

第 7 回 続高橋セミナー

正規分布を仮定した打ち切りデータを含む回帰分析入門

2018 年 7 月

高橋 行雄 BioStat 研究所(株)

要約： 打ち切りデータは、欠測値として扱う、または打ち切られた時の値などに置き換える等の便宜的な解析がしばしば行われる。データが正規分布すると仮定できる場合は、右側打ち切りデータに対しては上側確率、観測されたデータに対しては確率密度を与えることにより、Excel のソルバーで平均値および標準偏差を推定すること容易にできる。また、打ち切りデータを考慮した回帰分析でも回帰パラメータの推定が容易に行える。打ち切りデータを含む統計解析は、加速試験による製品寿命解析、癌などの生存時間解析、経済分野では、トップコーディングされたデータ（打ち切りデータ）に対する Tobit モデルなど、分野毎に発展してきた。新たに開発された JMP の「寿命の二変量」は、正規分布も包含するパラメトリックな汎用的な統計解析を可能としている。今回は、薬効薬理試験として行われた発毛試験データを用いて、打ち切りデータを含む回帰分析について詳しく解説する。

目 次

0. はじめに-----	1
1. 発毛実験で発生する打ち切りデータ-----	4
2. 打ち切りデータを考慮した平均値と標準偏差-----	8
3. 段階的な学習-----	13
4. 最小 2 乗法の復習-----	15
5. 正規分布の確率密度の活用-----	18
6. 正規分布の上側確率の活用-----	23
7. 打ち切りを含む回帰分析-----	28
8. JMP の出力を EXCEL で活用-----	36
9. JMP の寿命の二変量の活用-----	39
10. 無料の ONDEMAND SAS を用いた解析-----	46
11. 文献データに対する統計ソフトでの追試-----	49
12. Excel によるニュートン・ラフソン法の適用-----	54
13. Excel による Tobit 回帰分析-----	60
文 献-----	67
索引-----	68
EXCEL, JMP ファイル 一覧-----	71

目次 正規分布を仮定した打ち切りデータを含む回帰分析入門

0. はじめに	1
1. 発毛実験で発生する打ち切りデータ	4
2. 打ち切りデータを考慮した平均値と標準偏差	8
3. 段階的な学習	13
4. 最小 2 乗法の復習	15
5. 正規分布の確率密度の活用	18
6. 正規分布の上側確率の活用	23
7. 打ち切りを含む回帰分析	28
8. JMP の出力を EXCEL で活用	36
9. JMP の寿命の二変量の活用	39
10. 無料の ON DEMAND SAS を用いた解析	46
11. 文献データに対する統計ソフトでの追試	49
12. EXCEL によるニュートン・ラフソン法の適用	54
13. EXCEL による TOBIT 回帰分析	60
文献	67
索引	68
EXCEL, JMP ファイル 一覧	71

0. はじめに

2017 年 11 月の第 2 期医薬安全性研究会 第 21 回定例会での「じっくり勉強シリーズ」で、「正規分布を仮定した打ち切りデータを含む回帰分析入門」と題して教育講演を行った。さらに、午後の大正製薬の黒須真介氏の講演「探索的薬理試験データに対する種々の解析方法の適用と比較」に引き続き、「打ち切りデータを含む場合の新しい回帰分析の考え方」と題する講演を行った。これらの講演は、互いに関連するので合わせて第 7 回の続高橋セミナーとする。

スライド 1

じっくり勉強すれば身につく統計入門

**正規分布を仮定した
打ち切りデータを含む回帰分析入門**

BioStat研究所(株)
高橋 行雄

2017.11.18 高橋行雄 1

スライド 2

目 次	
1. 打ち切りデータの実例	3
2. 正規分布の活用	17
3. 段階的な学習	15
4. 最小2乗法の復習	19
5. 正規分布確率密度の活用	25
6. 正規分布の上側確率の活用	35
7. 打ち切りを含む回帰分析	46
8. JMPの出力を Excel で活用	59
まとめ	66
2017.11.18 高橋行雄 2	

じっくり勉強シリーズの「正規分布を仮定した打ち切りデータを含む回帰分析入門」は、Excel を主体にしており、第 8 節までであったが、黒須真介氏の講演に引く続いて「打ち切りデータを含む場合の新しい回帰分析の考え方」は、統計ソフト JMP による解析方法、結果の見方について解説したので、第 9 節で「JMP の寿命の二変量の活用」として付け加えた。

打ち切りデータは、欠測値として扱う、または打ち切られた時の値などに置き換える等の便宜的な解析がしばしば行われる。データが正規分布すると仮定できる場合は、その累積分布関数及び確率密度関数を用いることにより打ち切りデータを含む場合であっても平均値および標準偏差を推定することができる。

観測されたデータが、正規分布に従うと仮定した場合に、データを数直線上に並べてプロットし、その直線上にしかるべき正規分布の確率密度曲線を描くことを考えよう。厳密に描くためには、算術平均と標準偏差を用いて正規分布の確率密度を計算することにより実現できる。それぞれのデータがどのくらいの確率で起きるかは、確率密度に適当な幅を掛けることにより算出できるが、確率密度そのものでも相対的な比較は可能である。

右側打ち切りがある場合は、打ち切り点以上の大きな値をとることが確かなのであれば、その点以上の上側確率とみなすことが妥当である。また、打ち切りが左側ならば下側確率とすれば良く、区間打ち切りの場合には、その区間の正規分布の面積 $1 - (\text{下側確率} + \text{上側確率})$ とするのが妥当である。

このように考えるならば、打ち切りデータを含むような場合であっても正規分布のパラメータ（平均、標準偏差）を推定するために、それぞれのデータの確率密度、および上側確率などの積を最大にするような平均と標準偏差を推定することが可能である。

打ち切りデータがある場合に、それらを考慮した回帰分析を必要とする様々な分野がある。工業製品の場合であれば、各種の加速試験データを元に、通常の使用環境での寿命を推定し、製品説明書などでユーザーに示すことが一般化している。このために打ち切りデータを含む回帰分析の方法が発達してきた。この場合のような寿命データに対しては、正規分布でなくワイブル分布が標準的に用いられている。

経済統計の分野では、年間年収に対する調査データでは、2500 万円以上の場合には、匿名化のために 2500 万円以上する打ち切りが行われる場合がしばしばある。年間収入に対する他の要因を考慮した回帰分析を行う際に、打ち切りデータを含む回帰分析の適用が必要となる。誤差分布に正規分布を仮定した場合の打ち切りデータがある場合に解析法は、Tobit モデルとして知られている。

臨床試験の分野では、打ち切りデータがある場合の解析法として、各種の生存時間データ解析が開発されてきた。様々な疾患の生存時間の分布は多様で特定しがたいために、ノンパラメトリックな方法、Cox の比例ハザードモデルなどセミパラメトリックな方法が主体である。そのために、正規分布あるいはワイブル分布を仮定するようなパラメトリックな解析法は、適用されにくい。そのために生存時間解析の関連の成書では、打ち切りデータを含む回帰分析の適用事例を見出すことができない。

癌の生存時間解析の名著である Kalbfleisch, J.D. and Prentice, R.L. (2002) には、加速試験データの事例として、1.1.4 節に Motorette 試験（発電機のモータの絶縁組織の寿命）のデータが、示されている。そして、第 3.7.節 Illustrations In More Complex Data Sets で、対数正規分布を仮定した回帰分析の結果が示されている。このことから打ち切りデータを含む回帰分析は、難解な課題として認識されてきたようである。

(Table 1.3) Hours to Failure of Motorettes

150°C	All 10 motorettes without failure at 8064 hours
170°C	1764, 2772, 3444, 3542, 3780, 4860, 5196 3 motorettes without failure at 5448 hours
190°C	408, 408, 1344, 1344, 1440 5 motorettes without failure at 1680 hours
220°C	408, 408, 504, 504, 504 5 motorettes without failure at 528 hours

この Motorette 試験を題材に高橋 (2015b) では, LIFEREG プロシジャにより, ワイブル分布のあてはめを試みている. また, 人工データではあるが, 正規分布, および対数正規分布を仮定した右側打ち切りデータを含む回帰分析についても解説されている.

秋山ら (2016) は, 左側打ち切り (定量下限未満) を含むデータに対する要約統計量と各種の統計解析を紹介している. また, 高橋 (2016) は, 正規分布を仮定した左側打ち切り (定量下限未満) を含むデータについての最尤法について詳しく紹介している. 様々な左側打ち切りの統計解析については, Breen (1996) に広範な事例が紹介されている.

第 10 節以後は, 定例会では割愛した文献データについての追試を含め, 打ち切りデータを含む回帰分析の理論的背景を付け加えた.

第 10 節では, 無料の On-Demad SAS の LIFEREG プロシジャを用い, 打ち切りデータを含む発毛実験データについての解析事例を示す.

第 11 節では, 打ち切りデータを含んだ文献データを用いて, LIFEREG プロシジャ, 「寿命の一変量」, および, Excel のソルバーを用いて, 平均値と標準偏差の推定結果を示す.

第 12 節では, Excel の基本機能のみを用いて, 正規分布を仮定した打ち切りデータに対して, 平均値と標準偏差を求めるための対数尤度関数を定義し, その偏微分式からニュートン・ラフソン法によるパラメータの推定方法を示す.

第 13 節では, 打ち切りがある発毛データを用いた回帰分析を Excel の基本機能のみを用いたニュートン・ラフソン法により, 回帰パラメータ, および, 誤差分散の推定を行う. 対数尤度関数の導出には, 第 12 節での偏微分式を活用する方法を示す.

1. 発毛実験で発生する打ち切りデータ

マウスを用いた発毛試験データについて黒須(2017)は、種々の解析方法の適用と比較を行った。発毛試験は、観察を36日目で終了したために、36日目で発毛が観察されない打ち切りデータが含まれている。打ち切り日を含む発毛日のデータに対し、以下のノンパラメトリックの多重比較、パラメトリックな多重比較について検討されている。

- (a) ノンパラメトリック多重比較の方法である Steel 検定,
- (b) 観察終了36日目における発毛スコアを用いた Steel 検定,
- (c) Kaplan-Meier 曲線による図示, 生存時間解析における Dunnett 型の比較に応じた多重性調整 log-rank 検定,
- (d) スコア経時推移の曲線下面積に対するパラメトリックな Dunnett 検定により対照群と薬剤投与群との群間比較を行う。

スライド 3

1. 打ち切りデータの実例

2017.11.18 実験行進 3

スライド 4

発毛までの日数

- ◆ 実験の概要
 - 発毛までの日数を測定
 - 試験期間は, 36日間
 - 対照群と3用量の群間比較試験
 - 1群当たり10匹
- ◆ 反応変数
 - 36日目までに発毛した場合の日数
 - 36日に発毛していなければ, 36日(打ち切りあり)

2017.11.18 実験行進 4

論文は、J-STAGE で公開されており、詳細な実験結果も添付資料となっている。発毛試験は、マウスの背中を剃毛後、薬剤を塗布しながら 36 日目まで、2～3 日間隔でスコアを計 15 回、経時的に観察し、発毛した面積に応じた発毛スコアで評価している。

- 1：皮膚がピンク色,
- 2：皮膚が灰色（剃毛部の 30%未満）,
- 3：皮膚が灰色（剃毛部の 60%未満）または毛が伸長（剃毛部の 30%未満）
- 4：皮膚が灰色（剃毛部の 60%以上）または毛が伸長（剃毛部の 60%未満）
- 5：毛が伸長（剃毛部の 60%以上）

表 1 に対照群の 10 匹の観察結果を示す。動物番号 1 のラットは、36 日目で発毛スコア 5 に達している。動物番号 2 のラットは、36 日目で発毛スコア 4 であり、スコア 5 に達していない。スコア 5 に達した日数を発毛した日とするならば、動物番号 2 のラットは、36 日で打

ち切りデータとなる。ここでの打ち切りデータとは、36 日以後も薬剤を塗布して観察を続けているならば、スコア 5 に達する可能性があるが、その日数は不明と解釈される。

表 1 対照群の発毛データ

	動物番号	観察日															スコア 5 に達した日 (灰色セルは打ち切り)	
		2	5	7	9	12	14	16	19	21	23	26	28	30	33	36		
Control	1	1	1	1	1	1	1	1	1	1	1	1	1	2	4	5	36	
	2	1	1	1	1	1	1	1	1	1	1	2	2	2	3	4	36	
	3	1	1	1	1	1	1	1	1	1	1	1	2	2	5	5	33	
	4	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	36	
	5	1	1	1	1	1	1	1	1	1	1	2	4	5	5	5	30	
	6	1	1	1	1	1	1	1	1	1	2	2	3	5	5	5	30	
	7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	36	
	8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	36	
	9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	36
	10	1	1	1	1	1	1	1	1	1	1	2	2	2	3	5	5	33

(低濃度, 中濃度, 高濃度) は略

スライド 5 は, 低濃度群, 中濃度群, 高濃度群のデータも含めて発毛日について整理した表である。対照群の動物番号 1 のラットは, 36 日目で発毛スコア 5 に達しているので 36, 動物番号 2 のラットは, 36 日目でスコア 5 に達していないので 36+ としている。

スライド 5

実験結果・発毛までの日数

Group	群	1	2	3	4	5	6	7	8	9	10	打ち切り
1	対照群	36	36+	33	36+	30	30	36+	36+	36+	33	5匹
2	低濃度	36	28	36	36+	28	36+	36+	28	33	30	3匹
3	中濃度	19	23	26	33	36+	21	16	26	33	26	1匹
4	高濃度	14	28	12	21	23	14	12	14	26	21	0
36日で発毛(36), 36日で発毛せず(36+)												

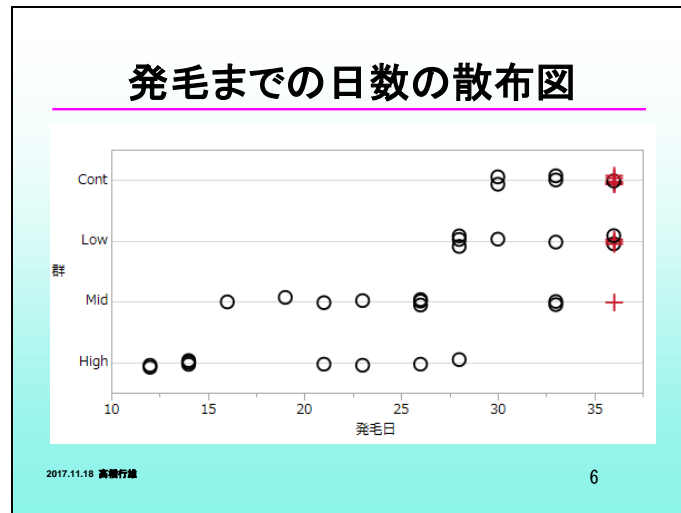
- ◆ 実験開始から反復塗布
- ◆ 2～3日毎に発毛状態を観察
- ◆ 36日目で発毛が観察される場合もある

2017.11.18 高橋行雄

5

対象群には, 36+ が 5 匹いるが, 濃度が高くなるにつれて, 打ち切りとなるラットは減り, 高濃度群では, 全て 36 日以前に発毛が観察されている。スライド 6 に群別に発毛日についての散布図を示す。打ち切りデータは, 36 日目で+の記号で示している。

スライド 6



実験結果に打ち切りデータがなければ、塗布濃度を説明変数 x としての回帰分析、あるいはシグモイド曲線をあてはめる非線形回帰分析が適用できる。打ち切りデータを便宜的に 36+ を 37 に読み替えて回帰分析を適用することは形式的には可能ではあるが、対照群、あるいは低濃度群の発現日数を過小評価することになってしまう。

スライド 7

回帰分析をしたいのだが

- ◆ 36日での打ち切りデータの扱いはどうすればよいのだろうか。
- ◆ 36日までに発毛したデータだけで回帰分析を行うしかないのだろうか。
- ◆ 半数が打ち切りられた群があるので、回帰分析は、あきらめなければならないのだろうか。
- ◆ なぜ、すべて発毛するまで観察しなかったのだろうか。

2017.11.18 高橋行雄 7

スライド 8

回帰分析はできる

- ◆ 打ち切りデータを含む回帰分析は、故障・寿命データの統計解析の分野で確立している。
- ◆ JMPの「寿命の二変量」を使えば簡単にできるようになっている。
 - ◆ ただし、打ち切りデータを含む統計解析の基礎的な素養を必要とする。
- ◆ 本日は、Excelを用いて、打ち切りデータを含む統計解析の基礎について学習する。

2017.11.18 高橋行雄 8

打ち切りデータを含む回帰分析については、寿命あるいは故障データの統計解析では標準的な解析方法であるが、他の分野に普及しているとは言い難い。SASのLIFEREGプロシジャを用いれば、打ち切りを含む回帰分析は簡単にできるようになっているが、そのことについて述べられている日本語の教科書は見当たらない。従って、自ら統計解析を実施している研究者が、参考となるような教科書あるいは論文がないかぎり、適用することは不可能であろう。

打ち切りデータがある場合の統計解析は、生存時間解析として臨床試験の統計解析の分野では、よく知られており多くの教科書・論文もあるが、打ち切りデータを含む回帰分析についての記述を見出すことができない。

SAS の LIFEREG プロシジャでできると述べたが、結果の解釈には相当な統計の素養が必要であり、一般には推奨することができない。幸い、SAS 社から提供されている JMP の「寿命の二変量」は、LIFEREG プロシジャに比べ、結果の解釈に役立つ図が多数用意されているだけでなく、LIFEREG プロシジャでの対応が困難な課題に対しても包括的な出力結果を簡単に得ることができる。ただし、打ち切りデータを含む統計解析について基礎的な素養を必要とするが、推奨できる入門書は全くない。そこで、Excel を用いて、打ち切りデータを含む統計解析の基礎の基礎についてレンガを積みがごとく解説し、打ち切りデータを含む回帰分析について、身近に感じられるようになることを念願している。

2. 打ち切りデータを考慮した平均値と標準偏差

実験結果については、何らかの統計量で要約を示すことが一般的である。対照群 10 匹中 5 匹の発毛日が 36+ のような打ち切りデータの場合にどのような要約統計量を示したらよいのであろうか。打ち切りデータが含まれていなければ、平均と標準偏差は、簡単に計算できる。打ち切りデータがごく少数ならば、そのまま使って平均値を「母平均」の推定値として良いかも知れない。また、報告書に示す統計量としては、中央値と 4 分位範囲を出せば十分かも知れない。

スライド 10

2. 打ち切りを考慮した統計量

2017.11.18 実験行進 10

スライド 9

群ごとの平均値と標準偏差

- ◆ 打ち切りデータが含まれていなければ、平均と標準偏差は、簡単に推定できる。
- ◆ 打ち切りデータが、ごく少数ならば、そのまま使っても、平均値を「母平均」の推定値としてよいかも。
- ◆ 統計量としては、中央値と4分位範囲を出せば十分かも知れない。

2017.11.18 実験行進 9

各群の要約統計量の提示ならば、打ち切りデータが含まれていたとしても、統計的に適切な中央値などを示せばよく、打ち切りデータを考慮した要約統計量などは必要としない。しかし、2 群間の平均値の差の検定、回帰分析などを行うためには、どのように対応したら良いのであろうか。

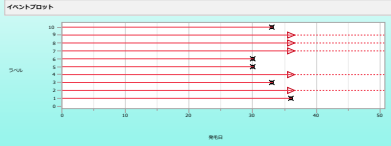
表 2 打ち切りを 1 とする JMP データセット

		群	No	Group	発毛日	打ち切り
○	1	Cont	1	1	36	0
+	2	Cont	2	1	36	1
○	3	Cont	3	1	33	0
+	4	Cont	4	1	36	1
○	5	Cont	5	1	30	0
○	6	Cont	6	1	30	0
+	7	Cont	7	1	36	1
+	8	Cont	8	1	36	1
+	9	Cont	9	1	36	1
○	10	Cont	10	1	33	0

スライド 11

打ち切りデータを考慮した平均

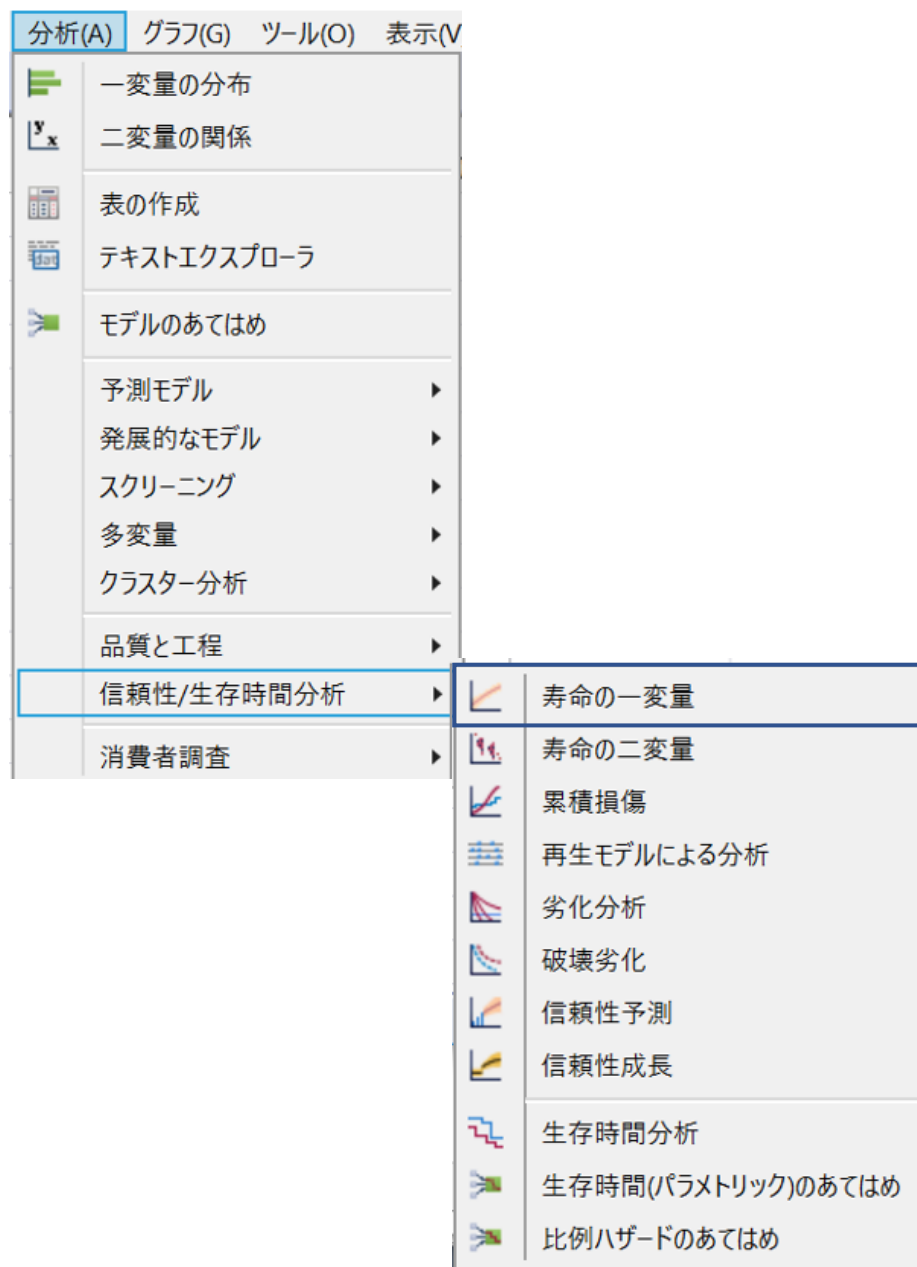
- ◆ 打ち切りデータを含む各用量群ごとの平均と標準偏差を求めたい。
- ◆ JMP「寿命の一変量」を使う



