があてはめられそうである．

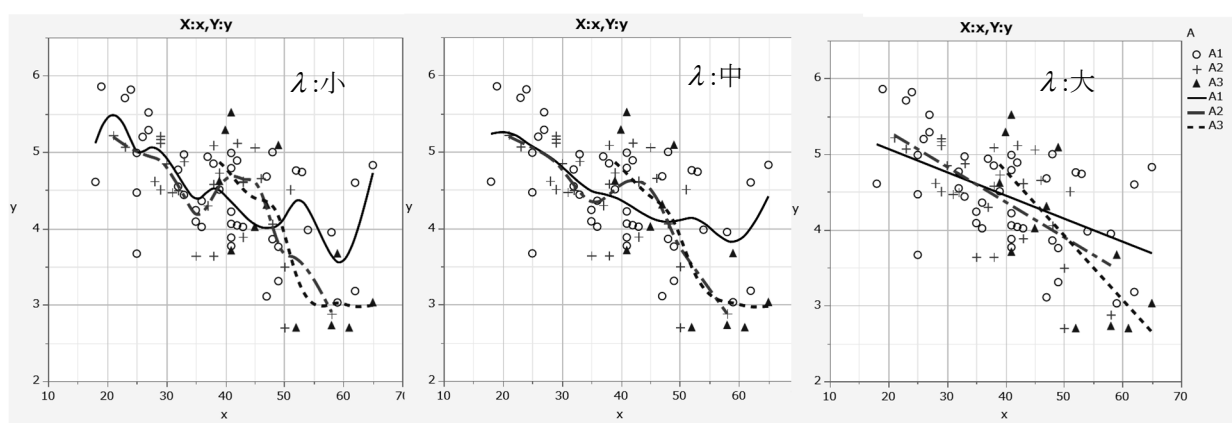


図 10.3 JMP のグラフビルダーによるスプライン曲線のあてはめ

10.2. 複数のロジスティック曲線の同時あてはめ

図 10.1 に示したように直線のあてはめは、得られたデータの範囲内での推定の問題に広く使われている。しかし、年齢と肺活量の関連は、成人となる年齢で最大化し、加齢と共に徐々に低下し、ある最小値に収束するようなシグモイド曲線をあてはめることが自然である。表 10.5 から加齢と共にカドミウムガスに 10 年以上曝露された作業員 A₃ 群の肺活量は、非曝露 A₁ 群に比べて 55 歳で 0.4943 リットル、65 歳で 1.0392 リットルと少なくなっている。外挿は問題があることは承知しているが、75 歳で 1.5 リットルの減少が予測されると言っているのだろうか。いずれにしても回帰直線をあてはめることは、便宜的な方法であると認識しなければならない。代表的なシグモイド曲線は、ロジスティック曲線であるが、他にも第 11 章で示すようにプロビット曲線、ワイブル曲線、ゴンペルツ曲線などもあるが、ここでは立ち入らない。

ロジスティック曲線のあてはめの基礎

ロジスティック曲線のあてはめは、2 値反応に対するロジスティック回帰を連想するが、量的な反応に対しても拡張できる。図 10.1 の散布図から、非曝露 A₁ 群の 18 歳～30 歳の肺活量の平均は、5.0 リットル程度で、60 歳で約 3.8 リットルである。肺活量が、年齢と共に直線的に減少するのではなく、図 10.3 に示したスプライン曲線のあてはめからも、ある最小値に漸近するロジスティック曲線となると想定される。

データ数 n_i に対して反応数 y_i の割合を p_i とした場合に、ロジスティック曲線は、

$$p_i = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} + \varepsilon_i, \quad \varepsilon_i \sim 2\text{項分布} \quad (10.13)$$

として定義され、最小の反応が 0.0、最大の反応が 1.0 となり、誤差に 2 項分布を仮定する。また、分母および分子を $\exp(\beta_0 + \beta_1 x_i)$ で除した式

$$p_i = \frac{1}{1 + \exp[-(\beta_0 + \beta_1 x_i)]} + \varepsilon_i, \quad \varepsilon_i \sim 2\text{項分布} \quad (10.14)$$

も広く使われている。

式 (10.13) を $(\beta_0 + \beta_1 x_i)$ について誤差を無視して解くと、

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_i \quad (10.15)$$

いわゆるロジット $\ln[p_i / (1-p_i)]$ に関する線形式が得られる。この式だけを見るとロジットに対して回帰分析が適用できるように思われるが、式 (10.15) の誤差に正規分布を仮定する

ことができないので、いわゆる反復重み付き回帰による最尤法を適用する必要がある。詳しくは、高橋 (2017), 「一般化線形モデルを Excel で極め活用する ―プロビット法・ロジット法・補 2 重対数法―」, 高橋 (2018), 「二値データの用量反応関係における新たな解析 ―プロビット・ロジット・補 2 重対数の使い分け―」を参照のこと。

式 (10.14) で $n_i=1$ とし、反応が (0, 1) ではない計量値 y_i のロジスティック曲線は、最大値を θ_0 , 最小値を θ_1 として

$$y_i = \theta_1 + \frac{\theta_0 - \theta_1}{1 + \exp[-(\beta_0 + \beta_1 x_i)]} + \varepsilon_i, \quad \varepsilon_i \sim \text{正規分布} \quad (10.16)$$

と定義される。さらに、式 (10.16) を拡張し、3 群共通の最大値を θ_0 , 各群の最小値を θ_k とするロジスティック曲線

$$y_i = \theta_k + \frac{\theta_0 - \theta_k}{1 + \exp[-(\beta_0 + \beta_1 x_i)]} + \varepsilon_i, \quad \varepsilon_i \sim \text{正規分布} \quad (10.17)$$

として拡張できる。ただし、 $\theta_k = (\theta_1 a_1 + \theta_2 a_2 + \theta_3 a_3)$, a_k は A_k 群の場合 1, それ以外は 0 とする (1, 1) 標示型ダミー変数とする。

Excel の回帰分析は、1 つの変数 (2 乗なども 1 つの変数) に 1 つのパラメータを持つ線形和

$$y_i = \sum_{k=0}^p \theta_k x_{k,i} + \varepsilon_i \quad (10.18)$$

として定義された線形モデル式にしか適用できない。式 (10.17) のように線形モデル式とならない場合に対しては、推定したいパラメータ $\hat{\theta}^{(0)}$ におおよその初期値を設定し、推定値 $\hat{y}_i^{(0)}$ を計算する。反応 y_i と推定値 $\hat{y}_i^{(0)}$ の残差 $\hat{\varepsilon}_i^{(0)} = y_i - \hat{y}_i^{(0)}$ の平方和を $S_e^{(0)}$ とし、何らかの方法で残差平方和 $S_e^{(0)}$ を小さくするような微小なパラメータ増減値 $\hat{\delta}^{(0)}$ を推定し、元のパラメータに加え $\hat{\theta}^{(1)} = \hat{\theta}^{(0)} + \hat{\delta}^{(0)}$ とする。再度、推定値 $\hat{y}_i^{(1)}$ を計算し、反応 y_i との残差 $\hat{\varepsilon}_i^{(1)} = \hat{y}_i^{(1)} - y_i$ の残差平方和を $S_e^{(1)}$ とし、さらに、残差平方和 $S_e^{(1)}$ を小さくするような微小なパラメータ増減値 $\hat{\delta}^{(1)}$ を推定する。線形モデルでない場合には、このように逐次的に残差平方和 $S_e^{(m)}$ を減少させ、増減値 $\hat{\delta}^{(m)}$ がほとんどゼロとなるまで、パラメータを変化させる方法が非線形回帰モデルとして定式化されている。本章の末尾の「Wikipedia のガウス・ニュートン法」に示した、非線形のカエリス・メンテン式のあてはめが、典型的な事例である。

かなり、厄介な方法であるが、多くの統計ソフトに、このような「線形モデルではない」という意味での「非線形モデル」のパラメータを推定する関数あるいはアプリケーションが組み込まれている。SAS では「NLIN プロシジャ」、JMP では「非線形回帰」、R では「nls 関数」、Excel でも「ソルバー」というアドイン・アプリケーションを使うことにより、残差平方和 $S_e^{(m)}$ を最小化するようなパラメータの推定が容易にできる。線形モデル式で表せない式

(10.17) の 6 パラメータを Excel の「ソルバー」を用いて推定することから始め、次節でロジスティック曲線のパラメータの 95%信頼区間の推定方法について導入する。なお、芳賀 (2016), 「医薬品開発のための統計解析, 第 3 部 非線形モデル改訂版」が、非線形モデルに関する必読の書である。

ソルバーによる残差平方和の最小化

表 10.6 に示すのは、表 10.1 に示した「カドミウムガス曝露による肺活量の減少」データに対し、式 (10.17) を適用し、Excel のソルバーを用いて推定したパラメータ $\hat{\theta}$ である。この推定値を得るため、表 10.6 右上段に示す初期値を与えて Excel のソルバーで、残差平方和 S_e が最小になるようにパラメータを変化させて得られた結果である。

各群の最小値を求めるために (1, 1) 型ダミー変数を $\mathbf{a}_k = [a_1, a_2, a_3]$ とし、対応するパラメータ $\hat{\theta}_k = [\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3]^T$ との積和、

$$\hat{\theta}_k = \mathbf{a}_k \hat{\theta}_k = \text{Mmult}(\mathbf{a}_k \text{ の範囲}, \hat{\theta}_k \text{ の範囲}) \quad (10.19)$$

によって各群の $\hat{\theta}_k$ を求める。これにより、表 10.6 の $\hat{\theta}_k$ の欄に、A₁ 群の 44 個のデータに対し $\hat{\theta}_1 = 3.4947$, A₂ 群に対し $\hat{\theta}_2 = 2.9612$, A₃ 群に対し $\hat{\theta}_3 = 2.8613$ が設定される。

表 10.6 Excel ソルバーによる 3 本のロジスティック曲線の同時あてはめ

パラメータの推定値										初期値			
		最大値		θ^{\wedge}_0	4.9924	変化 させる セル	誤差分散の推定値				5.0000		
		最小値		a_1	θ^{\wedge}_1		3.4947	$\hat{\sigma}^2 = \frac{29.2457}{84-6}$				4.0000	
				a_2	θ^{\wedge}_2		2.9621					3.5000	
				a_3	θ^{\wedge}_3		2.8613	$= 0.3749$				3.0000	
		回帰パラメータ		$\theta^{\wedge}_4: \beta^{\wedge}_0$	6.0867						4.0000		
				$\theta^{\wedge}_5: \beta^{\wedge}_1$	-0.1273					-0.1000			
		残差平方和		S_e	29.2457	最小化セル	ロジスティック曲線の推定値						
変数		肺活量		最小値	推定値	$y - y^{\wedge}$	非曝露		曝露<10	曝露>10			
No	A	a_1	a_2	a_3	x	y	θ^{\wedge}_k	y^{\wedge}	ε^{\wedge}	x	$y^{\wedge(A1)}$	$y^{\wedge(A2)}$	$y^{\wedge(A3)}$
1	A ₁	1	0	0	27	5.29	3.4947	4.8936	0.3964	15	4.9698		
2	A ₁	1	0	0	25	3.67	3.4947	4.9147	-1.2447	20	4.9503	4.9353	
:										25	4.9147	4.8870	
43	A ₁	1	0	0	62	3.18	3.4947	3.7065	-0.5265	30	4.8520	4.8021	
44	A ₁	1	0	0	59	3.03	3.4947	3.7858	-0.7558	35	4.7476	4.6605	4.6440
45	A ₂	0	1	0	29	5.21	2.9621	4.8229	0.3871	40	4.5885	4.4448	4.4176
46	A ₂	0	1	0	29	5.17	2.9621	4.8229	0.3471	45	4.3769	4.1580	4.1165
:										50	4.1407	3.8378	3.7805
71	A ₂	0	1	0	46	4.66	2.9621	4.0948	0.5652	55	3.9237	3.5436	3.4717
72	A ₂	0	1	0	58	2.88	2.9621	3.3988	-0.5188	60	3.7571	3.3179	3.2347
73	A ₃	0	0	1	39	4.62	2.8613	4.4695	0.1505	65	3.6461	3.1673	3.0766
74	A ₃	0	0	1	40	5.29	2.8613	4.4176	0.8724	70	3.5788	3.0761	2.9809
:										75	3.5404	3.0241	2.9263
83	A ₃	0	0	1	58	2.73	2.8613	3.3196	-0.5896	∞	3.4947	2.9621	2.8613
84	A ₃	0	0	1	59	3.67	2.8613	3.2755	0.3945		θ^{\wedge}_1	θ^{\wedge}_2	θ^{\wedge}_3

A₁ 群の年齢 $x_1 = 27$ における推定値 \hat{y}_1 は,

$$\left. \begin{aligned} \hat{y}_1^{(A_1)} &= \hat{\theta}_1 + \frac{\hat{\theta}_0 - \hat{\theta}_1}{1 + \exp[-(\hat{\beta}_0 + \hat{\beta}_1 x_1)]} \\ &= 3.4947 + \frac{4.9924 - 3.4947}{1 + \exp[-(6.0867 + (-0.1273) \times 27)]} = 4.8936 \end{aligned} \right\} \quad (10.20)$$

として計算されている. 残差 $\hat{\varepsilon}_1$ は,

$$\left. \begin{aligned} \hat{\varepsilon}_1 &= y_1 - \hat{y}_1 \\ &= 5.29 - 4.8936 = 0.3964 \end{aligned} \right\} \quad (10.21)$$

である. これらの計算を $i=1, 2, \dots, 84$ まで繰り返し, 残差平方和 S_e が,

$$\left. \begin{aligned} S_e &= \sum_i \hat{\varepsilon}_i^2 \\ &= \text{SumSq}(\hat{\varepsilon} \text{ の範囲}) = 29.2457 \end{aligned} \right\} \quad (10.22)$$

として計算されている.

表 10.6 右は, 推定されたパラメータ $\hat{\theta}$ を用いて A₁ 非曝露群では, 15 歳から 5 歳刻みで 75 歳までの推定値 $\hat{y}^{(A_1)}$ を計算し, A₂ 曝露<10 年群は 20 歳から, A₃ 曝露>10 年群では 35 歳からと各群の最小年齢を考慮して $\hat{y}^{(A_2)}$ および $\hat{y}^{(A_3)}$ を計算している. これは, 図 10.4 示す Excel の散布図上に推定曲線を重ね書きする際に, 年齢 x が小さい順となっていないと連続した曲線がきれいに描けないためである.

図 10.4 は, 反応 y_i の層別散布図に群ごとのロジスティック曲線を散布図上に重ね書きした結果である. 回帰直線をあてはめた図 10.1 で苦し紛れの解釈をしたのであるが, より現実的な解釈が可能となる. A₁ 非曝露群では, 20 歳で 4.95 リットル, 40 歳で 4.59 リットル, 65 歳で 3.65 リットルとシグモイド曲線状に減少し, 最小値 $\hat{\theta}_1 = 3.4947$ リットルに漸近する.

図 10.1 に示した A₂ 曝露<10 年群の直線回帰では, A₁ 非曝露群の直線の傾きに比べ傾斜が同程度で, カドミウムガスの曝露の影響が少ないとの印象を持つ. これは, 20 歳代での曝露年数が少ない人達も数多く含まれていることも影響しているのかも知れないが, 図 10.4 のロジスティック曲線のあてはめでは, A₃ 曝露>10 年群にほぼ一致した結果になっていて, 同じデータでの分析にもかかわらず, 推定の様子が全く異なる. これは, 年齢が 40 歳ぐらいまでは, カドミウムガスの曝露による肺活量に与える影響が相対的に少ないことの表れであり, A₂ 曝露<10 年群の 50 歳以上の 4 症例の肺活量の分布は, A₃ 曝露>10 年群に類似していることも, 同様のロジスティック曲線の推定になったと思われる.