2017.11.18 実験行進 11

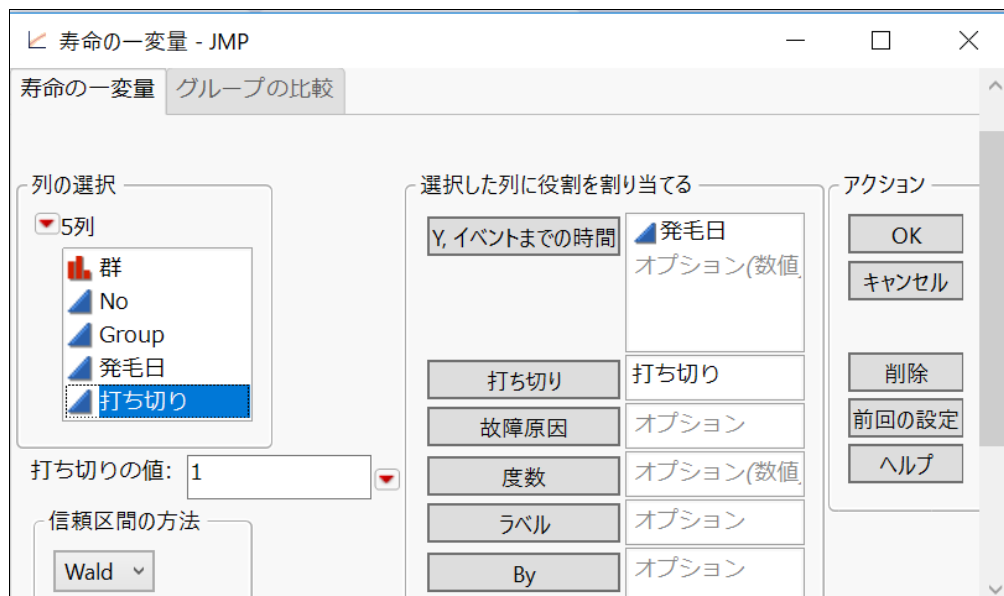
そこで、JMP「寿命の一変量」を使って、打ち切りデータを含んだ場合について、平均値と標準偏差を算出してみる。打ち切りデータであることを明示するために、変数名「打ち切り」を新たに作成し、打ち切りでなければ0、打ち切りであれば1とする。JMPの「分析」メニューの中から「信頼性/生存時間解析」を選択し、更に「寿命の一変量」を選択する。

図 1 JMP の分析メニュー (Ver. 14)



「寿命の一変量」の設定画面では、「発毛日」を「Y, イベントまでの時間」に、変数「打ち切り」を「打ち切り」に設定する。

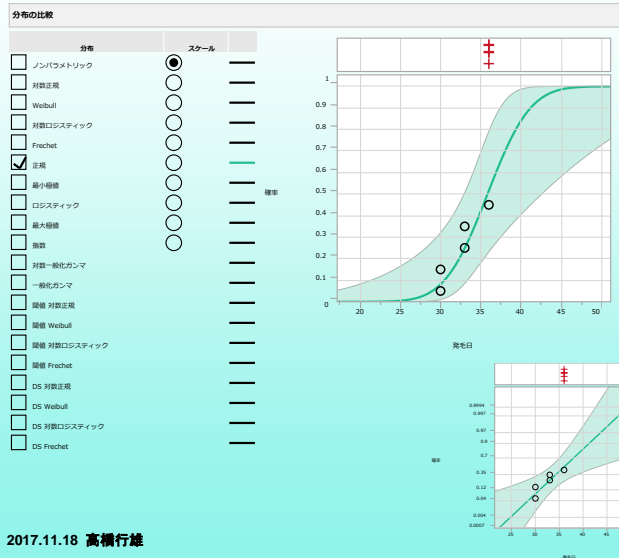
図 2 寿命の一変量



実行するとスライド 11 に示した結果の「イベントプロット」が表示される。スライド 12 の出力は、「分布の比較」の「分布」の欄で「正規」を選択し、「スケール」の欄で「ノンパラメトリック」を選択した場合の結果である。さらに、発毛日の X 軸をマウスで 20 日～50 日なるように整形し、確率 Y 軸を 0～1 となるように整形する。図の中の曲線は、累積正規分布およびその 95%信頼区間であり、打ち切りでない発毛日データ (30, 30, 33, 33, 36, 36) が ○印で表示され、5 個の打ち切りデータ 36+は、図の上方に+の重なりで示されている。

累積正規分布の下方にある図は、「スケール」を「ノンパラメトリック」から「正規」に変更した結果で、打ち切りデータを考慮した正規確率プロットである。

対照群に正規分布をあてはめる



スケールを正規とすれば、打ち切りデータを考慮した正規分位点プロットとなる。

12

スライド 13 には、データが正規分布に従うと仮定した場合の平均値と標準偏差が推定されている。打ち切りデータに対しても、正規分布に従うと仮定しているが、どのような方法で

対照群の平均値と標準偏差

パラメトリック推定 - 正規						
パラメータ	推定	標準誤	下側95%	上側95%	規準	
位置	35.835	1.623	32.654	39.015	(-2)*対数尤度	35.562677
尺度	4.172	1.484	1.264	7.080	AICc	41.276963
平均	35.835	1.623	32.654	39.015	BIC	40.167848

- ◆ パラメータ欄の位置は、平均を意味する。
- ◆ “ 尺度は、標準偏差 ”
- ◆ 平均値 35.835 は、中央値 36 に近い

1	2	3	4	5	6	7	8	9	10
30	30	33	33	36	36+	36+	36+	36+	36+

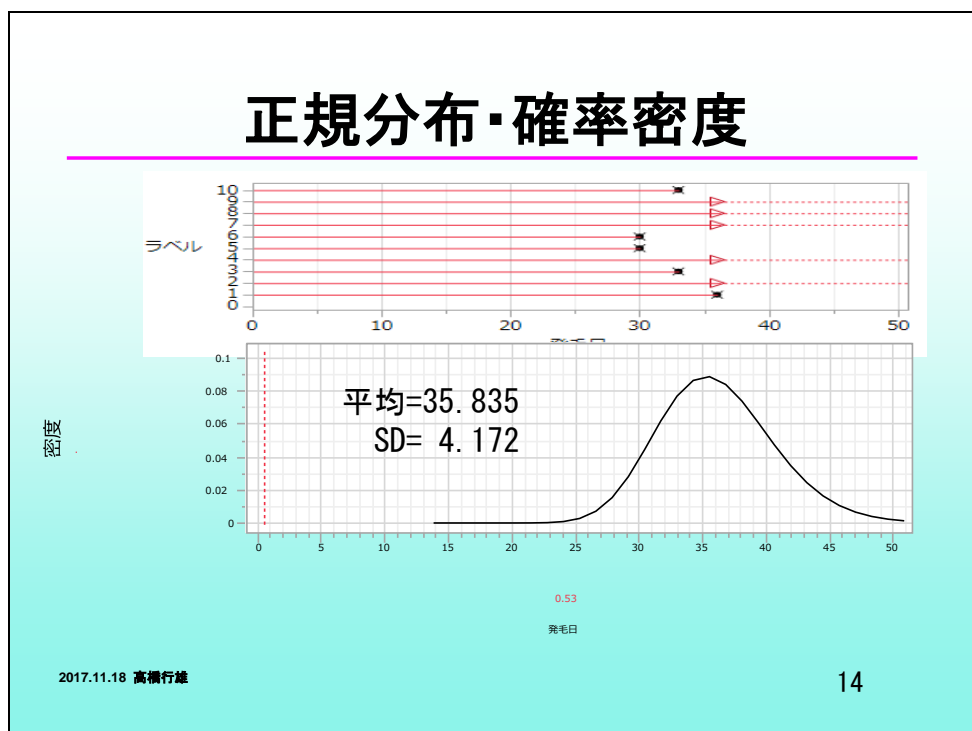
2017.11.18 高橋行雄

13

平均と標準偏差を求めているのであろうか。JMP の解説書にも記述がない。どのような方法なのかわからない場合に、統計ソフトの出力結果を報告書に使うことにためらいが生ずるであろう。尺度は標準偏差とスライドで説明はあるが、正しいのであろうか？ 平均値は中央値に近いとあるが、どのような計算方法なのかわからなければ報告書に使うのはためらわれる。

スライド 14 は、**プロファイル** 出力の **分布プロファイル** を整形して、「イベントプロット」の下に張り付けた結果である。平均=35.835 と SD=4.172 は、別途上書きしたもので、JMP の出力結果ではない。これは、単にスライド 12 に示した累積正規分布の代わりに確率密度曲線をつかった別表現である。

スライド 14



3. 段階的な学習

故 芳賀先生は、スライド 16 に示すように「説明は、レンガを積むがごとく段階的に行い、途中を省略しない」、「途中を省略しなければならないような説明はしない。見栄は張らない」、「統計ソフトの出力結果について Excel で追試できない場合には、その解析結果について、自信をもって説明できないので、説明を差し控える」ことを繰り返し繰り返し強調されていた。

スライド 15

3. 段階的な学習

2017.11.18 高橋行雄

15

スライド 16

故・芳賀先生の実践

- ◆ 説明は、レンガを積むがごとく段階的に行い、途中を省略しない。
- ◆ 途中を省略しなければならないような説明はしない。見栄は張らない。
- ◆ 統計ソフトの出力結果について Excel で追試できない場合には、その解析結果について、自信をもって説明できないので、説明を差し控える。

2017.11.18 高橋行雄

16

故 芳賀先生の 3 部作、通称グリーン本もこの考え方に貫かれて執筆されている。「Excel で理論的説明をし、その計算結果が、JMP の出力と一致することを常に確認している」ことが貫かれている。また、芳賀（2010）では、「不揃いな乱塊法などの実験は、混合効果モデルとなり、JMP の解析した場合の『REML』の結果は、グリーン本では、Excel で追試できない」ことを明記している。

スライド 17

通称グリーン本

- ◆ グリーン本（医薬品開発のための統計解析・3 部作）レンガを積むがごとく、行間を飛ばさず段階的な説明がされている。
- ◆ Excel で理論的説明をし、その計算結果が、JMP の出力と一致することを常に確認している。
- ◆ 不揃いな乱塊法などの実験は、混合効果モデルとなり、JMP の解析した場合の「REML」の結果は、グリーン本では、「Excel で追試できない」ことを明記している。

2017.11.18 高橋行雄

17

スライド 18

打ち切りデータを含む統計解析

- ◆ グリーン本には、含まれていない。
- ◆ グリーン本で用いている統計解析は、最小 2 乗法を基本としている。
- ◆ 線形モデルの場合には、Linest 関数による回帰分析を主体にしている。
- ◆ 非線形モデルの最小 2 乗法では、ソルバーを用いて誤差平方和を最小化して、パラメータを推定している。

2017.11.18 高橋行雄

18

打ち切りデータを含む統計解析は、残念ながらグリーン本には、含まれていない。これは、グリーン本で用いている統計解析は、最小 2 乗法を基本としていること、線形モデルの場合には、Excel の `Linest()` 関数による線形回帰分析を主体にしていること、非線形モデルの最小 2 乗法では、ソルバーを用いて誤差平方和を最小化してパラメータを推定している。なお、最尤法については、2 値反応のグリーン本第 3 部の「第 4 章 ロジスティック回帰分析」で取り上げられているのみである。

4. 最小 2 乗法の復習

打ち切りデータを含む統計解析には、最小 2 乗法でなく最尤法が使われている。最尤法を理解し、実践するためには段階的な学習が欠かせない。第 1 歩は、最小 2 乗法による各種の推定を Excel のソルバーを用いて実践することである。高橋(2013)、じっくり勉強シリーズ第 7 回 http://www.scientist-press.com/edit_html/green/seminar7.pdf では、「統計ソフトが対応していない生物統計の各種の課題を Excel でサクサク解こう」をテーマとした。

スライド 19

4. 最小2乗法の復習

2017.11.18 高橋行雄

19

スライド 20

じっくり勉強シリーズ 第7回

- ◆ 回帰分析・再入門
 - 統計ソフトが対応していない生物統計の各種の課題をExcelでサクサク解こう

回帰分析の一步進んだ活用として逆推定がある。未知検体の測定値 y から濃度 x を求める場合、回帰直線が求めれば y から x の点推定値は計算できるが、推定精度を知るために必要な信頼区間の算出方法に触れた資料は少ない。Y 軸方向の平均値や個別データの 95%信頼区間までは多くの統計ソフトで計算可能だが、逆推定の個別データの 95%信頼区間までは対応していない。本セッションでは、使う機会が多いがキチンと解説されているとはいえない回帰分析について、Excel を用いて理解することを目的とする。そのために様々な応用が可能な計算の方法をじっくりと解説し、数値例や手順の資料を提供する。そして、受講者が回帰分析の逆推定や重み付き回帰分析などの高度の解析を日常的に使いこなせるようになることを願っている。

2017.11.18 高橋行雄

20

最小 2 乗平均は、算術平均に一致することが知られていて、最小 2 乗平均を求めるために偏差平方和を最小にするような計算手順は、全く必要ない。しかし、打ち切りデータを含む

スライド 21

偏差平方和

- ◆ 各種の平均値の中で、最もよく使われるのは「算術平均」である。
- ◆ 最小 2 乗法では、推定平均を μ^{\wedge} としたときに、 μ^{\wedge} と各データとの偏差の合計(偏差平方和)が最小となるような μ^{\wedge} を推定する。
- ◆ 最小 2 乗推定量 μ^{\wedge} は、算術平均値に一致することが、グリーン本で強調されている。
注) μ^{\wedge} は、ミュー・ハットと読む。

2017.11.18 高橋行雄

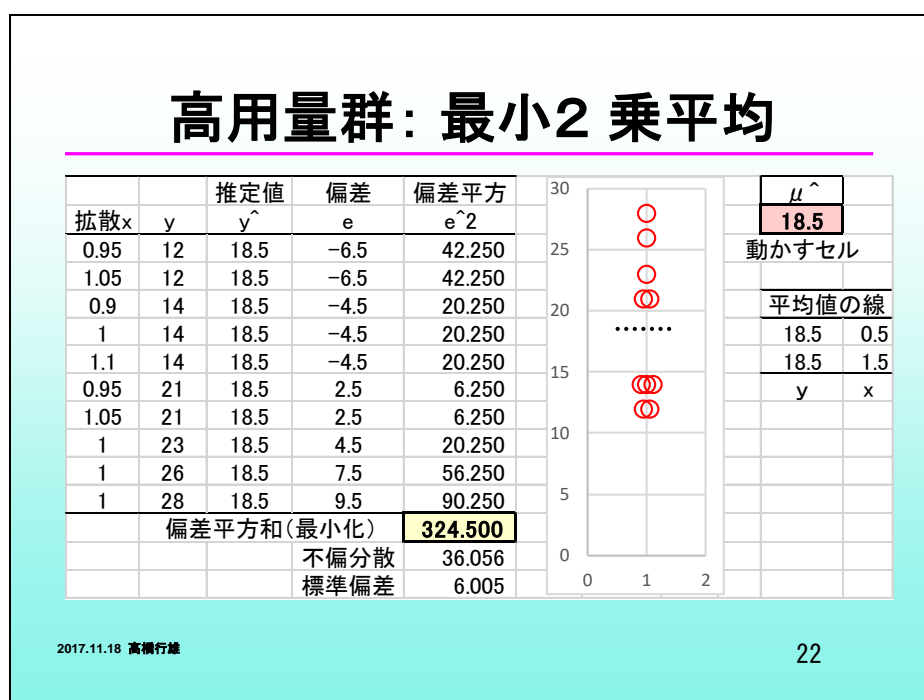
21

データに対して、算術平均を適用することはできない。まず、打ち切りがないデータに対して、偏差平方和を最小にするようなある値を推定して、それが算術平均に一致することを確認するのが、初めの一步である。

スライド 22 に示すように、発毛試験の高濃度群の小さい順位に並べたデータ (12, 12, 14, 14, 14, 21, 21, 23, 26, 28) に対して、最小 2 乗平均を求めてみよう。ある値を「動かすセル」 $\hat{\mu}=18.5$ に入力すると、推定値 \hat{y} の欄に $\hat{\mu}=18.5$ が、すべてにコピーされている。偏差 e の欄には、 $y_i - \hat{y}_i$ が計算されている。さらに、偏差 e の平方 e^2 が計算され、それらの和、偏差平方和が 324.5 と計算されている。

偏差平方和の下に不偏分散と標準偏差が計算されているが、不偏分散は、偏差平方和を自由度 ($n-1$) で割ったものであり、標準偏差は、不偏分散の平方根を取ったものである。このような Excel の計算シートを作成し、偏差平方和が最小となるように、ある値 $\hat{\mu}$ を試行錯誤的に動かして、ある $\hat{\mu}$ の場合に結果が最小になることが確認できれば、それが最小 2 乗法によって推定された最小 2 乗平均である。

スライド 22



偏差平方和が最小であることを確認するためには、 $\hat{\mu}$ を前後に動かして、偏差平方和が最小になることを確認すればよい。スライド 23 では、18.5 の前後 (18.0, 19.0) としているが、(18.49, 18.51) の様に狭めても、18.5 の場合の偏差平方和が小さいことが確認できる

このような探索的な作業は、Excel のソルバーが代行してくれる。偏差平方和の場合には、**目的セルの設定** には、「偏差平方和」のセルを貼り付け、**目標値** では **最小値** を選択し、**変数セルの変更** に「 $\hat{\mu}$ 」のセルを貼り付け、実行する。

スライド 23

平均を変化させた結果

$\hat{\mu}$	偏差平方和	
18	327.000	
18.5	324.500	最小値
19	327.000	

標準偏差は、偏差平方和を用いて計算
 $SD = \sqrt{324.5/(10-1)} = 6.005$

2017.11.18 高橋行雄 23

スライド 24

ソルバーの使い方

- ◆ 設定: Excelのファイルタブ>オプション>アドイン>ソルバーアドイン
- ◆ 呼び出し: データタブ>ソルバー

平方和など
 最小値を選択
 必要に応じて選択

2017.11.18 高橋行雄 24

図 3 ソルバーの設定画面

平方和など
 最小値を選択
 必要に応じて選択

Excel のソルバーによる各種の応用については、芳賀（2004）を参照してもらいたい。私も、これに刺激を受けて、ソルバーを活用するようになった。

5. 正規分布の確率密度の活用

最小 2 乗法は、統計解析の基本であり回帰分析、重回帰分析、分散分析、非線形回帰分析などで使われている統計解析のための標準的な各種のパラメータ推定のための最も基本的な考え方である。しかし、データが正規分布に従っていると仮定した場合であっても、その仮定は全く用いられていない。単に仮定するだけであり、(0, 1) 反応でベルヌーイ分布に従うようなデータでも、正規分布を仮定している最小 2 乗法による回帰分析を何の問題なく適用できる。このような最小 2 乗法の融通むげ的な柔軟性が、何のためらいもなく汎用的に使われ続けられているのであろう。

スライド 26 に示すように、打ち切りデータには、偏差平方和が定義できない。まず、打ち切りのない高濃度群のデータに対して、ある平均値 $\hat{\mu}$ とある標準偏差 $\hat{\sigma}$ を持つ正規分布を想定して、それぞれのデータに対する確率密度を計算してみよう。それぞれのデータの確率密度の積を考え、確率密度の積を最大にするような $\hat{\mu}$ と $\hat{\sigma}$ を推定してみる。

スライド 25

5. 正規分布確の率密度の活用

2017.11.18 高瀬行雄

25

スライド 26

最小2乗法とは異なる方法

- ◆ 打ち切りデータには、偏差平方和が定義できないので、ある平均値 μ^{\wedge} とある標準偏差 σ^{\wedge} を持つ正規分布を想定して、それぞれのデータに対する確率密度を計算する。
- ◆ それぞれの確率密度の積を考え、確率密度の積を最大にするような μ^{\wedge} と σ^{\wedge} を推定してみる。
- ◆ 打ち切りのない高用量群で計算してみる

2017.11.18 高瀬行雄

26

その前に、正規分布の確率密度関数と累積分布関数について復習する。平均値が $\mu=0.0$ で、標準偏差 σ を (1.0, 0.5, 0.3) と変化させたときに、正規分布の確率密度を計算する。正規分布の確率密度関数を $f(x; \mu, \sigma)$ とした時に、 $\sigma=1$ の場合 $f(x; 0, 1)=0.399$ となる。これは、Excel の関数の Norm.dist() を使うことにより計算できる。標準偏差を $\sigma=0.5$ と小さくすると $f(x; 0, 0.5)=0.798$ と確率密度は大きくなり、標準偏差を $\sigma=0.3$ とすると $f(x; 0, 0.3)=1.330$ の様に 1 を超える。

スライド 28 に標準偏差 σ を (1.0, 0.5, 0.3) と変化させた場合の確率密度曲線を示す。標準偏差 $\sigma=1$ の場合は標準正規であり、標準偏差を小さくすると曲線が尖ってくる。逆に大きくすると平べったくなる。

スライド 27

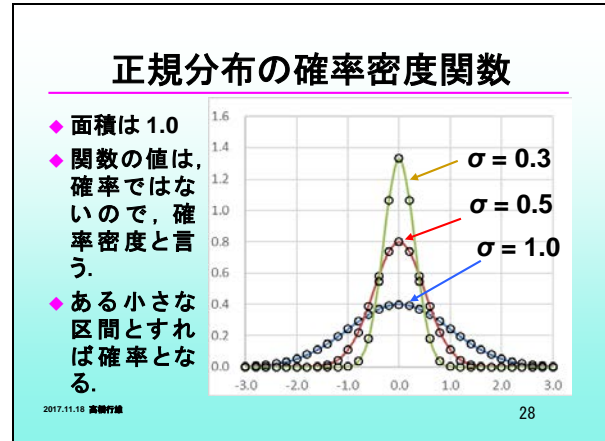
確率密度とは何か

$\mu=$	0	0	0	0	0	0
$\sigma=$	1	0.5	0.3	1	0.5	0.3
	確率密度関数 $f(x)$			累積分布関数 $F(x)$		
x	$N(0, 1)$	$N(0, 0.5)$	$N(0, 0.3)$	$N(0, 1)$	$N(0, 0.5)$	$N(0, 0.3)$
-1.0	0.242	0.108	0.005	0.159	0.023	0.000
0.0	0.399	0.798	1.330	0.500	0.500	0.500
1.0	0.242	0.108	0.005	0.841	0.977	1.000
3.0	0.004	0.000	0.000	0.999	1.000	1.000
	Norm.dist($x, \mu, \sigma, \text{false}$)			Norm.dist($x, \mu, \sigma, \text{true}$)		