A₃ 曝露>10 年群の最少年齢は 39 歳であり, 回帰直線のあてはめでは, 肺活量が 5 リットルから 65 歳まで 10 歳あたりの肺活量の減少は, 式 (10.3) で推定された傾き $\hat{\beta}_1^{(A_3)} = -0.0851$ に

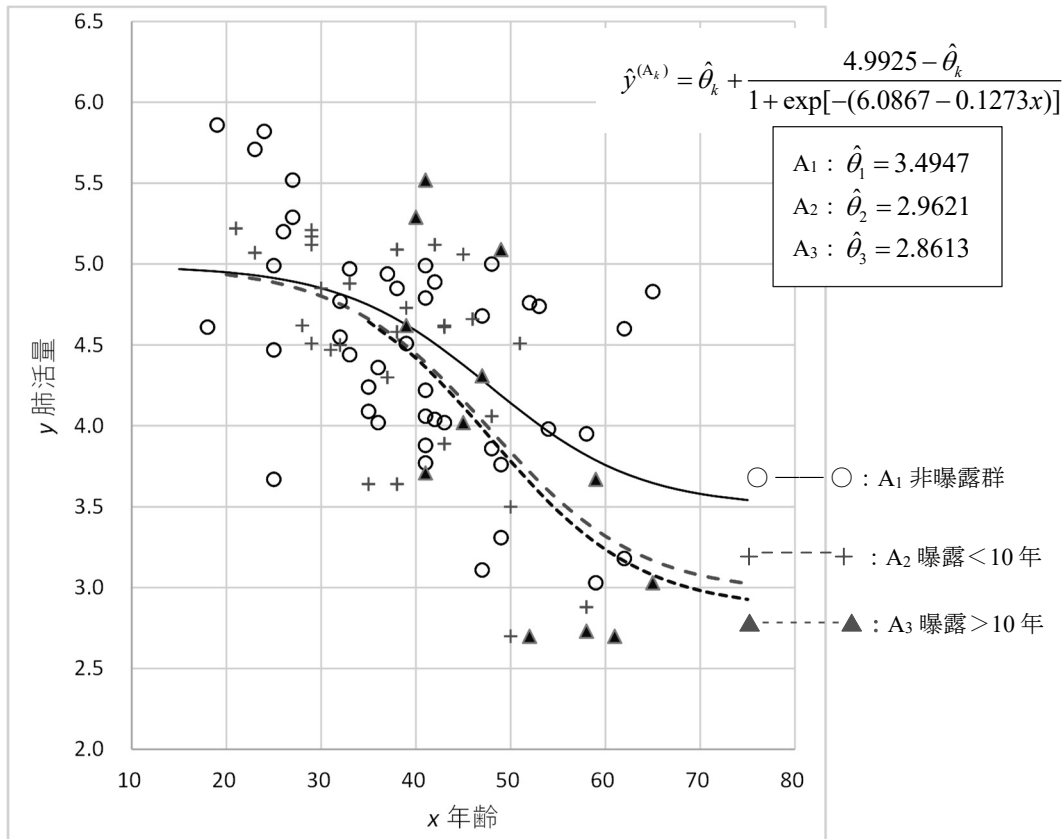


図 10.4 非曝露・曝露群に対するロジスティック曲線のあてはめ

10歳を掛けて -0.851 リットルの減少と推定されている。このこと自体は、データの存在範囲なので問題はないが、どこまで肺活量が落ちるのかについて外挿による推定は、危険である。ロジスティック曲線によるあてはめでは、パラメータとして最小値が $\hat{\theta}_3 = 2.8613$ リットルとして推定され、65歳で3.08リットル、75歳で2.93リットルと徐々に下限 $\hat{\theta}_3 = 2.8613$ リットルに漸近する。

これらの結果を総合して、カドミウムガスの曝露を受けた場合には、加齢と共に肺活量の減少が非曝露群に比べ大きくなり、

$$\left. \begin{aligned} \hat{\theta}_3 - \hat{\theta}_1 &= 2.8613 - 3.4947 = -0.6335 \\ \hat{\theta}_3 - \hat{\theta}_2 &= 2.8613 - 2.9621 = -0.5326 \end{aligned} \right\} \quad (10.23)$$

高年齢になった際に、約0.5リットル強の肺活量の減少と推測される。このような考察は、図10.1の回帰直線のあてはめではなく、元々の現象に忠実な統計モデルの適用によって可能となる。Excelのソルバーは、3本のロジスティック曲線のあてはめによるパラメータの推定が手軽に行えるのであるが、推定値の95%信頼区間を求めるために必要なパラメータの共分散行列の出力がない。そこで、第10.3節で、パラメータの共分散行列の算出方法、および、各種の95%信頼区間の推定方法について示す。

パラメータの初期値の探索的な推定

残差平方和 S_e をソルバーで最小にするため、「変化させるセル」には、何らかの「初期値」を必要とする。そのためには、図 10.1 をよく観察し、最大値 θ_0 は 5.0 ぐらいと推定し、 A_1 群の最小値 θ_1 は、4.0 ぐらい、 A_2 群は 3.5 ぐらい、 A_3 群は 3.0 ぐらい、とあたりが付けられる。

さて、 $\exp()$ 関数の中の線形式のパラメータ $\hat{\beta}_0$ とパラメータ $\hat{\beta}_1$ の初期値は、どのように設定したら良いのだろうか。最初は、 $\hat{\beta}_0=1$ 、 $\hat{\beta}_1=1$ から初めて、どちらかを前後に動かしてみる。形状パラメータ $\hat{\beta}_1$ を大きくしても $\hat{y}^{(A)}$ は、フラットのまま変化しない。 $\hat{\beta}_1=-1$ とすると下方に直線のまま移動、 $\hat{\beta}_1=-0.5$ にしても同じ、 $\hat{\beta}_1=-0.1$ にするとわずかにシグモイド曲線状となる。ソルバーでパラメータを変化させると、求める解が得られる。解が得られた状態で、パラメータを前後に動かしてシグモイド曲線がどの様に変化するか観察することにより、理解が深まるのであるが、場当たりのためであり推奨できない。

パラメータの初期値の設定の正攻法

初期値設定の正攻法は、各パラメータがロジスティック曲線の形状にどのように変化を与えるかを洞察することである。回帰分析における ($\hat{\beta}_0$ が切片、 $\hat{\beta}_1$ が傾き) にとらわれることなく、 $\exp[-(\beta_0 + \beta_1 x_i)]$ のふるまいを指数関数の引数として認識するために、次式のように μ_{LGS} を位置パラメータ、 σ_{LGS} を形状パラメータに変換することを考える。次式、

$$\left. \begin{aligned} -(\beta_0 + \beta_1 x_i) &= \frac{x_i - \mu_{LGS}}{\sigma_{LGS}} \\ -\beta_0 - \beta_1 x_i &= -\frac{\mu_{LGS}}{\sigma_{LGS}} + \frac{1}{\sigma_{LGS}} x_i \end{aligned} \right\} \quad (10.24)$$

の関係から、

$$\left. \begin{aligned} -\beta_0 &= -\frac{\mu_{LGS}}{\sigma_{LGS}}, & -\beta_1 &= \frac{1}{\sigma_{LGS}} \\ \beta_0 &= \frac{\mu_{LGS}}{\sigma_{LGS}}, & \beta_1 &= -\frac{1}{\sigma_{LGS}} \end{aligned} \right\} \quad (10.25)$$

が導き出される。ここで、 μ_{LGS} は、ロジスティック曲線の X 軸方向の中心位置のパラメータであり、 σ_{LGS} は正規分布の標準偏差と同様の形状パラメータである。

表 10.1 に示されている年齢 x のおよその平均から $\hat{\mu}_{LGS} = 40$ とし、図 10.1 の散布図から、X 軸方向のデータの広がりの約 4 分の 1 から $\hat{\sigma}_{LGS} = 10$ と見積れば、

$$\hat{\beta}_0 = \frac{\hat{\mu}_{LGS}}{\hat{\sigma}_{LGS}} = \frac{40}{10} = 4, \quad \hat{\beta}_1 = -\frac{1}{\hat{\sigma}_{LGS}} = -\frac{1}{10} = -0.10 \quad (10.26)$$

が得られる．これらの推測された初期値を設定した結果を図 10.5 に示す．まずまずのあてはめとなっている．

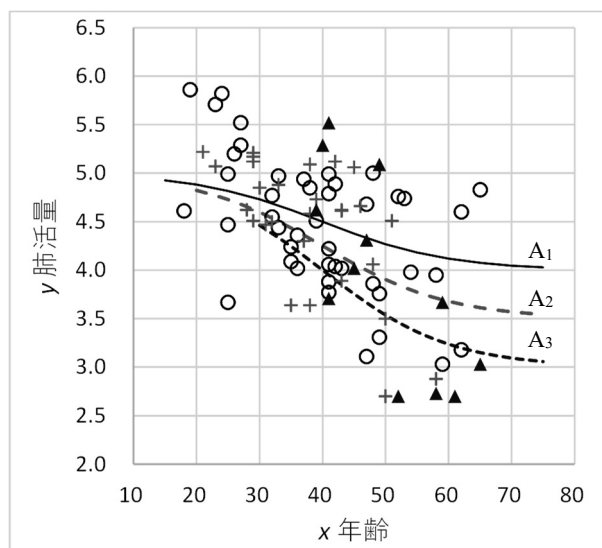


図 10.5 初期パラメータに対するロジスティック曲線

表 10.7 パラメータの初期値

			推定値	
最大値		θ_0^{\wedge}	5.0000	変化 させる セル
最小値	A ₁	θ_1^{\wedge}	4.0000	
	A ₂	θ_2^{\wedge}	3.5000	
	A ₃	θ_3^{\wedge}	3.0000	
回帰パラメータ	$\theta^{\wedge}_4 : \beta^{\wedge}_0$		4.0000	
	$\theta^{\wedge}_5 : \beta^{\wedge}_1$		-0.1000	
残差平方和	S_e		32.8667	最小化セル

ソルバーで S_e を最小にするように $\hat{\theta}$ を変化させると表 10.6 の結果が得られる

Excel のソルバーの活用法

Excel ソルバーは、各種の統計的な課題の解決に役に立つ．表 10.7 に示したのは、ある初期値に対し、取り敢えず計算された残差平方和 S_e が“最小”になるようにパラメータを変化させ、非線形モデルに対する“最小 2 乗解”を求めている．Excel の「回帰分析」を使わなくとも、残差平方和 S_e の計算式 $S_e = \sum_{i=1}^n [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x)]^2$ を Excel シート上で設定すれば、どのような初期値 ($\hat{\beta}_0$, $\hat{\beta}_1$) を与えても、ソルバーで最小 2 乗解を得ることができる．第 12 章の表 12.5, 表 12.7, 表 12.9 に示すのは、Excel の回帰分析の代わりにソルバーで、パラメータの推定を行っている事例である．表 12.2 では、非線形モデルの典型例として知られているミカエリス・メンテン式のパラメータ推定での事例である．

Excel ソルバーは、第 1 章でも使われている．表 1.27 では、回帰直線の 95%信頼区間が、X 軸を横切る x の推定にソルバーが用いられている．これは、 x を与えて回帰の推定値の上側 95%信頼区間の推定値が“0”となるように“ x ”を変化させた結果である．第 5 章の表 5.10 では、“0”ではなく“目標値 6”としている．第 7 章の表 7.9, 表 7.17 も同様の事例である．

10.3. ロジスティック曲線の 95%信頼区間

線形モデルに対する回帰分析に際して、各種の 95%信頼区間の推定に必要なパラメータの共分散行列 $\Sigma(\hat{\beta})$ あるいは $\Sigma(\hat{\theta})$ を用いた活用事例を数多く示してきた。Excel の回帰分析から得られる誤差分散の推定値 $\hat{\sigma}^2$ とデザイン行列 X を用いて、パラメータの共分散行列

$$\Sigma(\hat{\beta}) = (X^T X)^{-1} \hat{\sigma}^2 \quad (10.27)$$

を求めることができた。このような行列計算でパラメータの共分散行列を推定できるのは、式 (10.18) で示した線形モデルの場合に限定されている。式 (10.17) で示したロジスティック曲線は、線形モデルではないので、式 (10.27) に示したデザイン行列 X を用いたパラメータの共分散行列 $\Sigma(\hat{\beta})$ あるいは $\Sigma(\hat{\theta})$ を求めることは、残念ながらできない。

ロジスティック曲線式のパラメータに関する偏微分

線形モデルで表わすことができない非線形モデルの場合には、与えられたモデル式に対し、各パラメータで偏微分した微分係数を新たな変数 z_k とし、 z_k に関する微係数行列を Z とする。これをデザイン行列 X と同様に扱い、

$$\Sigma(\hat{\theta}) = (Z^T Z)^{-1} \hat{\sigma}^2 \quad (10.28)$$

により、パラメータの共分散行列 $\Sigma(\hat{\theta})$ が得られる。線形モデルの $\Sigma(\hat{\beta})$ あるいは $\Sigma(\hat{\theta})$ と同様に非線形の $\Sigma(\hat{\theta})$ を、ロジスティック曲線式のパラメータの共分散行列として用いることにより、各種の 95%信頼区間の推定を行なうことができる。と言うのは簡単だが、実際に Excel で実行するためにはロジスティック曲線の式 (10.17) をパラメータで偏微分をしなければならないし、参考になる書籍も見出すことができない。最近の統計ソフトは、式を与えれば、内部で自動的に偏微分する機能が備わっており、偏微分した結果を出力することもできる。なお、JMP と SAS を用いた場合については、第 10.4 節および第 10.5 節で例示する。

ソルバーであてはめたロジスティック曲線の式 (10.17) にダミー変数を加えた式は、

$$y_i = (\theta_1 a_1 + \theta_2 a_2 + \theta_3 a_3) + \frac{\theta_0 - (\theta_1 a_1 + \theta_2 a_2 + \theta_3 a_3)}{1 + \exp[-(\beta_0 + \beta_1 x_i)]} + \varepsilon_i \quad (10.29)$$

である。この式に対し、全てのパラメータ ($\theta_0, \theta_1, \theta_2, \theta_3, \beta_0, \beta_1$) で偏微分する。このままでは、見通しが悪いので、

$$f(x_i) = \theta_k + (\theta_0 - \theta_k)[1 + g(x_i)]^{-1} \quad (10.30)$$

$$\text{ただし, } \theta_k = \theta_1 a_1 + \theta_2 a_2 + \theta_3 a_3$$

$$g(x_i) = \exp[-(\beta_0 + \beta_1 x_i)]$$

とすると見通しが良くなる。

パラメータに関する偏微分式を以下に示す.

$$\left. \begin{aligned} \frac{\partial f(x_i)}{\partial \theta_0} &= [1 + g(x_i)]^{-1} \\ \frac{\partial f(x_i)}{\partial \theta_1} &= a_1 - a_1 [1 + g(x_i)]^{-1} \\ \frac{\partial f(x_i)}{\partial \theta_2} &= a_2 - a_2 [1 + g(x_i)]^{-1} \\ \frac{\partial f(x_i)}{\partial \theta_3} &= a_3 - a_3 [1 + g(x_i)]^{-1} \end{aligned} \right\} \quad (10.31)$$

$$\left. \begin{aligned} \frac{\partial f(x_i)}{\partial \beta_0} &= (\theta_0 - \theta_k) g(x_i) [1 + g(x_i)]^{-2} \\ \frac{\partial f(x_i)}{\partial \beta_1} &= \frac{\partial f(x_i)}{\partial \beta_0} x_i \end{aligned} \right\} \quad (10.32)$$

これらの偏微分式に対する数値計算は、表 10.8 に示すように Excel を用いればストレスなく計算でき、また計算された微係数行列 \mathbf{Z} を用いた $(\mathbf{Z}^T \mathbf{Z})^{-1}$ も容易に計算することができる.

表 10.8 に示した計算シートは、込み入っているが、列方向に表の下部に示した計算式を第 1 行目に順次入力し、第 84 行目までフィルハンドルで計算式をコピーする. そのためには、 $(\theta_1, \theta_2, \theta_3)$ などダミー変数 a_k のパラメータの参照には、セル位置が変わらないように

$$\begin{aligned} \hat{\theta}_k &:= \text{Mmult} (\text{D14:F14}, \$\text{J\$6}:\text{J\$8}) \\ a_1 \sim a_3, \quad \hat{\theta}_1 \sim \hat{\theta}_3 \end{aligned}$$

のように参照するセル位置にファンクションキー「F4」により「\$」を付ける. この計算は、 $\hat{\theta}_k$ 変数とし A_1 群の場合に $\hat{\theta}_1 = 3.495$ が、 A_2 群の場合に $\hat{\theta}_2 = 2.962$ が、 A_3 群の場合に $\hat{\theta}_3 = 2.861$ が、設定される. 引き続き、 \hat{y}_1 を

$$g(x_1) = \exp[-(\hat{\beta}_0 + \hat{\beta}_1 x_1)] = 0.071,$$

$$\text{推定値 } \hat{y}_1 = \hat{\theta}_k + (\hat{\theta}_0 - \hat{\theta}_k) / [1 + g(x_1)] = 4.894,$$

推定し、残差 $\hat{\varepsilon}_1 = y_1 - \hat{y}_1 = 0.396$ を計算する. 式 (10.31) および式 (10.32) の偏微分式は、Excel の表示上の制約から「 $\partial_ \theta_0$ 」のように接頭語「 $\partial_$ 」にパラメータ名をつけて偏微分した結果の変数名としている.

微係数行列 \mathbf{Z} は、サイズが 84×6 であり、 6×6 の積和行列の逆行列 $(\mathbf{Z}^T \mathbf{Z})^{-1}$ は、Excel の行列関数を用いて次式

$$(\mathbf{Z}^T \mathbf{Z})^{-1} := \text{Minverse} (\text{Mmult} (\text{Transpose} (\mathbf{Z} \text{ の範囲}), \mathbf{Z} \text{ の範囲}))$$

で計算する.

表 10.8 Excel による偏微分式の計算および積和行列の逆数

						θ^{\wedge}	推定値											
						最大値	θ^{\wedge}_0	4.9925	変化 させる セル									
						最小値	A ₁	θ^{\wedge}_1		3.4947								
							A ₂	θ^{\wedge}_2		2.9621								
							A ₃	θ^{\wedge}_3		2.8613								
						回帰パラメータ		$\theta^{\wedge}_4;\beta^{\wedge}_0$		6.0862								
								$\theta^{\wedge}_5;\beta^{\wedge}_1$		-0.1273								
						残差平方和		S_e	29.2457	最小化セル		$\sigma^2=$ 0.3749		$=S_e/(84-6)$				
群		ダミー		年齢		肺活量		— 中間 —		推定値	残差	—————		微係数列行 Z		—————		
No	A	a_1	a_2	a_3	x_i	y_i	θ_k^{\wedge}	$g(x_i)$	y_i^{\wedge}	ε_i^{\wedge}	$\partial \theta_0$	$\partial \theta_1$	$\partial \theta_2$	$\partial \theta_3$	$\partial \beta_0$	$\partial \beta_1$		
1	A ₁	1	0	0	27	5.29	3.495	0.071	4.894	0.396	0.934	0.066	0.000	0.000	0.092	2.492		
2	A ₁	1	0	0	25	3.67	3.495	0.055	4.915	-1.245	0.948	0.052	0.000	0.000	0.074	1.843		
:																		
44	A ₁	1	0	0	59	3.03	3.495	4.145	3.786	-0.756	0.194	0.806	0.000	0.000	0.235	13.838		
45	A ₂	0	1	0	29	5.21	2.962	0.091	4.823	0.387	0.917	0.000	0.083	0.000	0.155	4.506		
46	A ₂	0	1	0	29	5.17	2.962	0.091	4.823	0.347	0.917	0.000	0.083	0.000	0.155	4.506		
:																		
72	A ₂	0	1	0	58	2.88	2.962	3.649	3.399	-0.519	0.215	0.000	0.785	0.000	0.343	19.880		
73	A ₃	0	0	1	39	4.62	2.861	0.325	4.469	0.151	0.755	0.000	0.000	0.245	0.395	15.392		
74	A ₃	0	0	1	40	5.29	2.861	0.369	4.418	0.872	0.730	0.000	0.000	0.270	0.420	16.792		
:																		
84	A ₃	0	0	1	59	3.67	2.861	4.145	3.276	0.394	0.194	0.000	0.000	0.806	0.334	19.690		
		$\theta_k^{\wedge}=Mmult(a \text{ の範囲}, \theta^{\wedge} \text{ の範囲})$								$\partial \theta_0=1/(1+g(x_i))$			$\partial \theta_3=a_3-a_3/(1+g(x_i))$					
		$g(x_i)=\exp(-(\beta_0^{\wedge}+\beta_1^{\wedge}x_i))$								$\partial \theta_1=a_1-a_1/(1+g(x_i))$			$\partial \beta_0=(\theta_k^{\wedge}-\theta_k^{\wedge})g(x_i)/(1+g(x_i))^2$					
		$y_i^{\wedge}=\theta_k^{\wedge}+(\theta_0^{\wedge}-\theta_k^{\wedge})/(1+g(x_i))$								$\partial \theta_2=a_2-a_2/(1+g(x_i))$			$\partial \beta=(\partial \beta_0)*x_i$					

図 10.6 左にシグモイド曲線の共通の最大値および各群の最小値の推定値と 95%信頼区間を示す。最大値は、各群で共通であり 4.99 (4.44, 5.54) リットルと推定されている。各群の最小値の推定値の 95%信頼区間は、極めて広く A_1 非曝露群と A_3 曝露>10 年群の間に全く差が見いだされない。これは、高年齢層のデータが少なく、最大 65 歳までなので、最小値は外挿された結果となり 95%信頼区間が大きく広がっているためである。図 10.6 左には、 A_1 非曝露群と A_2 群および A_3 群の差についての 95%信頼区間を示している。

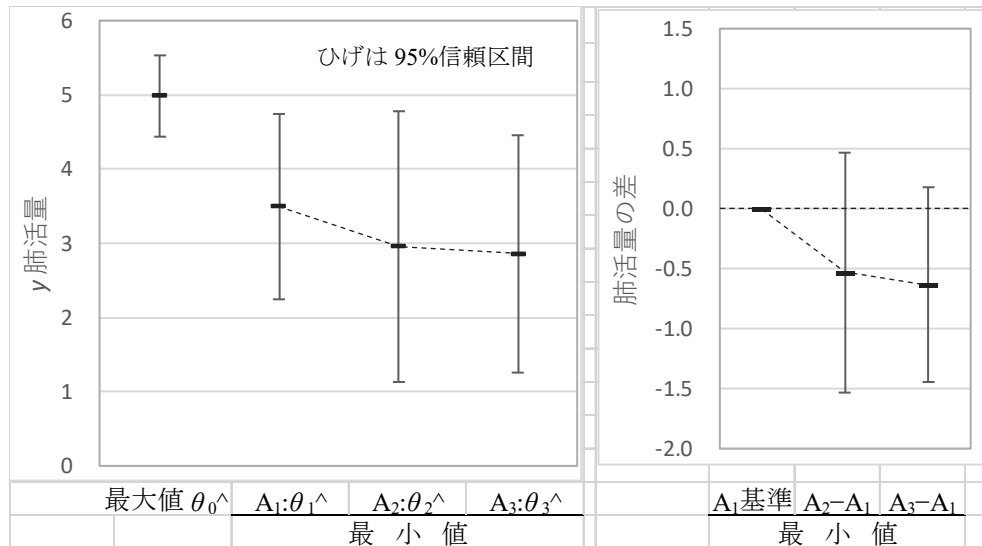


図 10.6 シグモイド曲線の共通の最大値および各層の最小値

パラメータ間の差の推定値は、簡単に求められるが、分散は、表 10.9 に示したパラメータの共分散行列 $\Sigma(\hat{\theta})$ を用いて合成分散の一般式

$$\begin{aligned} Var(\hat{\theta}_2 - \hat{\theta}_1) &= Var(\hat{\theta}_2) - 2Cov(\hat{\theta}_2, \hat{\theta}_1) + Var(\hat{\theta}_1) \\ &= 0.8387 - 2 \times 0.4899 + 0.3957 = 0.2548 \end{aligned} \quad (10.33)$$

$$Var(\hat{\theta}_3 - \hat{\theta}_1) = Var(\hat{\theta}_3) - 2Cov(\hat{\theta}_3, \hat{\theta}_1) + Var(\hat{\theta}_1) = 0.1681 \quad (10.34)$$

によって求める。これらの推定された分散を用いて、表 10.10 に示すように 95%信頼区間を計算し、この結果を図 10.6 右のひげとして示した。

表 10.10 パラメータの差の 95%信頼区間

θ 間の差	推定値	分散	SE	$t_{0.05} \times SE$	L 95%	U 95%
$\theta_1^{\wedge} - \theta_1^{\wedge}$	0	0	0	0.0000	0.0000	0.0000
$\theta_2^{\wedge} - \theta_1^{\wedge}$	-0.5326	0.2545	0.5045	1.0043	-1.5369	0.4718
$\theta_3^{\wedge} - \theta_1^{\wedge}$	-0.6334	0.1681	0.4100	0.8163	-1.4497	0.1829

ロジスティック曲線の最小値の群間差の 95%信頼区間は、群ごとの分散に対し相対的に小さくなり、 A_1 非曝露群に対し有意な差とは言えないものの、さらなるデータの取集により、有意な差となることが期待される。

ロジスティック曲線の 95%信頼区間

回帰直線における 95%信頼区間の表示と同様にロジスティック曲線の 95%信頼区間に挑戦する。図 10.4 に示した 3 本のロジスティック曲線は、式 (10.30) に表 10.6 で推定されたパラメータを代入すると次式となり、

$$y^{(A_k)} = \hat{\theta}_k + \frac{4.9925 - \hat{\theta}_k}{1 + \exp[-(6.0867 - 0.1273x)]} \quad (10.35)$$

$$A_1 : \hat{\theta}_1 = 3.4947, \quad A_2 : \hat{\theta}_2 = 2.9621, \quad A_3 : \hat{\theta}_3 = 2.8613$$

年齢 x を適当な範囲と刻みで設定し肺活量 \hat{y} を推定し、散布図上に重ね書きして描いた。

推定値 \hat{y} の分散 $Var(\hat{y})$ は、表 10.8 で示した微係数行列 \mathbf{Z} から求められたパラメータの共分散行列 $\mathbf{\Sigma}(\hat{\theta})$ に関する、 \hat{y} についての微係数ベクトル \mathbf{z} による 2 次形式

$$Var(\hat{y}) = \mathbf{z} \mathbf{\Sigma}(\hat{\theta}) \mathbf{z}^T \quad (10.36)$$

によって求めることができる。

表 10.11 に A_1 群に対するロジスティック曲線の 95%信頼区間を推定するために、年齢 x を 10 歳から 5 歳刻みで 80 歳まで変化させた場合の推定値 \hat{y} と分散 $Var(\hat{y})$ を式 (10.36) によって

表 10.11 各種のロジスティック曲線の 95%信頼区間

群	—変数—				推定値			z_0	z_1	z_2	z_3	z_4	z_5				
A	a_1	a_2	a_3	x	θ_k	$g(x)$	y^\wedge	$\partial \theta_0$	$\partial \theta_1$	$\partial \theta_2$	$\partial \theta_3$	$\partial \beta_0$	$\partial \beta_1$	$Var(y^\wedge)$	SE	$L95\%$	$U95\%$
A_1	1	0	0	10	3.49	0.01	4.98	0.99	0.01	0.00	0.00	0.01	0.12	0.06	0.24	4.50	5.46
	1	0	0	15	3.49	0.02	4.97	0.98	0.02	0.00	0.00	0.02	0.33	0.05	0.22	4.54	5.40
				:													
	1	0	0	80	3.49	60.0	3.52	0.02	0.98	0.00	0.00	0.02	1.93	0.29	0.54	2.45	4.59
A_2	0	1	0	10	2.96	0.01	4.98	0.99	0.00	0.01	0.00	0.02	0.16	0.05	0.23	4.52	5.43
	0	1	0	15	2.96	0.02	4.96	0.98	0.00	0.02	0.00	0.03	0.45	0.04	0.20	4.56	5.36
				:													
	0	1	0	80	2.96	60.0	3.00	0.02	0.00	0.98	0.00	0.03	2.62	0.65	0.80	1.40	4.60
A_3	0	0	1	10	2.86	0.01	4.98	0.99	0.00	0.00	0.01	0.02	0.17	0.05	0.23	4.52	5.43
	0	0	1	15	2.86	0.02	4.96	0.98	0.00	0.00	0.02	0.03	0.48	0.04	0.20	4.57	5.36
				:													
	0	0	1	80	2.86	60.0	2.90	0.02	0.00	0.00	0.98	0.03	2.75	0.46	0.68	1.54	4.25
A_3	-1	0	1	0	-0.63	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	-0.01	0.01
$-A_1$	-1	0	1	10	-0.63	0.01	-0.01	0.00	0.0	0.00	0.01	0.01	0.05	0.00	0.02	-0.04	0.03
				:													
	-1	0	1	100	-0.63	765	-0.63	0.00	-1.0	0.00	1.00	0.00	0.08	0.17	0.41	-1.44	0.18
A_2	-1	1	0	0	-0.53	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	-0.01	0.01
$-A_1$	-1	1	0	10	-0.53	0.01	0.00	0.00	0.0	0.01	0.00	0.00	0.04	0.00	0.01	-0.03	0.02
				:													
	-1	1	0	100	-0.53	765	-0.53	0.00	-1.0	1.00	0.00	0.00	0.07	0.25	0.50	-1.53	0.47

計算し、95%信頼区間を求めた結果である。計算には、表 10.9 に示したパラメータの推定値および共分散行列を用い、表 10.8 の下段に示した計算式を用いている。

表 10.11 で計算されたロジスティック曲線の推定値 \hat{y} ，および、 $L95\%$ ， $U95\%$ を A_1 非曝露群の散布図上に重ね書きした結果を図 10.7 左に示す。年齢を 10 歳に外挿した場合の 95%信頼区間は (4.50, 5.46) であり、図 10.6 の最大値の 95%信頼区間に漸近し、年齢を 80 歳に外挿した場合の (2.445, 4.593) は、最小値の A_1 群に漸近する。高年齢への外挿に対して、95%信頼区間の増大が著しいことがわかる。同様に、 A_2 曝露<10 年群および A_3 曝露>10 年群についても 95%信頼区間を計算しそれぞれの散布図上に重ね書きした結果を示す。

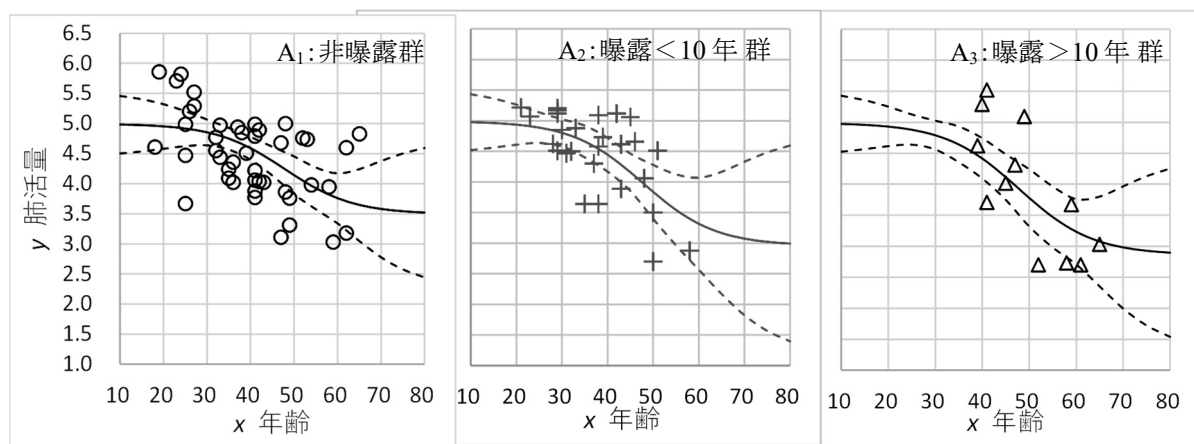


図 10.7 ロジスティック曲線 95%信頼区間

ロジスティック曲線の差の 95%信頼区間

A_1 非曝露群のロジスティック曲線は、表 10.8 から $\hat{\theta}_1 = 3.4947$ に収束し、その分散は、表 10.9 に示したパラメータの共分散行列 $\Sigma(\hat{\theta})$ の対角要素から $Var(\hat{\theta}_1) = 0.3957$ であり、95%信頼区間は (2.2423, 4.7470) となる。 A_3 群の場合は、 $\hat{\theta}_3 = 2.8612$ (1.2595, 4.4628) であり、 A_1 非曝露群と A_3 曝露>10 年群のロジスティック曲線の収束値の差は、表 10.10 から

$$\hat{\theta}_3 - \hat{\theta}_1 = 2.8612 - 3.4947 = -0.6334$$

である。その分散は、

$$Var(\hat{\theta}_3 - \hat{\theta}_1) = Var(\hat{\theta}_3) + 2Cov(\hat{\theta}_3, \hat{\theta}_1) + Var(\hat{\theta}_1) = 0.1681$$

なので、95%信頼区間は (-1.4498, 0.1829) と 0 を超え、有意な差とは言えない。

A_1 群と A_2 群および A_3 群のロジスティック曲線の差は、表 10.11 下段に示すように、(1, 1) 標示型ダミー変数の差として求められる。 A_2 群の場合は、 $(a_1 \ a_2 \ a_3) = (0 \ 1 \ 0)$ で、 A_1

群の場合は, $(a_1 \ a_2 \ a_3) = (1 \ 0 \ 0)$ なので, それらの差 $(A_2 - A_1)$ は, $(a_1 \ a_2 \ a_3) = (-1 \ 1 \ 0)$ となり,

$$\left. \begin{aligned} y^{(A_2 - A_1)} &= \hat{\theta}_3 + \frac{4.9925 - \hat{\theta}_2}{1 + \exp[-(6.0867 - 0.1273x)]} - \hat{\theta}_1 - \frac{4.9925 - \hat{\theta}_1}{1 + \exp[-(6.0867 - 0.1273x)]} \\ &= (\hat{\theta}_2 - \hat{\theta}_1) + \frac{0.0000 - (\hat{\theta}_2 - \hat{\theta}_1)}{1 + \exp[-(6.0867 - 0.1273x)]} \end{aligned} \right\} \quad (10.37)$$

$$A_1 : \hat{\theta}_1 = 3.4947, \quad A_2 : \hat{\theta}_2 = 2.9621, \quad A_3 : \hat{\theta}_3 = 2.8613$$

が Excel シート上で計算されている. このロジスティック曲線は, 最大値が 0.0000, 最小値が $(\hat{\theta}_2 - \hat{\theta}_1) = -0.5326$ のロジスティック曲線となることが理解される. A_3 群と A_1 群の場合は, 最大値は同様に 0.0000, 最小値は, $(\hat{\theta}_3 - \hat{\theta}_1) = -0.6334$ となる.