正規分布の確率密度関数: $f(x)$ or $\phi(x)$
 " 累積分布関数: $F(x)$ or $\Phi(x)$
 (下側確率)

2017.11.18 高橋行雄 27

スライド 28



事象 A が起きる確率を $p_A = 0.3$ ，事象 B が起きる確率を $p_B = 0.2$ とした場合に，互いに独立の場合に事象 A と事象 B が同時に起きる確率は， $p_{A \cap B} = 0.3 \times 0.2 = 0.06$ の様に確率の積となる。正規分布の場合 $(x-0.001) \sim (x+0.001)$ などの様に小さな区間を設定すれば，面積になり，その間に起きる確率となる。確率密度は，「確率」ではないが，確率と同様の性質を持っている。

スライド 29

- ### なぜ 積 なのか
- ◆ 事象 A が起きる確率: 0.3
 - ◆ 事象 B が起きる確率: 0.2
 - ◆ 互いに独立の場合に, 事象 A と 事象 B が同時に起きる確率 $0.3 \times 0.2 = 0.06$
 - ◆ $(x-0.001)$ から $(x+0.001)$ などの様に小さな区間を設定すれば, 面積になり, その間に起きる確率となる。確率密度は, 「確率」ではないが, 確率と同様の性質を持っている。
- 2017.11.18 高橋行雄 29

スライド 30 に高濃度群の 10 匹の発毛日についての正規分布の確率密度の積の計算結果を示す．表の右端では，平均 $\hat{\mu}=18.5$ ， $\hat{\sigma}=3$ とした場合に発毛日 $y_1=12$ の確率密度は， $L_1=f(12,18.5,3)=0.0127$ と計算されている．それらの 10 個の積は，

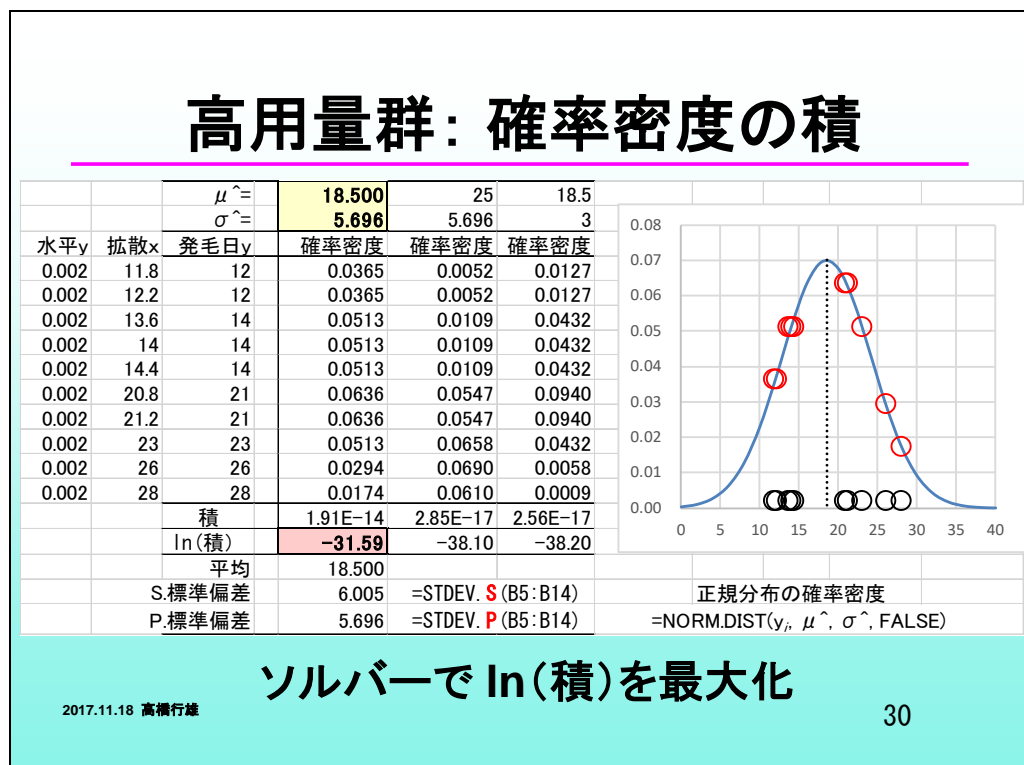
$$L=\prod_{i=1}^{10} L_i=\prod_{i=1}^{10} f(y_i,18.5,3)=2.56\times 10^{-17}$$

であり，対数を取ると $\ln L=\ln(2.56\times 10^{-17})=-38.20$ となる．この $\ln(\text{積})=\ln L$ を最大化するように，平均 $\hat{\mu}$ ，標準偏差 $\hat{\sigma}$ を変化させてソルバーで $\ln L$ を最大化させると $\ln L=-31.59$ ， $L=1.91\times 10^{-14}$ となり， $\hat{\mu}=18.5$ ， $\hat{\sigma}=5.696$ が得られる．

スライド 30 の正規分布をあてはめたグラフは， $\hat{\mu}=18.5$ ， $\hat{\sigma}=5.696$ として発毛日を 0 から 40 まで小さな刻みで正規分布の確率密度を Excel シートの下方で計算した表を用いて線グラフ化したものである．

X 軸上の発毛日の○印は，Y 軸の値を 0.002 とし X 軸を発毛日でプロットしたものである．同じ発毛日が重ならないように発毛日を微調整している．確率密度曲線上の○印は，微調整した発毛日を X 軸，それぞれの発毛日の確率密度を Y 軸としてプロットしたものである．

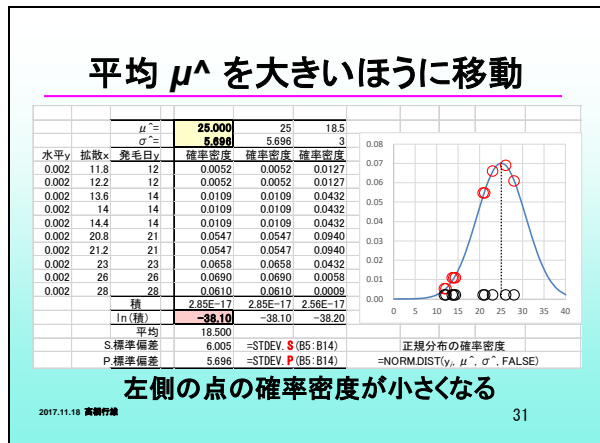
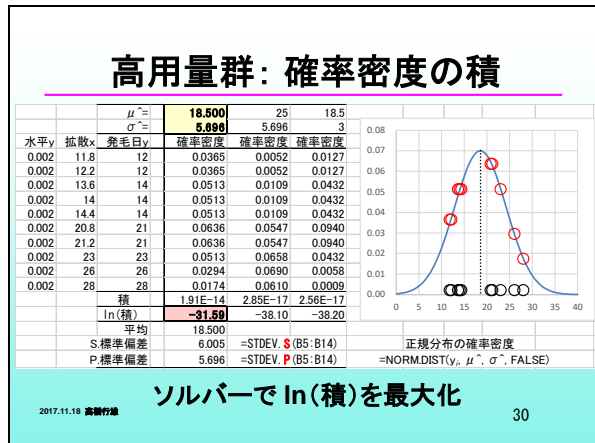
スライド 30



スライド 31 は、標準偏差は 5.686 のままで、平均を 25 と大きくした場合で、発毛日 12 の場合の確率密度が 0.0052 と小さくなっている。発毛日 28 の場合は確率密度が 0.0172 から 0.0610 と大きくなっているが、確率密度の積は $L = 2.85 \times 10^{-17}$ と小さくなっている。

スライド 30 再掲

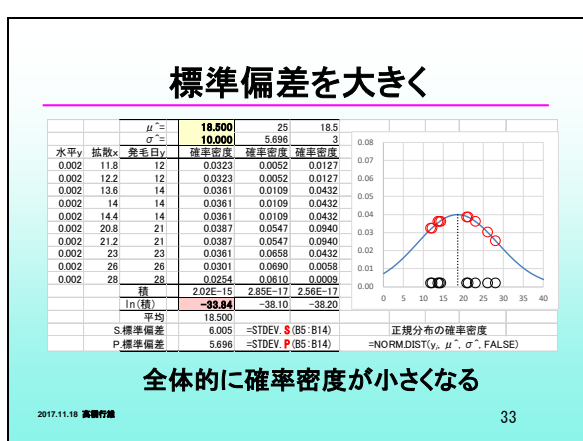
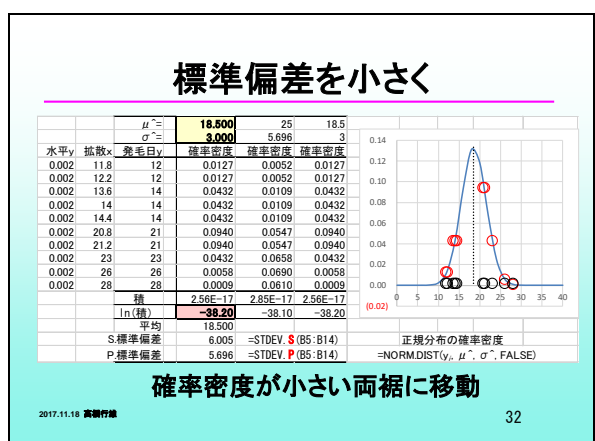
スライド 31



スライド 32 は、平均は 18.5 のままで、標準偏差を 3 と小さくした場合であり、発毛日 21 日の確率密度は 0.940 と 0.0636 より大きくなるが、確率密度の積は $L = 2.56 \times 10^{-17}$ と小さくなっている。スライド 33 は、平均は 18.5 のままで、標準偏差を 10 と大きくした場合であり、確率密度曲線は平坦化し、確率密度も小さくなっている、確率密度の積は、 $L = 2.02 \times 10^{-15}$ と小さくなる。

スライド 32

スライド 33



スライド 34 は、正規分布のパラメータと確率密度の積についてこれまで比較した結果をまとめたものである。確率密度の積を最大化する平均は、算術平均 18.5 に一致するが、標準偏差は 5.696 と通常の方法

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{324.5}{9}} = 6.005$$

で計算された 6.005 に一致しない。

推定された 5.696 は、偏差平方和を自由度 ($n-1$) で割るのではなく、

$$\tilde{\sigma} = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}} = \sqrt{\frac{324.5}{10}} = 5.696$$

とデータ数 n で割った結果となる。これは、得られたデータを「母集団」と見なした標準偏差となっている。

スライド 34

確率密度の ln(積) の最大化

$\hat{\mu} =$	18.500	25	18.5
$\hat{\sigma} =$	5.696	5.696	3
積	1.91E-14	2.85E-17	2.56E-17
ln(積)	-31.59	-38.10	-38.20

最大値

$\hat{\sigma}$ は、 $SD = \sqrt{324.5/10} = 5.696$ に等しい

偏差平方和/データ数 n

平均	18.500	
S.標準偏差	6.005	=STDEV. S (B5:B14)
P.標準偏差	5.696	=STDEV. P (B5:B14)

2017.11.18 高橋行雄

34

Excel の関数 Stdev.s() の s は、標本 sample (標本) の意味で、データが母集団からサンプリングされたと仮定する場合で、標準偏差を計算する場合に母平均が不明なので、サンプリングされた平均 \bar{y} で代用している。そのため、平均からの差のデータは、自由度が 1 減るので $n-1$ で割り算をしている。Stdev.p() の p は、population (母集団) の意味で、10 個のデータそのものを母集団とみなした場合であり \bar{y} は母平均 μ と等しくなり、自由度の減少を考慮する必要がなくなる。

6. 正規分布の上側確率の活用

スライド 36 に示すように対照群では、発毛日が観察されたのは、5 匹で (30, 30, 33, 33, 36 日) である。残りの 5 匹は 36 日には発毛が観察されなかったため (36+) と表記する。中央値は、36 と 36+ の間なので、36+ を 37 とみなせば、36.5 となり、観察を続けて発毛日が得られた場合でも同じであり正確である。しかし、四分位範囲は、(32.5, 36+) と苦し紛れの結果となる

スライド 35

6. 正規分布の上側確率の活用

2017.11.18 高橋行雄

35

スライド 36

対照群のデータ

- ◆ 対照群では、発毛日が観察されたのは5匹で(30, 30, 33, 33, 36), 残りの5匹は36日(36+)には発毛が観察されなかった(36+).
- ◆ 中央値は、36と36+の間なので、36+を37とみなせば、36.5となる.
- ◆ 四分位範囲は、(32.5, 36+)と苦し紛れの結果となる.

1	2	3	4	5	6	7	8	9	10
30	30	33	33	36	36+	36+	36+	36+	36+

2017.11.18 高橋行雄

36

データが正規分布に従うと仮定した場合に、打ち切りデータがない場合には、正規分布の確率密度を用いて最尤法によって平均と標準偏差を推定した。しかし、打ち切りデータがある場合には、この考え方をすることができない。打ち切りデータ 36+ の場合は、発毛日が

スライド 37

打ち切りデータの取り扱いの基本

- ◆ データが正規分布に従うと仮定する.
- ◆ 正規分布のパラメータは、平均 μ と標準偏差 σ である.
- ◆ 観測されたデータは、正規分布の確率密度としたのだが、36+ の場合は、発毛日が36日以上なので、確率で考えると36日の右側全体、すなわち正規分布の上側確率とするのが妥当であろう.

2017.11.18 高橋行雄

37

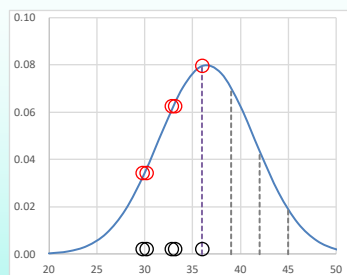
36 日以上であることは確かであるが、何日目であるかはわからない。この場合に、確率で考えると正規分布の 36 日以上は、正規分布の右側全体、すなわち正規分布の上側確率とするのが妥当であろう。

対照群の場合について平均を 36.5、標準偏差を 5 とした場合の上側確率をスライド 38 に示す。Excel の正規分布の分布関数は、下側確率を計算するので上側確率は、1 から引けばよい。上側確率は、 $1 - \text{Norm.dist}(36, 36.5, 5, \text{true}) = 0.5398$ と計算されている。発毛日が 36+ のデータは 5 匹なので、それらの積は、 4.58×10^{-2} であり、(30, 30, 33, 33, 36) の確率密度の積は、 3.64×10^{-7} である。それらの積の積は、 $(4.58 \times 10^{-2}) \times (3.64 \times 10^{-7}) = 1.67 \times 10^{-8}$ と計算されている。その対数は、 $\ln(1.67 \times 10^{-8}) = -17.91$ となる。

スライド 38

対照群：確率密度と上側確率

		$\mu^{\wedge} =$	36.500	
		$\sigma^{\wedge} =$	5.000	
水平y	拡散x	発毛日	確率密度	上側確率
0.002	29.8	30	0.0343	
0.002	30.2	30	0.0343	
0.002	32.8	33	0.0625	
0.002	33.2	33	0.0625	
0.002	36	36	0.0794	
0.002		36 +		0.5398
0.002		36 +		0.5398
0.002		36 +		0.5398
0.002		36 +		0.5398
0.002		36 +		0.5398
		積	3.64E-07	4.58E-02
		積の積	1.67E-08	
		ln(積の積)	-17.91	



$$\mu^{\wedge} = 36.5$$

$$\sigma^{\wedge} = 5$$

$$\text{上側確率: } 1 - \text{Norm.Dist}(x_i, \mu^{\wedge}, \sigma^{\wedge}, \text{true})$$

2017.11.18 高橋行雄

38

スライド 40 に示すように、 $\ln(\text{積の積})$ を最大化するためにソルバーで $\hat{\mu} = 36.5$ と $\hat{\sigma} = 5$ を変化させると、 $\hat{\mu} = 35.835$ と $\hat{\sigma} = 4.172$ となり積の積は 1.90×10^{-8} と 1.67×10^{-8} と大きくなる。

スライド 41 に示すように、平均値を 32.5 と小さいほうに動かすと上側確率が小さくなり、積の積は 1.26×10^{-9} と小さくなる。スライド 42 に示すように、標準偏差を 10 と大きくすると確率密度が全体的に小さくなり、積の積は 1.94×10^{-9} と小さくなる。

Excel のソルバーで積の積の対数を最大化した結果に対して、平均値と標準偏差を変化させた場合をスライド 44 に示す。平均が 35.854、標準偏差が 4.172 の場合に $\ln(\text{積} \times \text{積}) = -17.78$ であるが、表の中央に示すように標準偏差を 5 とすると $\ln(\text{積} \times \text{積}) = -17.92$ と小さくなる。表の左に示すように平均を 36.5 と大きくすると $\ln(\text{積} \times \text{積}) = -17.89$ と小さくなることがわかる。

スライド 44

ln(積の積)が最大であることの確認

$\hat{\mu} =$	35.835		35.835		36.500	
$\hat{\sigma} =$	4.172		5.000		4.172	
発毛日	確率密度	上側確率	確率密度	上側確率	確率密度	上側確率
30	0.0360		0.0404		0.0284	
30	0.0360		0.0404		0.0284	
33	0.0759		0.0679		0.0673	
33	0.0759		0.0679		0.0673	
36	0.0956		0.0797		0.0949	
36 +		0.4842		0.4868		0.5477
36 +		0.4842		0.4868		0.5477
36 +		0.4842		0.4868		0.5477
36 +		0.4842		0.4868		0.5477
36 +		0.4842		0.4868		0.5477
積	7.12E-07	2.66E-02	6.00E-07	2.73E-02	3.47E-07	4.93E-02
積の積	1.90E-08		1.64E-08		1.71E-08	
ln(積の積)	-17.78		-17.92		-17.89	

最大値

2017.11.18 高橋行雄

44

スライド 45 に示すように、中濃度の場合は、打ち切りが 1 匹であり、積の積の対数を最大化するのは、 $\hat{\mu} = 26.202$ と $\hat{\sigma} = 6.761$ である。

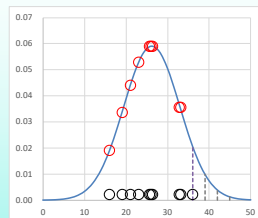
スライド 45

中用量群：打ち切りが1匹

	$\hat{\mu} =$	26.202	
	$\hat{\sigma} =$	6.761	
水平v	拡散x	発毛日	確率密度 上側確率
0.002	16	16	0.0189
0.002	19	19	0.0335
0.002	21	21	0.0439
0.002	23	23	0.0527
0.002	25.6	26	0.0590
0.002	26	26	0.0590
0.002	26.4	26	0.0590
0.002	32.8	33	0.0356
0.002	33.2	33	0.0356
0.002	36	36 +	0.0736
積		3.81E-13	7.36E-02
積の積		2.80E-14	
ln(積の積)		-31.21	
平均		25.900	
S.標準偏差		6.506 =STDEV. S(B5:B14)	
P.標準偏差		6.172 =STDEV. P(B5:B14)	

$\hat{\mu} = 26.202$

$\hat{\sigma} = 6.761$



$\hat{\mu} = 26.202$

$\hat{\sigma} = 6.761$

算術平均は、25.5 と過少評価となる

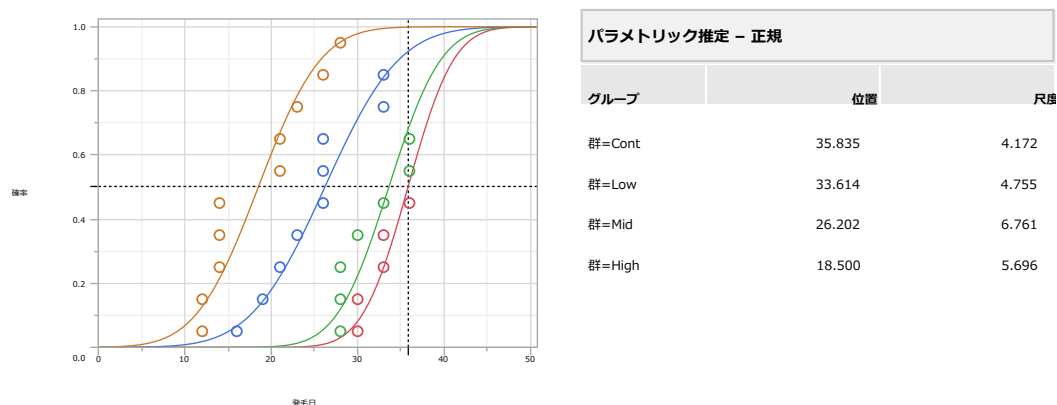
2017.11.18 高橋行雄

45

JMP の「一変量の分布」の場合は、群ごとに計算するのが基本であるが、群を指定する機能もあり、一括して平均値（位置）と尺度（標準偏差）を推定することができる。Cont 群の平均（位置）は、35.835 日、標準偏差（尺度）は、4.172 日であり、左の散布図上に各群の正規分布の累積分布関数が重ね書きされていて、右端が Cont 群である。

スライド 45 で示した Excel で計算した中用量 Mid 群の平均 26.202 日、標準偏差 6.761 日についても再現されている。

図 4 JMP の寿命の一変量による累積正規分布のあてはめと統計量



7. 打ち切りを含む回帰分析

対照群を 1, 低濃度群を 2, 中濃度群を 3, 高濃度群を 4 として, 打ち切りをデータ含む回帰分析に拡張してみよう. 打ち切りがなければ JMP の「二変量の関係」を使うことにより容易に計算できる. 芳賀のグリーン本では, Linest 関数を用いた回帰分析を標準としているが, Excel のソルバーを用いてもスライド 47 に示す手順によって最小 2 乗法による回帰パラメータを推定することができる.

スライド 46

スライド 47

7. 打ち切りを含む回帰分析

2017.11.18 高橋行雄

46

最小2乗法による回帰分析

- ◆ 切片と傾きを (β_0, β_1)
- ◆ データを $(x_1, y_1), \dots, (x_n, y_n)$ とする.
- ◆ $y_i^{\wedge} = \beta_0^{\wedge} + \beta_1^{\wedge} x_i$
- ◆ 偏差平方和: $SS = \sum (y_i - y_i^{\wedge})^2$ を最小化するような $(\beta_0^{\wedge}, \beta_1^{\wedge})$ を求める.
- ◆ なお, Excel の Linest 関数でも簡単に計算できる.

2017.11.18 高橋行雄

47

スライド 48 に打ち切りデータを全て 36 日とした場合の JMP の「二変量の関係」を用いた結果を示す.

スライド 48

JMPの二変量の関係

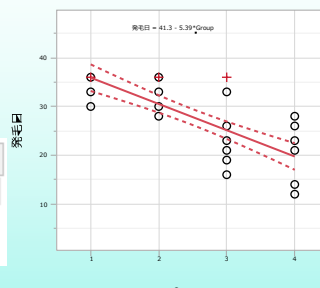
分散分析

要因	自由度	平方和	平均平方	F値
モデル	1	1452.6050	1452.61	55.8033
誤差	38	989.1700	26.03	
全体(修正済み)	39	2441.7750		

p値(Prob>F) <.0001*

パラメータ推定値

項	推定値	標準誤差	t値	p値(Prob> t)
切片	41.3	1.976011	20.90	<.0001*
Group	-5.39	0.721537	-7.47	<.0001*



- ◆ 打ち切りを無視し, そのままの発現日をデータとして使用.
- ◆ 回帰直線の勾配は, 平坦化する.

2017.11.18 高橋行雄

48

得られた回帰直線は、 $y = 41.3 - 5.39x$ であり、直線の勾配の t 統計量は、 $t_{slope} = -7.47$ と高度に有意な結果となっているが、実験を継続して全ての発現日が得られたとした場合に比べ、勾配は平坦化の方向へのバイアスとなる。

群ごとの平均と標準偏差の推定に、正規分布の確率密度の積を最大化する方法を示してきた。打ち切りデータに対しては、正規分布の上側確率とし、確率密度の積と上側確率の積の積を最大化してきた。回帰分析の場合も同様に、推定値 \hat{y}_i を平均とみなした正規分布を考える。

スライド 49

打ち切りデータを含む回帰分析

- ◆ 群ごとの平均と標準偏差の推定に、正規分布の確率密度の積を最大化する方法を示してきた。
- ◆ 打ち切りデータに対しては、正規分布の上側確率とし、確率密度の積と上側確率の積の積を最大化してきた。
- ◆ 回帰分析の場合も同様に、推定値 \hat{y}_i を平均とみなした正規分布を考える。

2017.11.18 高橋行雄

49

スライド 50 に一般的な回帰分析を Excel のソルバーを使うことを前提にした回帰パラメータの推定方法を示した。打ち切りがある場合には、打ち切られた発毛日 y_i について回帰パ

スライド 50

Excel での対応 1

- ◆ データを $(x_1, y_1, \delta_1), \dots, (x_n, y_n, \delta_n)$ とする。
 - 観測データあり: $\delta_i = 0$
 - 打ち切るデータ: $\delta_i = 1$
- ◆ 回帰係数の初期値 $\beta_0^{\wedge}, \beta_1^{\wedge}, \sigma^{\wedge}$ に適当な値を設定する。
- ◆ 回帰の推定値の計算: $\hat{y}_i = \beta_0^{\wedge} + \beta_1^{\wedge} x_i$
- ◆ \hat{y}_i を平均値とみなして y_i に対する確率密度を計算する。

2017.11.18 高橋行雄

50

スライド 51

Excel での対応 2

- ◆ δ_i の値により正規分布の確率密度か上側確率を選択し計算する。
 - $\delta_i = 0$: Norm.dist($y_i, \hat{y}_i, \sigma^{\wedge}, \text{false}$)
 - $\delta_i = 1$: 1 - Norm.dist($y_i, \hat{y}_i, \sigma^{\wedge}, \text{true}$)
- ◆ これらの対数を取り、全ての i について加える。
 - 群平均の場合は、全ての積を計算し、その対数を計算したが、それぞれについて対数を取り、加えたほうが、計算精度の面から望ましい。

2017.11.18 高橋行雄

51

未知のパラメータとして、切片 $\hat{\beta}_0$ 、傾き $\hat{\beta}_1$ 、標準偏差 $\hat{\sigma}$ を動かして、(上側確率の積) \times (確率密度の積) を最大化させる。一群の場合は、平均 $\hat{\mu}$ と標準偏差 $\hat{\sigma}$ の 2 つのパラメータであったが、回帰分析の場合は、平均 $\hat{\mu}$ の代わりに $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ を使う。

$$1 - F(36; \hat{y}_i, \hat{\sigma}) = 1 - F(36; 33.334, 5.889) = 0.654$$
$$1 - F(36; \hat{y}_i, \hat{\sigma}) = 1 - F(36; 25.895, 5.889) = 0.040$$

打ち切りでないデータの場合は、確率密度であり、対照群の 30 は、

$$f(30; \hat{y}_i, \hat{\sigma}) = f(30; 25.895, 5.889) = 0.025$$

スライド 52

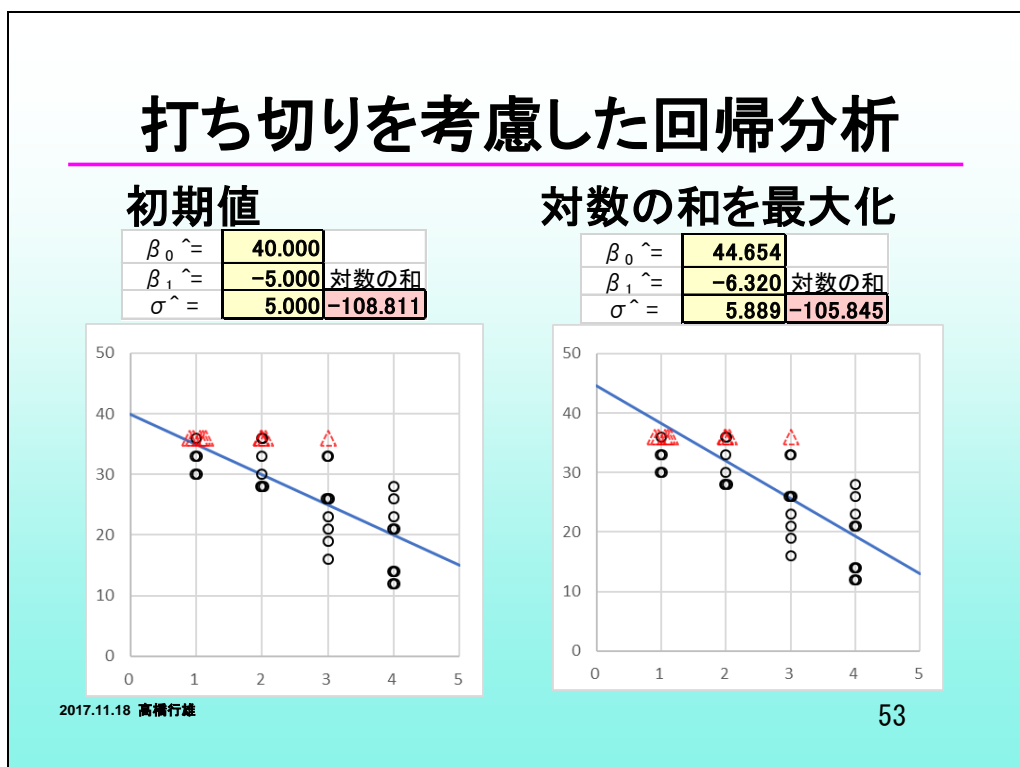
						$\beta_0 \hat{=}$	44.654						
						$\beta_1 \hat{=}$	-6.320						対数の和
						$\sigma^2 \hat{=}$	5.889						-105.845
拡散x群	No	Group	発毛日	打ち切り	y [^]	確率密度	上側確率		対数				
0.90 Cont	2	1	36	1	38.334		0.654		-0.425				
0.95 Cont	4	1	36	1	38.334		0.654		-0.425				
1.05 Cont	7	1	36	1	38.334		0.654		-0.425				
1.10 Cont	8	1	36	1	38.334		0.654		-0.425				
1.15 Cont	9	1	36	1	38.334		0.654		-0.425				
1.98 Low	4	2	36	1	32.015		0.249		-1.389				
2.00 Low	6	2	36	1	32.015		0.249		-1.389				
2.05 Low	7	2	36	1	32.015		0.249		-1.389				
3.00 Mid	5	3	36	1	25.695		0.040		-3.217				
0.99 Cont	5	1	30	0	38.334	0.025			-3.693				
1.01 Cont	6	1	30	0	38.334	0.025			-3.693				
0.99 Cont	3	1	33	0	38.334	0.045			-3.102				
4.00 High	9	4	26	0	19.375	0.036			-3.325				
4.00 High	2	4	28	0	19.375	0.023			-3.764				

2017.11.18 高橋行雄

52

スライド 53 に示すようにパラメータの初期値として、切片 $\hat{\beta}_0 = 40$ ，傾き $\hat{\beta}_1 = -5$ ，標準偏差 $\hat{\sigma} = 5$ とした場合の対数の和は $\ln L = -108.811$ と計算されている．対数の和を最大にするようにソルバーで切片 $\hat{\beta}_0$ ，傾き $\hat{\beta}_1$ ，標準偏差 $\hat{\sigma}$ を動かすと対数の和は $\ln L = -105.849$ と大きくなる．Excel の結果が正しいか JMP を使って検証する．

スライド 53

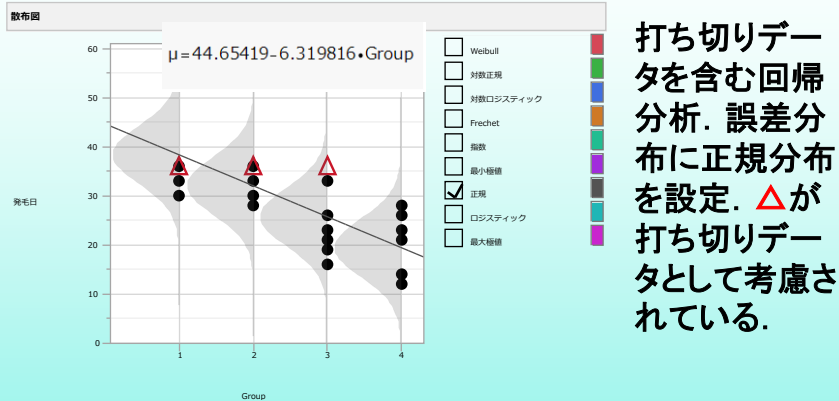


スライド 54 に JMP の寿命の二変量を使った結果を示す．使い方の詳細は，39 ページ以後で示すので，ここでは結果のみを用いる．散布図上の \triangle 印が打ち切りデータであり， \bullet 印が発毛日の観察がされている場合である．求められた傾きは -6.32 と打ち切りを無視したスライド 48 での -5.38 よりも急勾配となっている．

スライド 48：打ち切りを無視 $\hat{y}_i = 41.3 - 5.38x$

スライド 53：打ち切りを考慮 $\hat{y}_i = 44.7 - 6.32x$

JMP「寿命の二変量」散布図



- ◆ 打ち切りを考慮した回帰分析.
- ◆ 直線の勾配は, -6.32 傾きが急になる.

2017.11.18 高橋行雄

54

スライド 55 に回帰パラメータの推定値を示す. 切片 $\hat{\beta}_0 = 44.654$, 傾き $\hat{\beta}_1 = -6.320$, 標準偏差 $\hat{\sigma} = 5.889$ となり, スライド 52 の Excel のソルバーの結果と一致している.

JMP「寿命の二変量」推定値

推定値				
パラメータ	推定値	標準誤差	下側95%	上側95%
β_0	44.65419	2.5439852	39.66807	49.64031
β_1	-6.31982	0.8945875	-8.07318	-4.56646
σ	5.88952	0.7705576	4.37925	7.39978

$$\mu = 44.65419 - 6.319816 \cdot \text{Group}$$

- ◆ 通常の回帰分析の結果につきものの t 値と P 値が出力されない.
- ◆ 標準偏差 σ も同時に推定されている.

2017.11.18 高橋行雄

55

通常の回帰分析の結果に付きものの t 値と P 値は、母集団データと見なしているので出力されない。その代わりに、95%信頼区間が出力されている。この計算のために信頼限界値は、標準正規分布の両側 5%点である 1.96 が用いられている。

「寿命の二変量」では、代わりの統計量として対数尤度を用いた尤度比検定が用いられている。これまで、(確率密度の積) × (上側確率の積) の対数、または、それぞれの対数の和などと述べていたのが、統計用語では対数尤度と称されている。スライド 56 には、(-2) 対数尤度は、パラメータを変化させ最大化した対数尤度の負を 2 倍したものである。

スライド 53 の対数の和を最大化した場合に -105.845 が得られているので (-2) 倍すると 211.690 となる。スライド 56 では、「モデル」の欄の「回帰」の行にある 211.69 に等しい。「効果なし」は、群番号を無視した場合、すなわち、パラメータとして総平均と標準偏差の 2 つを用いた場合の (-2) 倍の対数尤度 = 245.7339 である。

スライド 56

代わりの統計量: (-2) 対数尤度

モデル					
診断統計量	モデル	(-2)*対数尤度	AICc	BIC	パラメータ数
<input type="checkbox"/>	効果なし	245.7339	250.0583	253.1117	2
	回帰	211.69	218.3567	222.7567	3
<input checked="" type="checkbox"/>	別々の位置	209.7529	221.5176	228.1973	5
<input checked="" type="checkbox"/>	別々の位置と尺度	208.1914	228.8366	237.7024	8

- ◆ これまでのExcel の計算シートでは、 $\ln(\text{積の積})$ と表記していたが、対数尤度が一般的な用語である。
- ◆ 「効果なし」は、全体平均のあてはめ
- ◆ 「別々の位置」は、群ごとの平均値を用いた場合
- ◆ 「別々の位置と尺度」は、群毎の平均と標準偏差を用いた場合

スライド 57 の「モデル比較の検定」の説明の欄の「効果なし vs. 回帰」の「尤度比カイ 2 乗」欄 34.0439 は、効果なしの(-2)倍の対数尤度から回帰の(-2)倍の対数尤度を引いた結果である。

$$\text{尤度比カイ 2 乗} = 245.7339 - 211.69 = 34.044$$

パラメータ数についても、効果なしの 2 パラメータから回帰のパラメータ数 3 と 1 増えている。この増分が自由度となり、(-2)倍の対数尤度の差が自由度 1 のカイ 2 乗分布に従うことを利用して検定が行える。「効果なし」は、全ての群をプールして総平均と標準偏差を推定した場合であり、「回帰」は、回帰直線上の推定値を平均とし、各群に共通の標準偏差を推定した場合である。したがって、「効果なし vs. 回帰」は、実験データに対して回帰直線のあてはめに対し、統計的に意味があるかの検定統計量となっている。

スライド 57

尤度比検定

モデル比較の検定			
説明	尤度比カイ2乗	自由度	p値(Prob>ChiSq)
効果なし vs. 回帰	34.04391	1	<.0001*
回帰 vs. 別々の位置	1.937138	2	0.3796
別々の位置 vs. 別々の位置と尺度	1.561486	3	0.6682

◆「効果なし vs. 回帰」は、それぞれの (-2)対数尤度の差であり、これが、カイ2乗分布に従うことから検定を行っている。

モデル	(-2)*対数	パラメータ数	差	自由度差	カイ2乗 P値
効果なし	245.734	2	34.044	1	0.0000
回帰	211.690	3	基準	基準	-
別々の位置	209.753	5	1.937	2	0.3796
別々の位置と尺度	208.191	8	1.562	3	0.6682

2017.11.18 高橋行雄 57

スライド 58 に示すように、「別々の位置」は、回帰直線のあてはめではなく、各群の平均を通る折れ線のあてはめで、共通の標準偏差を推定している。したがって、「回帰 vs. 別々の位置」は、回帰直線のあてはめに対して直線以外の別の曲線のあてはめの可能性についての検定統計量となっている。

(-2)倍の対数尤度の差 1.937 が、自由度 2 のカイ 2 乗分布に従うことから、統計的には有意ではなく、直線以外の曲線のあてはめの可能性がないことが示されている。

「別々の位置と尺度」は、各群の平均と標準偏差を別々に推定した場合であり、「別々の位置 vs. 別々の位置と尺度」は、各群の標準偏差が異なるかの検定統計量となっている。
 (-2)倍の対数尤度の差 1.561 が、自由度 3 のカイ 2 乗分布に従うことから、統計的には有意ではなく、各群に共通な標準偏差をあてはめることが支持されている。

スライド 58

尤度比検定の使い方

モデル比較の検定				
	説明	尤度比カイ2乗	自由度	p値(Prob>ChiSq)
(1)	効果なし vs. 回帰	34.04391	1	<.0001*
(2)	回帰 vs. 別々の位置	1.937138	2	0.3796
(3)	別々の位置 vs. 別々の位置と尺度	1.561486	3	0.6682

- (1) **p<0.0001** なので、傾きが有意と判定
- (2) 別々の群平均(σ は共通)を考える必要はない.
- (3) 群ごとの σ を考える必要はない.

8. JMP の出力を Excel で活用

現実の問題に対して統計ソフトを活用することは当然のことではある。しかし、統計ソフトの結果だけを使えば、どのような考え方で計算されているかまでは理解する必要がないのではないかと、理解しようとしても関連する成書でも理論の説明はあっても数値計算の事例は、統計ソフトの結果が示されるだけで、理論と実際の上に大きな断絶が存在し、統計ソフトの使い方さえわかれば良いのではないのだろうか。このような考え方は、多くの実務家の共通の思いかもしれない。そのために、だれもが使えるオープンな Excel を使い、新しい統計的方法の導入を試み、統計ソフトを適切に活用できるような配慮をしてきた。

統計ソフトは、多くの要望に応えるように日々進歩しているが、対応できていない課題もある。そのような課題に対して、すべて Excel で解決しようとするのではなく、統計ソフトの一部の計算結果を活用することによって、課題を解決できる。一例として、回帰直線の 95% 信頼区間の問題を取り上げる。

回帰直線を引いただけでは、結果の表示として不十分である。「二変量の関係」で、回帰直線の 95% 信頼区間の場合と同様に、打ち切りがある場合の回帰直線の 95% 信頼区間を表示したいが、JMP「寿命の二変量」では、回帰直線の 95% 信頼区間の出力は現在のない。そこで、JMP の分散共分散行列を Excel に取り込み回帰直線の 95% 信頼区間を計算してみよう。

スライド 59

8. JMPの出力を Excel で活用

2017.11.18 高橋行雄

59

スライド 60

回帰直線の95%信頼区間

- ◆ 回帰直線を引いただけでは、不十分である。
- ◆ 打ち切りがある場合の回帰直線の95%信頼区間を表示したい。
- ◆ JMP「寿命の二変量」では、回帰直線の95%信頼区間の出力はない。
- ◆ JMPの分散共分散行列を Excel に取り込み回帰直線の95%信頼区間を計算し、表示する。

2017.11.18 高橋行雄

60

回帰直線の 95% 信頼区間を描くためには、回帰パラメータの分散と共分散を必要とする。JMP の出力に「共分散行列」がある。なお、Excel のソルバーを使っても「共分散行列」を計算することは、尤度関数について 2 階の偏微分式の計算が必要となり、やってできなくはな

いが推奨できない．統計ソフトの結果の活用すべきである．どのくらい大変なのかは，第 13 節の表 13，および 表 15 を参照することで理解できよう．

スライド 61 に示すように，「共分散行列」から回帰パラメータの分散と共分散を抜き出す．回帰直線の推定値 \hat{y} の分散 $Var(\hat{y})$ の計算式をスライド 62 に示す．

スライド 61

分散共分散行列

パラメータ	β_0	β_1	σ
β_0	6.4719	-2.1001	0.3662
β_1	-2.1001	0.8003	-0.0986
σ	0.3662	-0.0986	0.5938

- ◆ 対角要素が分散
 - $Var(\beta_0) = 6.4719$, $Var(\beta_1) = 0.8003$
- ◆ 共分散
 - $Cov(\beta_0, \beta_1) = -2.1001$

61

スライド 62

\hat{y} の分散

$$Var(\hat{y}) = Var(\beta_0 + \beta_1 x)$$

$$= Var(\beta_0) + 2Cov(\beta_0, \beta_1)x + Var(\beta_1)x^2$$

$x = [1, x]$ $\Sigma =$

パラメータ	β_0	β_1
β_0	6.4719	-2.1001
β_1	-2.1001	0.8003

と置いたときに，次のような行列計算でも計算できる．

$$Var(\hat{y}) = x \Sigma x^T$$

62

スライド 63 に x に対する回帰直線の推定値 \hat{y} ，その分散 $Var(\hat{y})$ ，95%信頼区間の計算結果が示されている．スライド 64 には，作図した結果が示されている．

スライド 63

$Var(\hat{y})$ および 95% 信頼区間

パラメータ	β_0	β_1	σ
β_0	6.471861	-2.10006	0.366203
β_1	-2.10006	0.800287	-0.09864
σ	0.366202	-0.09864	0.593759

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

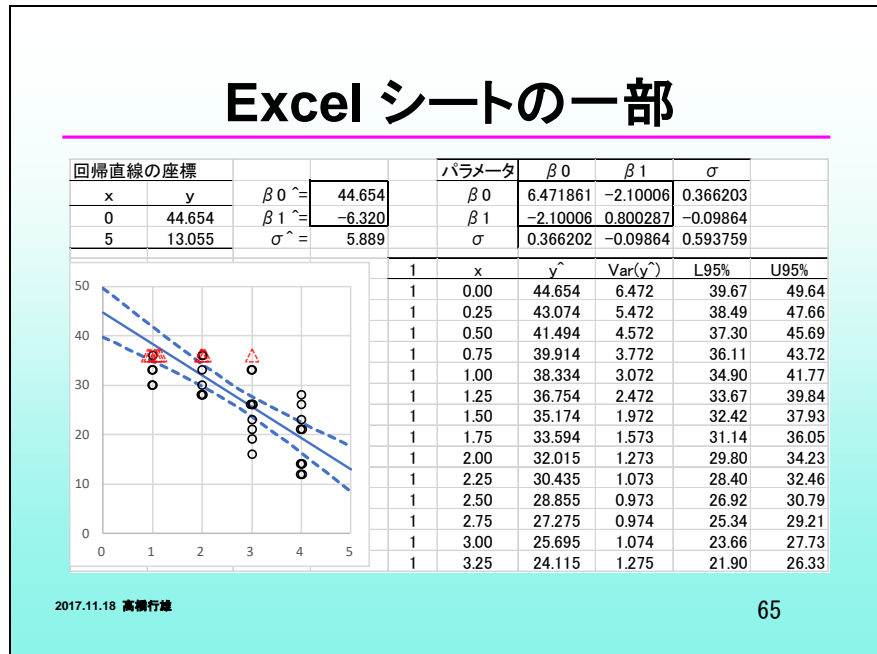
	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

	β_0	β_1	σ
$\hat{\beta}_0$	44.654		
$\hat{\beta}_1$	-6.320		
$\hat{\sigma}$	5.889		

スライド 65 には、Excel シート上ででの 95%信頼区間の計算結果と、それを利用した作図の結果が示されている。なお、この結果は、スライド 55 で示した結果に一致する。

スライド 65 (スライド 63+64)



誤差が正規分布に従うことを積極的に活用することにより、打ち切りデータがある場合でも、回帰分析が行えることを示した。JMP の「寿命の二変量」は、誤差分布を与えることにより、打ち切りデータを考慮した各種の回帰分析が行える。Excel を用いて、打ち切りデータを含む回帰分析の実際を示した。スライド 67 は予告であり、第 9 節で、詳細に説明する。

スライド 66

まとめ

- ◆ 誤差が正規分布に従うことを積極的に活用することにより、打ち切りデータがある場合でも、回帰分析が行えることを示した。
- ◆ JMP の「寿命の二変量」は、誤差分布を与えることにより、打ち切りデータを考慮した各種の回帰分析が行える。
- ◆ Excel を用いて、打ち切りデータを含む回帰分析の実際を示した。