表 10.11 下段に年齢 x に 0 歳から 10 歳ごとに 100 歳までの差の推定値と 95%信頼区間が計算されているので, 図 10.8 に重ね書きした結果を示す. これらから, カドミウムガスの曝露により, 肺活量は加齢と共に約 0.6 リットル減少すると推定されるが, 高齢者のデータが少ないために有意な差とは言い切れない. 図 10.2 の直線回帰の差からは, 55 歳以上で有意な肺活量の減少となっていたのと対照的である.

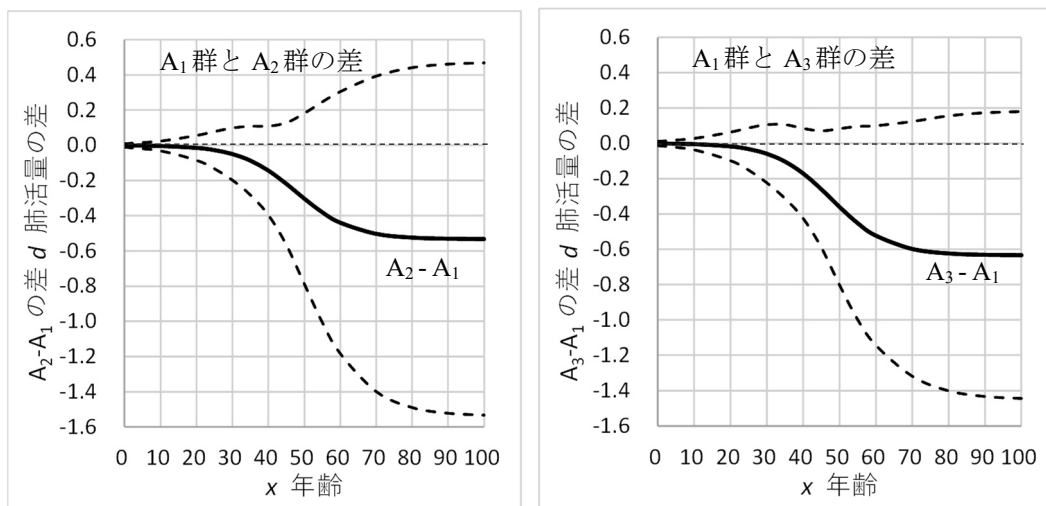


図 10.8 群間の差に対するロジスティック曲線のおよび 95%信頼区間

10.4. JMP による 3 本のロジスティック曲線の同時あてはめ

JMP によるロジスティック曲線のあてはめ

JMP によるロジスティック曲線のあてはめは、「非線形回帰」を用いる。表 10.12 に示す JMP データセットの形式は、表 10.6 に示した Excel のファイルの形式と同様である。JMP の場合も表 10.8 に示したようにパラメータに対し初期値を設定し、あらかじめ肺活量 y_i に対する予測値 \hat{y}_i を JMP ファイル上に設定する必要がある。予測値 \hat{y}_i の設定は、計算式エディタのパラメータの新規作成によって設定し、それらを引用しつつ予測値 \hat{y}_i の計算式を GUI により設定する。パラメータは太字で、変数はイタリックで区別して表示されている。JMP の非線形回帰の詳しい使い方については、芳賀（2016）を参照のこと。

表 10.12 JMP データファイルおよび予測値 \hat{y} の計算式の設定

	No	A	a1	a2	a3	x	y	y^
○	1	1 A1	1	0	0	27	5.29	4.7858
○	2	2 A1	1	0	0	25	3.67	4.8176
:								
○	44	44 A1	1	0	0	59	3.03	4.1301
+	45	45 A2	0	1	0	29	5.21	4.6254
+	46	46 A2	0	1	0	29	5.17	4.6254
:								
+	72	72 A2	0	1	0	58	2.88	3.7128
▲	73	73 A3	0	0	1	39	4.62	4.0500
▲	74	74 A3	0	0	1	40	5.29	4.0000
:								
▲	83	83 A3	0	0	1	58	2.73	3.2837
▲	84	84 A3	0	0	1	59	3.67	3.2602

初期値 \hat{y} :

パラメータ

パラメータの新規作成...

$\theta_0 = 5$
 $\theta_1 = 4$
 $\theta_2 = 3.5$
 $\theta_3 = 3$
 $\beta_0 = 4$
 $\beta_1 = -0.1$

$$\frac{\theta_0 + \theta_1 \cdot a_1 + \theta_2 \cdot a_2 + \theta_3 \cdot a_3}{1 + \text{Exp} \left(- \left(\beta_0 + \beta_1 \cdot x \right) \right)}$$

JMP による非線形回帰

図 10.9 に示すように JMP の「非線形回帰」を起動し、「Y, 応答変数」に「y」を、「X, 予測式列」に「 \hat{y} 」をセットし、実行する。デフォルトの反復回数（60 回）に達しても収束しないので、「実行」を 3 回繰り返す。

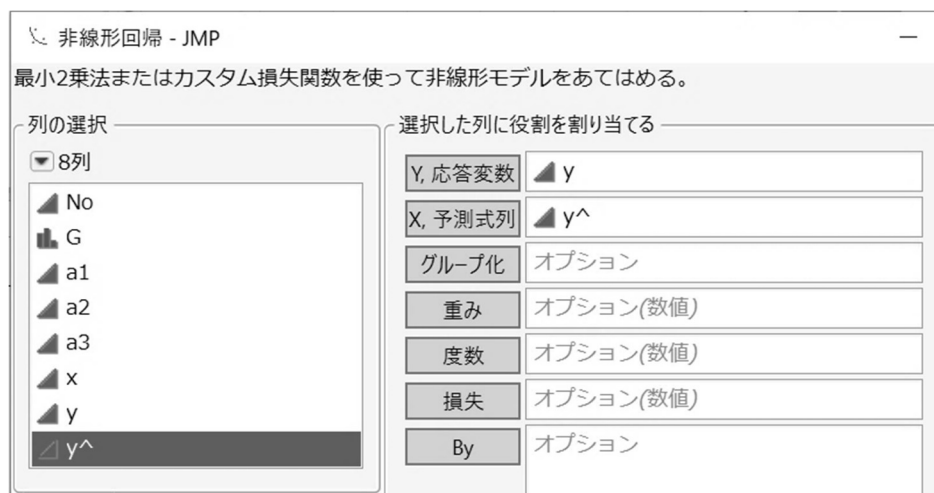


図 10.9 JMP の「非線形回帰」の起動

表 10.13 に JMP の非線形回帰で収束した結果を示す。パラメータの推定値は、表 10.6 に示した Excel のソルバーで得られた結果に少数以下 3 桁まで一致し、標準誤差 SE が出力されている。JMP の非線形回帰で出力された SE は、表 10.9 に示した微分係数行列 \mathbf{Z} を用いて導出されたパラメータの共分散行列 $\Sigma(\hat{\theta})$ の対角要素の平方根をとることにより求められた SE に少数以下 3 桁まで一致する。少数以下 4 桁まで一致しないのは、表 10.14 に示すようにパラメータ間の相関が 1.0 に近い $\text{Corr}(\hat{\beta}_0, \hat{\beta}_1) = -0.9866$ などがあ

表 10.13 JMP の非線形回帰による 3 本のロジスティック回帰のパラメータ推定

解				
SSE	DFE	MSE	RMSE	R2乗
29.2457	78	0.3749	0.6123	0.3861
パラメータ	推定値	近似標準誤差	推定値	SE
θ_0	4.9925	0.2753	4.9925	0.2753
θ_1	3.4947	0.6291	3.4947	0.6290
θ_2	2.9621	0.9159	2.9621	0.9158
θ_3	2.8612	0.8045	2.8613	0.8045
β_0	6.0859	3.8721	6.0862	3.8722
β_1	-0.1272	0.0919	-0.1273	0.0919
解法: 解析 Gauss-Newton			表 10.9 抜粋	

表 10.9 抜粋

り，残差平方和 S_e を最小化するパラメータがなかなか確定できないためであること，また，統計ソフトによるパラメータ探索の打ち切り条件が完全に一致しないためでもある．

表 10.14 パラメータ（推定値）の相関行列

推定値の相関						
	θ_0	θ_1	θ_2	θ_3	β_0	β_1
θ_0	1.0000	-0.6018	-0.4743	-0.5458	-0.8657	0.8082
θ_1	-0.6018	1.0000	0.8505	0.8643	0.7785	-0.8487
θ_2	-0.4743	0.8505	1.0000	0.8437	0.6637	-0.7559
θ_3	-0.5458	0.8643	0.8437	1.0000	0.7623	-0.8376
β_0	-0.8657	0.7785	0.6637	0.7623	1.0000	-0.9866
β_1	0.8082	-0.8487	-0.7559	-0.8376	-0.9866	1.0000

表 10.15 に示すように，表 10.9 で示したパラメータの共分散行列 $\Sigma(\hat{\theta})$ も JMP のバージョン 16（2021 年リリース）から出力されるようになったので，これらの結果を Excel に取り込むことにより，各種の推定が容易に行えるようになったことは喜ばしい．表 10.14 に示したパラメータの相関行列は，多重共線性の確認のために必要である．しかし，パラメータの共分散行列は，内部の計算のためには大活躍しているが，それ自体を見ても役立つ情報に欠けているために，出力されてこなかったと思われる．

表 10.15 パラメータの共分散行列 $\Sigma(\hat{\theta})$

推定値の共分散						
	θ_0	θ_1	θ_2	θ_3	β_0	β_1
θ_0	0.0758	-0.1042	-0.1196	-0.1209	-0.9228	0.0204
θ_1	-0.1042	0.3957	0.4900	0.4374	1.8962	-0.0491
θ_2	-0.1196	0.4900	0.8388	0.6216	2.3537	-0.0636
θ_3	-0.1209	0.4374	0.6216	0.6473	2.3748	-0.0619
β_0	-0.9228	1.8962	2.3537	2.3748	14.9930	-0.3511
β_1	0.0204	-0.0491	-0.0636	-0.0619	-0.3511	0.0084

JMP に限ったわけではないが，現代の統計ソフトでの「非線形回帰」は，厄介なパラメータに関する偏微分式の設定が，内部に組み込まれた自動微分の機能により，必要としなくなっている．JMP では，「予測プロファイル」により，図 10.10 に示すようにシグモイド曲線の 95%信頼区間の表示も手軽に行える．予測プロファイルには，任意の年齢に対する推定値および 95%信頼区間の表示ができるようになっている． A_1 群の 50 歳での予測値は，($a_1=1$, $a_2=0$, $a_3=0$, $x=50$) と設定することにより，図 10.10 左に予測値が 4.14 と表示され，さらに

95%信頼区間が (3.83, 4.45) と示されている. この JMP の「予測プロファイル」で作成したロジスティック曲線の 95%信頼区間は, Excel で作成した図 10.7 左に対応する.

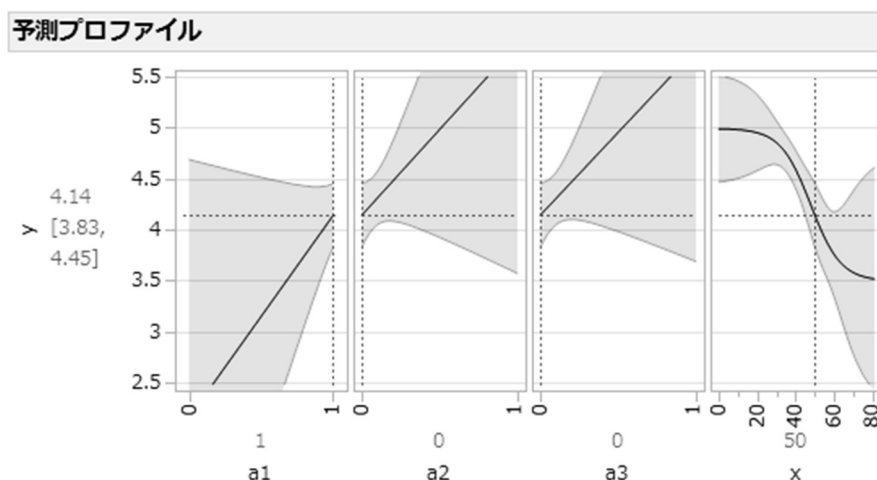


図 10.10 A_1 群のシグモイド曲線と 95%信頼区間の表示

ロジスティック曲線の 95%信頼区間

「非線形回帰」がサポートしているグラフ表示は, 他にも沢山あるが, 図 10.4 および図 10.7 で示したような元のデータ y と予測値 \hat{y} を重ね書きする機能はないので, 予測値 \hat{y} , SE , および, 95%信頼区間を元の JMP データファイルに追加して書き出し, 「グラフビルダー」などで別途作成しても良いし, Excel に予測値 \hat{y} を取り出して, 編集機能が豊富な Excel の散布図を用いることもできる. 図 10.11 に示すように予測値 \hat{y} などの JMP ファイルへの出力に際し, 予測式 (計算式) 付きを選択することにより, 内部の計算式を JMP ファイルで可視化できる. これにより, JMP 内部での計算方法を知ることができる.