2017.11.18 高橋行雄

66

スライド 67

JMP の「寿命の二変量」

- ◆ 使い方は、かなり高級である。
- ◆ 実際の使い方については、定例会にて示す。

最初の画面に変数名を入れ、このまま実行すると、とんでもない結果となる。

多くの選択肢の中から適切なものを選ばなければならない。

2017.11.18 高橋行雄

67

9. JMP の寿命の二変量の活用

JMP の「寿命の二変量」は、打ち切りデータ含む場合の回帰分析のためのプラットフォームである。

スライド 1

打ち切りデータを含む場合の
新しい回帰分析の考え方

BioStat研究所(株)
高橋 行雄

2017.11.18 高橋行雄 1

スライド 2

JMPの「寿命の二変量」

◆ 昔のJMP

- 生存時間分析
 - 生存時間(パラメトリック)のあてはめ
 - 比例ハザードのあてはめ

◆ 現在のJMP
新たに追加

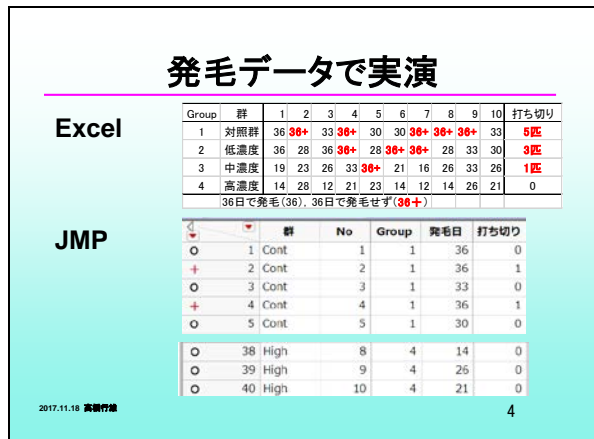
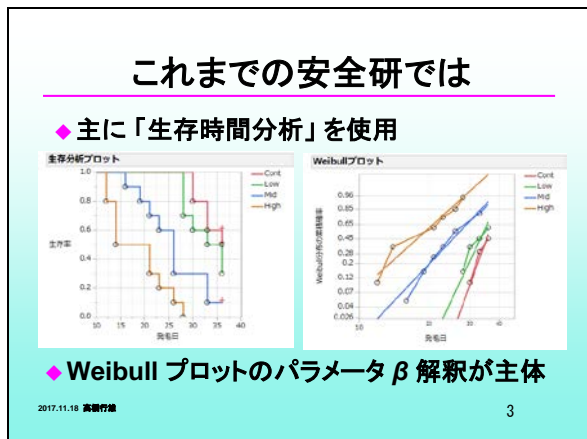
- 寿命の一変量
- 寿命の二変量
- 累積損傷
- 再生モデルによる分析
- 劣化分析
- 破壊劣化
- 信頼性予測
- 信頼性成長

2017.11.18 高橋行雄 2

従前からのノンパラメトリックな「生存時間解析」, 「生存時間(パラメトリック)のあてはめ」, 「比例ハザードのあてはめ」に対し, 新たに追加されたプラットフォームに「寿命の一変量」, 「寿命の二変量」などが追加された。従前からの「一変量の分布」, 「二変量の関係」は, 伝統的な最小2乗法を基盤としているが, 「寿命の一変量」, 「寿命の二変量」は, 最小2乗法ではなく, 最尤法を基盤としている。なお, 最尤法については, Excel を用いて分布関数の確率密度と上側確率などを使う方法として, これまで詳細に示してきた。

打ち切りがあるデータは, 最小2乗法では扱うことができないが, 最尤法では取り扱えることが特長である。さらに最小2乗法と異なるのは, 平均だけでなく標準偏差も併せて推定の対象としている。このことにより, 比較したい群間で分散が異なる場合, 共通とする場合などを設定することが可能であり, これまでとは異なる新しい統計解析が可能となる。

これまで第2期医薬安全性研究会で打ち切りがある場合の統計解析として, スライド3に示すように, カプラン・マイヤー曲線, ワイブルプロット止まりであり, 回帰分析の適用などは手が届かなかった。



打ち切りデータがある場合の統計解析は、「」がメジャーで多くの成書が出版されているが、スライド 5 で示した成書には、正規分布を前提にした回帰分析について触れていない。打ち切りデータを扱う分野とし、各種の工業製品の寿命試験データの解析があり、ラットを使った発毛試験も同様の考え方が適用できる。

生存時間解析の書物

- ◆ 大橋靖男・浜田知久馬(1995)
 - 生存時間解析・SASによる生物統計
- ◆ 大橋靖男・浜田知久馬・魚住龍史(2016)
 - 生存時間解析[応用編]・SASによる生物統計
- ◆ 五所正彦監訳(2014)
 - 生存時間解析入門

注)ワイブル回帰について限定的な解説

- Cox回帰(比例ハザードモデル)が主体

2017.11.18 高橋行雄 5

「寿命の二変量」の特徴的な使い方について概観してみる。多くの統計解析では、データの分布が正規分布に従うと仮定するとの前置きがされるが、仮定するだけで正規分布に従わない場合でも頓着なく統計解析が行われている。「寿命の二変量」では、正規分布に限らず、用いる分布を明示することが必須である。

スライド 6 に示したように、打ち切りデータを含む統計解析の理解の第一歩は、正規分布を前提にした解析である。寿命データの統計解析は、ワイブル分布を前提にしており、このことが、正規分布でも考え方は全く同じであるとの発想を阻害してる。誤差が正規分布に従うことを前提にすれば、

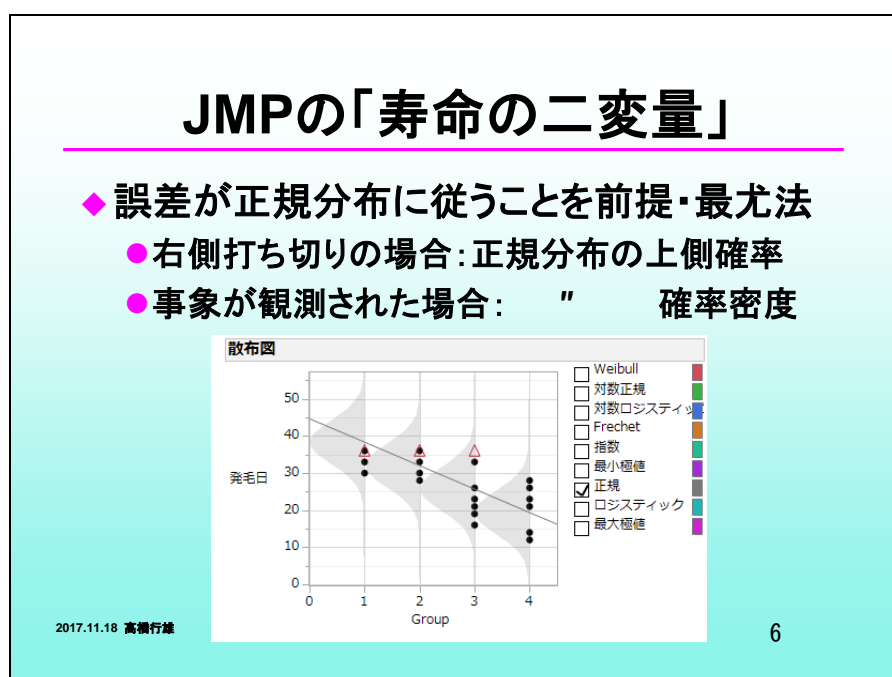
右側打ち切りの場合 : 正規分布の上側確率

事象が観測された場合 : " 確率密度

直前の観察日から現時点 " 区間確率

のように基礎知識の積み上げで記述することができる。

スライド 6



そして、スライド 7 に示したように Excel で視覚化するために、任意の平均と標準偏差に對して確率密度関数と累積分布関数を Excel で描いてみる。右側打ち切りは、その時点以後にどこかで事象が起きると理解する。その場合、打ち切られた時点以後に起きる確率は、正規分布の上側確率となる。事象が観測された場合は、直前の観察日から現時点までの区間の確率と認識することもできる。

スライド 8 に示すように単回帰分析も基礎の基礎であり、Excel のソルバーを用いた解析になれることが世界を広げることができる。最小 2 乗法の原理原則である偏差平方和を Excel のソルバーで実感する。打ち切りデータに対しては、回帰の推定値に対する偏差平方の計算に無理があると認識をする。それに代わる方法として、正規分布を積極的に使用した統計的方法を適用すれば、との発想の転換が救いの神である。

正規分布は、基礎の基礎

- ◆ 任意の平均と標準偏差に対して確率密度関数と累積分布関数を Excel で描く。
- ◆ 右側打ち切りは、その時点以後にどこかで事象が起きると理解する。
 - その場合、打ち切られた時点以後に起きる確率は、正規分布の上側確率となる。
- ◆ 事象が観測された場合は、直前の観察日から現時点までの確率と認識する。

2017.11.18 高橋行雄 7

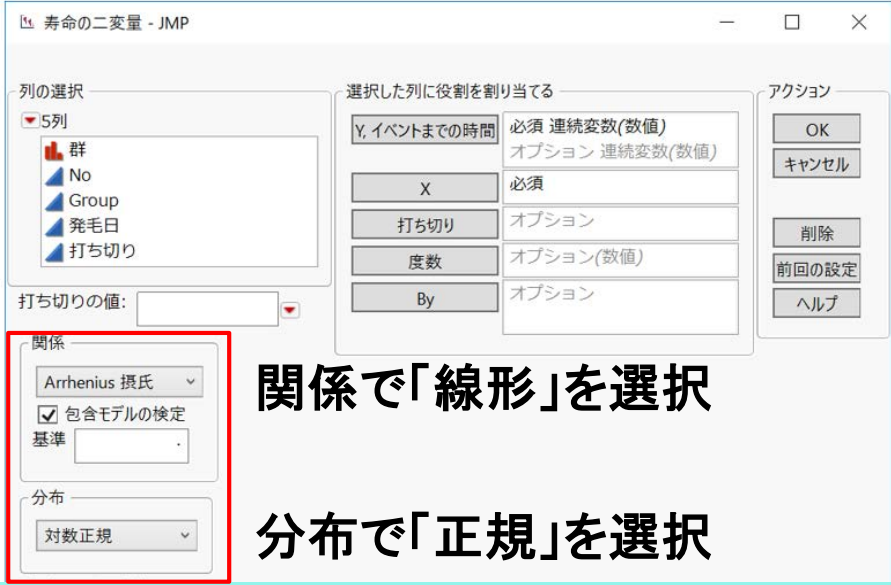
単回帰分析も基礎の基礎

- ◆ 最小2乗法の原理原則である偏差平方和を Excel のソルバーで実感する。
- ◆ 打ち切りデータに対しては、回帰の推定値に対して偏差平方が定義できないと認識をする。
- ◆ それに代わる方法として、正規分布を積極的に使用した統計的方法を適用すれば、との発想の転換をする。

2017.11.18 高橋行雄 8

スライド9に「寿命の二変量」の設定パネルを示す。標準で設定されるのは、「関係」では、「Arrhenius 摂氏」という寿命データの統計解析ではお馴染みの温度に関する変換式であるので、プルダウンメニューから「線形」を選ぶ。さらに、「分布」が「対数正規」となっているので「正規」を選択する。

寿命の二変量

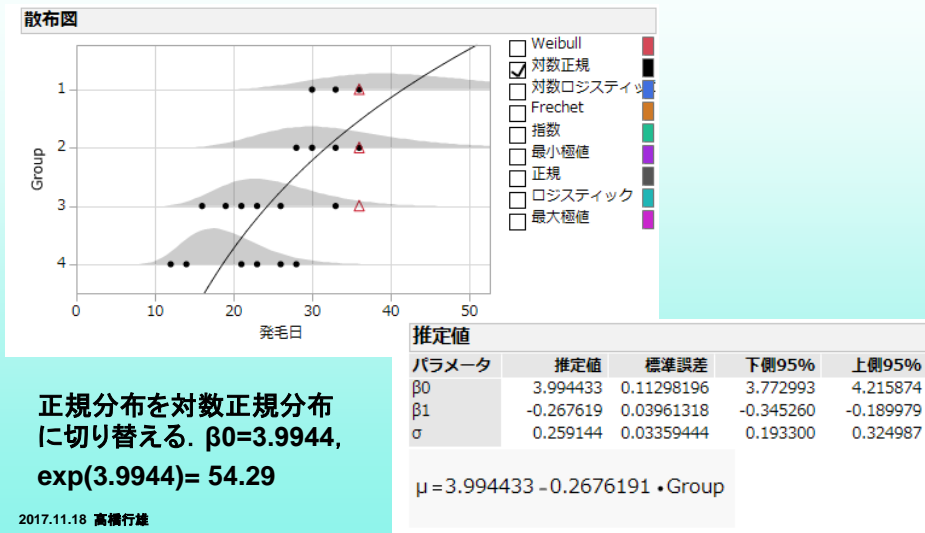


関係で「線形」を選択

分布で「正規」を選択

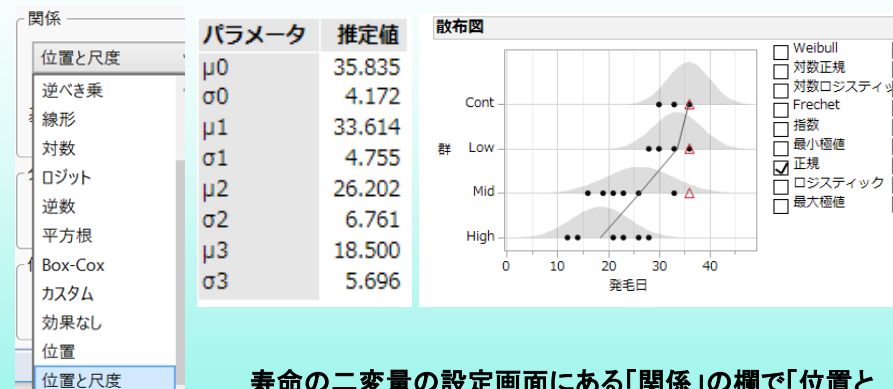
2017.11.18 高橋行雄 9

対数正規分布のあてはめ



スライド 15 は、「関係」で、「位置と尺度」を選んだ場合の結果である。群ごとに打ち切りを含むデータについて平均値と標準偏差が推定されている。

群ごとのあてはめ



スライド 16 に示した「寿命二変量」の背景は、製品寿命が極めて長い製品に対して、製品寿命を推定したいが、通常の使用環境下での経過観察は極めて困難である。過酷な温度・湿度などを設定した加速試験を行い、通常の使用環境下での寿命の推定方式が研究され、その結果として JMP の「寿命の二変量」が登場した。これについては、高橋（2015a）、続高橋セミナー第 4 回：「寿命試験データの統計解析」が詳しい。

<http://www.yukms.com/biostat/takahasi2/rec/004.htm>

スライド 16

「寿命二変量」の背景

- ◆ 製品寿命が極めて長い製品に対して、製品寿命を推定したいが、通常の使用環境下での経過観察は極めて困難である。
- ◆ 過酷な温度・湿度などを設定した加速試験を行い、通常の使用環境下での寿命の推定方式が研究され、その結果として JMP の「寿命の二変量」が登場した。

第4回 続高橋セミナー：寿命試験データの統計解析

<http://www.yukms.com/biostat/takahasi2/rec/004.htm>

2017.11.18 高橋行雄

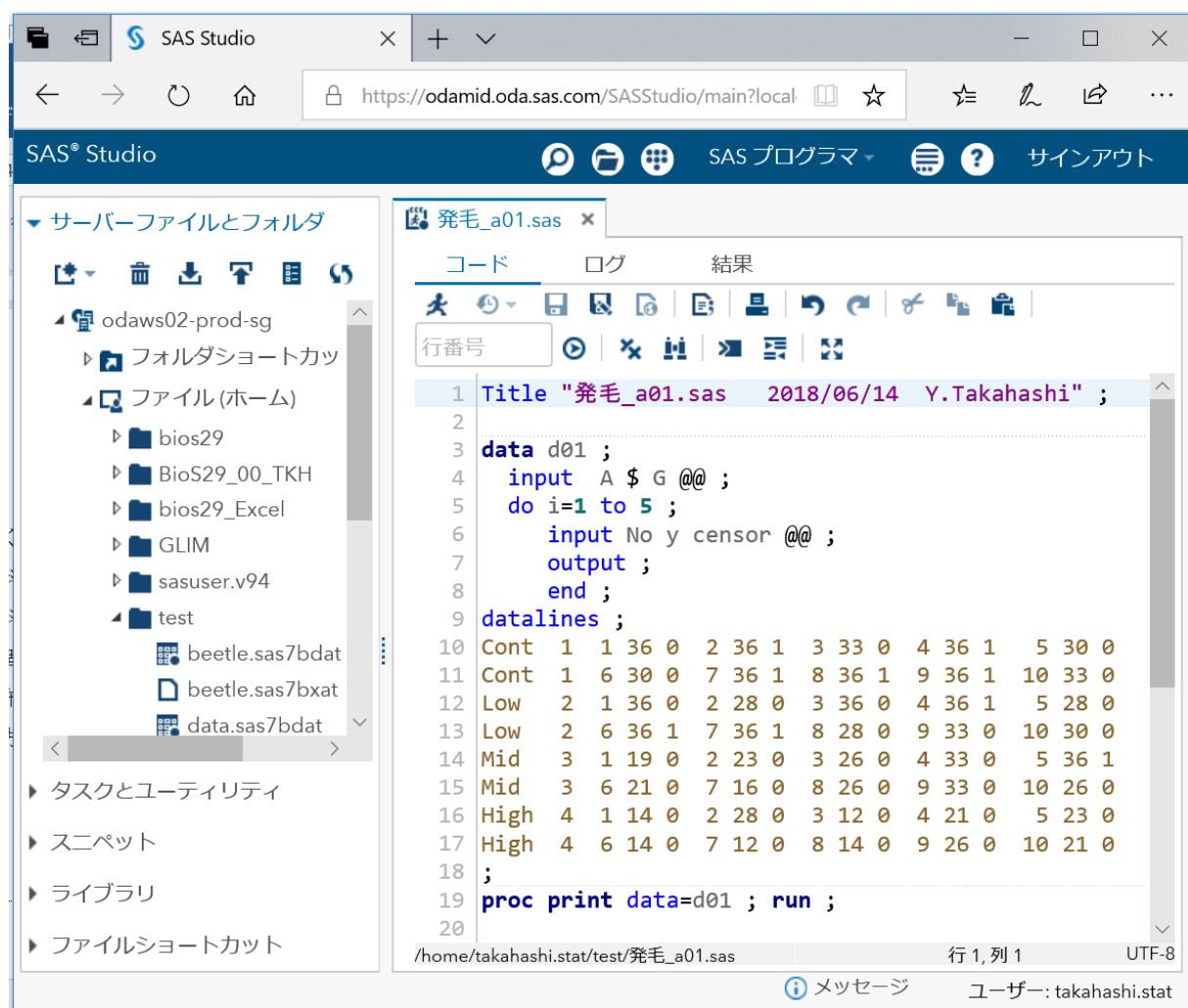
16

寿命データの統計解析については、奥野ら(1988) が詳しい。また、JMP の「寿命の二変量」の説明書、SAS Institute(2012)、「品質管理および信頼性/生存時間」は操作法が主体であるので、理異論的な説明は、Meeker and Escobar(1998)を参照のこと。

10. 無料の On Demand SAS を用いた解析

多くの統計ソフトの中で SAS は、世界中で断トツの評価を得てはいるものの毎年の高額な使用料の負担に耐えられなくなった日本の教育関連施設での利用が壊滅状態となり、代わりに無料で使える多くの統計解析関数を包含する R 言語が広く使われるようになってきた。近年、個人的な研究あるいは学習用の場合に、無料で SAS が使えるようになった。そこで、正規分布を仮定した打ち切りデータの統計解析について例示したい。なお、OnDemand SAS の使い方については、高波洋平ら（2016）を参照されたい。

図 5 OnDemand SAS の画面で発毛データの SAS データセット化のためのコード



OnDemand SAS で、正規分布を仮定した打ち切りデータを含む回帰分析を行うのは、SAS/STAT のパラメトリックな生存時間解析として知られている LIFEREG プロシジャおよび SAS/QC の信頼性データの解析を目的とした RELIABILITY プロシジャである。

パラメトリックな生存時間解析では、ワイブル分布を用いるのが基本であるが、正規分布も使えるようになっている。

LIFEREG プロシジャによる回帰分析の SAS コードは、

```
proc lifereg data=d01 ;
    model y*censor(1) = G / dist=normal ;
run;
```

であり、出力結果は以下に示すように、第 10 章のスライド 11 と同じ結果が得られる。

表 3 打ち切りデータを含む回帰分析

Analysis of Maximum Likelihood Parameter Estimates							
Parameter	自由度	推定値	標準誤差	95% Confidence Limits		カイ 2 乗	Pr > ChiSq
Intercept	1	44.6542	2.5440	39.6681	49.6403	308.10	<.0001
G	1	-6.3198	0.8946	-8.0732	-4.5665	49.91	<.0001
Scale	1	5.8895	0.7706	4.5574	7.6111		

Fit Statistics	
-2 Log Likelihood	211.690
AIC (smaller is better)	217.690
AICC (smaller is better)	218.357
BIC (smaller is better)	222.757

各群毎の打ち切りを考慮した平均と標準偏差を求める SAS コードは、

```
proc lifereg data=d01 ;
    class g ;
    model y*censor(1) = G / dist=normal ;
    by g ;
run;
```

であり、出力結果は以下に示すように、第 10 章のスライド 15 と同じ結果が得られる。

表 4 <G1: 対照群>

Analysis of Maximum Likelihood Parameter Estimates								
Parameter		自由度	推定値	標準誤差	95% Confidence Limits		カイ 2 乗	Pr > ChiSq
Intercept		1	35.8347	1.6228	32.6541	39.0152	487.64	<.0001
G	1	0	0.0000
Scale		1	4.1719	1.4836	2.0780	8.3758		

<G2：低用量群>

Analysis of Maximum Likelihood Parameter Estimates								
Parameter		自由度	推定値	標準誤差	95% Confidence Limits		カイ 2 乗	Pr > ChiSq
Intercept		1	33.6137	1.6052	30.4676	36.7599	438.51	<.0001
G	2	0	0.0000
Scale		1	4.7547	1.3588	2.7157	8.3249		

<G3：中用量群>

Analysis of Maximum Likelihood Parameter Estimates								
Parameter		自由度	推定値	標準誤差	95% Confidence Limits		カイ 2 乗	Pr > ChiSq
Intercept		1	26.2018	2.1575	21.9732	30.4304	147.49	<.0001
G	3	0	0.0000
Scale		1	6.7608	1.6405	4.2020	10.8778		

<G4：高用量群>

Analysis of Maximum Likelihood Parameter Estimates								
Parameter		自由度	推定値	標準誤差	95% Confidence Limits		カイ 2 乗	Pr > ChiSq
Intercept		1	18.5000	1.8014	14.9693	22.0307	105.47	<.0001
G	4	0	0.0000
Scale		1	5.6965	1.2738	3.6751	8.8296		

JMP では、群毎に標準偏差（尺度）を設定できる機能があり、それを用いて群毎の平均と標準偏差を簡便に得ることができた。SAS の LIFEREG プロシジャでは、尺度はモデル全体で一つとしているために、by ステートメントで群毎のモデルとして群毎の標準偏差（尺度）を求めた。

11. 文献データに対する統計ソフトでの追試

岩崎ら(2006)は、正規分布を仮定した場合に右側打ち切りがある場合に、正規分布のパラメータ μ と σ の推定を EM アルゴリズムによって求めている。

標準正規分布 $N(0,1)$ の確率密度関数および累積分布関数をそれぞれ $\phi(z)$ 、 $\Phi(z)$ とすると、 $N(\mu, \sigma^2)$ の確率密度関数は $f(x) = \phi((x - \mu)/\sigma)/\sigma$ 、累積分布関数は $F(x) = \Phi((x - \mu)/\sigma)$ と表わされる。 $N(\mu, \sigma^2)$ からの n 個の無作為標本中の m 個 x_1, \dots, x_m が c 以下で観測され、 $n - m$ 個が c を越えて打ち切りになったとすると、対数尤度関数は

$$\ln L = \sum_{i=1}^m \ln \left[\frac{1}{\sigma} \phi \left(\frac{x_i - \mu}{\sigma} \right) \right] + \sum_{i=m+1}^n \ln \left[1 - \Phi \left(\frac{c - \mu}{\sigma} \right) \right]$$

となる。パラメータ μ 、 σ の最尤推定値は、原理的には $\ln L$ を μ 、および σ で 2 階の偏微分し、ニュートン・ラフソン法で $\ln L$ を最大にするようなパラメータ μ, σ を求めればよい。

中学生の肺活量のデータで EM アルゴリズムを利用した計算結果が、 $\hat{\mu} = 2202.62$ 、 $\hat{\sigma}^2 = 94205.591 = 306.93^2$ と示されている。

表 5 中学生の肺活量データ

ID	1	2	3	4	5	6	7	8	9	10
肺活量	1700	1850	2000	2100	2150	2200	2300	2400	2500	2500
打ち切り									+	+

岩崎(2002)には、同じデータで EM アルゴリズムの具体的な計算方法が述べられ、17 回の反復計算で収束したと述べられている。EM アルゴリズムによる最尤法は、計算機のパワーが少ない時代に考案された計算方法であるが、現代の統計ソフトでは、ほとんどが対数尤度の 2 階の偏微分式を用いたニュートン・ラフソン法が使われている。

肺活量のデータを用いて LIFEREG プロシジャを用いて追試を行う。第 10 節に準じて、10 人分の肺活量に打ち切りコードをセットする。LIFEREG プロシジャでは、model ステートメントで model y*censor(1) により肺活量データに打ち切りデータを組み合わせる。変数リストは、1 群しかないので空白とし「切片」のみのモデルがセットされる。オプションの dist=normal で、正規分布を指定し、計算過程を出力するための itprint も含めておく。

```

===== SAS コード =====
Title "肺活量_a01.sas 2018/06/16 Y.Takahashi" ;

data d01 ;
  do i=1 to 5 ;
    input ID y censor @@ ; output ;
  end ;
datalines ;
1 1700 0 2 1850 0 3 2000 0 4 2100 0 5 2150 0
6 2200 0 7 2300 0 8 2400 0 9 2500 1 10 2500 1
;
proc print data=d01 ; run ;

proc lifereg data=d01 ;
  model y*censor(1) = / dist=normal itprint ;
run;

```

SAS の LIFEREG プロシジャは、対数尤度関数をパラメータで 2 階の偏微分を使うニュートン・ラフソン法が使われている。これは、出力に「Last Evaluation of the Negative of the Hessian」とヘシアン行列（2 階の偏微分行列）が明示されていることで確認される。反復の過程は、「Iteration History for Parameter Estimates」で確認でき、5 回の反復で収束している。Intercept および Scale の収束値が、 $\hat{\mu} = 2202.620$ 、 $\hat{\sigma} = 306.93$ となり、岩崎らの EM アルゴリズムを用いた最尤法に一致している。なお、棚橋ら(2009)も、岩崎らのデータを用いて、LIFEMREG プロシジャによる結果を示している。

===== 表 6 SAS 出力 =====

肺活量_a01.sas 2018/06/16 Y.Takahashi

Obs	i	ID	y	censor
1	1	1	1700	0
2	2	2	1850	0
3	3	3	2000	0
4	4	4	2100	0
5	5	5	2150	0
6	1	6	2200	0
7	2	7	2300	0
8	3	8	2400	0
9	4	9	2500	1
10	5	10	2500	1

表 6 SAS 出力 (続き)

肺活量_a01.sas 2018/06/16 Y.Takahashi

The LIFEREG Procedure

Iteration History for Parameter Estimates				
Iter	Ridge	Loglikelihood	Intercept	Scale
0	0	-59.492036	2170	267.91375063
1	0	-59.311356	2195.234023	295.37683802
2	0	-59.298866	2202.0428503	305.84649371
3	0	-59.298767	2202.6151326	306.91955559
4	0	-59.298767	2202.6196103	306.92928972
5	0	-59.298767	2202.6196103	306.92928972

Last Evaluation of the Negative of the Hessian		
	Intercept	Scale
Intercept	0.0001018459	-0.000015452
Scale	-0.000015452	0.0001548696

Fit Statistics	
-2 Log Likelihood	118.598
AIC (smaller is better)	122.598
AICC (smaller is better)	124.312
BIC (smaller is better)	123.203

Analysis of Maximum Likelihood Parameter Estimates							
Parameter	自由度	推定値	標準誤差	95% Confidence Limits		カイ 2 乗	Pr > ChiSq
Intercept	1	2202.620	99.8483	2006.921	2398.319	486.63	<.0001
Scale	1	306.9293	80.9710	183.0135	514.7468		

これまで示したきた JMP の「寿命の一変量」で簡単に求めることができるが、SAS の LIFEREG プロシジャに対応する「生存時間のあてはめ(パラメトリック)」を使った結果を示す。パラメトリックな生存時間分析では、ワイブル分布を仮定するのが一般的であり、正規分布を仮定した解析もできることは、ほとんど知られていない。

SAS の LIFEREG に対応する JMP の「生存時間のあてはめ(パラメトリック)」による結果を示す。GUI での設定は省略し、JMP のスクリプトを次に示す。

===== JMP のスクリプト =====

```
Fit Model (
  Censor ( :打ち切り ),
  Censor Code ( "1" ),
  Y ( :肺活量 ),
  Effects,
  Personality ( "Parametric Survival" ),
  Distribution ( "Normal" ),
  Run ( Likelihood Ratio Tests ( 1 ) )
);
```

JMP の結果を次に示す. 「寿命の一変量」は, 多彩な出力ができるが, 「生存時間 (パラメトリック) のあてはめ」は, シンプルな出力に限定されている. もちろん, 結果は $\hat{\mu} = 2202.620$, $\hat{\sigma} = 306.93$ と一致している.

===== 表 7 肺活量についての JMP の出力 =====

生存時間のあてはめ(パラメトリック)				
イベントまでの時間: 肺活量	AICc	124.3118	使用されたオブザベーション	10
分布: 正規	BIC	123.2027	非打ち切りの個数	8
打ち切り変数: 打ち切り	(-2)*対数尤度	118.5975	右側打ち切りの個数	2
パラメータ推定値				
項	推定	標準誤	下側95%	上側95%
切片	2202.61961	99.8483	2006.9205	2398.3187
σ	306.92929	80.970961	148.22912	465.62946

信頼区間はWald法による

Excel のソルバーを用いたパラメータの推定は, 第 6 節のスライド 38 以後で詳細に示してきた. 肺活量データについて同様の Excel シートを表 8 に示す.

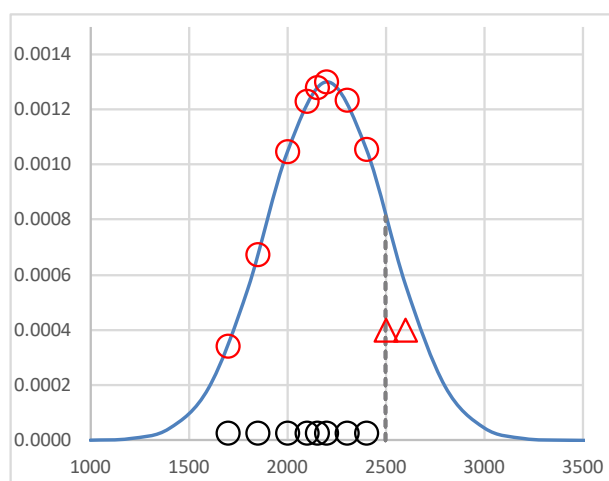
Excel のソルバーで積の積の対数を最大化した結果に対して, 平均値と標準偏差を変化させた場合, 平均が 2203.62, 標準偏差が 306.93 の場合に $\ln(\text{積} \times \text{積}) = -59.30$ であるが, 表の中央に示すように標準偏差を 400.0 とすると $\ln(\text{積} \times \text{積}) = -59.73$ と小さくなる. 表の左に示すように平均を 2000.0 と小さくすると $\ln(\text{積} \times \text{積}) = -61.40$ と小さくなることわかる.

表 8 肺活量データに対する Excel ソルバーを用いた平均と標準偏差の推定

$\mu^{\wedge}=$	2202.62		2202.618		2000.000	
$\sigma^{\wedge}=$	306.93	94204.8	400.000		306.928	
肺活量	確率密度	上側確率	確率密度	上側確率	確率密度	上側確率
1700	0.0003		0.0005		0.0008	
1850	0.0007		0.0007		0.0012	
2000	0.0010		0.0009		0.0013	
2100	0.0012		0.0010		0.0012	
2150	0.0013		0.0010		0.0012	
2200	0.0013		0.0010		0.0011	
2300	0.0012		0.0010		0.0008	
2400	0.0011		0.0009		0.0006	
2500 +		0.1663		0.2286		0.0517
2500 +		0.1663		0.2286		0.0517
積	6.38E-25	2.77E-02	2.19E-25	5.23E-02	8.10E-25	2.67E-03
積の積	1.77E-26		1.14E-26		2.16E-27	
ln(積の積)	-59.30		-59.73		-61.40	

図 6 は、打ち切りデータを含む確率密度、および上側確率を可視化するために作成した。X 軸は、肺活量で、軸の上の○印が、観測されたデータで、平均が 2202.62、標準偏差が 306.93 の正規分布の確率密度曲線上の赤○は、それぞれのデータに対応する確率密度である。X 軸の 2500 mL が打ち切り値で、2つの赤△が、打ち切りデータを示している。それらの上側確率は、2500 mL 以上の面積である。

図 6 確率密度および上側確率の可視化



12. Excel によるニュートン・ラフソン法の適用

第 11 節の肺活量のデータについてソルバーを用いずに、対数尤度の偏微分式を用いたニュートン・ラフソン法による計算を試みる。ニュートン・ラフソン法は、対数尤度関数のパラメータについて 2 階の偏微分を必要とし、反復計算による煩わしさがあるために、ほとんどの成書で計算過程が略され、統計ソフトでの結果を使った記述になっている。

ニュートン・ラフソン法を一枚の Excel シート上で、VBA (Visual Basic for Applications) を使わずに基本の関数のみで実現することは、最尤法を前提とする統計モデルの理論を学習するために欠かせない。

正規分布 $N(\mu, \sigma^2)$ からの n 個の無作為にサンプリングした標本の中で m 個 y_1, \dots, y_m が c 以下で観測され、 $n - m$ 個が c を越えて打ち切りになったとすると、対数尤度関数は

$$\ln L = \sum_{i=1}^m \ln \left[\frac{1}{\sigma} \phi \left(\frac{y_i - \mu}{\sigma} \right) \right] + \sum_{i=m+1}^n \ln \left[1 - \Phi \left(\frac{c - \mu}{\sigma} \right) \right]$$

となる。正規分布の確率密度関数と分布関数の和となっているので、別々にパラメータ μ と σ について偏微分を行う。なお、 c は固定値でなくとも良い。

対数尤度関数を偏微分する際に、一般的には確率密度関数の対数について整理した式を用いて偏微分するのが一般的であるので、参考までに本節の末尾に示す。確率密度関数および分布関数計算が関数として組み込まれている Excel を使う場合には、そのままの関数として偏微分することにより見通しが良くなる。標準正規分布の確率密度関数は、

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{z^2}{2} \right), \quad \text{ただし, } z = \frac{y - \mu}{\sigma}$$

であり、 $\phi(z)$ を、 z で微分すると

$$\frac{\partial \phi(z)}{\partial z} = \frac{-z}{\sqrt{2\pi}} \exp \left(-\frac{z^2}{2} \right) = -z\phi(z)$$

と元の確率密度関数を含む形で簡略化できる。対数を取った場合は、 $\phi(z)$ がキャンセルされ、

$$\frac{\partial \ln \phi(z)}{\partial z} = \frac{-z\phi(z)}{\phi(z)} = -z$$

の様に、さらに簡略化できる。標準正規分布の分布関数

$$\Phi(z) = \int_{-\infty}^z \phi(z) dz$$

の偏微分は、定義に従って

$$\frac{\partial \Phi(z)}{\partial z} = \varphi(z)$$

と確率密度関数 $\varphi(z)$ となり，対数を取った場合には，

$$\frac{\partial \ln \Phi(z)}{\partial z} = \frac{\varphi(z)}{\Phi(z)}$$

となり，上側確率 $1 - \Phi(z)$ の場合は，

$$\frac{\partial \ln[1 - \Phi(z)]}{\partial z} = \frac{-\varphi(z)}{1 - \Phi(z)}$$

となる．これらの関係を用いて， μ および σ について偏微分を行う．

対数を取った確率密度関数 $\ln f(y)$ を偏微分しやすいような式

$$\ln f(y) = \ln \left[\frac{1}{\sigma} \varphi \left(\frac{y - \mu}{\sigma} \right) \right] = \ln \left[\varphi \left(\frac{y}{\sigma} - \frac{\mu}{\sigma} \right) \right] - \ln \sigma$$

にして， μ についての偏微分をすると，

$$\frac{\partial \ln f(y)}{\partial \mu} = \left(\frac{-1}{\sigma} \right) \cdot \left(\frac{-(y - \mu)}{\sigma} \right) = \frac{z}{\sigma}$$

となる． σ についての偏微分は，

$$\frac{\partial \ln f(y)}{\partial \sigma} = \left[\frac{-(y - \mu)}{\sigma^2} \right] \cdot \left[\frac{-(y - \mu)}{\sigma} \right] - \frac{1}{\sigma} = \frac{z^2 - 1}{\sigma}$$

となる．これらを用いて 2 階の微分は，以下の通りである．

$$\frac{\partial^2 \ln f(y)}{\partial \mu \partial \mu} = \frac{-1}{\sigma^2}, \quad \frac{\partial^2 \ln f(y)}{\partial \mu \partial \sigma} = \frac{-2(y - \mu)}{\sigma^3} = \frac{-2z}{\sigma^2}, \quad \frac{\partial^2 \ln f(y)}{\partial \sigma \partial \sigma} = \frac{-3(y - \mu)^2}{\sigma^4} + \frac{1}{\sigma^2} = \frac{1 - 3z^2}{\sigma^2}$$

分布関数の上側確率についての偏微分は，以下の通り．

$$\ln[1 - \Phi(z)] = \ln \left[1 - \Phi \left(\frac{y - \mu}{\sigma} \right) \right]$$

$$\frac{\partial \ln[1 - \Phi(z)]}{\partial \mu} = \frac{1}{\sigma} \cdot \frac{\varphi(z)}{1 - \Phi(z)} = \frac{Q(z)}{\sigma}$$

$$\frac{\partial \ln[1 - \Phi(z)]}{\partial \sigma} = \frac{(y - \mu)}{\sigma^2} \cdot \frac{\varphi(z)}{1 - \Phi(z)} = \frac{z}{\sigma} \cdot \frac{\varphi(z)}{1 - \Phi(z)} = \frac{zQ(z)}{\sigma}$$

$$\text{ただし, } z = \frac{y - \mu}{\sigma}, \quad Q(z) = \frac{\varphi(z)}{1 - \Phi(z)}$$

2 階の微分は，1 階の偏微分式に含まれる $Q(z) = \varphi(z) / (1 - \Phi(z))$ について z での微分を必要となる．微分した結果は，

$$\frac{\partial Q(z)}{\partial z} = \frac{-z\varphi(z)}{1-\Phi(z)} + \left(\frac{\varphi(z)}{1-\Phi(z)} \right)^2 = -zQ(z) + Q(z)^2$$

となる．この結果を使って， μ および σ について 2 階の偏微分を求める．

$$\frac{\partial^2 \ln[1-\Phi(z)]}{\partial \mu \partial \mu} = \frac{(y-\mu)}{\sigma^3} \frac{\varphi(z)}{1-\Phi(z)} - \frac{1}{\sigma^2} \left(\frac{\varphi(z)}{1-\Phi(z)} \right)^2 = \frac{zQ(z) - Q(z)^2}{\sigma^2}$$

$$\frac{\partial^2 [1-\Phi(z)]}{\partial \mu \partial \sigma} = \frac{z^2-1}{\sigma^2} \cdot \frac{\varphi(z)}{1-\Phi(z)} - \frac{z}{\sigma^2} \left(\frac{\varphi(z)}{1-\Phi(z)} \right)^2 = \frac{(z^2-1)Q(z) - zQ(z)^2}{\sigma^2}$$

$$\frac{\partial^2 \ln[1-\Phi(z)]}{\partial \sigma \partial \sigma} = \frac{(z^3-2z)}{\sigma^2} \frac{\varphi(z)}{1-\Phi(z)} - \frac{z^2}{\sigma^2} \left(\frac{\varphi(z)}{1-\Phi(z)} \right)^2 = \frac{(z^3-2z)Q(z) - z^2Q(z)^2}{\sigma^2}$$

これらの 1 階のおよび 2 階の偏微分式を Excel シートに埋め込み，ニュートン・ラフソン法により対数尤度を最大化する μ および σ を推定することができる．打ち切りがあれば 1，なければ 0 となる変数 δ_i とし，Excel の if() 関数を用いてそれぞれの y_i に対し，尤度 L_i を計算し，対数尤度 $\ln L$ を計算する．

$$L_i = \text{if} [\delta_i = 0, (\varphi(z_i)/\hat{\sigma}), (1-\Phi(z_i))],$$

$$\delta_i = 0, \quad \delta_i \neq 0$$

$$\ln L = \sum_i \ln L_i$$

$$\text{ただし, } z_i = (y_i - \hat{\mu}) / \hat{\sigma}$$

それぞれの y_i に対し 1 階の偏微分式を次式で計算し，

$$\partial \mu_i = \text{if} [\delta_i = 0, z_i / \hat{\sigma}, Q(z_i) / \hat{\sigma}],$$

$$\partial \sigma_i = \text{if} [\delta_i = 0, (z_i^2 - 1) / \hat{\sigma}, z_i Q(z_i) / \hat{\sigma}]$$

$$\text{ただし, } Q(z_i) = \frac{\varphi(z_i)}{1-\Phi(z_i)}$$

i について合計し，スコアベクトル U とする．

$$U = \begin{bmatrix} \sum_i \partial \mu_i \\ \sum_i \partial \sigma_i \end{bmatrix}$$

それぞれの x_i に対し 2 階の偏微分式を次式で計算し，

$$\partial \mu_i \partial \mu_i = \text{if} \{ \delta_i = 0, -1 / \hat{\sigma}^2, [z_i Q(z_i) - Q(z_i)^2] / \hat{\sigma}^2 \}$$

$$\partial \mu_i \partial \sigma_i = \text{if} \{ \delta_i = 0, -2z_i / \hat{\sigma}^2, [(z_i^2 - 1)Q(z_i) - z_i Q(z_i)^2] / \hat{\sigma}^2 \}$$

$$\partial \sigma_i \partial \sigma_i = \text{if} \{ \delta_i = 0, (1 - 3z_i^2) / \hat{\sigma}^2, [(z_i^3 - 2z_i)Q(z_i) - z_i^2 Q(z_i)^2] / \hat{\sigma}^2 \}$$

$$\text{ただし, } \partial \sigma_i \partial \mu_i = \partial \mu_i \partial \sigma_i$$

i について合計し，ヘシアン行列 H とする．

$$H = \begin{bmatrix} \sum_i \partial \mu_i \partial \mu_i & \sum_i \partial \sigma_i \partial \mu_i \\ \sum_i \partial \mu_i \partial \sigma_i & \sum_i \partial \sigma_i \partial \sigma_i \end{bmatrix}$$

初期値として岩崎(2002)のEMアルゴリズムでの初期値と同様に、観測された8個のデータの平均=2087.5, 標準偏差=231.07(序数は8)を元のパラメータとし、変化量 $(-H)^{-1}U$ を加える.

$$\begin{array}{c} \text{新たなパラメータ} \end{array} \begin{bmatrix} \hat{\mu}_{new} \\ \hat{\sigma}_{new} \end{bmatrix} = \begin{array}{c} \begin{bmatrix} \hat{\mu}_{old} \\ \hat{\sigma}_{old} \end{bmatrix} \\ \text{元のパラメータ} \end{array} + \begin{array}{c} (-H)^{-1}U \\ \text{変化量} \end{array}$$

表 9 より

変化量	負の逆行列	1階の
$U(-H)^{-1}$	$(-H)^{-1}$	偏微分 U
75.14	5712.31 -1137.05	0.0212
32.43	-1137.05 1397.04	0.0405

Excel による行列計算 = Mmult(Minverse(-H), U)

表 9 より

新たな パラメータ	=	元の パラメータ	+	変化量 $(-H)^{-1}U$
2162.64		2087.50		75.14
248.58		216.15		32.43

新たなパラメータをコピーして、元のパラメータに「値」のみ貼り付け、1回目の反復とする. どのように2回目の反復, 3回目, と変化量が0になるまで繰り返す.