予測信頼限界の保存	
個別信頼限界の保存	
計算式の保存	
	予測式の保存
	予測値の標準誤差の保存
	個々の標準誤差の保存
	残差計算式の保存
	予測信頼限界の計算式の保存
	個別信頼限界の計算式の保存
	逆推定計算式の保存
	解の計算式を保存

予測値に対する信頼区間の計算式を含んだ新しいデータ列を作成する。この信頼区間は、期待値に対する推定値の分散から算出される。

図 10.11 JMP の計算式を含む予測値の 95%信頼区間の保存

表 10.16 JMP データファイルに出力された予測値, SE , 95%信頼区間

	No	A	a1	a2	a3	x	y	y [^]	予測値 y [^]	予測値の標準誤差 y [^]	下限M y [^]	上限M y [^]
○	1	1 A1	1	0	0	27	5.29	4.7858	4.8937	0.1293	4.6363	5.1510
○	2	2 A1	1	0	0	25	3.67	4.8176	4.9147	0.1459	4.6243	5.2052
⋮												
○	44	44 A1	1	0	0	59	3.03	4.1301	3.7858	0.1970	3.3937	4.1779
+	45	45 A2	0	1	0	29	5.21	4.6254	4.8230	0.1110	4.6020	5.0439
+	46	46 A2	0	1	0	29	5.17	4.6254	4.8230	0.1110	4.6020	5.0439
⋮												
+	72	72 A2	0	1	0	58	2.88	3.7128	3.3988	0.3385	2.7250	4.0727
▲	73	73 A3	0	0	1	39	4.62	4.0500	4.4695	0.1609	4.1491	4.7898
▲	74	74 A3	0	0	1	40	5.29	4.0000	4.4176	0.1631	4.0930	4.7423
⋮												
▲	83	83 A3	0	0	1	58	2.73	3.2837	3.3197	0.2488	2.8243	3.8150
▲	84	84 A3	0	0	1	59	3.67	3.2602	3.2755	0.2543	2.7693	3.7818

新たな列として出力された「予測値 y^{\wedge} 」について、計算式エディタを起動して内容を確認する。表 10.16 に示されている数値の裏には、図 10.12 に示すような計算式が埋め込まれていて、最初のデータ ($a_1=1$, $a_2=0$, $a_3=0$, $x=27$) を式に代入した結果とし、 $\hat{y}_1 = 4.8937$ が表示されている。

a1 • 3.4946507199
 + a2 • 2.9620974482
 + a3 • 2.8612283464

$$\hat{y}_i = (\hat{\theta}_1 a_{1,i} + \hat{\theta}_2 a_{2,i} + \hat{\theta}_3 a_{3,i}) + \frac{\hat{\theta}_0 - (\hat{\theta}_1 a_{1,i} + \hat{\theta}_2 a_{2,i} + \hat{\theta}_3 a_{3,i})}{1 + \exp[-(\hat{\beta}_0 + \hat{\beta}_1 x_i)]}$$

$$+ \frac{\left(4.9924878393 - \left(a1 \cdot 3.4946507199 + a2 \cdot 2.9620974482 + a3 \cdot 2.8612283464 \right) \right)}{\left(1 + \text{Exp} \left(- \left(6.0859239471 + -0.127248413 \cdot x \right) \right) \right)}$$

図 10.12 「予測値 y^{\wedge} 」に埋め込まれている計算式

「予測値の標準誤差 y^{\wedge} 」に含まれている計算式を確認した結果を図 10.13 に示す。ページ全体をおおう目もくらむような計算式となっている。左端に $\sqrt{\quad}$ を縦に長く引き伸ばした表示があるので、予測値 y^{\wedge} の分散の平方根を取り、 SE の計算がされている。「Vec Quadratic ([6×6], Matrix(…))」となっているのは、以下の 2 次形式による計算を、パラメータの共分散行列 $\Sigma(\hat{\theta})$ を用い、予測値 y^{\wedge} についてのパラメータによる偏微分式に推定されたパラ

<p>Vec Quadratic Matrix</p>	<p>2 次形式の計算の関数 [6×6] は、表 10.15 に示した パラメータの共分散行列</p>	<div style="border: 1px solid black; padding: 10px; margin-bottom: 10px;"> $\frac{\partial f(x_i)}{\partial \theta_0} = [1 + g(x_i)]^{-1}$ $\frac{\partial f(x_i)}{\partial \theta_1} = a_{1,i} - a_{1,i}[1 + g(x_i)]^{-1}$ $\frac{\partial f(x_i)}{\partial \theta_2} = a_{i,2} - a_{i,2}[1 + g(x_i)]^{-1}$ $\frac{\partial f(x_i)}{\partial \theta_3} = a_{i,3} - a_{i,3}[1 + g(x_i)]^{-1}$ $\theta_k = \theta_1 a_{1,i} + \theta_2 a_{2,i} + \theta_3 a_{3,i}$ $g(x_i) = \exp[-(\beta_0 + \beta_0 x_i)]$ </div> <div style="border: 1px solid black; padding: 10px;"> $\frac{\partial f(x_i)}{\partial \beta_0} = (\theta_0 - \theta_k)g(x_i)[1 + g(x_i)]^{-2}$ $\frac{\partial f(x_i)}{\partial \beta_1} = \frac{\partial f(x_i)}{\partial \beta_0} x_i$ </div>
$\begin{pmatrix} \frac{1}{1 + \exp\left(-\left(6.0859263112 + -0.127248475 \cdot x\right)\right)} \\ a1 + \frac{-a1}{1 + \exp\left(-\left(6.0859263112 + -0.127248475 \cdot x\right)\right)} \\ a2 + \frac{-a2}{1 + \exp\left(-\left(6.0859263112 + -0.127248475 \cdot x\right)\right)} \\ a3 + \frac{-a3}{1 + \exp\left(-\left(6.0859263112 + -0.127248475 \cdot x\right)\right)} \\ -1 \\ \cdot \exp\left(-\left(6.0859263112 + -0.127248475 \cdot x\right)\right) \\ - \left(4.992487734 - \left(a1 \cdot 3.4946511645 + a2 \cdot 2.9620980914 + a3 \cdot 2.8612289403\right)\right) \\ \cdot \left(1 + \exp\left(-\left(6.0859263112 + -0.127248475 \cdot x\right)\right)\right)^{-2} \\ \left(\exp\left(-\left(6.0859263112 + -0.127248475 \cdot x\right)\right)\right) \\ \cdot -x \\ - \left(4.992487734 - \left(a1 \cdot 3.4946511645 + a2 \cdot 2.9620980914 + a3 \cdot 2.8612289403\right)\right) \\ \cdot \left(1 + \exp\left(-\left(6.0859263112 + -0.127248475 \cdot x\right)\right)\right)^{-2} \end{pmatrix}$		

図 10.13 「予測値の標準誤差 y^{\wedge} 」に埋め込まれている計算式

メータ代入した計算式が示されている．実際には内部での計算なので，表 10.17 に示す Excel での計算結果と対比されたい．

表 10.17 JMP データファイルに出力された予測値，SE，95%信頼区間（表 10.8 抜粋）

A ₁	—変数—				推定値		z_0	z_1	z_2	z_3	z_4	z_5					
No	a_1	a_2	a_3	x	θ_k	$g(x)$	y^\wedge	$\partial \theta_0$	$\partial \theta_1$	$\partial \theta_2$	$\partial \theta_3$	$\partial \theta_4$	$\partial \theta_5$	$Var(y^\wedge)$	SE	L95%	U95%
1	1	0	0	27	3.49	0.07	4.89	0.93	0.07	0.00	0.00	0.09	2.49	0.02	0.13	4.64	5.15
$\theta_k^\wedge = \text{Mmult}(a \text{ の範囲}, \theta^\wedge \text{ の範囲})$								$\partial \theta_0 = 1/(1+g(x_i))$				$\partial \theta_3 = a_3 - a_3/(1+g(x_i))$					
$g(x_i) = \exp(-(\beta_0^\wedge + \beta_1^\wedge x_i))$								$\partial \theta_1 = a_1 - a_1/(1+g(x_i))$				$\partial \beta_0 = (\theta_0^\wedge - \theta_k^\wedge)g(x_i)/(1+g(x_i))^2$					
$y_i^\wedge = \theta_k^\wedge + (\theta_0^\wedge - \theta_k^\wedge)/(1+g(x_i))$								$\partial \theta_2 = a_2 - a_2/(1+g(x_i))$				$\partial \beta_1 = (\partial \beta_0) * x_i$					

標準誤差 SE の計算は，偏微分した結果を

$$z = [0.93 \quad 0.07 \quad 0.00 \quad 0.00 \quad 0.09 \quad 2.49]$$

と微分係数ベクトル z とし，表 10.15 に示したパラメータの共分散行列 $\Sigma(\hat{\theta})$ を用い

$$SE = \sqrt{z \Sigma(\hat{\theta}) z^T} = 0.1293 \quad (10.38)$$

として計算している．

表 10.16 に示した予測値を用い，図 10.4 を JMP のグラフビルダーを用いて再現してみよう．図 10.14 に示すように Y 軸に肺活量 y ，X 軸に年齢 x をセットし，予測値 y^\wedge を Y 軸にさらに加える．因子 A を「重ね合わせ」にセットし，記号を小さなドットに変更する．予測値を外挿するために，表 10.16 に手作業でデータを追加し，予測値 y^\wedge の計算結果を得ている．

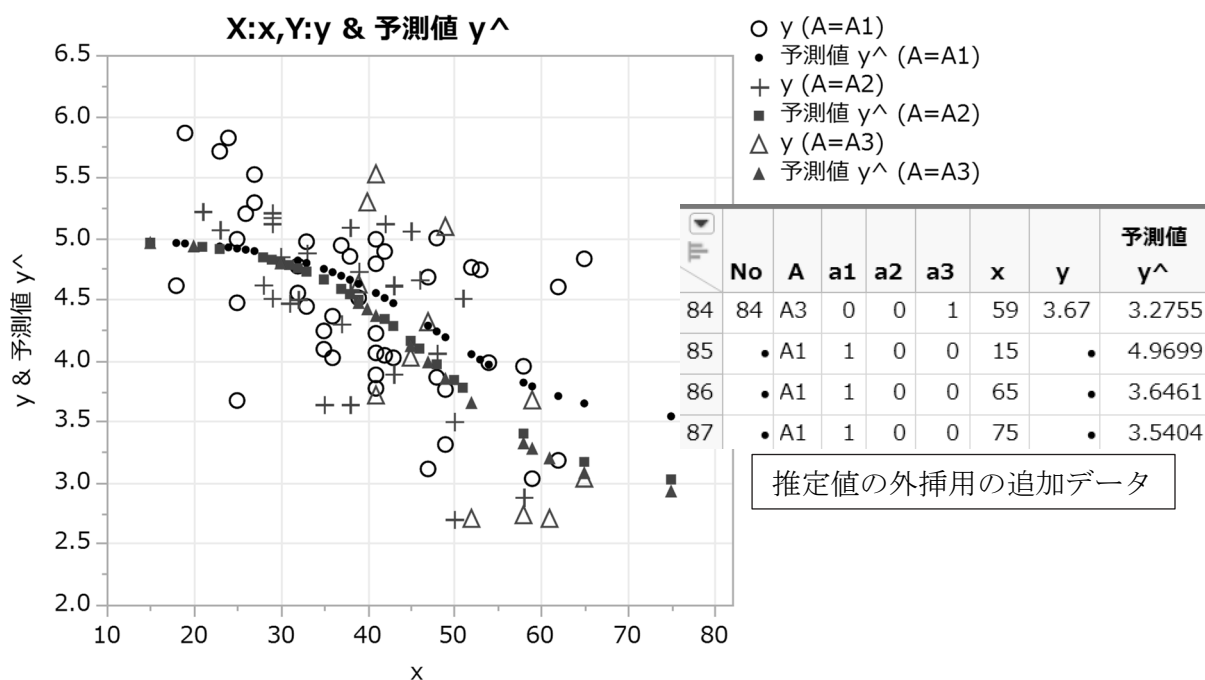
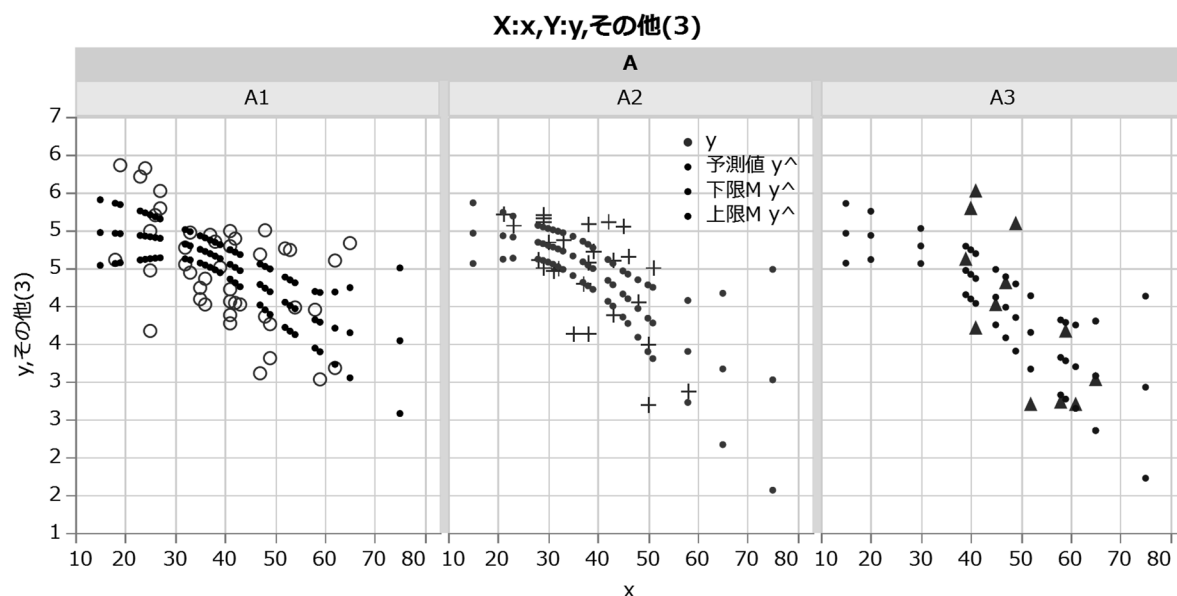


図 10.14 散布図へロジスティック曲線の重ね合わせ

図 10.14 で 95%信頼区間を重ね書きしても互いに重なり合って、判別がつかなくなるので、図 10.15 に示すように、X 軸方向で因子 A の 3 群を分割表示し、95%信頼区間を重ね書きした結果を示す。



Where(11行が除外されています。)

図 10.15 グラフビルダーによる 95%信頼区間入りロジスティック曲線のあてはめ

JMP によるロジスティック曲線式のパラメータに関する偏微分

最近の統計ソフトでは、非線形回帰による解析に不可欠なパラメータに関する偏微分を自動的に行う機能が組み込まれており、非線形回帰が、線形回帰と同様に気楽に使えるようになったことはうれしい限りである。JMP の「非線形回帰」では、どのように偏微分が行われているかをログに出力する機能があるので、以下に示す。

予測式

$$\begin{aligned} & \theta_1 * a_1 + \theta_2 * a_2 + \theta_3 * a_3 + (T\#54 = \theta_0 - (\theta_1 * a_1 + \theta_2 * a_2 + \theta_3 * a_3)) / \\ & \text{First}(T\#51 = 1 + \text{First}(T\#52 = \text{Exp}(-(\beta_0 + \beta_1 * :x)), \\ & T\#55 = T\#52 * -:x), T\#53 = -1 * T\#52 \end{aligned}$$

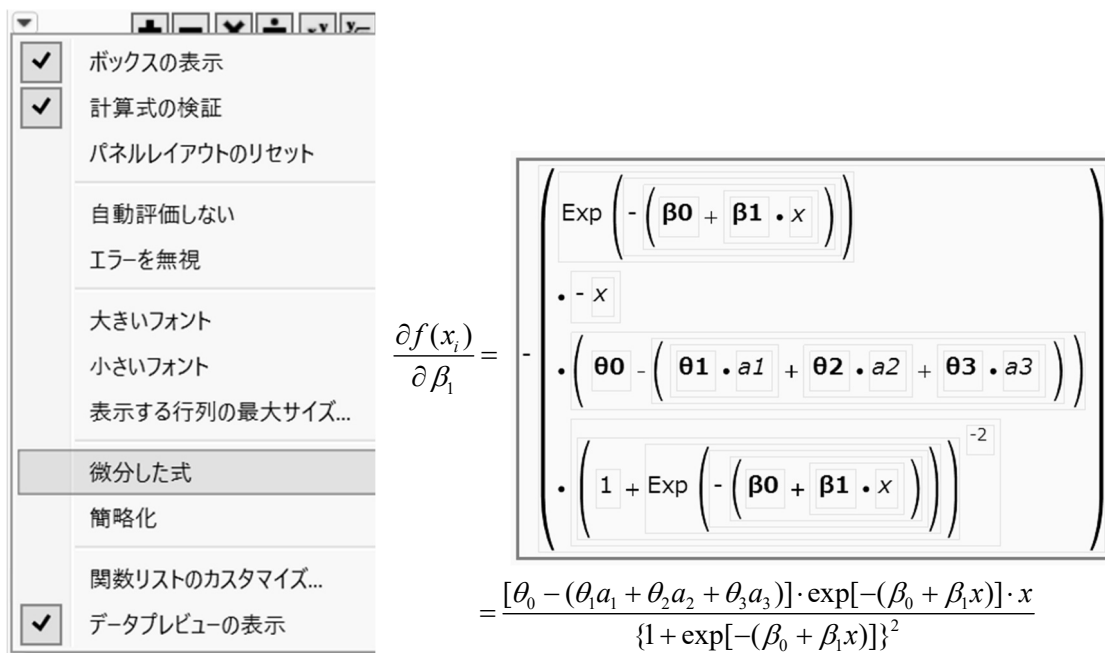
予測式をパラメータで微分した式

$$\begin{aligned} & \{ 1 / T\#51, & \rightarrow \theta_0 \text{ で偏微分した式} \\ & :a_1 + -:a_1 / T\#51, & \rightarrow \theta_1 \text{ で偏微分した式} \\ & :a_2 + -:a_2 / T\#51, & : \\ & :a_3 + -:a_3 / T\#51, & : \end{aligned}$$