表 9 初期値に対するニュートン・ラフソン法による計算結果

正規分布				元の		変化量	新たな	1階の		2階の偏微分		負の逆行列	
Newton-Raphson法				パラメータ		$(-H)^{-1}U$	パラメータ		偏微分 U		H	$(-H)^{-1}$	
				$\mu^{\wedge} =$	2087.50	75.14	2162.64		0.0212	-2.09E-04	-1.70E-04	5712.04	-1136.96
				$\sigma^{\wedge} =$	216.15	32.43	248.58		0.0405	-1.70E-04	-8.54E-04	-1136.96	1396.90
				$\sigma^{\wedge 2} =$	46719		$\ln L =$	-61.498	$-2 \ln L =$	123.00			
No	y	$+\delta$	z	$f(z)$	$1-\Phi(z)$	$\varphi z / (1-\Phi z)$	Li	$\ln Li$	$\partial \mu$	$\partial \sigma$	$\partial \mu \partial \mu$	$\partial \mu \partial \sigma$	$\partial \sigma \partial \sigma$
1	1700	0	-1.793	0.0004	0.9635	0.0830	0.0004	-7.9019	-8.29E-03	1.02E-02	-2.14E-05	7.67E-05	-1.85E-04
2	1850	0	-1.099	0.0010	0.8641	0.2525	0.0010	-6.8986	-5.08E-03	9.59E-04	-2.14E-05	4.70E-05	-5.61E-05
3	2000	0	-0.405	0.0017	0.6572	0.5593	0.0017	-6.3768	-1.87E-03	-3.87E-03	-2.14E-05	1.73E-05	1.09E-05
4	2100	0	0.058	0.0018	0.4769	0.8351	0.0018	-6.2966	2.68E-04	-4.61E-03	-2.14E-05	-2.48E-06	2.12E-05
5	2150	0	0.289	0.0018	0.3862	0.9906	0.0018	-6.3367	1.34E-03	-4.24E-03	-2.14E-05	-1.24E-05	1.60E-05
6	2200	0	0.520	0.0016	0.3014	1.1561	0.0016	-6.4303	2.41E-03	-3.37E-03	-2.14E-05	-2.23E-05	4.01E-06
7	2300	0	0.983	0.0011	0.1628	1.5116	0.0011	-6.7782	4.55E-03	-1.55E-04	-2.14E-05	-4.21E-05	-4.07E-05
8	2400	0	1.446	0.0006	0.0741	1.8927	0.0006	-7.3400	6.69E-03	5.04E-03	-2.14E-05	-6.19E-05	-1.13E-04
9	2500	1	1.908	0.0003	0.0282	2.2924	0.0282	-3.5696	1.06E-02	2.02E-02	-1.88E-05	-8.50E-05	-2.56E-04
10	2500	1	1.908	0.0003	0.0282	2.2924	0.0282	-3.5696	1.06E-02	2.02E-02	-1.88E-05	-8.50E-05	-2.56E-04
							計		2.12E-02	4.05E-02	-2.09E-04	-1.70E-04	-8.54E-04

表 10 に反復した場合のパラメータの変化を示す．EM アルゴリズムの 17 回よりも少ない 5 回で，変化量が 0 となり収束した．表 11 に第 5 反復した時点での計算シートを示す．

表 10 繰り返し計算による推定値の変化

繰返し	μ	σ
0	2087.50	216.15
1	2162.64	248.58
2	2190.00	283.92
3	2200.75	302.92
4	2202.57	306.80
5	2202.62	306.93

表 11 収束した第 5 反復でのニュートン・ラフソン法による計算結果

正規分布				元の	変化量	新たな			1階の	2階の偏微分		負の逆行列	
Newton-Raphson法				パラメータ	$(-H)^{-1}U$	パラメータ			偏微分 U	H		$(-H)^{-1}$	
				$\mu^{\wedge} =$	2202.62	0.00	2202.62		0.0000	-1.02E-04	1.55E-05	9969.67	994.74
				$\sigma^{\wedge} =$	306.93	0.00	306.93		0.0000	1.55E-05	-1.55E-04	994.74	6556.28
				$\sigma^{\wedge 2} =$	94206		$\ln L =$	-59.299	$-2 \ln L =$	118.60			
No	y	$+\delta$	z	$f(z)$	$1-\Phi(z)$	$\frac{\phi z}{(1-\Phi z)}$	Li	$\ln Li$	$\partial \mu$	$\partial \sigma$	$\partial \mu \partial \mu$	$\partial \mu \partial \sigma$	$\partial \sigma \partial \sigma$
1	1700	0	-1.638	0.0003	0.9492	0.1100	0.0003	-7.9864	-5.34E-03	5.48E-03	-1.06E-05	3.48E-05	-7.48E-05
2	1850	0	-1.149	0.0007	0.8747	0.2357	0.0007	-7.3055	-3.74E-03	1.04E-03	-1.06E-05	2.44E-05	-3.14E-05
3	2000	0	-0.660	0.0010	0.7454	0.4304	0.0010	-6.8635	-2.15E-03	-1.84E-03	-1.06E-05	1.40E-05	-3.26E-06
4	2100	0	-0.334	0.0012	0.6309	0.5979	0.0012	-6.7014	-1.09E-03	-2.89E-03	-1.06E-05	7.10E-06	7.06E-06
5	2150	0	-0.171	0.0013	0.5681	0.6920	0.0013	-6.6603	-5.59E-04	-3.16E-03	-1.06E-05	3.64E-06	9.68E-06
6	2200	0	-0.009	0.0013	0.5034	0.7925	0.0013	-6.6456	-2.78E-05	-3.26E-03	-1.06E-05	1.81E-07	1.06E-05
7	2300	0	0.317	0.0012	0.3755	1.0102	0.0012	-6.6959	1.03E-03	-2.93E-03	-1.06E-05	-6.74E-06	7.41E-06
8	2400	0	0.643	0.0011	0.2601	1.2474	0.0011	-6.8523	2.10E-03	-1.91E-03	-1.06E-05	-1.37E-05	-2.55E-06
9	2500	1	0.969	0.0008	0.1663	1.5003	0.1663	-1.7940	4.89E-03	4.74E-03	-8.46E-06	-2.41E-05	-3.88E-05
10	2500	1	0.969	0.0008	0.1663	1.5003	0.1663	-1.7940	4.89E-03	4.74E-03	-8.46E-06	-2.41E-05	-3.88E-05
								計	3.04E-09	2.06E-08	-1.02E-04	1.55E-05	-1.55E-04

新たなパラメータを元のパラメータに値のみの貼り付けを繰り返す方法を示したのであるが，Excel のソルバーを用いて，対数尤度 $\ln L$ を最大にするように，元のパラメータを変数セルに設定しても解が得られる．あえて，面倒な 2 階の偏微分式を求め，Excel シート上でニュートン・ラフソン法による反復計算の過程を示したのは，統計ソフトが実際に行っている計算方式を可視化するためである．

正規分布の確率密度関数の対数尤度関数についての偏微分式は，高橋(2105a) に示されているものを引用する．正規分布は，パラメータを μ ，形状パラメータを σ としその確率密度関数，および対数を取った場合は，

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right), \quad \ln f(y) = -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(y-\mu)^2}{2\sigma^2}$$

なので、 $\ln f(y)$ を μ , (σ^2) で偏微分して、次の結果を得る.

$$\frac{\partial \ln f(y)}{\partial \mu} = \frac{(y-\mu)}{(\sigma^2)}, \quad \frac{\partial \ln f(y)}{\partial (\sigma^2)} = -\frac{1}{2(\sigma^2)} + \frac{(y-\mu)^2}{2(\sigma^2)^2}$$

さらに2階の偏微分を計算する.

$$\frac{\partial \ln f(y)}{\partial \mu \partial \mu} = -\frac{1}{(\sigma^2)}, \quad \frac{\partial \ln f(y)}{\partial \mu \partial (\sigma^2)} = \frac{(y-\mu)}{(\sigma^2)^2}, \quad \frac{\partial \ln f(y)}{\partial (\sigma^2) \partial (\sigma^2)} = \frac{1}{2(\sigma^2)^2} - \frac{(y-\mu)^2}{2(\sigma^2)^3}$$

$\ln f(y)$ を μ , σ で偏微分した場合は、次の結果を得る.

$$\frac{\partial \ln f(y)}{\partial \mu} = \frac{(y-\mu)}{\sigma^2}, \quad \frac{\partial \ln f(y)}{\partial \sigma} = -\frac{1}{\sigma} + \frac{(y-\mu)^2}{\sigma^3}$$

さらに2階の偏微分を計算すると

$$\frac{\partial \ln f(y)}{\partial \mu \partial \mu} = -\frac{1}{\sigma^2}, \quad \frac{\partial \ln f(y)}{\partial \mu \partial \sigma} = -\frac{2(y-\mu)}{\sigma^3}, \quad \frac{\partial \ln f(y)}{\partial \sigma \partial \sigma} = \frac{1}{\sigma^2} - \frac{3(y-\mu)^2}{\sigma^4}$$

が得られ、 $z = (y-\mu)/\sigma$ とおけば、 $\ln(\phi(z)/\sigma)$ をそのまま使った結果とももちろん一致する.

正規分布の上側確率の対数 $\ln(1-\Phi(z))$ についての偏微分は、高橋(2107) で2値反応に対するプロビット法をニュートン・ラフソン法によって Excel で計算した際に導出したが、プロビット曲線としての累積正規分布を使うだけなので $z = \beta_0 + \beta_1 y$ とした線形式で対応した.

今回は、 μ と σ を直接推定するために $z = (y-\mu)/\sigma$ の形式で対応したために、 $z = \beta_0 + \beta_1 y$ の場合よりも偏微分式が煩雑になる. 偏微分式を簡素化するために、 $\beta_0 = -\mu/\sigma$, および $\beta_1 = 1/\sigma$ とすることもできるが、回帰分析に際して、 $\mu = \alpha + \beta x$ とするために、別の煩雑さが入り込みことになり、 $z = (y-\mu)/\sigma$ のままで対応することにした.

次節に、 $z = (y-\alpha-\beta x)/\sigma$ とした回帰分析の場合についての偏微分式は、本節で示した偏微分式に x を掛けるか、 x^2 を掛けるか、だけで済ますことができる. 更に、単回帰式でなく、重回帰式の場合でも、偏微分式にそれぞれの変数を組み合わせて掛けて対応ができる.

ニュートン・ラフソン法を使う必要のある様々な統計モデルに対応するために、本節で示した、正規分布の確率密度関数、および累積分布関数にたいする偏微分の定式化は、見通しの良い方法と思われる. 私にとっても、以前のセミナーテキストの煩雑な偏微分式は、読者にとっても避けたいような式であったと反省している.

13. Excel による Tobit 回帰分析

打ち切りデータを含む回帰分析で、誤差に正規分布を仮定する場合は、計量経済学の分野で Tobit モデルと言われている。牧ら(1997)では、第4章で「Probit, Logit, Tobit」が論じられている。Probit, および Logit は、2 値データの解析方法であるが、Tobit は、打ち切りのある計量データの解析であり、負の値を取ることができない耐久消費財への支出の分析に Tobin によって初めて用いられたことから、Tobit モデルと呼ばれるようになったそうである。Tobit モデルと同様な状況は、微量の化学物質の計量する場合に測定限界値が生ずる場合にもある。

発毛試験データの場合には、打ち切りデータが、ある値以上となっていて、Tobit モデルとは、言い難い面もある。しかし、計量経済学の分野でも年間収入などの場合にある金額以上を打ち切りデータ（トップコーディング）とする場合も Tobit モデルという場合もある。ある値以下ならば、正規分布の下側確率となり、ある値以上ならば上側確率となるだけで、どちらでも同じ統計モデルと理解できる。

打ち切りデータを含む回帰分析について、ほとんどの成書は、統計ソフトを使った例示であり、数値計算によって自ら計算を試みようとしたときに必要な偏微分式が省略されているため、断念してしまうことになりがちである。幸い、第12節で示した偏微分式を用いれば、比較的容易に「Tobit」モデルを数値計算レベルで解くことができる。

発毛実験データで、打ち切りデータを含む回帰分析について、Excel のソルバー、JMP の「寿命の二変量」、OnDemand SAS の LIFEREG プロシジャでの解析結果を示してきたが、Excel でのニュートン・ラフソン法による計算の提示は、避けてきた。これは、前節で示したように、対数尤度関数を偏微分することは、難儀なことでありその偏微分式が正しいかの検証は、実際にニュートン・ラフソン法を適用して、きちっと解が求まることで示すしかない。偏微分に際し、プラス・マイナスの付け間違いなど、しばしば見逃してしまい、解が求まらないことで、偏微分式のミスに気が付く。

前節での偏微分式を用いることにより、打ち切りのある回帰分析の偏微分式を導出し、Excel のソルバーに頼らず発毛データを用いた回帰分析を行ってみよう。前節では、正規分布 $N(\mu, \sigma^2)$ からの n 個の無作為にサンプリングした標本の中で m 個 y_1, \dots, y_m が c 以下で観測され、 $n - m$ 個が c を越えて打ち切りになったとすると、対数尤度関数は

$$\ln L = \sum_{i=1}^m \ln \left[\frac{1}{\sigma} \phi \left(\frac{y_i - \mu}{\sigma} \right) \right] + \sum_{i=m+1}^n \ln \left[1 - \Phi \left(\frac{c - \mu}{\sigma} \right) \right]$$

となる。正規分布の密度関数と分布関数の和となっているので、別々にパラメータ μ と σ について偏微分を行うことにする。

一般的な回帰式分析では、 $y_i = \alpha + \beta x_i + \varepsilon_i$, $\varepsilon_i \sim N(\varepsilon_i | 0, \sigma^2)$ として最小 2 乗法によりパラメータを推定しているが、打ち切りデータがある場合には適用できない。そこで、最尤法による回帰式を

$$\mu_i = \alpha + \beta x_i + \sigma \varepsilon_i$$

とする。ここで、誤差項に含めた σ も推定対象であることを明示する必要がある。回帰の対数尤度関数は、 μ を線形式 $\alpha + \beta x_i$ に置き換えた次式となる。

$$\ln L = \sum_{i=1}^m \ln \left[\frac{1}{\sigma} \varphi \left(\frac{y_i - \alpha - \beta x_i}{\sigma} \right) \right] + \sum_{i=m+1}^n \ln \left[1 - \Phi \left(\frac{c - \alpha - \beta x_i}{\sigma} \right) \right]$$

前節と同様に、標準正規分布の密度関数は、 z の中身が、

$$\varphi(z_i) = \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{z_i^2}{2} \right), \quad \text{ただし, } z_i = \frac{y_i - \alpha - \beta x_i}{\sigma}$$

と異なるだけである。

対数を取った密度関数 $\ln f(y)$ を偏微分しやすいような式（添え字は省略）

$$\ln f(y) = \ln \left[\frac{1}{\sigma} \varphi \left(\frac{y - \alpha - \beta x}{\sigma} \right) \right] = \ln \left[\varphi \left(\frac{y}{\sigma} - \frac{\alpha}{\sigma} - \frac{\beta x}{\sigma} \right) \right] - \ln \sigma$$

にして、 μ の代わりに α で偏微分をしても結果は同じ

$$\frac{\partial \ln f(y)}{\partial \alpha} = \frac{z}{\sigma}$$

であり、 β で偏微分した場合は、 x が付いた式

$$\frac{\partial \ln f(y)}{\partial \beta} = \frac{z}{\sigma} x$$

となる。 σ についての偏微分も $(y_i - \mu) / \sigma$ の場合

$$\frac{\partial \ln f(y)}{\partial \sigma} = \frac{z^2 - 1}{\sigma}$$

と同じ結果になる。これらを用いて 2 階の偏微分は、 β で偏微分した場合は、 $(y_i - \mu) / \sigma$ の場合に x を更に付けた式となる。 β で 2 階偏微分した場合は、 x^2 を付けた式となる。

$$\begin{aligned} \frac{\partial^2 \ln f(y)}{\partial \alpha \partial \alpha} &= \frac{-1}{\sigma^2}, & \frac{\partial^2 \ln f(y)}{\partial \alpha \partial \beta} &= \frac{-1}{\sigma^2} x, & \frac{\partial^2 \ln f(y)}{\partial \alpha \partial \sigma} &= \frac{-2z}{\sigma^2} \\ \frac{\partial^2 \ln f(y)}{\partial \beta \partial \beta} &= \frac{-1}{\sigma^2} x^2, & \frac{\partial^2 \ln f(y)}{\partial \beta \partial \sigma} &= \frac{-2z}{\sigma^2} x, & \frac{\partial^2 \ln f(y)}{\partial \sigma \partial \sigma} &= \frac{1 - 3z^2}{\sigma^2} \end{aligned}$$

分布関数の上側確率についての偏微分は、同様に x または x^2 を付与した式となる。

$$\ln[1-\Phi(z)] = \ln\left[1-\Phi\left(\frac{y-\alpha-\beta x}{\sigma}\right)\right]$$

$$\frac{\partial \ln[1-\Phi(z)]}{\partial \alpha} = \frac{Q(z)}{\sigma}, \quad \frac{\partial \ln[1-\Phi(z)]}{\partial \beta} = \frac{Q(z)}{\sigma}x, \quad \frac{\partial \ln[1-\Phi(z)]}{\partial \sigma} = \frac{zQ(z)}{\sigma}$$

$$\text{ただし, } Q(z) = \frac{\varphi(z)}{1-\Phi(z)}$$

2 階の偏微分は、 β で偏微分した場合は、 x をさらに付けた式となる。

$$\frac{\partial^2 \ln[1-\Phi(z)]}{\partial \alpha \partial \alpha} = \frac{zQ(z)-Q(z)^2}{\sigma^2}, \quad \frac{\partial^2 \ln[1-\Phi(z)]}{\partial \alpha \partial \beta} = \frac{zQ(z)-Q(z)^2}{\sigma^2}x,$$

$$\frac{\partial^2 \ln[1-\Phi(z)]}{\partial \alpha \partial \sigma} = \frac{(z^2-1)Q(z)-zQ(z)^2}{\sigma^2}, \quad \frac{\partial^2 \ln[1-\Phi(z)]}{\partial \beta \partial \beta} = \frac{zQ(z)-Q(z)^2}{\sigma^2}x^2$$

$$\frac{\partial^2 \ln[1-\Phi(z)]}{\partial \beta \partial \sigma} = \frac{(z^2-1)Q(z)-zQ(z)^2}{\sigma^2}x, \quad \frac{\partial^2 \ln[1-\Phi(z)]}{\partial \sigma \partial \sigma} = \frac{(z^3-2z)Q(z)-z^2Q(z)^2}{\sigma^2}$$

これらの 1 階のおよび 2 階の偏微分式を Excel シートに埋め込み、ニュートン・ラフソン法により対数尤度を最大化する μ および σ を推定することができる。打ち切りがあれば 1、なければ 0 となる変数 δ_i とし、Excel の if() 関数を用いてそれぞれの y_i に対し、尤度 L_i を計算し、対数尤度 $\ln L$ を計算する。

$$L_i = \text{if} [\delta_i = 0, (\varphi(\hat{z}_i)/\hat{\sigma}), (1-\Phi(\hat{z}_i))],$$

$$\delta_i = 0, \quad \delta_i \neq 0$$

$$\ln L = \sum_i \ln L_i, \quad \text{ただし, } \hat{z}_i = \frac{y_i - \hat{\alpha} - \hat{\beta}x_i}{\hat{\sigma}}$$

それぞれの y_i に対し 1 階の偏微分式を次式で計算し、

$$\partial \alpha_i = \text{if} [\delta_i = 0, \hat{z}_i/\hat{\sigma}, Q(\hat{z}_i)/\hat{\sigma}],$$

$$\partial \beta_i = \text{if} [\delta_i = 0, \hat{z}_i x_i/\hat{\sigma}, Q(\hat{z}_i)x_i/\hat{\sigma}],$$

$$\partial \sigma_i = \text{if} [\delta_i = 0, (\hat{z}_i^2-1)/\hat{\sigma}, \hat{z}_i Q(\hat{z}_i)/\hat{\sigma}]$$

$$\text{ただし, } Q(z_i) = \frac{\varphi(z_i)}{1-\Phi(z_i)}$$

i について合計し、スコアベクトル U とする。

$$U = \begin{bmatrix} \sum_i \partial \alpha_i \\ \sum_i \partial \beta_i \\ \sum_i \partial \sigma_i \end{bmatrix}$$

それぞれの x_i に対し 2 階の偏微分式を次式で計算し、

$$\begin{aligned}\partial\alpha_i\partial\alpha_i &= \text{if } \{\delta_i=0, -1/\hat{\sigma}^2, [\hat{z}_i Q(\hat{z}_i) - Q(\hat{z}_i)^2]/\hat{\sigma}^2\} \\ \partial\alpha_i\partial\beta_i &= \text{if } \{\delta_i=0, -x_i/\hat{\sigma}^2, [\hat{z}_i Q(\hat{z}_i) - Q(\hat{z}_i)^2]x_i/\hat{\sigma}^2\} \\ \partial\alpha_i\partial\sigma_i &= \text{if } \{\delta_i=0, -2\hat{z}_i/\hat{\sigma}^2, [(\hat{z}_i^2-1)Q(\hat{z}_i) - \hat{z}_i Q(\hat{z}_i)^2]/\hat{\sigma}^2\} \\ \partial\beta_i\partial\beta_i &= \text{if } \{\delta_i=0, -x_i^2/\hat{\sigma}^2, [\hat{z}_i Q(\hat{z}_i) - Q(\hat{z}_i)^2]x_i^2/\hat{\sigma}^2\} \\ \partial\beta_i\partial\sigma_i &= \text{if } \{\delta_i=0, -2\hat{z}_i x_i/\hat{\sigma}^2, [(\hat{z}_i^2-1)Q(\hat{z}_i) - \hat{z}_i Q(\hat{z}_i)^2]x_i/\hat{\sigma}^2\} \\ \partial\sigma_i\partial\sigma_i &= \text{if } \{\delta_i=0, (1-3z_i^2)/\hat{\sigma}^2, [(z_i^3-2z_i)Q(z_i) - z_i^2 Q(z_i)^2]/\hat{\sigma}^2\}\end{aligned}$$

i について合計し、ヘシアン行列 H とする。

$$H = \begin{bmatrix} \sum_i \partial\alpha_i\partial\alpha_i & \sum_i \partial\alpha_i\partial\beta_i & \sum_i \partial\alpha_i\partial\sigma_i \\ \sum_i \partial\alpha_i\partial\beta_i & \sum_i \partial\beta_i\partial\beta_i & \sum_i \partial\beta_i\partial\sigma_i \\ \sum_i \partial\alpha_i\partial\sigma_i & \sum_i \partial\beta_i\partial\sigma_i & \sum_i \partial\sigma_i\partial\sigma_i \end{bmatrix}$$

初期値として、打ち切りを無視した回帰分析のパラメータ $\alpha=39.537$, $\beta=-5.077$, $\sigma^2=5.094^2$ を使い、元のパラメータとし、変化量 $(-H)^{-1}U$ を加える。

表 12 回帰分析の結果

要因	自由	平方	平均平方	項	推定	標準誤
モデル	1	913.1155	913.115	切片	39.537341	2.544757
誤差	29	752.5619	25.950	x	-5.077413	0.855957
全体(修正済み)	30	1665.6774				

$$\begin{array}{c} \text{新たなパラメータ} \\ \begin{bmatrix} \hat{\alpha}_{new} \\ \hat{\beta}_{new} \\ \hat{\sigma}_{new} \end{bmatrix} \end{array} = \begin{array}{c} \begin{bmatrix} \hat{\alpha}_{old} \\ \hat{\beta}_{old} \\ \hat{\sigma}_{old} \end{bmatrix} \\ \text{元のパラメータ} \end{array} + \begin{array}{c} (-H)^{-1}U \\ \text{変化量} \end{array}$$

表 13 の結果を用いて

変化量	変化量	=	負の逆行列			1階の
	$(-H)^{-1}U$		$(-H)^{-1}$			偏微分 U
	4.2268		4.8270	-1.5391	-0.3805	2.5402
	-1.0107		-1.5391	0.5831	0.0848	4.6156
	0.2074		-0.3805	0.0848	0.3199	2.4467

Excel による行列計算 = Mmult (Minverse(-H) , U)

$$\begin{array}{|c|} \hline \text{新たな} \\ \hline \text{パラメータ} \\ \hline 43.7638 \\ -6.0877 \\ 5.3014 \\ \hline \end{array} = \begin{array}{|c|} \hline \text{元の} \\ \hline \text{パラメータ} \\ \hline 39.5370 \\ -5.0770 \\ 5.0940 \\ \hline \end{array} + \begin{array}{|c|} \hline \text{変化量} \\ \hline (-H)^{-1}U \\ \hline 4.2268 \\ -1.0107 \\ 0.2074 \\ \hline \end{array}$$

新たなパラメータをコピーして、元のパラメータに「値」のみ貼り付け、1回目の反復とする。同様に2回目の反復、3回目、と変化量が0になるまで繰り返す。

表 13 初期値に対するニュートン・ラフソン法による計算結果

正規分布				元の		変化量	新たな		1階の		2階の偏微分		負の逆行列					
Newton-Raphson法				パラメータ		$(-H)^{-1}U$	パラメータ		偏微分U		H		$(-H)^{-1}$					
				$\alpha^{\wedge} =$	39.5370	4.2268	43.7638		2.5402	-1.4600	-3.7453	-0.7439	4.8270	-1.5391	-0.3805			
				$\beta^{\wedge} =$	-5.0770	-1.0107	-6.0877		4.6156	-3.7453	-11.3913	-1.4354	-1.5391	0.5831	0.0848			
				$\sigma^{\wedge} =$	5.0940	0.2074	5.3014		2.4467	-0.7439	-1.4354	-3.6307	-0.3805	0.0848	0.3199			
				$\sigma^{\wedge 2} =$	26		$\ln L =$	-109.80	$-2 \ln L =$	219.61								
No	x	y	$\frac{+}{\delta}$	z	f(z)	$1-\Phi(z)$	$\frac{\varphi z}{(1-\Phi z)}$	Li	$\ln Li$	$\partial \alpha$	$\partial \beta$	$\partial \sigma$	$\partial \alpha \partial \alpha$	$\partial \alpha \partial \beta$	$\partial \alpha \partial \sigma$	$\partial \beta \partial \beta$	$\partial \beta \partial \sigma$	$\partial \sigma \partial \sigma$
								計		2.5402	4.6156	2.4467	-1.4600	-3.7453	-0.7439	-11.3913	-1.4354	-3.6307
1	1	36	1	0.3023	0.0748	0.3812	0.9998	0.3812	-0.9644	0.1963	0.1963	0.0593	-0.0269	-0.0269	-0.0467	-0.0269	-0.0467	-0.0258
2	1	36	1	0.3023	0.0748	0.3812	0.9998	0.3812	-0.9644	0.1963	0.1963	0.0593	-0.0269	-0.0269	-0.0467	-0.0269	-0.0467	-0.0258
3	1	36	1	0.3023	0.0748	0.3812	0.9998	0.3812	-0.9644	0.1963	0.1963	0.0593	-0.0269	-0.0269	-0.0467	-0.0269	-0.0467	-0.0258
4	1	36	1	0.3023	0.0748	0.3812	0.9998	0.3812	-0.9644	0.1963	0.1963	0.0593	-0.0269	-0.0269	-0.0467	-0.0269	-0.0467	-0.0258
5	1	36	1	0.3023	0.0748	0.3812	0.9998	0.3812	-0.9644	0.1963	0.1963	0.0593	-0.0269	-0.0269	-0.0467	-0.0269	-0.0467	-0.0258
6	2	36	1	1.2990	0.0337	0.0970	1.7695	0.0970	-2.3333	0.3474	0.6947	0.4512	-0.0321	-0.0642	-0.1099	-0.1283	-0.2197	-0.2313
7	2	36	1	1.2990	0.0337	0.0970	1.7695	0.0970	-2.3333	0.3474	0.6947	0.4512	-0.0321	-0.0642	-0.1099	-0.1283	-0.2197	-0.2313
8	2	36	1	1.2990	0.0337	0.0970	1.7695	0.0970	-2.3333	0.3474	0.6947	0.4512	-0.0321	-0.0642	-0.1099	-0.1283	-0.2197	-0.2313
9	3	36	1	2.2956	0.0056	0.0108	2.6375	0.0108	-4.5238	0.5178	1.5533	1.1886	-0.0347	-0.1042	-0.1814	-0.3127	-0.5442	-0.6498
10	1	30	0	-0.8755	0.0534	0.8094	0.3360	0.0534	-2.9303	-0.1719	-0.1719	-0.0458	-0.0385	-0.0385	0.0675	-0.0385	0.0675	-0.0501
11	1	30	0	-0.8755	0.0534	0.8094	0.3360	0.0534	-2.9303	-0.1719	-0.1719	-0.0458	-0.0385	-0.0385	0.0675	-0.0385	0.0675	-0.0501
12	1	33	0	-0.2866	0.0752	0.6128	0.6248	0.0752	-2.5881	-0.0563	-0.0563	-0.1802	-0.0385	-0.0385	0.0221	-0.0385	0.0221	0.0290
13	1	33	0	-0.2866	0.0752	0.6128	0.6248	0.0752	-2.5881	-0.0563	-0.0563	-0.1802	-0.0385	-0.0385	0.0221	-0.0385	0.0221	0.0290
14	1	36	0	0.3023	0.0748	0.3812	0.9998	0.0748	-2.5927	0.0593	0.0593	-0.1784	-0.0385	-0.0385	-0.0233	-0.0385	-0.0233	0.0280
15	2	28	0	-0.2715	0.0755	0.6070	0.6335	0.0755	-2.5839	-0.0533	-0.1066	-0.1818	-0.0385	-0.0771	0.0209	-0.1541	0.0419	0.0300
38	4	23	0	0.7403	0.0595	0.2296	1.3213	0.0595	-2.8210	0.1453	0.5813	-0.0887	-0.0385	-0.1541	-0.0571	-0.6166	-0.2282	-0.0248
39	4	26	0	1.3292	0.0324	0.0919	1.7947	0.0324	-3.4304	0.2609	1.0437	0.1505	-0.0385	-0.1541	-0.1024	-0.6166	-0.4098	-0.1657
40	4	28	0	1.7218	0.0178	0.0426	2.1293	0.0178	-4.0294	0.3380	1.3520	0.3857	-0.0385	-0.1541	-0.1327	-0.6166	-0.5308	-0.3042

表 14 に反復した場合のパラメータの変化を示す。第 5 回目で、変化量が 0 となり収束した。表 15 に第 5 反復した時点での計算シートを示す。

表 14 繰り返し計算による推定値の変化

	反復 0	1	2	3	4	5
$\alpha^{\wedge} =$	39.5370	43.7638	44.4781	44.6454	44.6542	44.6542
$\beta^{\wedge} =$	-5.0770	-6.0877	-6.2738	-6.3175	-6.3198	-6.3198
$\sigma^{\wedge} =$	5.0940	5.3014	5.7571	5.8822	5.8895	5.8895

新たなパラメータを元のパラメータに値のみの貼り付けを繰り返す方法を示したのであるが、Excel のソルバーを用いて、対数尤度 $\ln L$ を最大にするように、元のパラメータを変数

表 16 収束した第 5 反復での共分散行列

負の逆行列			共分散行列				負の逆行列		
$(-H)^{-1}$			パラメータ	β_0	β_1	σ	$(-H)^{-1}$		
6.4718	-2.1000	0.3662	β_0	6.4719	-2.1001	0.3662	6.4719	-2.1001	0.3662
-2.1000	0.8003	-0.0986	β_1	-2.1001	0.8003	-0.0986	-2.1001	0.8003	-0.0986
0.3662	-0.0986	0.5937	σ	0.3662	-0.0986	0.5938	0.3662	-0.0986	0.5938

Excel 反復 5

JMP

Excel 反復 6

この原因は、Excel では、5 回の反復で収束したと Excel の有効数字の範囲内で判断したのであるが、JMP では、更に反復を繰り返したための相違である。Excel でもう 1 反復した場合の共分散は、JMP の結果と一致する。

共分散行列を活用することにより、統計ソフトでは対応していない種々の推定を行うことができる。多くの入門的な統計の教科書では、「行列」表記を避けているために、回帰直線の 95%信頼区間、および個別データの 95%信頼区間は、「この式で与えられる」と天下りの記述をしている。

打ち切りがある場合でも、無い場合でも回帰直線の 95%信頼区間の求め方は、得られた共分散行列を用いれば、全く同じ式となる。スライド 61 とスライド 62 を再掲する。

スライド 61 (再掲)

分散共分散行列

パラメータ	β_0	β_1	σ
β_0	6.4719	-2.1001	0.3662
β_1	-2.1001	0.8003	-0.0986
σ	0.3662	-0.0986	0.5938

- ◆ 対角要素が分散
 - ◆ $\text{Var}(\beta_0) = 6.4719$, $\text{Var}(\beta_1) = 0.8003$
- ◆ 共分散
 - ◆ $\text{Cov}(\beta_0, \beta_1) = -2.1001$

61

スライド 62 (再掲)

\hat{y} の分散

$$\text{Var}(\hat{y}) = \text{Var}(\beta_0 \hat{+} \beta_1 \hat{x})$$

$$= \text{Var}(\beta_0 \hat{+}) + 2\text{Cov}(\beta_0 \hat{+}, \beta_1 \hat{x}) + \text{Var}(\beta_1 \hat{x})$$

$$\mathbf{x} = [1, x] \quad \Sigma = \begin{array}{|c|cc|} \hline \text{パラメータ} & \beta_0 & \beta_1 \\ \hline \beta_0 & 6.4719 & -2.1001 \\ \beta_1 & -2.1001 & 0.8003 \\ \hline \end{array}$$

と置いたときに、次のような行列計算でも計算できる。

$$\text{Var}(\hat{y}) = \mathbf{x} \Sigma \mathbf{x}^T$$

62

参考文献

- Breen, R.H. (1996), Regression Models :Censored Sample Selected or Truncated Data Using Minitab and R., Sage University.
- Kalbfleisch, J.D. and Prentice, R.L. (2002), The Statistical Analysis of Failure Time Data 2nd ed., Wiley.
- Meeker and Escobar (1998), Statistical Method for Reliability Data, Wiley.
- SAS Institute (2012), 品質管理および信頼性/生存時間, SAS Institut Inc.
- 岩崎学 (2002), 不完全データの統計解析, 3.2.3 節 値打ち切り, エコノミスト社.
- 岩崎学, 阿部貴行 (2006), 打ち切りおよびトランケーションの下でのパラメータ推定に及ぼす切断点の影響評価, 応用統計学, Vol.35, No.1 (2006), 49-60 .
- https://www.jstage.jst.go.jp/article/jappstat1971/35/1/35_1_49/_pdf/-char/ja .
- 奥野忠一監訳, 柴田義貞, 藤野和健, 鎌倉稔成訳 (1988), 寿命データの解析, 日科技連出版社.
- 黒須真介 (2017), 探索的薬理試験データに対する種々の解析方法の適用と比較, 日薬理誌, 150, 4-9. https://www.jstage.jst.go.jp/article/fpi/150/1/150_4/_pdf/-char/ja
- 高橋行雄 (2013), 回帰分析・再入門ー統計ソフトが対応していない生物統計の各種の課題を Excel でサクサク解こうー, 医薬安全研定例会 第 7 回 じっくり勉強セミナー.
- http://www.scientist-press.com/edit_html/green/seminar7.pdf
- 高橋行雄 (2015a), 続高橋セミナー第 4 回:ー寿命試験データの統計解析,
- <http://www.yukms.com/biostat/takahasi2/rec/004.htm>
- 高橋行雄 (2015b), LIFEREG プロシジャを用いた打ち切りデータを含むワイブル回帰, SAS ユーザー総会論文集:123-132.
- 秋山 功, 富山 茂巳, 高橋 行雄 (2016), 定量下限未満を含むデータの要約統計量と各種の統計解析, 第 18 回医薬安全研定例会資料, https://biostat.jp/archive_teireikai_2.php .
- 高橋行雄 (2016), 定量下限未満を含むデータの要約統計量と各種の統計解析ー最尤法について の補足ー, 第 18 回医薬安全研定例会資料, https://biostat.jp/archive_teireikai_2.php .
- 高橋行雄 (2017), 続高橋セミナー第 6 回:一般化線形モデルを Excel で極め活用するープロビット法・ロジット法・補2重対数法ー. <http://www.yukms.com/biostat/takahasi2/rec/006.htm> .
- 高波洋平, 舟尾暢男 (2016), SAS Studio によるやさしい統計データ分析, オーム社.
- 棚橋昌也ら (2009), LIFEREG プロシジャによる正規分布からの打ち切り標本に対するパラメータ推定と二標本有意差検定, SAS ユーザー総会論文集:219-228.
- <http://www.sascom.jp/download/pdf/SUGJ2009.pdf> .
- 芳賀敏郎 (2004), 最小 2 乗法, 最尤法, 線形モデル, 非線形モデル,
- <http://www.yukms.com/biostat/haga/download/archive/likelihood/Likelihood.pdf>
- 芳賀敏郎 (2010), 医薬品開発のための統計解析 第 3 部 非線形モデル, サイエンティスト社.
- 牧厚志, 宮内環, 浪花貞夫, 縄田数満 (1997), 応用計量経済学 II, 多賀出版.











索引

Excel	if()		56	文献 高橋 (2105a)		58
Excel	Linest()関数	線形回帰分析	14	文献 高橋 (2107)		59
Excel	Minverse()	逆行列	57	文献 高波ら (2016)		46
Excel	Minverse()	逆行列	63	文献 黒須 (2017)		4
Excel	Mmult()	行列の積	37	文献 秋山ら (2016)		3
Excel	Mmult()	行列の積	57	文献 棚橋ら (2009)		50
Excel	Mmult()	行列の積	63	文献 芳賀 (2004)		17
Excel	Norm.dist()	正規分布	18	文献 芳賀 (2010)		13
Excel	Norm.dist()	正規分布	24	文献 牧ら (1997)		60
Excel	Stdv.p()	母集団	22	1階の偏微分	スコアベクトル	56
Excel	Stdv.s()	標本	22	2階の微分	μ	55
Excel	Transpose()	転置行列	37	2階の微分	σ	55
Excel	逆行列	Minverse()	57	2階の偏微分	β	61
Excel	逆行列	Minverse()	63	2階の偏微分	ニュートン・ラフソン法	49
Excel	行列の積	Mmult()	37	2階の偏微分行列	ヘシアン行列	50
Excel	行列の積	Mmult()	57	2階の偏微分行列	ヘシアン行列	64
Excel	行列の積	Mmult()	63	2階の偏微分式	ヘシアン行列	56
Excel	ソルバー	目的セルの設定	17	-2倍の対数尤度		33
Excel	転置行列	Transpose()	37	95%信頼区間	回帰直線	36
Excel	正規分布	Norm.dist()	24	95%信頼区間	共分散行列	36
Excel	線形回帰分析	Linest()関数	14	Arrhenius摂氏	設定パネル	42
JMP	スクリプト		52	EMアルゴリズム		49
JMP	一変量の分布	群	27	Excel	ニュートン・ラフソン法	54
JMP	一変量の分布		39	LIFEREG	正規分布	47
JMP	寿命の一変量	イベントプロット	10	$\ln L$	対数尤度	20
JMP	寿命の一変量		9	Motorette試験	発電機モータ寿命	2
JMP	寿命の二変量		9	OnDemand SAS	無料	46
JMP	寿命の二変量		39	Tobitモデル	計量経済学	60
JMP	信頼性/生存時間解析		9	α	偏微分	61
JMP	生存時間(パラメトリック)のあてはめ		39	β	2階の偏微分	61
JMP	生存時間のあてはめ(パラメトリック)		52	β	偏微分	61
JMP	生存時間解析		39	μ	2階の微分	55
JMP	設定画面	寿命の一変量	10	μ	偏微分	54
JMP	二変量の関係		28	σ	2階の微分	55
JMP	二変量の関係		39	σ	偏微分	54
JMP	比例ハザードのあてはめ		39	σ^2	偏微分	59
JMP	分析メニュー		9	σ	偏微分	59
SAS	LIFEREGプロシジャ	パラメトリックな生存時間解析	46	$\phi(z)$	標準正規分布	49
SAS	LIFEREGプロシジャ		7	$\Phi(z)$	標準正規分布	49
SAS	RELIABILITYプロシジャ	信頼性データの解析	46	位置	平均	12
文献	Breen (1996)		3	医薬安全性研究会	じっくり勉強シリーズ	1
文献	Kalbfleisch (2002)		2	カプラン・マイヤー曲線		39
文献	Meeker and Escobar (1998)		45	グリーン本	芳賀	13
文献	SAS Institute (2012)		45	じっくり勉強シリーズ	医薬安全性研究会	1
文献	奥野ら (1988)		45	スコアベクトル	1階の偏微分	56
文献	岩崎 (2002)		49	スコアベクトル		62
文献	岩崎ら (2006)		49	ソルバー	設定画面	17
文献	高橋 (2013)		15	ソルバー	肺活量データ	53
文献	高橋 (2015a)		45	ソルバー	変数セルの変更	17
文献	高橋 (2015b)		3	対数尤度	$\ln L$	20
文献	高橋 (2016)		3	探索的	薬理試験	1

トップコーディング	打ち切り	60	寿命の二変量	JMP	39
ニュートン・ラフソン法	2階の偏微分	49	寿命の二変量	共分散行列	36
ニュートン・ラフソン法	Excel	54	寿命の二変量	設定パネル	42
ニュートン・ラフソン法	プロビット法	59	上側確率	可視化	53
ニュートン・ラフソン法	変化量	57	上側確率	偏微分	55
ニュートン・ラフソン法	変化量	63	上側確率	偏微分	63
ノンパラメトリック	多重比較	4	上側確率の積	対数尤度	33
発電機モータ寿命	Motorette試験	2	正規確率プロット	打ち切り	10
パラメトリック	多重比較	4	正規分布	LIFEREG	47
プロビット法	ニュートン・ラフソン法	59	正規分布	Norm.dist()	18
ヘシアン行列	2階の偏微分行列	50	正規分布	確率密度関数	18
ヘシアン行列	2階の偏微分行列	64	正規分布	打ち切り	11
ヘシアン行列	2階の偏微分式	56	正規分布	累積分布関数	18
ヘシアン行列		63	正規分布上側確率		23
薬理試験	探索的	1	生存時間(パラメトリ	JMP	39
レンガを積む	段階的	13	生存時間解析	JMP	39
ワイブルプロット		39	生存時間解析	書物	40
ワイブル分布	寿命データ	42	生存時間解析	打ち切り	7
位置	平均値	27	設定パネル	Arrhenius摂氏	42
位置と尺度	設定パネル	43	設定パネル	位置と尺度	43
一変量の分布	JMP	39	設定画面	ソルバー	17
加速試験	過酷な温度・湿度	45	測定限界値	打ち切り	60
可視化	確率密度	53	多重比較	ノンパラメトリック	4
可視化	上側確率	53	多重比較	パラメトリック	4
過酷な温度・湿度	加速試験	45	打ち切り	トップコーディング	60
回帰	対数尤度関数	61	打ち切り	回帰分析	6
回帰	尤度比検定	34	打ち切り	回帰分析	29
回帰直線	95%信頼区間	36	打ち切り	最尤法	15
回帰直線	分散	37	打ち切り	正規確率プロット	10
回帰分析	打ち切り	6	打ち切り	正規分布	11
回帰分析	打ち切り	29	打ち切り	生存時間解析	7
回帰分析	打ち切りデータ	47	打ち切り	測定限界値	60
回帰分析	発毛データ	60	打ち切り	打ち切り	4
確率密度	可視化	53	打ち切り	対数尤度関数	49
確率密度	積	19	打ち切り	対数尤度関数	54
確率密度の積	対数尤度	33	打ち切り	肺活量データ	49
確率密度関数	正規分布	18	打ち切り	標準偏差	9
共分散行列	95%信頼区間	36	打ち切り	標準偏差	47
共分散行列	寿命の二変量	36	打ち切り	平均	47
共分散行列		64	打ち切り	平均値	9
計量経済学	Tobitモデル	60	打ち切りデータ	回帰分析	47
検証	偏微分式	60	対数正規分布	あてはめ	43
最小2乗平均	算術平均	16	対数尤度	確率密度の積	33
最尤推定値		49	対数尤度	上側確率の積	33
最尤法	打ち切り	15	対数尤度関数	回帰	61
算術平均	最小2乗平均	16	対数尤度関数	打ち切り	49
尺度	標準偏差	12	対数尤度関数	打ち切り	54
尺度	標準偏差	27	対数尤度関数	偏微分	54
寿命データ	ワイブル分布	42	段階的	レンガを積む	13
寿命の一変量	設定画面	10	二変量の関係	JMP	39

肺活量データ	ソルバー	53	平均値	位置	27
肺活量データ	打ち切り	49	平均値	打ち切り	9
発毛スコア		4	偏差平方和が最小		16
発毛データ	回帰分析	60	偏微分	α	61
発毛試験	打ち切り	4	偏微分	β	61
比例ハザードのあてはめ	JMP	39	偏微分	μ	54
微分	標準正規分布	54	偏微分	σ	54
標準正規分布	$\phi(z)$	49	偏微分	上側確率	55
標準正規分布	$\Phi(z)$	49	偏微分	上側確率	63
標準正規分布	微分	54	偏微分	対数尤度関数	54
標準偏差	尺度	12	偏微分	分散 σ^2	59
標準偏差	尺度	27	偏微分	σ	59
標準偏差	打ち切り	9	偏微分式	検証	60
標準偏差	打ち切り	47	変化量	ニュートン・ラフソン法	57
分散	回帰直線	37	変化量	ニュートン・ラフソン法	63
分散 σ^2	偏微分	58	尤度比カイ2乗		24
平均	位置	12	尤度比検定	回帰	34
平均	打ち切り	47	尤度比検定		33
			累積分布関数	正規分布	18

Excel, JMP ファイル一覧

名前	サイズ
 正規打ち切り_スライド中の図表_05-45	66 KB
 正規打ち切り_スライド中の図表_52-3,63-5_回帰	26 KB
 正規打ち切り_肺活量	2 KB
 正規打ち切り_肺活量_正規分布のあてはめ	40 KB
 正規打ち切り_肺活量データ	2 KB
 正規打ち切り_肺活量データsas	1 KB
 正規打ち切り_発毛_回帰分析	85 KB
 正規打ち切り_発毛データ_対照群	12 KB
 正規打ち切り_発毛データ_4群	31 KB
 正規打ち切り_発毛データ_4群sas	1 KB

非売品，無断複製を禁ずる

第 7 回 続高橋セミナー

正規分布を仮定したち切りデータを含む

回帰分析入門

BioStat 研究所(株)

〒105-0014 東京都 港区 芝 1-12-3

2018 年 7 月 17 日 高橋 行雄

takahashi.stat@nifty.com , FAX : 03-342-8035