$-(T\#53 * T\#54 * T\#51 ^{(-2)}), \quad \rightarrow \beta_0 \text{ で偏微分した式}$
 $-(T\#55 * T\#54 * T\#51 ^{(-2)})\} \quad \rightarrow \beta_1 \text{ で偏微分した式}$

ログに出力された予測式および偏微分式は、内部計算の過程を示しているの、難解であり、数式としての微分式にするのは煩わしい。そこで、JMP データファイル上で GUI による偏微分する機能を使うことを推奨する。

- 手順 1) 偏微分したいパラメータについて新たな変数 ∂_{β_1} を列に追加する。
- 手順 2) 予測式 \hat{y} の計算式をコピーする。
- 手順 3) 変数 ∂_{β_1} のブランクの計算式を開き、ペーストする。
- 手順 4) ペーストされた予測式 \hat{y} のパラメータ β_1 を選択する。
- 手順 5) プルダウンメニューから「微分した式」を選択する。
- 手順 6) 計算式エディターに微分した式が表示され、変数 ∂_{β_1} 列に微分した計算結果が埋め込まれる。



$$\frac{\partial f(x_i)}{\partial \beta_1} = - \left(\begin{array}{l} \text{Exp} \left(- \left(\beta_0 + \beta_1 \cdot x \right) \right) \\ \cdot -x \\ \cdot \left(\theta_0 - \left(\theta_1 \cdot a_1 + \theta_2 \cdot a_2 + \theta_3 \cdot a_3 \right) \right) \\ \cdot \left(1 + \text{Exp} \left(- \left(\beta_0 + \beta_1 \cdot x \right) \right) \right)^{-2} \end{array} \right)$$

$$= \frac{[\theta_0 - (\theta_1 a_1 + \theta_2 a_2 + \theta_3 a_3)] \cdot \exp[-(\beta_0 + \beta_1 x)] \cdot x}{\{1 + \exp[-(\beta_0 + \beta_1 x)]\}^2}$$

図 10.16 JMP の「微分した式」で得た $\hat{y} = f(x_i)$ の β_1 に関する偏微分式

図 10.16 に示した $\hat{y} = f(x_i)$ の β_1 に関する偏微分式は、図 10.13 に示した「予測値の標準誤差 \hat{y} 」の計算式の中で用いられている。

10.5. SAS の非線形 NLIN プロシジャによる解析

SAS には、非線形回帰のための NLIN プロシジャが古くからサポートされていたが、シグモイド曲線の式に対する偏微分式も明示的に自ら与える必要があり、手軽に使える状況ではなかった。その後、JMP と同様に自動偏微分の機能が加わり、気楽に使えるようになった。ここに示すのは、無償で継続的に使える OnDemand SAS による結果である。

非線形モデルの解析法については、自由にダウンロードできる SAS の NLIN プロシジャの解説書 SAS Institute (2014), 「The NLIN Procedure」が最も優れている。全 118 ページで、「Getting Started: NLIN Procedure」節に、線形モデルと非線形モデルの識別方法に引き続き、サンプルプログラムと実行結果が詳細に示されている。また、非線形モデルの解析法についても丁寧な説明があり、非線形モデルについて学習するために適している。

SAS/NLIN プロシジャによる解析

表 10.1 で示した Excel ファイル上のデータをスペース区切りのテキストを SAS の DATA ステップで、SAS データセット化する。データセット化は、一行ごとに A₁ 群の 3 組の (x_i, y_i) , A₂ 群の 2 組, A₃ 群の 1 組の順に読込、15 行目でデータリストが尽きるまで読み込む。13 行目以後に欠測値「.」もデータセットに書き込まれるが、解析に際しては無視される。

SAS プログラム 10.1

```
TITLE1 '層別シグモイド.SAS << カドミウムガスの曝露 >>' ;
```

```
DATA d01 ;
  input n @ ;
  A ="A1"; a1=1; a2=0; a3=0;
  do i=1,2,3 ;
    input x y @ ; output ; end;
```

```
  A ="A2"; a1=0; a2=1; a3=0;
  do i=1,2 ;
    input x y @ ; output ; end;
```

```
  A ="A3"; a1=0; a2=0; a3=1;
  input x y @ ; output ;
```

```
/* ----- A1 ----- ----- A2 ----- -- A3 -- */
/* x y x y x y x y x y x y x y */
datalines ;
1 27 5.29 35 4.09 48 3.86 29 5.21 43 4.61 39 4.62
2 25 3.67 35 4.24 47 4.68 29 5.17 39 4.73 40 5.29
3 24 5.82 41 3.88 53 4.74 33 4.88 38 4.58 41 5.52
4 32 4.77 38 4.85 49 3.76 32 4.50 42 5.12 41 3.71
5 23 5.71 41 4.79 54 3.98 31 4.47 43 3.89 45 4.02
```

```

6  25  4.47  36  4.36  48  5.00  29  5.12  43  4.62  49  5.09
7  32  4.55  36  4.02  49  3.31  29  4.51  37  4.30  52  2.70
8  18  4.61  41  3.77  47  3.11  30  4.85  50  2.70  47  4.31
9  19  5.86  41  4.22  52  4.76  21  5.22  50  3.50  61  2.70
10 26  5.20  37  4.94  58  3.95  28  4.62  45  5.06  65  3.03
11 33  4.44  42  4.04  62  4.60  23  5.07  48  4.06  58  2.73
12 27  5.52  39  4.51  65  4.83  35  3.64  51  4.51  59  3.67
13 33  4.97  41  4.06  62  3.18  38  3.64  46  4.66  .  .
14 25  4.99  43  4.02  59  3.03  38  5.09  58  2.88  .  .
15 42  4.89  41  4.99  .  .  .  .  .  .  .  .
;
proc print data=d01 ; run;

```

NLIN プロシジャの設定は、DATA ステップで作成された SAS データセット「d01」を使い、最大の反復回数 maxiter を 200 回、list オプションで偏微分式の出力を設定している。図 10.5 で設定したパラメータの初期値の設定を parms ステートメントで設定している。モデル式は、式 (10.30) に準じて model ステートメントで設定している。

SAS プログラム 10.2 3 本のロジスティック曲線の同時あてはめ

```

proc nlin data=d01 maxiter=200 list outest=cov1 ;
  parms      θ0=5  θ1=4  θ2=3.5  θ3=3  β0=4  β1=-0.1 ;
            θ_k = (θ1*a1 + θ2*a2 + θ3*a3) ;
            g_x = exp(-(β0 + β1*x)) ;
  model      y = θ_k + (θ0 - θ_k) * (1 + g_x)**(-1) ;
run ;

proc print data=cov1 ; run;

```

表 10.18 に分散分析表とパラメータの推定値を示す。JMP で計算した表 10.13 に一致することが確かめられる。表 10.21 示すパラメータ間の相関行列から、 $\hat{\beta}_0$ と $\hat{\beta}_1$ の相関が 0.9866 となり、多重共線性すれすれであり、このことがなかなか収束しない理由でもある。OnDemand SAS の標準出力は、HTML 形式なので、結果を Excel シートに張り付け、書式を整え Word 上に「図（拡張メタファイル）」形式でコピー&ペーストした結果である。

表 10.18 SAS/NLIN プロシジャによるパラメータの推定結果

要因	自由度	平方和	平均平方	F 値	Pr > F
Model	5	18.3953	3.6791	9.81	<.0001
Error	78	29.2457	0.3749		
C. Total	83	47.6410			
パラメータ	推定値	近似標準誤差	近似 95% 信頼限界		
θ0	4.9925	0.2753	4.4444	5.5406	
θ1	3.4946	0.6291	2.2421	4.7471	
θ2	2.9621	0.9159	1.1386	4.7855	
θ3	2.8612	0.8046	1.2594	4.4630	
β0	6.0858	3.8720	-1.6228	13.7944	
β1	-0.1272	0.0919	-0.3102	0.0557	

NLIN プロシジャでは、パラメータの相関行列がデフォルトで出力される。ただし、パラメータの共分散行列は、「OUTEST=cov1」オプションで外部ファイルに出力し、その結果を PRINT プロシジャで出力する必要がある。表 10.19 に結果を示すが、反復計算の過程が出力され、「FINAL」が、最終的に得られたパラメータであり、「COVB」の行にパラメータ共分散行列が出力されている。

表 10.19 パラメータの共分散行列 $\Sigma(\hat{\theta})$

OBS	_TYPE_	_STATUS_	_NAME_	θ_0	θ_1	θ_2	θ_3	β_0	β_1
125	COVB	0 Converged	θ_0	0.0758	-0.1042	-0.1196	-0.1209	-0.9229	0.0205
126	COVB	0 Converged	θ_1	-0.1042	0.3958	0.4901	0.4375	1.8964	-0.0491
127	COVB	0 Converged	θ_2	-0.1196	0.4901	0.8389	0.6217	2.3538	-0.0636
128	COVB	0 Converged	θ_3	-0.1209	0.4375	0.6217	0.6474	2.3749	-0.0619
129	COVB	0 Converged	β_0	-0.9229	1.8964	2.3538	2.3749	14.9926	-0.3511
130	COVB	0 Converged	β_1	0.0205	-0.0491	-0.0636	-0.0619	-0.3511	0.0085

表 10.20 に示すように NLIN プロシジャの「list」オプションにより、内部で計算されている偏微分した結果を出力することができる。「@MODEL.y/@ β_0 」と「@MODEL.y/@ β_1 」の式は組み入っているが、式 (10.32) に示したように、

表 10.20 内部で計算された偏微分式の結果

Listing of Compiled Program Code		
Stmt	Line:Col	Statement as Parsed
1	71:08:00	$\theta_k = \theta_1 * a_1 + \theta_2 * a_2 + \theta_3 * a_3;$
2	72:08:00	$g_x = \text{EXP}(-(\beta_0 + \beta_1 * x));$
2	72:08:00	$@g_x/@\beta_0 = -g_x;$
2	72:08:00	$@g_x/@\beta_1 = -x * g_x;$
3	73:05:00	$\text{MODEL.y} = \theta_k + (\theta_0 - \theta_k) * (1 + g_x) ** -1;$
3	73:05:00	$@\text{MODEL.y}/@\theta_0 = (1 + g_x) ** -1;$
3	73:05:00	$@\text{MODEL.y}/@\theta_1 = a_1 + -a_1 * (1 + g_x) ** -1;$
3	73:05:00	$@\text{MODEL.y}/@\theta_2 = a_2 + -a_2 * (1 + g_x) ** -1;$
3	73:05:00	$@\text{MODEL.y}/@\theta_3 = a_3 + -a_3 * (1 + g_x) ** -1;$
3	73:05:00	$@\text{MODEL.y}/@\beta_0 = (\theta_0 - \theta_k) * @g_x/@\beta_0 * -(1 + g_x) ** -2;$
3	73:05:00	$@\text{MODEL.y}/@\beta_1 = (\theta_0 - \theta_k) * @g_x/@\beta_1 * -(1 + g_x) ** -2;$

$$f(x_i) = \theta_k + (\theta_0 - \theta_k)[1 + g(x_i)]^{-1}$$

$$\text{ただし, } \theta_k = \theta_1 a_{i,1} + \theta_2 a_{i,2} + \theta_3 a_{i,3}$$

$$g(x_i) = \exp[-(\beta_0 + \beta_1 x_i)]$$

とおくと、 θ_0 から θ_4 までの偏微分は、式 (10.31) と同じであるが、 β_0 と β_1 の偏微分式には、他の偏微分式が引用されているが、整理すると式 (10.32) に一致する。

$$z_4 = \frac{\partial f(x_i)}{\partial \beta_0} = (\theta_0 - \theta_j) \frac{\partial g(x_i)}{\partial \beta_0} \{-[1 + g(x_i)]^{-2}\} = (\theta_0 - \theta_j) g(x_i) [1 + g(x_i)]^{-2}$$

$$z_5 = \frac{\partial f(x_i)}{\partial \beta_1} = (\theta_0 - \theta_j) \frac{\partial g(x_i)}{\partial \beta_0} \{-[1 + g(x_i)]^{-2}\} = (\theta_0 - \theta_j) \cdot x_i \cdot g(x_i) [1 + g(x_i)]^{-2} = \frac{\partial f(x_i)}{\partial \beta_0} x_i$$

式 (10.32) に一致する.

パラメータの共分散行列とパラメータの相関行列は、表 10.21 の Excel シート示すように、定義に従った簡単な行列計算で片方が得れば、もう片方を計算することができる。なお、SE と示したのは、表 10.18 に示したパラメータの推定結果の「近似標準誤差」であり、有効数字が少数定以下 4 桁なので、若干の計算誤差が出ている。

表 10.21 パラメータの相関行列からパラメータの共分散行列の求め方

近似相関行列							
	θ_0	θ_1	θ_2	θ_3	β_0	β_1	SE
θ_0	1	-0.6018	-0.4743	-0.5458	-0.8657	0.8082	0.2753
θ_1	-0.6018	1	0.8505	0.8643	0.7785	-0.8487	0.6291
θ_2	-0.4743	0.8505	1	0.8437	0.6637	-0.7559	0.9159
θ_3	-0.5458	0.8643	0.8437	1	0.7623	-0.8376	0.8046
β_0	-0.8657	0.7785	0.6637	0.7623	1	-0.9866	3.8720
β_1	0.8082	-0.8487	-0.7559	-0.8376	-0.9866	1	0.0919
Excel の行列計算によるパラメータの共分散行列の算出 $\Sigma(\hat{\theta})$							
	θ_0	θ_1	θ_2	θ_3	β_0	β_1	
θ_0	0.0758	-0.1042	-0.1196	-0.1209	-0.9228	0.0204	
θ_1	-0.1042	0.3958	0.4901	0.4375	1.8963	-0.0491	
θ_2	-0.1196	0.4901	0.8389	0.6217	2.3538	-0.0636	
θ_3	-0.1209	0.4375	0.6217	0.6474	2.3750	-0.0619	
β_0	-0.9228	1.8963	2.3538	2.3750	14.9924	-0.3511	
β_1	0.0204	-0.0491	-0.0636	-0.0619	-0.3511	0.0084	
=(相関行列の範囲)*(SEの範囲)*Transpose (SEの範囲)							

NLIN プロシジャの出力結果を Excel の折れ線グラフで

NLIN プロシジャの出力結果を Excel に取り込むことにより、これまで示してきた Excel の線グラフおよび散布図を用いた各種のグラフ表示に繋げることができる。表 10.18 に示した 3 本のロジスティック曲線の最大値および最小値に関して 95 信頼区間のひげ付きグラフ、A₁ 非曝露群と A₂ および A₃ 曝露群の差についても同様に髭付きグラフに示したい。すでに、図 10.6 で示してあるが、SAS/NLIN プロシジャの出力結果から、同様のグラフ表示をする方法を示すことにより、SAS と Excel の連携法の例示とする。

表 10.22 に示すのは、表 10.18 のパラメータの推定結果から $\hat{\theta}_0 \sim \hat{\theta}_3$ を抜き出し、さらに、表 10.21 からパラメータの共分散行列から $\hat{\theta}_1 \sim \hat{\theta}_3$ を抜き出した結果である。Excel の線グラフ

で 95%信頼区間の幅を付けるためには、推定値に対し $t \times SE$ の上下幅が必要なので、(上限 - 下限) / 2 による計算を付け加えてある。

A₁ 非曝露群との差は、推定値の差から求められが、その分散は

$$\begin{aligned} Var(\hat{\theta}_2 - \hat{\theta}_1) &= Var(\hat{\theta}_2) - 2Cov(\hat{\theta}_2, \hat{\theta}_1) + Var(\hat{\theta}_1) \\ &= 0.3958 - 2 \times 0.4901 + 0.8389 = 0.2545 \end{aligned}$$

によって計算されている。図 10.17 は、図 10.6 と同じ線グラフを別途作成した表 10.22 の結果を用いて作成したものである。

表 10.22 パラメータの推定値および共分散行列から信頼区間の幅の算出

パラメータ	推定値		近似標準誤差	近似 95% 信頼限界		t*SE 幅
θ0	4.9925		0.2753	4.4444	5.5406	0.5481
θ1	3.4946		0.6291	2.2421	4.7471	1.2525
θ2	2.9621		0.9159	1.1386	4.7855	1.8235
θ3	2.8612		0.8046	1.2594	4.4630	1.6018
差	推定値	分散	SE	L 95%	U 95%	t*SE 幅
$\theta_1 - \theta_1$	0					0
$\theta_2 - \theta_1$	-0.5325	0.2545	0.5045	-1.5369	0.4719	1.0044
$\theta_3 - \theta_1$	-0.6334	0.1682	0.4101	-1.4498	0.1830	0.8164
パラメータの共分散行列				$t_{0.05}(84-6)=$	1.9908	
	θ_1	θ_2	θ_3			
θ_1	0.3958	0.4901	0.4375			
θ_2	0.4901	0.8389	0.6217			
θ_3	0.4375	0.6217	0.6474			

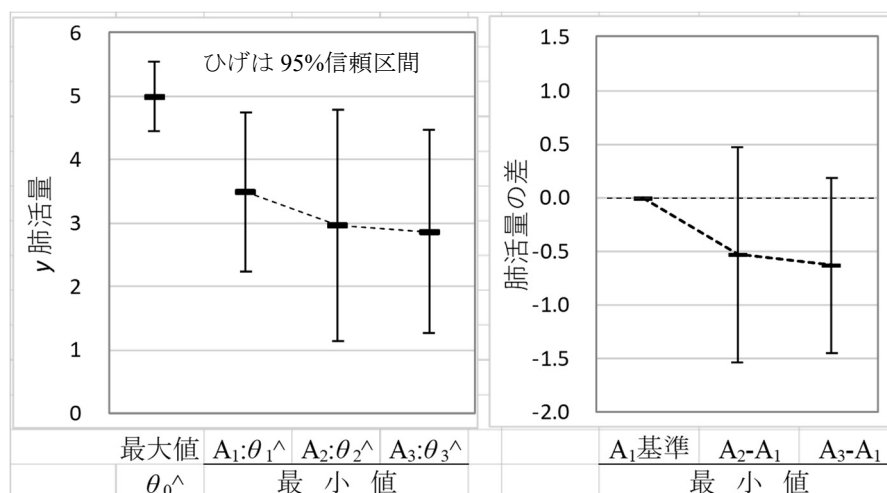


図 10.17 シグモイド曲線の共通の最大値および各層の最小値 (図 10.6 再掲)

10.6. 残差線形化法によるロジスティック曲線のパラメータ推定

第 10.2 節では、ロジスティック曲線のパラメータ推定をするために Excel のソルバーを用いた。第 10.3 節では、ロジスティック曲線式をパラメータで偏微分し、微分係数行列 \mathbf{Z} からパラメータの共分散行列 $\Sigma(\hat{\theta})$ を求めることにより、各種の推定値に対し 95%信頼区間が得られることを示した。ただし、初期値からどのような手順でパラメータの推定値を得るのかについては、“ソルバー”まかせであった。また、第 10.4 節の JMP の「非線形回帰」、第 10.5 節の SAS の NLIN プロシジャを用いた場合についても、同様にブラック・ボックスのままであった。そこで、本節では、残差 $\hat{\varepsilon}_i = y_i - \hat{y}_i$ に対し、微分係数行列 \mathbf{Z} をデザイン行列と見なした残差線形化法によるパラメータの推定値法を示す。

残差を反応とした回帰分析の反復

ロジスティック曲線のパラメータ推定は、パラメータの初期値によって推定された $\hat{y}_i^{(0)}$ と y_i の残差 $\hat{\varepsilon}_i^{(0)} = y_i - \hat{y}_i^{(0)}$ を新たな反応とみなし、デザイン行列 \mathbf{X} に変えて微分係数行列 \mathbf{Z} を用いて回帰分析を繰り返し行うことから、残差線形化法とも言われている。表 10.23 に示すのは、表 10.6 に示した初期値 ($\hat{\theta}_0^{(0)} = 5.0$, $\hat{\theta}_1^{(0)} = 4.0$, $\hat{\theta}_2^{(0)} = 3.5$, $\hat{\theta}_3^{(0)} = 3.0$, $\hat{\beta}_0^{(0)} = 4.0$, $\hat{\beta}_1^{(0)}$)

表 10.23 初期値に対する残差 $\hat{\varepsilon}_i^{(0)}$ および微分係数行列 $\mathbf{Z}^{(0)}$ の計算結果

					θ	$\theta^{\wedge(m-1)}$	$\delta^{\wedge(m-1)}$	$\theta^{\wedge(m)}$		$(Z^T Z)^{-1}$						$Z^T \varepsilon^{\wedge}$
				最大値	θ_0	5.000	0.046	5.046		0.683	-0.447	-0.449	-0.536	-5.194	0.120	5.343
				最小A ₁	θ_1	4.000	-0.647	3.353		-0.447	0.606	0.696	0.796	4.301	-0.117	-2.368
				A ₂	θ_2	3.500	-0.674	2.826		-0.449	0.696	1.126	1.097	4.806	-0.140	1.262
				A ₃	θ_3	3.000	-0.360	2.640		-0.536	0.796	1.097	1.429	6.013	-0.171	2.006
				切片	β_0	4.000	1.124	5.124		-5.194	4.301	4.806	6.013	47.645	-1.165	2.976
				傾き	β_1	-0.100	0.006	-0.094		0.120	-0.117	-0.140	-0.171	-1.165	0.030	103.8
				残差平方和	S_e	32.867										

$= -0.10$) を用いて, 式 (10.29) により $\hat{y}_i^{(0)}$ を推定し, さらに, 残差 $\varepsilon_i^{(0)}$ を計算している. 式 (10.31) および式 (10.32) に示したパラメータで偏微分した式 z_k を計算し, 84×6 の微係数行列 $\mathbf{Z}^{(0)}$ を計算した結果である.

残差 $\varepsilon_i^{(0)}$ を反応として, 微係数行列 $\mathbf{Z}^{(0)}$ をデザイン行列 \mathbf{X} と見なした回帰分析により, パラメータ $\delta_i^{(0)}$ を線形モデル $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ と同様に

$$\hat{\boldsymbol{\delta}}^{(0)} = (\mathbf{Z}^{(0)T} \mathbf{Z}^{(0)})^{-1} \mathbf{Z}^{(0)T} \boldsymbol{\varepsilon}^{(0)} \quad (10.39)$$

によりパラメータ $\hat{\boldsymbol{\delta}}^{(0)}$ を推定する. 表 10.23 右上段に, $(\mathbf{Z}^{(0)T} \mathbf{Z}^{(0)})^{-1}$ および $\mathbf{Z}^{(0)T} \boldsymbol{\varepsilon}^{(0)}$ の計算結果が示され, 表 10.23 右上段にパラメータの推定値としての増減値 $\hat{\boldsymbol{\delta}}^{(0)}$ の推定結果が示されている. さらに, 推定された増減値 $\hat{\boldsymbol{\delta}}^{(0)}$ を初期パラメータ $\hat{\boldsymbol{\theta}}^{(0)}$ に加え

$$\hat{\boldsymbol{\theta}}^{(0)} + \hat{\boldsymbol{\delta}}^{(0)} = \hat{\boldsymbol{\theta}}^{(1)} \quad (10.40)$$

次のように, パラメータ $\hat{\boldsymbol{\theta}}^{(0)}$ を更新した $\hat{\boldsymbol{\theta}}^{(1)}$

$$m=1: \begin{array}{|c|} \hline \boldsymbol{\theta}^{(m-1)} \\ \hline 5.000 \\ 4.000 \\ 3.500 \\ 3.000 \\ 4.000 \\ -0.100 \\ \hline \end{array} + \begin{array}{|c|} \hline \boldsymbol{\delta}^{(m-1)} \\ \hline 0.046 \\ -0.647 \\ -0.674 \\ -0.360 \\ 1.124 \\ 0.006 \\ \hline \end{array} = \begin{array}{|c|} \hline \boldsymbol{\theta}^{(m)} \\ \hline 5.046 \\ 3.353 \\ 2.826 \\ 2.640 \\ 5.124 \\ -0.094 \\ \hline \end{array}$$

が計算されている. 推定された $\hat{\boldsymbol{\theta}}^{(1)}$ をコピーし, 第 2 回目の反復 $m=2$ とし, $\hat{\boldsymbol{\theta}}^{(m-1)}$ に値のみをペーストする. すると, Excel の自動計算機能により, パラメータを $\hat{\boldsymbol{\theta}}^{(1)}$ として全ての計算が更新される. 更新された増減値 $\hat{\boldsymbol{\delta}}^{(1)}$ をパラメータ $\hat{\boldsymbol{\theta}}^{(1)}$ に加えパラメータ $\hat{\boldsymbol{\theta}}^{(2)}$ が

$$m=2: \begin{array}{|c|} \hline \boldsymbol{\theta}^{(m-1)} \\ \hline 5.046 \\ 3.353 \\ 2.826 \\ 2.640 \\ 5.124 \\ -0.094 \\ \hline \end{array} + \begin{array}{|c|} \hline \boldsymbol{\delta}^{(m-1)} \\ \hline -0.210 \\ 1.708 \\ 2.093 \\ 2.305 \\ 2.883 \\ -0.098 \\ \hline \end{array} = \begin{array}{|c|} \hline \boldsymbol{\theta}^{(m)} \\ \hline 4.836 \\ 5.062 \\ 4.919 \\ 4.946 \\ 8.007 \\ -0.192 \\ \hline \end{array}$$

得られる. このような繰り返して, 増減値 $\hat{\boldsymbol{\delta}}^{(m-1)}$ 全体が 0 とみなせるまで繰り返す.

経験的には, 何回かの繰り返して増減値 $\hat{\boldsymbol{\delta}}^{(m-1)}$ が 0 となるのだが, ここに示した事例では途中で残差平方和 S_e が減少するのではなく, 逆に増大し解が求められなくなってしまう. このような場合には, 無理をせずソルバーで残差平方和 S_e を最小にするように, パラメータを変化させつつ増減値 $\hat{\boldsymbol{\delta}}^{(m-1)}$ が 0 となることを確認するとよい. ソルバーには, 残差平方和 S_e が増大しないような計算上の工夫がされていると思われるが, その方法についてはブラック・ボックスである. ただし, とんでもない初期値に対しては, 計算不能となってしまうことも

ある。また、なかなか収束しない場合は、「実行を一時中止しました」とのメッセージが出るが、残差平方和 S_e が減少していれば、「続行」することにより、表 10.24 に示すように増減値 $\hat{\delta}^{(m-1)}$ が 0.0 となる。

表 10.24 3本のシグモイド曲線の同時推定の結果

					θ	$\theta^{\wedge(m-1)}$	$\delta^{\wedge(m-1)}$	$\theta^{\wedge(m)}$			$(Z^T Z)^{-1}$						$Z^T \varepsilon^{\wedge}$
					最大値 θ_0	4.992	0.000	4.992			0.202	-0.278	-0.319	-0.322	-2.461	0.055	0.000
					最小 $A_1 \theta_1$	3.495	0.000	3.495			-0.278	1.055	1.307	1.166	5.057	-0.131	0.000
					$A_2 \theta_2$	2.962	0.000	2.962			-0.319	1.307	2.237	1.658	6.277	-0.170	0.000
					$A_3 \theta_3$	2.861	0.000	2.861			-0.322	1.166	1.658	1.726	6.333	-0.165	0.000
					切片 β_0	6.087	0.000	6.087			-2.461	5.057	6.277	6.333	39.992	-0.936	0.000
					傾き β_1	-0.127	0.000	-0.127			0.055	-0.131	-0.170	-0.165	-0.936	0.023	0.007
					平方和 S_e	29.246											
											----- 微分係数行列 Z -----						
	群	ダミー 年齢			肺活量				推定値	残差	z_0	z_1	z_2	z_3	z_4	z_5	
No	A	a_1	a_2	a_3	x	y	θ_k	$g(x)$	y^{\wedge}	ε^{\wedge}	$\partial \theta_0$	$\partial \theta_1$	$\partial \theta_2$	$\partial \theta_3$	$\partial \theta_4$	$\partial \theta_5$	
1	A_1	1	0	0	27	5.29	3.495	0.071	4.894	0.396	0.934	0.066	0.000	0.000	0.092	2.491	
2	A_1	1	0	0	25	3.67	3.495	0.055	4.915	-1.245	0.948	0.052	0.000	0.000	0.074	1.842	
:																	
44	A_1	1	0	0	59	3.03	3.495	4.145	3.786	-0.756	0.194	0.806	0.000	0.000	0.235	13.84	
45	A_2	0	1	0	29	5.21	2.962	0.091	4.823	0.387	0.917	0.000	0.083	0.000	0.155	4.504	
46	A_2	0	1	0	29	5.17	2.962	0.091	4.823	0.347	0.917	0.000	0.083	0.000	0.155	4.504	
:																	
72	A_2	0	1	0	58	2.88	2.962	3.649	3.399	-0.519	0.215	0.000	0.785	0.000	0.343	19.88	
73	A_3	0	0	1	39	4.62	2.861	0.325	4.469	0.151	0.755	0.000	0.000	0.245	0.395	15.39	
74	A_3	0	0	1	40	5.29	2.861	0.369	4.418	0.872	0.730	0.000	0.000	0.270	0.420	16.79	
:																	
84	A_3	0	0	1	59	3.67	2.861	4.145	3.275	0.395	0.194	0.000	0.000	0.806	0.334	19.69	

推定されたパラメータを用いて各群の推定値を求め、散布図上に重ね書きした結果が、すでに示した図 10.4 である。カドミウムガス非曝露 A_1 群は、最大肺活量の $\hat{\theta}_0=4.9925$ リットルから徐々に肺活量が減少し $\hat{\theta}_1=3.4947$ リットルに収束している。曝露 A_2 群は、 $\hat{\theta}_2=2.9621$ リットルと曝露 A_3 群の $\hat{\theta}_3=2.8612$ リットルに近い推定値が得られている。これらから、カドミウムガスの曝露により、徐々に肺活量が低下し、最終的に非曝露群に対し、約 0.6 リットルの減少となると推測される。

Wikipedia のガウス・ニュートン法

ロジスティック曲線のあてはめにかぎらず、簡単ではあるが典型的な事例を示し、きちっとした解析法の考え方を導入するのが、段階的な進め方である。この章では、いきなり高度な応用事例から初めてしまった。本書のテーマである「層別因子を含む探索的な回帰分析入門」の趣旨として、「層別因子を含むロジスティック曲線」を最初に取り上げたのは、推定結果の統計的な解釈にどのように寄与するのかを提示したいと考えたからである。

非線形回帰の典型例は、第 12 章で取り上げるミカエリス・メンテン式である。Web 上の Wikipedia「非線形回帰」にも登場する。さらに、Wikipedia「ミカエリス・メンテン式」を参照すると、「線形化によるパラメータ推定は、非線形回帰に劣る」とあるが、非線形回帰の数値例が見当たらない。このような状況を改善するために適切な計算事例の提示が必要と考えている。

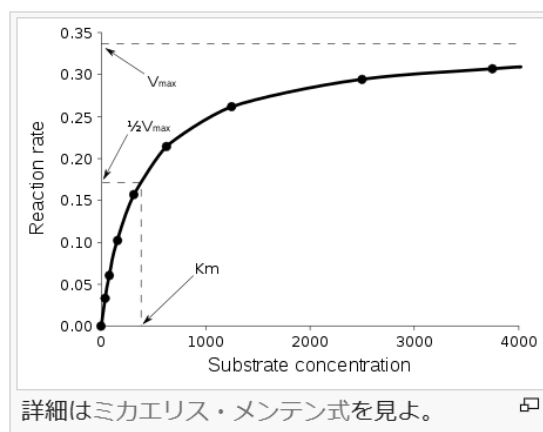


図 10.18 Wikipedia「非線形回帰」

数値例は、Wikipedia「ガウス・ニュートン法」で見出すことができる。掲載されている数値を使い残差線形化法（ガウス・ニュートン法の計算方法）を用いた事例を表 10.25 に示す。Wikipediaで示されてる初期値（ $\hat{\beta}_1 = 0.9$, $\hat{\beta}_2 = 0.5$ ）を与え、反復を繰り返すと、5 回目で Wikipedia に示されている（ $\hat{\beta}_1 = 0.362$, $\hat{\beta}_2 = 0.556$ ）が得られる。

ただし、推定されてパラメータ $\hat{\beta}_1 = 0.362$ は、最大値 V_{max} であり、 $\hat{\beta}_2 = 0.556$ は最大値の 1/2 となる濃度 K_m である。さらなる解説は、第 12 章を参照のこと。

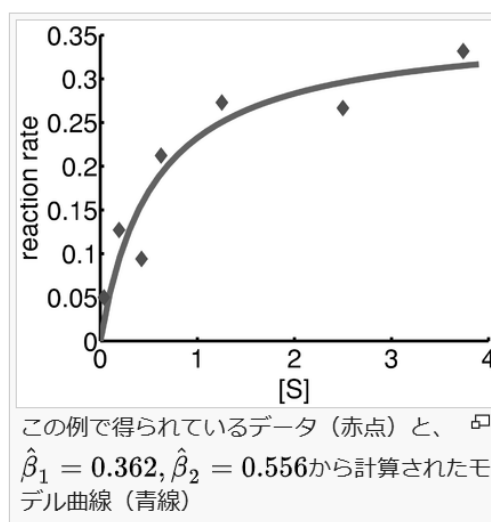


図 10.19 「ガウス・ニュートン法」

表 10.25 Wikipedia の「ガウス・ニュートン法」のデータ・反復 5 で収束

初期値			$\beta^{(m)}$	δ	$\beta^{(m+1)}$	$(Z^T Z)^{-1}$	
0.9	$\beta_1^{\wedge} = V_{max}^{\wedge}$		0.362	0.000	0.362	1.521	6.341
0.2	$\beta_2^{\wedge} = K_m^{\wedge}$		0.556	0.000	0.556	6.341	36.172
		$S_e =$	0.008				
	[S]	V	V^{\wedge}	$\varepsilon^{\wedge} = V - V^{\wedge}$		$\partial V_{max}^{\wedge}$	∂K_m^{\wedge}
No	x	y	y^{\wedge}	$\varepsilon^{\wedge} = y - y^{\wedge}$		z_1	z_2
1	0.038	0.050	0.023	0.027		0.064	-0.039
2	0.194	0.127	0.094	0.033		0.259	-0.125
3	0.425	0.094	0.157	-0.063		0.433	-0.160
4	0.626	0.2122	0.192	0.021		0.530	-0.162
5	1.253	0.2729	0.251	0.022		0.693	-0.139
6	2.500	0.2665	0.296	-0.029		0.818	-0.097
7	3.740	0.3317	0.315	0.017		0.871	-0.073
$V^{\wedge} = V_{max}^{\wedge} [S] / ([S] + K_m^{\wedge})$ $z_1 = [S] / ([S] + K_m^{\wedge})$ $z_2 = -V_{max}^{\wedge} [S] / ([S] + K_m^{\wedge})^2$							
			$\delta = (Z^T Z)^{-1} Z^T \varepsilon^{\wedge}$	$\beta^{(m+1)} = \beta^{(m)} + \delta$			

第10章 文献 索引

アーミテージら著, 椿・椿訳(2001) - 医学研究のための統計的方法, 第9.4節 群内の回帰	327
SAS Institute(2014) - The NLIN Procedure, SAS/STAT® 13.2 User's Guide.	358
高橋(2017) - 一般化線形モデルをExcelで極め活用する - プロビット法・ロジット法・補2重対数法 -	336
高橋(2018) - 二値データの用量反応関係における新たな解析 - プロビット・ロジット・補2重対数の使い分け -	336
芳賀(2016) - 医薬品開発のための統計解析, 第3部 非線形モデル 改訂版	337











第10章 索引

あ アーミテージら(2001) - カドミウムガス	327	か 95%信頼区間 - パラメータに関する	344
- 共分散分析	327	- パラメータの差	345
- 群内の回帰	327	- 回帰直線の差	332
OUTEST= - パラメータの共分散行列	360	- 予測プロファイル	351
あてはめの基礎 - ロジスティック曲線	335	- ロジスティック曲線	342, 346, 352
(1, 1)標示型 - ダミー変数	336	- ロジスティック曲線の差	347
位置パラメータ - 形状パラメータ	340	共通の最大値 - そこそこの95%信頼区間	345
Wikipedia - ガウス・ニュートン法	336	共分散行列 $\Sigma(\theta^{\wedge})$ - パラメータ(推定値)	351
- ガウス・ニュートン法	366	共分散分析 - アーミテージら(2001)	327
- 非線形回帰	366	グラフビルダー - 95%信頼区間	356
- ミカエリス・メンテン式	366	- スプライン曲線	334
埋め込まれている計算式 - 予測値 y^{\wedge}	353	- 予測値 y^{\wedge} の重ね合わせ	355
上側95%信頼区間が「0」 - ソルバー	334	苦し紛れの解釈 - 回帰直線	338
exp() 関数の中の線形式 - 初期値	340	群内の回帰 - アーミテージら(2001)	327
Excel - 積和行列の逆数	344	計算式 - 予測値の標準誤差 y^{\wedge}	353
- ソルバーの活用	341	計算式エディター - 微分した式	357
- 偏微分式の計算	344	計算式の設定 - 予測値 y^{\wedge}	349
Excel の線グラフ - SAS と Excel の連携法	361	計算式を含む - 予測値の95%信頼区間	352
HTML形式 - SAS の標準出力	359	形状パラメータ - 位置パラメータ	340
NLINプロシジャ - SAS の非線形	358	交互作用 - 層別因子を含む回帰分析	327
- 自動偏微分	358	交互作用の分離 - 残差平方和の差	331
か 回帰直線 - 苦し紛れの解釈	338	交互作用の平方和 - 残差平方和の差	330
- 便宜的な方法	335	交互作用モデル - 主効果モデル	330
回帰直線の傾きの差 - t 検定	330	合成分散の一般式 - パラメータの共分散行列	330
回帰直線の差 - 95%信頼区間	332	- パラメータ間の差	345
回帰分析/ Excel - ソルバー	341	- 差 d^{\wedge} の分散	332
回帰分析の反復 - 残差を反応	363	さ (最小値から最大値) に拡張 - (0 から 1) の範囲	327
ガウス・ニュートン法 - Wikipedia	336	最大値・最小値 - ロジスティック曲線	336
- Wikipedia	365	SAS と Excel の連携法 - Excel の線グラフ	361
- Wikipedia	366	SAS の非線形 - NLINプロシジャ	358
- 数値例	366	SAS の標準出力 - HTML形式	359
- ミカエリス・メンテン式	336	SAS プログラム - 3本のロジスティック曲線	359
各群の最小値 - 95%信頼区間	345	SAS/ OnDemand - 無償で継続的	358
下限に漸近 - ロジスティック曲線	339	SASデータセット化 - DATAステップ	358
重ね書き - 3本の回帰直線	329	SAS Institute(2013) - 非線形モデル	358
傾きが異なる回帰直線の差 - t 検定	333	差 d^{\wedge} の分散 - 合成分散の一般式	332
傾きに統計的に差 - どのように説明	331	残差を反応 - 回帰分析の反復	363
カドミウムガス - アーミテージら(2001)	327	残差線形化法 - パラメータ推定	363
- 加齢と共に約 0.6リットル減少	348	残差平方和 - 微小なパラメータ増減値	336
- 肺活量の減少	327, 328, 337	残差平方和の差 - 交互作用の分離	331
加齢と共に約 0.6リットル減少 - カドミウムガス	348	- 交互作用の平方和	330
95%信頼区間 - 回帰直線の差	332	残差平方和の最小化 - ソルバー	337
- 各群の最小値	345	散布図上に重ね書き - ロジスティック曲線	338, 347
- グラフビルダー	356	3本のロジスティック曲線 - SASプログラム	359
- 線形モデル	342	- JMP	349

さ	3本の回帰直線 - 重ね書き	329	な	内部で計算 - 偏微分式	360
	- デザイン行列 X	328		2次形式 - 微係数ベクトル z	346
	GUI による偏微分する機能 - JMP	357		- Vec Quadratic () 関数	354
	自動的に偏微分 - 統計ソフト	342		2値反応 - ロジスティック回帰	327, 335
	自動微分 - 非線形回帰/ 統計ソフト	351		年齢 x が小さい順 - 推定曲線を重ね書	338
	自動偏微分 - NLIN プロシジャ	358	は	肺活量の減少 - カドミウムガス	327, 328, 337
	JMP - 3本のロジスティック曲線	349		芳賀敏郎 (2016) - 非線形モデル	337
	- GUI による偏微分する機能	357		- 非線形回帰に関する必読の書	337
	- パラメータに関する偏微分	356		- 非線形回帰/ JMP	339
	- 非線形回帰	350		パラメータ (推定値) - 共分散行列 $\Sigma(\theta^{\wedge})$	351
	- Vec Quadratic () 関数	354		- 相関行列	351
	JMP の非線形回帰 - 標準誤差	350		パラメータで偏微分 - モデル式	342
	主効果モデル - 交互作用モデル	330		- ロジスティック曲線の式	342
	初期パラメータ - ロジスティック曲線	341		パラメータに関する - 95%信頼区間	344
	初期値 - exp () 関数の中の線形式	340		- 偏微分式	343
	初期値を最小化 - ソルバー	341		パラメータに関する偏微分 - JMP	356
	推定曲線を重ね書 - 年齢 x が小さい順	338		パラメータの共分散行列 - OUTEST=	360
	数値例 - ガウス・ニュートン法	366		- 合成分散の一般式	330
	スプライン曲線 - グラフビルダー	334		- パラメータの相関行列	361
	- 直線のあてはめの妥当性	334		パラメータの共分散行列 $\Sigma(\theta^{\wedge})$ -ロジスティクス曲線	344
	正攻法/ 設定 - パラメータの初期値	340		パラメータの差 - 95%信頼区間	345
	積和行列の逆行列 $(Z^T Z)^{-1}$ - 微係数行列 Z	343		パラメータの初期値 - 正攻法/ 設定	340
	積和行列の逆数 - Excel	344		- 探索的な推定	340
	(0 から 1) の範囲 - (最小値から最大値) に拡張	327		パラメータの相関行列 - デフォルトで出力	360
	線形モデル - 95%信頼区間	342		- パラメータの共分散行列	361
	線形モデルではない - 非線形モデル	336		パラメータ間の差 - 合成分散の一般式	345
	線形近似 - 層別散布図/ Excel	328		パラメータ推定 - 残差線形化法	363
	線形式 - ロジット	335		反応が量的変数 - ロジスティック曲線	327
	相関行列 - パラメータ (推定値)	351		微係数ベクトル z - 2次形式	346
	層別因子 - ロジスティック曲線	327		微係数行列 Z - 積和行列の逆行列 $(Z^T Z)^{-1}$	343
	層別因子を含む回帰分析 - 交互作用	327		- デザイン行列 X	342, 364
	層別散布図/ Excel - 線形近似	328		微小なパラメータ増減値 - 残差平方和	336
	そこそこの95%信頼区間 - 共通の最大値	345		非線形モデル - SAS Institute (2013)	358
	ソルバー - 上側95%信頼区間が「0」	334		- 線形モデルではない	336
	- 回帰分析/ Excel	341		- 芳賀 (2016)	337
	- 残差平方和の最小化	337		非線形回帰 - Wikipedia	366
	- 初期値を最小化	341		- ガウス・ニュートン法	365
	- ミカエリス・メンテン式	341		- JMP	350
	- 目標値	341		- ミカエリス・メンテン式	366
	ソルバーの活用法 - Excel	341		- ロジスティック曲線	349
	- 残差平方和の最小化	337		非線形回帰/ JMP - 芳賀 (2016)	339
た	ダミー変数 - (1, 1) 標示型	336		非線形回帰/ 統計ソフト - 自動微分	351
	探索的な推定 - パラメータの初期値	340		非線形回帰に関する必読の書 - 芳賀 (2016)	337
	直線のあてはめの妥当性 - スプライン曲線	334		1 つのパラメータ - 1 つの変数	336
	t 検定 - 傾きが異なる回帰直線の差	333		1 つの変数 - 1 つのパラメータ	336
	- 回帰直線の傾きの差	330		非曝露・曝露群 - ロジスティック曲線	339
	デフォルトで出力 - パラメータの相関行列	360		微分した式 - 計算式エディター	357
	DATA ステップ - SAS データセット化	358		標準誤差 - JMP の非線形回帰	350
	デザイン行列 X - 3本の回帰直線	328		Vec Quadratic () 関数 - JMP	354
	- 微係数行列 Z	342, 364		- 2次形式	354
	統計ソフト - 自動的に偏微分	342		便宜的な方法 - 回帰直線	335
	同時あてはめ - ロジスティック曲線/ 3本	337		偏微分式 - 内部で計算	360
	- ロジスティック曲線/ 複数	335		- パラメータに関する	343
	どのように説明 - 傾きに統計的に差	331		- list オプション	360

は	偏微分式の計算 - Excel	344	ら	ロジスティック回帰 - 2値反応	327, 355
ま	ミカエリス・メンテン式 - Wikipedia	366		ロジスティック曲線 - あてはめの基礎	335
	- ガウス・ニュートン法	336		- 下限に漸近	339
	- ソルバー	341		- 95%信頼区間	342, 346, 352
	- 非線形回帰	366		- 最大値・最小値	336
	無償で継続的 - SAS/ OnDemand	358		- 散布図上に重ね書き	338, 347
	目標値 - ソルバー	341		- 初期パラメータ	341
	モデル式 - パラメータで偏微分	342		- 層別因子	327
や	予測プロファイル - 95%信頼区間	351		- 反応が量的変数	327
	予測値 y^{\wedge} - 埋め込まれている計算式	353		- 非線形回帰	349
	- 計算式の設定	349		- 非曝露・曝露群	339
	予測値 y^{\wedge} の重ね合わせ - グラフビルダー	355		- 量的な反応	335
	予測値の95%信頼区間 - 計算式を含む	352	ら	ロジスティック曲線/ 3本 - 同時あてはめ	337
	予測値の標準誤差 y^{\wedge} - 計算式	353		ロジスティック曲線/ 複数 - 同時あてはめ	335
ら	listオプション - 偏微分式	360		ロジスティック曲線の差 - 95%信頼区間	347
	量的な反応 - ロジスティック曲線	335		ロジスティック曲線の式 - パラメータで偏微分	342
	ロジスティクス曲線-パラメータの共分散行列 $\Sigma(\theta^{\wedge})$	344		ロジット - 線形式	335

第 10 章 解析用ファイル一覧

サイズ	名前	種類
 66 KB	第10章_01_1_3本の直線	Microsoft Excel ワークシート
 7 KB	第10章_01_2_カドミウムガス_スプライン	JMP Data Table
 68 KB	第10章_02_3本のロジスティック曲線	Microsoft Excel ワークシート
 132 KB	第10章_03_ロジスティック曲線95%	Microsoft Excel ワークシート
 23 KB	第10章_04_1_カドミウムガス_ロジスティック曲線	JMP Data Table
 68 KB	第10章_04_2_ロジスティック曲線計算式	Microsoft Excel ワークシート
 2 KB	第10章_05_1_非線形回帰_SAS	テキストドキュメント
 43 KB	第10章_05_2_SAS_out	Microsoft Excel ワークシート
 162 KB	第10章_06_1_ロジ_残差線形化	Microsoft Excel ワークシート
 51 KB	第10章_06_2_ミカエリス_メンテン_Wikipedia	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 10 章 層別因子を含むロジスティック曲線のあてはめ

BioStat 研究所(株)
〒105-0014 東京都 港区 芝 1-12-3 の1005
2024 年 2 月 16 日 高橋 行雄
takahashi.stat@nifty.com , FAX : 03-342-8035